NEXT-GENERATION SEQUENCING AS A USEFUL TOOL IN CANINE TUMOUR RESEARCH

The benefit of Next-Generation Sequencing in canine tumour research for early detection, new prognostic tests and cancer therapy by using target sequencing.

Nora Krafft

Supervisor: Dr. Péter Vajdovich

University of Veterinary Medicine Budapest

2023

Supervisor:

.....

Dr. Péter Vajdovich

Department of Clinical Pathology and Oncology

University of Veterinary Medicine Budapest

November 2023

Abstract

The use of Next-Generation Sequencing and its technology and methods in canine oncology is established in this paper. There are many methods for tumour treatment in the veterinary field, however, target sequencing specifies therapy options, as well as earlier diagnosis and more precise prognosis.

This paper contains the preparation methods and the extension techniques up to the appropriate control of samples and their quality and quantity inspection before sequencing. Additionally, it displays the various techniques for sequencing and its analytical methods. Furthermore, the application of Next-Generation Sequencing is portrayed to clarify the identification, testing and treatment of cancer. Moreover, the importance of sequencing is shown in canine tumour research for the early diagnosis of genomic alterations.

Abstract in Hungarian

Szakdolgozatomban bemutatom az ún. “Next-Generation Sequencing” alkalmazását, technológiáját és módszereit, különös tekintettel a kutyák daganatos betegségeiben. Az állatgyógyászatban számos módszer létezik a daganatok kezelésére, azonban a célszekvenálás meghatározza a terápiás lehetőségeket, valamint a korábbi diagnózist és a pontosabb prognózist.

Sazkdolgozatom tartalmazza az előkészítési módszereket és az ún. “extension”, vagyis kiterjesztési technikákat egészen a minták megfelelő ellenőrzéséig és a szekvenálás előtti minőségi és mennyiségi ellenőrzésig. Ezenkívül bemutatja a különböző szekvenálási technikákat és annak analitikai módszereit. Továbbá bemutatják a következő generációs szekvenálás alkalmazását a daganatok azonosításának, tesztelésének és kezelésének tisztázása érdekében. Ezen túlmenően szakdolgozatomban megemlítem a szekvenálás fontosságát, amely kimutatható a kutyák daganatainak kutatásában, így a genomiális elváltozások korai diagnosztizálásában.

Table of Contents

1. Useful Terms

Next Generation Sequencing:

Next Generation Sequencing, or NGS, is a sequencing method where millions of sequencing reactions are carried out in parallel, increasing the sequencing throughput.

Reads:

The output of an NGS sequencing reaction. A read is a single uninterrupted series of nucleotides representing the sequence of the template.

Read Length:

The length of each sequencing read. This variable is always represented as an average read length since individual reads have varying lengths.

Coverage:

Next Generation Sequencing, or NGS, is a sequencing method where millions of sequencing the number of times a particular nucleotide is sequenced. Due to the error-prone sequencing reactions, random errors could occur. Therefore, 30x coverage is typically required to ensure each nucleotide sequence is accurate.

Deep Sequencing:

Sequencing where the coverage is greater than 30x. This is used in cases dealing with rare polymorphisms in which only a subset of the sample expresses the mutation. This method increases the range, complexity, sensitivity, and accuracy of the result.

Adapter:

Unique sequences are used to cap the ends of a fragmented DNA. The adapter's functions are as follows: 1) allow hybridization to the solid surface; 2) provide a priming location for both amplification and sequencing primers; and 3) provide barcoding for multiplexing different samples in the same run.

Library:

A collection of DNA fragments with adapters ligated to each end. Library preparation is required before a sequencing run.

Alignment:

Mapping a sequence read to a known reference genome.

Reference sequence/genome:

A fully sequenced and mapped genome is used for the mapping of sequence reads.

Specificity:

The percentage of sequences that map to the intended targets out of total bases per run. [40]

2. Abbreviations

Next-Generation Sequencing	NGS
Base Pairs	bp
Unique Molecular Identifiers	UMI
Polyethene Glycol	PEG
Emulsion PCR	ePCR
Quality Control and Quantification	QCQ
Quantitative PCR	qPCR
Copy DNA	cDNA
Genomic DNA	gDNA
Deoxyribonucleoside Triphosphate	dNTP
Promyelocytic leukaemia-retinoic acid receptor alpha	PML-RARA
Epidermal Growth Factor Receptor	EGFR
Non-small cell lung cancer	NSCLC
Sanger Sequencing	SS
Human epidermal growth factor receptor	HER2
Estrogen receptor	ER
Progesterone receptor	PR
Breast cancer gene 1	BRCA1
Breast cancer gene 2	BRCA2
Estrogen receptor modulator	SERM
Tumour, Node, Metastasis	TNM
Internal tandem Duplications	ITDs
Copy Number Variants	CNVs
Single-Nucleotide Variants	SNVs

3. Introduction

Next Generation Sequencing is a modernized technology that can identify an enormous number of nucleotides of genetic material by adding fluorescent nucleotides with DNA polymerase to a growing DNA template strand. With the use of Next Generation Sequencing, scientists can use fragmented DNA or RNA and simultaneously sequence hundreds of thousands of genes at once and therefore can detect unusual or rare variations in genetic material in a short amount of time. It allows us to have insight into understanding diagnosing diseases, such as cancer, but also to determine nonspecific genetic diseases and treatment options.

Next-Generation Sequencing is an improved DNA sequencing method and became more popular in the early 2000s. It is based on Sanger Sequencing, also called First Generation Sequencing, which was a method used to detect a complete human genome and selected animal and plant genomes.

After Frederick Sanger identified chain termination in 1975, he was able to replicate the DNA by duplicating the nucleotides. Initially, he labelled the synthetic nucleotides with radioactive material but then changed it to fluorescent, which is a safer way to identify the nucleotides. [2]

Sanger Sequencing started to be internationally acknowledged in 1990 when scientists started the Human Genome Project. It took until 2003 to sequence 85% of the human genome, but it only reached completion in 2022, meaning that it took 32 years to sequence the whole human genome, which consists of 3.2 billion base pairs. As this discovery had an impact on developing the healthcare system it also proved to be very time-consuming and inefficient. This led to a more advanced and accurate technology called Next Generation Sequencing, which can sequence the human genome within a day. [3]

With Next Generation Sequencing it is possible to sequence billions of strands simultaneously within a day, whereas Sanger Sequencing could only do one at a time. NGS does not specify DNA only but can also read RNA, which then is transcribed backwards to DNA. The DNA or RNA is collected from either a blood or tissue sample or cell cultures and then purified and centrifuged.

The next step is the library preparation. The short DNA fragments are collected from long DNA stretches, which then get cut into a specific size. After the DNA is cut by enzymes,

so-called adaptors connect to the DNA fragments. These adaptors contain the necessary information for sequencing but also contain an index that aids in identifying the sample. The sample must have a precise size and a high enough concentration for the sequencing.

Following the library preparation, the sequencing can start. On a glass surface of a flow cell the short pieces of DNA, also called oligonucleotides, bind to the surface of the flow cell. The oligonucleotide matches the adaptors of the library. This happens because the library gets denatured to form a single DNA strand, which then gets added to the flow cell and connects with the oligonucleotides. This will result in a strand called the forward strand. After that, the reverse strand is replicated, and the forward strand is washed away. This means that the library is now bound to the flow cell. Then fluorescence and PRC are added to bind and identify the neighbouring oligonucleotides which then form a connection. This process is also called Annealing. After that, the Extension happens, where the strands get copied and then denatured or melted. This process is repeated. These localized clusters are made, and the reverse strand gets cut off and washed away. With that, the forward strand stays connected to the flow cell and the sequencing can start.

The first step is that the sequencing primer binds to the forward strand. Following that the fluorescence nucleotides, guanine (G), adenine (A), cytosine (C), and thymine (T) get added to the flow cell with the DNA polymerase enzyme. Each nucleotide has a different colour or fluorescence tag and a terminator, therefore only one nucleotide gets sequenced at a time. Following that the first complementary base binds to the sequence which then gets read by the camera and records the colours of each cluster. A new solution is added and with that, the terminator is flushed away. Another solution with fluorescence nucleotides and DNA polymerase is added again and another nucleotide gets sequenced. This cycle is continued as often as it is set on the camera or sequencing system. At the end of the cycle, the sequences and the first index get washed away. In case more sequencing is needed the paired end of the bridge forms again and the oligonucleotide is used as a primer which then is sequenced as a second index and reverse strand of the library. This is called the unique dual indices which allows up to 384 samples in one flow cell. The DNA polymerase makes the reverse strand, and the forward strand is cut but also washed away. This allows the sequencing of the reverse strand.

Once the sequencing is complete, the system continues to filter and map out the unreadable DNA such as overlapping, leader sequences, or which are of low intensity.

After the filtering process, the samples are identified and sorted by reading the indices and referenced to the genome. This is also called mapping. Longer stretches of DNA or RNA can be analysed better by showing the correct alignment. The data is then demultiplexed which sorts the sequence reads into separate files for each sample in a sequenced run.

Sequencing aims to have coverage across the target DNA without missing any areas. When reading the nucleotides, the number of reads is 30x. The read depth is suitable for detecting mutation events in cancer at an average of 1500x.

Not only is Next Generation Sequencing used for diagnosing cancer and rare diseases but can also serve to find the optimal and specific treatment for cancer. It is also applied in various fields such as ecology, botany, and medical sciences. [4]

4. Library Preparation

4.1 Fragmentation

For the genomic material to be sequenced, it is necessary to extract the nucleic acids and fragment them into smaller pieces to make it readable for the sequencing technologies. This is commonly done by mechanical force by using nebulization where the air pressure can break down the DNA between 700-1330 bp within a few minutes. Due to high pressure, the mist is formed by pushing the DNA solution into a small hole, which then gets collected in the nebulizing unit. The size of the DNA can be changed by altering the gas pressure, the viscosity of the solution and the temperature. [5]

Another method of fragmenting DNA is by using sonication. The DNA is exposed to short durations of sonication, which causes hydrodynamic shearing. The DNA in the solution is degraded by breaking the hydrogen bonds which ruptures the helix. The disadvantage of sonication is that most machines cannot shear more than 300-500 bp. [6]

An effective alternative to mechanical fragmentation is enzymatic digestion. This method is less cost-effective and processes samples quickly at the same time and therefore reduces sample loss. However, the major issue with this method is sequence bias, since many enzymes that cleave the DNA have a recognition sequence or sequence preferences. Commonly used enzymes are transposases which create nicks on the DNA strand or move segments of the DNA, leading to fragmentation. [7]

4.2 Adaptor Ligation

To make the library complete, ligase enzymes covalently connect the adaptor to the DNA fragments. Adaptors have multiple functions such as attaching the sequences to a flow cell or making the platform compatible with the sequencing system. Adaptors may contain indexes that function like a barcode for identification and enrichment of samples. For better sampling adaptors can contain unique molecular identifiers (UMIs) which provide variant identification but correct errors and increase the accuracy by uniquely tagging each molecule in the library sample. To match the oligonucleotides of the flow cell, the adaptors connect to both ends of the DNA. This step is needed to bind the library to the flow cell.

4.3 Size Selection

Following adaptor ligation, the nucleic acid fragments need to be arranged according to size. This is a profound step to increase the quality of the data leading to increased capacity and improving the workflow of the sequencing. This cost-effective step eliminates the inadequate fragments from the library before the sequencing starts. This ensures that the library fragments are within the ideal size range for the sequencing instruments, being between 200-700 bp. Low molecular weight fragments, undesired enzymes, nucleotides, or buffer components are concurrently eliminated as selected fragments are placed into certain size ranges. Size selection should not be confused with fragmentation, which is a previous step of the library preparation, where mechanical force is used to shear the DNA. The most common ways used for size selection are gel-based size selection and magnetic bead-based size selection.

4.3.1 Gel Based Size Selection

This approach has a limited throughput but is dependable and efficient for size selection and therefore a preferred option. The fragments of certain sizes are manually removed from an agarose gel. The fragments are divided by using a molecular weight marker which helps select the preferred size. Automated gel-based methods are also used and offer improved reproducibility and are less time-consuming. With gel-based size selection larger fragments from the DNA sample bigger than 100 bp or smaller RNA fractions from RNA samples, which are very close in size to adaptor dimers, can be removed. [8]

4.3.2 Magnetic Bead-Based Size Selection

Using inert beads with polystyrene cores that are covered in magnetite and a layer of carboxyl molecules, magnetic bead-based size selection is a practical method. Nucleic acids bind to the beads reversibly when polyethene glycol (PEG) and salt are present. By adjusting the proportion of PEG, salt and beads to the nucleic acids, it is possible to regulate the sizes of fragments that bind to the beads. The magnetic force is applied after the nucleic acids bind to the beads. This allows the separation of the beads and therefore the bound fragments from the rest of the material. The fragments can then be obtained according to their size by either direct collection in the supernatant, which is typically the

case for smaller fragments, or indirectly from the beads via elution brought on by a change in buffer, which is typically the case for larger fragments. The advantage of magnetic bead-based size selection is that is automatic and has a high throughput. This method is suggested for fragments of over 100 bp in length and above. [9]

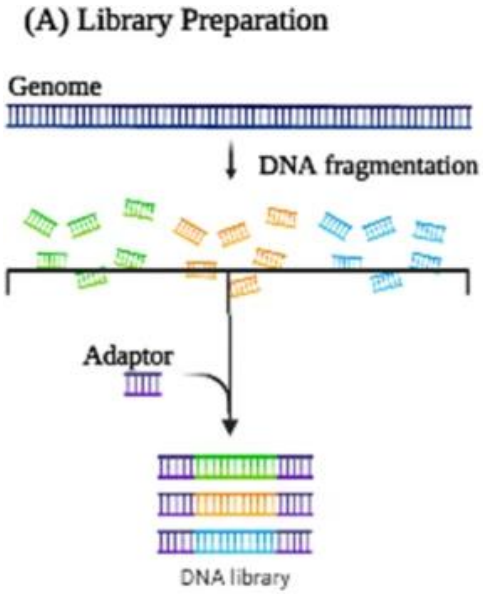

Diagram 1: Library Preparation

The genome is fragmented into smaller pieces by mechanical force or sonication.

Adaptors are added to the DNA fragments through enzymatic reactions with DNA ligase.

5. Amplification

After the library preparation, the DNA fragments are amplified on a solid phase, which helps increase the detection during sequencing. The DNA fragments are bound to a flow cell and then PCR is performed to create a set of identical fragments. To produce enough DNA to be sequenced, amplification is required to accurately quantify the fragments to obtain the optimal cluster density. [10] These fragments are also called cluster generation, where millions of copies of single-stranded DNA are produced. Amplification can be done by using bridge PCR or emulsion PCR. [12]

5.1 Bridge Amplification

For bridge PCR, the generation of DNA clusters takes place on a flow cell. On the flow cell, two types of oligonucleotides can be found. The oligonucleotides are complementary to the adaptor region of the fragment strand. The strand from the library, which is the forward strand, attaches to the oligonucleotide of the flow cell. After adding the polymerase, it makes the reverse strand of the fragment. This forms a double-stranded molecule, which gets denatured, and the forward strand is washed away. The strand is

amplified through bridge PCR and can now fold over to the adaptor region of the oligonucleotide next to it on the flow cell. Following this step, the polymerase forms the complementary strand, forming a double-stranded bridge. After denaturing the bridge, two single-stranded copies of the molecule are fixed on the flow cell. This procedure is carried out for each cluster on the flow cell at the same time. After the amplification, the reverse strands are removed from the flow cell and washed off, leaving only the forward strand.

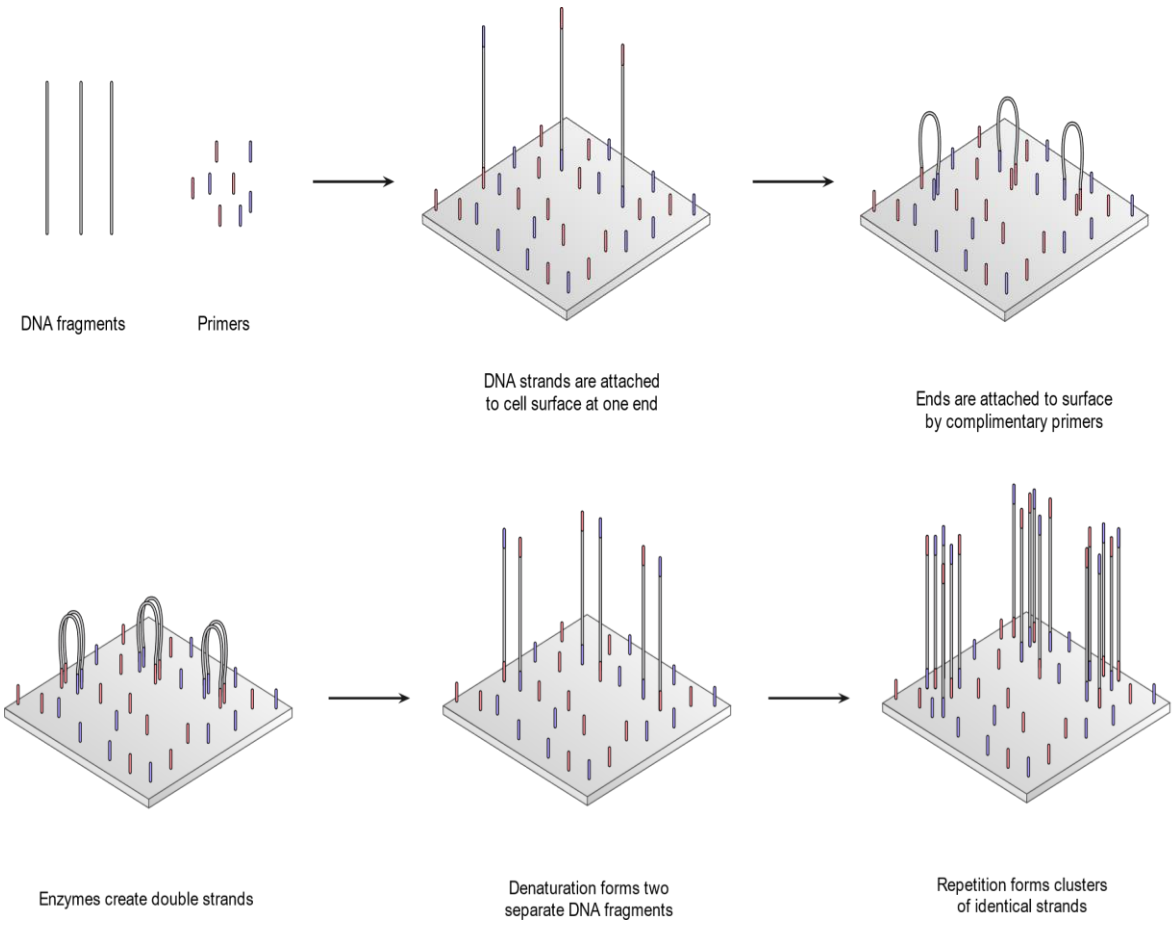

Diagram 2: Bridge Amplification

5.2 Emulsion PCR

When using the emulsion PCR method, the DNA fragments are mixed with beads that have forward primers attached to them. Beside the beads and the DNA fragments, there is a buffer which gets mixed with oil creating an emulsion. The DNA double-strand gets denatured, and the single-strand anneals with the primer of the bead. The other strand connects with the reverse primer which is also added to the emulsion. Following the Annealing, the primers extend to form new strands with the help of the polymerase enzyme. As this cycle continues, the DNA fragments are amplified on the bead. For sequencing, the bead can be used with the attached fragments. [13]

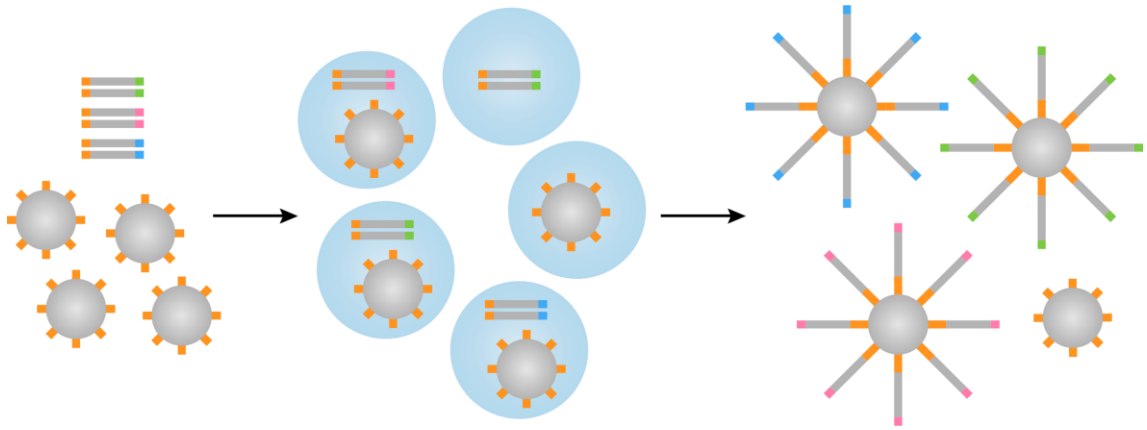

Diagram 3: Emulsion PRC

The beads, containing primers, and the DNA fragments are mixed with a buffer and oil, creating an emulsion. The DNA get denatured and connects to the primer of the bead.

6. Quality Control and Quantification

This is a crucial step before starting to sequence. It determines the library size and the distribution and concentration of the sample but can also reveal whether the samples are deteriorated or contain potential contaminants.

The quality control gives information about the size and distribution of the library, Larger DNA fragments are more likely to be sequenced more often than smaller DNA fragments, therefore this is a crucial procedure for larger DNA fragments as more polymerases can be attached compared to smaller fragments. Poor quality libraries can be signs of potential contaminants or preparation-related artefacts.

The amount of DNA in the sample is measured during the quantity control procedure to determine whether the concentration is accurate for sequencing since under-clustering could result in additional costs. On the other hand, excessive clustering in a highly concentrated library can damage the flow cell and result in sequencing failures, which might lead to resending the sample.

When preparing the libraries, there are three primary methods for the quality and quantity control. Microfluidics is used for quality analysis, while the quantity of the library is checked with quantitative PCR and fluorometric methods.

6.1 Quantitative PCR (qPCR)

A polymerase chain reaction called quantitative PCR or in short qPCR is used to amplify and measure fragments of the copied DNA and genomic DNA. Real-time monitoring of DNA amplification is possible with qPCR and is finished after measuring the DNA concentration. Because the adaptor attached to the DNA fragments already contains primers, that can be used in qPCR, it is practical to include the qPCR in your library preparation. At the same time, you can quantify only the adaptor-ligated DNA. Since qPCR only measures templates containing both adaptor sequences, it can give information about the concentration and whether the library size is appropriate. In addition, qPCR is a very sensitive way to measure DNA in comparison to other methods. In the case of PRC-free library kits, the qPCR allows to amplify and quantification of the DNA, whereas other methods such as fluorometric would not be able to increase the DNA concentration.

6.2 Fluorometric techniques

The fluorescence dyes which bind specifically to the double-stranded DNA molecules can be detected with the fluorometric technique. However, these methods are unspecific as they can bind to incomplete DNA fragments or residual primers. As such, fluorometric systems only give information about the concentration of the DNA but cannot tell anything about the purity of the samples. [14][15] Common dyes used for this method include Pico Green and Hoechst dyes as they have better affinity for DNA but little affinity for RNA. This means, that is possible to see if the DNA samples are contaminated with RNA, as they show increased emission when bound to DNA.

6.3 Microfluidic techniques

For the evaluation of the quality of the library size and distribution, the microfluidic technique is used, which is an electrophoresis system. The library size is significant because it enables sequencers to focus on the DNA fragments which are most complaisant to generate the longest reads. Smaller fragments would downstream the analysis, making it more difficult to assemble. Although this method has low accuracy with broad-sized library distribution, this technique can also be used for quantification, but only for narrow-sized libraries, for example for ligation-based RNA libraries. This method is recommended for assessing sample quality because it provides accurate information about the concentration and quality of the samples. [16]

7. Sequencing

7.1 Sequencing by Synthesis

At the end of the amplification, the reverse strands are washed off, leaving only the forward strand on the flow cell. Then a primer is added which can attach to the adaptor primer binding site of the forward strand. The primer is needed to initiate the DNA synthesis, where the polymerase adds a nucleotide which is tagged with fluorescence dye. The fluorophore acts as a blocking agent, so the enzyme can only connect one base per round to the forward strand. For every nucleotide, a different colour is used, creating a unique emission to each of the four base pairs. If a base is added, no other nucleotide can attach to the DNA strand until the fluorophore, which is also used as a terminator, is cut off. [17] After the template is terminated the DNA is loaded onto the sequencing machine and gets electrophoresed. This sorts the DNA according to size, having the smaller fragments at the bottom and the larger ones at the top of the capillary. Now the DNA sequence can be read through the fluorescence emission, which flows through the gel. The Sequencer can identify the colour of the dye with the help of a laser, and therefore records which nucleotide was added. Modern systems can analyse around 8-96 sequencing reactions at once. [18] After the read is completed, the sequenced complementary strand is washed off. In the following step the first index read primer is introduced to the DNA fragment. This allows the nucleotides to bind to the DNA again and generate another sequence. Then the read is washed off, allowing the DNA fragment to fold over and build a bridge to the oligo on the flow cell. This way index 2 can continue and the polymerase can extend the second strand. After the formation of the double-stranded DNA gets denatured and the primer end is blocked. The original forward strand is cut off and washed away. Now the second read can begin with the sequencing. This is repeated until the length of the read is achieved. The product of the second read is also washed away. The process is repeated until all the fragments are represented. [19]

7.2 Pyrosequencing

This technique can only achieve chain lengths of around 300-500 nucleotide base pairs but is more affordable and commercially available, compared to other sequencing methods. Millions of identical copies of each fragment are produced by PCR and distributed throughout thousands of wells, with just one kind of DNA fragment per well. The enzymes DNA polymerase, ATP sulfurylase, apyrase, and substrates luciferin and

adenosine 5´ phosphosulfate are added to the DNA fragments. One of the four types of nucleotides, which make up the DNA, are put into the wells. The DNA polymerase starts to integrate the nucleotides into each single-strand DNA template, which results in pyrophosphate production. The pyrophosphate is then transformed into adenosine triphosphate (ATP) with the help of adenosine 5´ phosphosulfate and ATP sulfurylase. The next step is the conversion of luciferin to oxyluciferin by luciferase with the involvement of ATP. This reaction emits light which can be detected and relates to the quantity of ATP involvement in the conversion. Apyrase is produced when unused nucleotides and ATP break down, enabling the reaction to continue with another nucleotide. Until the synthesis is finished, this process is repeated by adding each nucleotide one at a time. The detector picks up the light emission and determines the kind of quantity of nucleotides that are added. The light produced from the identical fragment will be three times brighter if three cytosine nucleotides are added consecutively than from DNA fragments that only have one cytosine added. If adding cytosine results in no release of light, then the next complementary base in the single-strand DNA template must be one of the other three nucleotides.

Pyrosequencing is used to obtain a segment of the DNA´s genetic code but also for identification of nucleotide polymorphisms, insertion deletions and other alterations. It is possible to further narrow down genomic areas linked to a genetic condition to discover certain genes or variants or to distinguish hereditary adverse responses to a given treatment before prescribing it. [20]

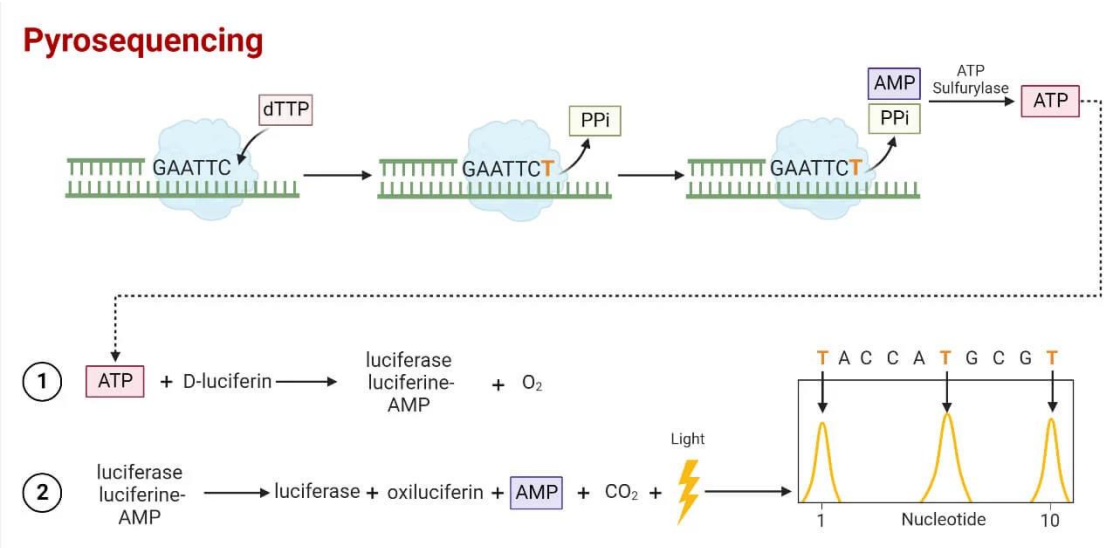

Diagram 4. Pyrosequencing

In Pyrosequencing ATP is formed which is needed for the conversion of luciferin to oxyluciferin. This process emits light which can be detected.

7.3 Sequencing by Ligation

This method is different to the previously mentioned techniques as it does not include the DNA polymerase enzyme, but instead the DNA ligase enzyme. The enzyme is used to identify the DNA sequence´s specific position. To discover the underlying sequence of the wanted DNA molecule the sensitivity of the ligase enzyme is used. [21]

The DNA ligase enzyme can connect two ends of the DNA molecule and therefore reacts to base-pairing mismatches. For the preparation of the sample emulsion PRC is used and their beads are put on a glass surface. The more beads are present the better the throughput of the technique. For sequencing a library of oligonucleotide probes with four different fluorescent dyes attached to their end, are ligated to each other. Six degenerate bases and two specific bases are included in each probe. To start the sequencing process, the primer binds to the adaptor sequence and synthesises the correct probe to the DNA fragment, meaning that the unknown DNA is matched with the bases. The enzyme then ligates the sequence to the primer. The remaining oligonucleotides are washed away, and the fluorescent signal is recorded. After the signal detection, the degenerate bases are cleaved, so that new probes can be added, and the next cycle can begin. The ligation continues for seven cycles until the DNA strand is denatured and another primer is added to repeat the sequencing until the reading of the DNA fragment is complete.

The primary disadvantage of Sequencing by Ligation is, that it is only suitable for short read length, which is not appropriate for many applications. [22]

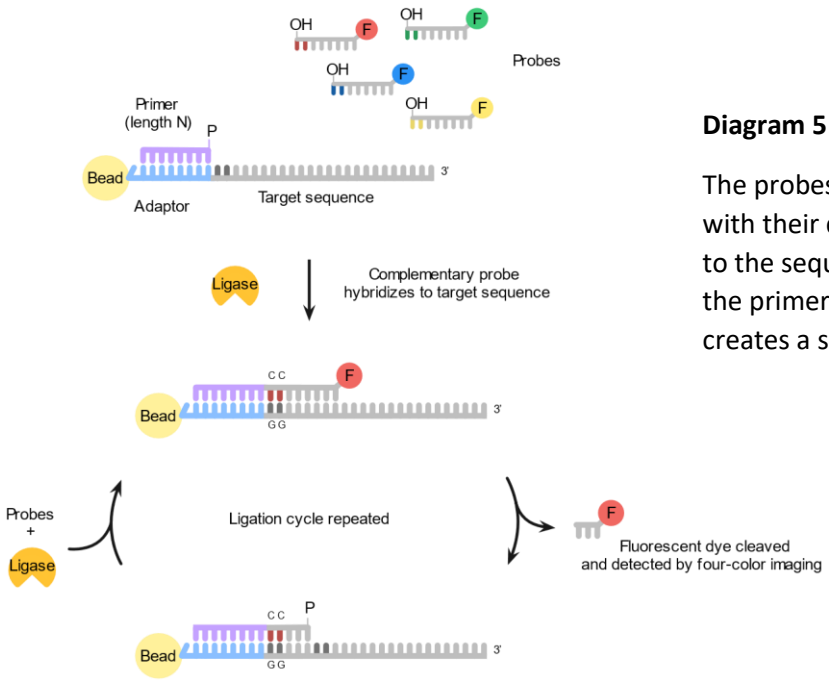

Diagram 5. Sequencing by Ligation

The probes contain the oligonucleotides with their different dye. The primer binds to the sequence and the DNA Ligase fuses the primer and the probe together. This creates a signal which can be recorded.

7.4 Ion Semiconductor Sequencing

Ion semiconductor sequencing uses the hydrogen ion release that occurs throughout the sequencing operation to identify a cluster's sequence. Every cluster is situated precisely above a semiconductor transistor that can identify variations in the solution's pH. A single H+ is released into the solution during nucleotide integration, and the semiconductor detects this. This method is based on the reaction of Pyrosequencing but is less cost-sensitive than Pyrosequencing itself. [23]

Naturally, when the deoxyribonucleoside triphosphate is added to a growing DNA strand a covalent bond is formed which releases a pyrophosphate and a hydrogen ion with a positive charge. The ion semiconductor can signal if a positive hydrogen ion is released after a deoxyribonucleoside triphosphate is added to the reaction. Single-stranded- DNA samples and DNA polymerase are added to microwells on a semiconductor. Once the nucleotides A, C, G or T are added, the polymerase incorporates these, and a biochemical reaction takes place. As a result, the hydrogen ion is released and the pH of the solution changes. All nucleotides that were not incorporated are washed out and with the introduction of new nucleotides, sequencing can continue. This is detected by the conductor, which contains an electronic chip and can be translated to a computer system. Therefore, it is not necessary to label the nucleotides with dyes, as the signal will be processed directly to the software.

This method proves to have increased sequencing speed and little operational and upfront expenses. The sequencing rate is limited by the cycling rate of nucleotide substrate that goes through the system. [24]

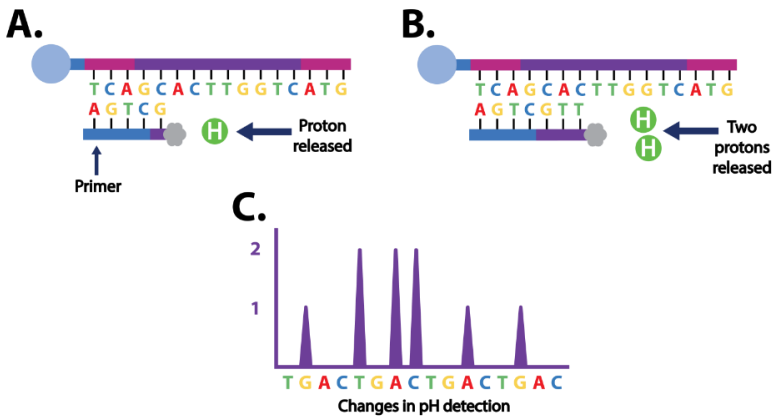

Diagram 6. Ion-Semiconductor Sequencing

Primer attachment releases positively charged hydrogen which alters the pH. This can be measured with the ion semiconductor.

8. Filtering and Mapping

Not all sequences are of importance and sometimes only a few can be used for analysis. To reduce downstream analysis and requirements the unwanted sequences are removed so only the desired sequences are utilised. [25] The cause for bad reads is low quality, reads that do not align due to mismatched bases, or alignment in the wrong place of the genome. Another reason for filtering the sequence sample is to remove the adaptors and reduce technical errors that might lead to a lower quality of the reads. [26]

To fully analyse the data, the genetic phenotype relationship must be distinguished. First, the chromosomes are broken up into smaller pieces that may be analysed and replicated. Next, the fragments are arranged in a certain order to their specific positions on the chromosomes. [27]

The method where NGS reads are allocated to their sample of origin based on their matching molecular barcode is known as demultiplexing. Demultiplexing creates a single sequence by selecting the appropriate sequence from each file, forming them into a single sequence. [28] This step includes a certain software which reads the barcode and sorts the sequences into files. Only identifiable barcodes are processed and not too like each other, to avoid any mismatch.

The readings overlap each other when they are matched to the reference genome. Two sequencing reads are produced from the same library using paired-end sequencing. The algorithm recognizes that these readings belong together during sequence alignment. Greater reliance may be placed on the accuracy of the alignment when analysing longer segments of DNA or RNA.

9. Metrics

One crucial sequencing statistic is read depth, which is the quantity of reads for a nucleotide. An ideal read depth for whole genome sequencing is 30x. For example, a 1500x average read depth can be used for uncommon mutation events in cancer.

Coverage is an additional critical measure in sequencing. The goal is to ensure that the target DNA has no gaps, for the completion of the sequence. [29]

10. Application of Next-Generation Sequencing in Clinical Oncology

10.1 Identification of Cancer Mutations using NGS

Uncommon somatic mutations can be efficiently and accurately detected by NGS methods. Numerous tumours such as bladder cancer, renal cell carcinoma, small cell lung cancer, prostate cancer, acute myelogenous leukaemia, and chronic lymphocytic leukaemia have been effectively treated with NGS to find new mutations. A unique genetic recombination known as PML-PARA was successfully detected using whole-genome sequencing in individuals suffering from a rare type of acute promyelocytic leukaemia, which was not identifiable through conventional cytogenic approaches.[30] Numerous novel genetic abnormalities and related potential therapeutic targets have been discovered in several malignancies by using whole-genome sequencing. These discoveries have shed an important light on the process behind carcinogenesis.

The most common genetic changes affecting the Epidermal Growth Factor Receptor (EGFR) in non-small cell lung cancer (NSCLC) are microdeletions at exon 19, which are closely linked to the patient's reaction to tyrosine kinase inhibitor therapy. Deep Next Generation Sequencing was performed on a set of 116 NSCLC DNA samples that were examined by Sanger Sequencing (SS), of which 106 samples had exon EGFR deletions and 10 samples, which were control samples had no deletions. However, NGS was able to discover deletions in thirteen instances (12 %) which SS was unable to define. Furthermore, NGS revealed the existence of complicated (double/multiple) frameshift deletions that resulted in an alteration in 21 (20%) of cases. In addition to that, NGS showed the existence of subpopulations of DNA molecules with EGFR deletions distinct from the primary one in 46 (43%) of tumours with possible pathogenetic and clinical implications. [31]

10.2 Genomic Testing for Hereditary Cancer Syndrome

Around five to ten per cent of malignancies are inherited. Next Generation Sequencing is now the most used genetic testing technique and is regarded as a gold standard for finding mutations in the United States and Europe. Numerous types of changes including single-nucleotide substitutions, tiny insertions and deletions or huge genomic duplications might be found. [32] Commercial testing for breast cancer in the US and Europe is based on Sequencing, and polymerase chain reaction amplification of individual exons. NGS is a useful method for identifying uncommon variants. The variant detection rate is significantly increased since it permits the testing of several genes simultaneously. Although many hereditary cancer patients have tested negative for genetic abnormalities, NGS makes it simpler to identify the mutated genes that cause the disease. In a study, 52 probands out of 300 high-risk breast cancer families had mutations that had not been discovered before. [33] The earliest and most direct step towards NGS´s introduction into clinical practice will be its application in genetic testing for hereditary cancer disorders. Clinical trials have employed whole-genome or whole-exome sequencing of malignant tumours for individualised treatment.

10.3 Next-Generation Sequencing for Individualised Cancer Treatment

NGS is used not only to detect novel somatic and genetic changes but also to enhance personalised medical care. Numerous studies have shown that individual treatment of cancer can be done by using NGS.

Research has found that pancreatic cancer poses several technological challenges for genetic studies. The utilisation of Next-Generation Sequencing has demonstrated efficacy in investigating diverse cancer forms by revealing the intricate relationship between distinct DNA and RNA alterations inside the tumour. Despite the various obstacles presented by pancreatic tumours specifically clinical sequencing-based diagnostics, NGS and analysis are now applicable in in-patient treatment-related research to find therapy choices to enhance the patient's outcome. [34]

The most common genetic changes affecting the Epidermal Growth Factor Receptor (EGFR) in non-small cell lung cancer (NSCLC) are microdeletions at exon 19, which are closely linked to the patient's reaction to tyrosine kinase inhibitor therapy. Deep Next

Generation Sequencing was performed on a set of 116 NSCLC DNA samples that were examined by Sanger Sequencing (SS), of which 106 samples had exon EGFR deletions and 10 samples, which were control samples had no deletions. However, NGS was able to discover deletions in thirteen instances (12 %) which SS was unable to define. Furthermore, NGS revealed the existence of complicated (double/multiple) frameshift deletions that resulted in an alteration in 21 (20%) of cases. In addition to that, NGS showed the existence of subpopulations of DNA molecules with EGFR deletions distinct from the primary one in 46 (43%) of tumours with possible pathogenetic and clinical implications. [35]

After 510 breast cancer exomes were sequenced by the Cancer Genome Atlas Network, four different subtypes were found. These were categorised into basal-like (ER-, PR-, HER2-), Human epidermal growth factor receptor 2-enriched (ER-, PR-, HER2+), luminal B (ER+ and/or PR+, HER2+) and luminal A (ER+ and/or PR+, HER2-). 40% of luminal A cancers had a PIK3CA gene mutation and among luminal B tumours, TP53 and PIK3CA were mutated in 29%. With TP53 mutated in 80% of instances, the basal subtype showed a more consistent pattern of mutation than either the luminal A or luminal B subtypes. HER2 amplification which is present in 80% of these tumours, is what defines the HER2-enriched subtype. Studies suggest that tumours of the basal type often have the highest frequency of mutations, whereas tumours of luminal A type usually have the lowest rates of mutations. [36]

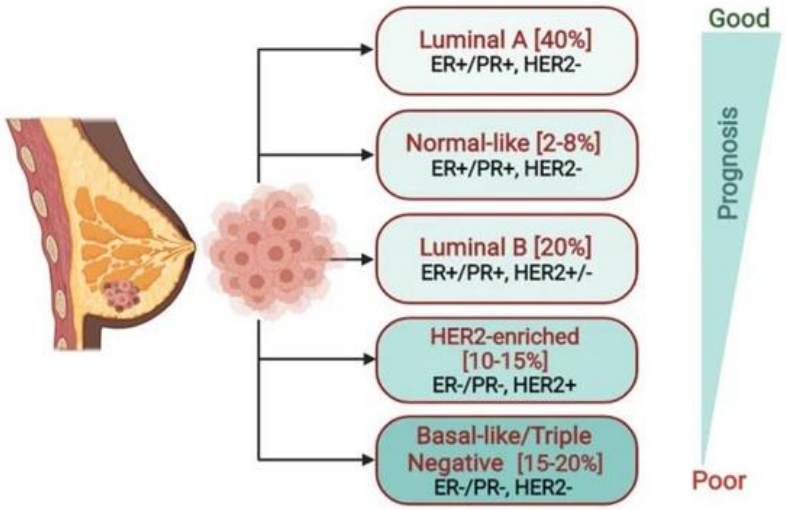

Diagram 7. Molecular Subtypes of Breast Cancer
Categorization of the subtypes based on their hormone receptor expression.

Next Generation Sequencing technologies provide information that helps clinicians to make better decisions about diagnosis and treatment. For instance, a physical examination, histology, and mammography have been used to diagnose breast cancer. Genetics became a significant factor after the identification of breast cancer gene 1 (BRCA1), breast cancer gene 2 (BRCA2) and other biomarkers.[37] More precise treatment is possible based on the genetic markers including HER2, ER and PR. Beneficial to diagnostics, the discovery of genes linked to cancer has prompted the creation of drugs that target particular molecules, including trastuzumab, which is one of the first treatments for HER2+ breast tumours. [38] Trastuzumab binds to an extracellular domain as a monoclonal antibody, preventing the signalling of HER2. Moreover, it may promote antibody-dependent cellular cytotoxicity, which kills HER2-expressing cells. [39] Tamoxifen demonstrates estrogenic agonist and antagonist properties. It is a selective estrogen receptor modulator (SERM) that is patient-specific due to its dual activity of preferentially binding to estrogen receptors and exerting both estrogenic and anti-estrogenic effects. It produces antiestrogenic and anticancer actions in breast tissue by antagonistically competing with estrogen-binding sites. By inhibiting subsequent intracellular mechanisms that slow down the cell cycle it is categorized as cytostatic. [40]

10.3.1 The Use of NGS in Target Therapy

For diagnosis, the tumour subgroups that were formerly identified by morphologic criteria are now identified by genetic alteration. For instance, an oncogenic fusion between DNAJB1 and PRKACA in 15/15 patients in a research examination of fibrolamellar hepatocellular carcinoma. [41]

The second is identifying a suitable therapy which targets these genetic alterations, as more and more treatments have indications derived from the results of DNA sequencing. Inappropriately targeted medicines may fail to be beneficial for patients and may cause damage.

Next-Generation Sequencing can also benefit patients when they stop responding to their target therapy. This could be due to an issue known as resistance mutation. Resistance mutations may occasionally be restricted to one or a few loci. For instance, a single point mutation is usually the cause of resistance to EGFR-targeted therapy in cancer, and this resistance may be addressed by simply using an alternative medication. [42]

However, through a complex process of epigenetic regulation, glioblastoma can develop resistance to EGFR-targeted therapy. Compared to a single gene assay, NGS provides a more thorough picture of tumour dynamics and has a higher chance of illuminating idiopathic resistance mechanics. Clonal EGFR mutations in extrachromosomal DNA reappear after the medication is discontinued. These findings point out that a very particular, dynamic, and adaptable mechanism via malignancies might sidestep treatments that aim to eliminate oncogenes that are on extrachromosomal DNA. [43]

Drug	Disease	DNA mutation	Action
Imatinib, Dasatinib, Nilotinib, Bosutinib	Chronic myelogenous leukemia	*BCR-ABL1* fusion	Indication for therapy
Ponatinib	Chronic myelogenous leukemia	*BCR-ABL1* fusion	Only indicated for T315I mutations
		T315I resistance mutation	
Erlotinib, Afatinib	Lung adenocarcinoma	*EGFR*	Indication for therapy
		Exon 19 deletions	
		L858R	
Vemurafenib, Dabrafenib	Melanoma	*BRAF* V600E	Indication for therapy
Tramatenib	Melanoma	*BRAF* V600E/K	Indication for therapy
Crizotinib	Lung cancer	*ALK* gene fusions	Indication for therapy
Cetuximab	Colon cancer	*KRAS* codon 12, 13	Contraindication to therapy
Olaparib	Ovarian cancer	*BRCA1* and *BRCA2* mutations	Indication for therapy

Diagram 8: Food and Drug Administration-approved drugs

Each medication has a unique genetic result which is included in the list of approved uses.

11. Next-Generation Sequencing in Canine Oncology

11.1 Importance of NGS in Canine Oncology

Tumours in dogs are as common as human tumours. Tumours in dogs not only develop spontaneously to those in people, but many of them also exhibit biological and histological characteristics. Both the clinical appearance and certain elements of clinical pathology are frequently comparable between dogs and humans. [44] There is comparatively significant homogeneity in dogs since many pure breeds have confined populations. Due to the absence of genetic variation, pure breeds require fewer harmful alleles to cause cancer and as a result mapping these alleles, identifying which are causative and contrasting them with the human genome that has the same cancer type are made easy. Furthermore, the lifespan of dogs is much shorter compared to people, therefore the clinical trials and research may be faster than in human oncology. [45]

In the past twenty years, the areas of genomics have grown rapidly, especially after the Human Genome Project was completed and after that, in 2005 the first canine genome was made public. Since then, a great deal of progress has been made in the fields of genomics which allowed us to sequence and comprehend the genetic composition of the genome and its encoded proteins. Breeding decisions, species categorization, disease resistance, pathophysiology, diagnostics, and biomarker identification have all been improved by the utilization of sequencing in the veterinary industry. In both humans and dogs predisposing factors such as environmental but also hereditary factors contribute to the explanation of why, in the US 1.6 million new cases are registered per year, in a population of 330 million people, whereas 4-6 million dogs of a population less than 90 million receive a cancer diagnosis annually. Similar to humans, dogs are more likely to develop cancer as they are and therefore up to 50% of dogs over 10 years of age develop cancer for the remaining lifespan. [46] Because many tumours are discovered at a late stage, when there has been microscopic or macroscopic dissemination, treatment is no longer possible and therefore canine cancer poses a considerable mortality risk. The enormous cost of canine cancer extends beyond physical consequences for the animal but also adds to the emotional and financial impact on the owner.

11.2 Differentiation of Genomic Alteration

Most genomic medical advancements have not yet been used in veterinary care, and development in canine genomics has not been nearly as quick as in human genomics. Very little of the scientific advances in human oncology have been applied to the field of cancer in dogs. For the standards of cancer care in veterinary medicine to catch up with standards in human medicine, there is an urgent need for additional research and development regarding the genetic predispositions underlying canine cancer syndromes as well as the detection, characterisation, and management of cancer in dogs.

Many somatic changes are quickly fixed by intracellular DNA repair mechanisms or, if severe enough cause the cell to die with no negative effects on the organism. However, when these changes happen at sites in the genome and are left uncorrected, a series of events result in the development of cancer. These changes provide the affected cells an advantage in terms of growth and/or survival by either inducing enhanced cell replication or by blocking the mechanism that blocks cell division. Over time the accumulation of additional somatic mutations might accelerate the development of tumours. This leads to cancer cells replicating more quickly, invading surrounding tissues, reaching distant organs through lymphatic vascular pathways, and evading the immune system's surveillance and control mechanism. [47]

The malignant mass is approximately 1 cm in size and weighs approximately when the number of cancer cells reaches 1 billion. At this point, the mass is usually detectable through physical and imaging examinations and might have already begun to cause clinical symptoms. The tumour spread is defined by the TNM (Tumour, Node, Metastasis) staging system. However, at its core, cancer is a genetic illness since it arises from genomic changes directly and cannot occur in the absence of such modifications. [48]

Changes linked to cancer can be either somatic, which are acquired after birth and only found in a portion of the body cells, or germline which are inherited and found in all cells. Humans are more susceptible to breast cancer due to germline changes that result in a predisposition to cancer, such as variations in the BRCA1 and BRCA2 genes. Dogs with mammary tumours have also been shown to have similar gene changes. [49]

Proactive cancer screening which can identify such malignancies at earlier stages is beneficial for patients with germline changes since they are typically diagnosed at a

younger age. Proactive screening in younger dogs, guided by the existence of germline mutations, is expected to exhibit growing therapeutic value and improve clinical outcomes. [50]

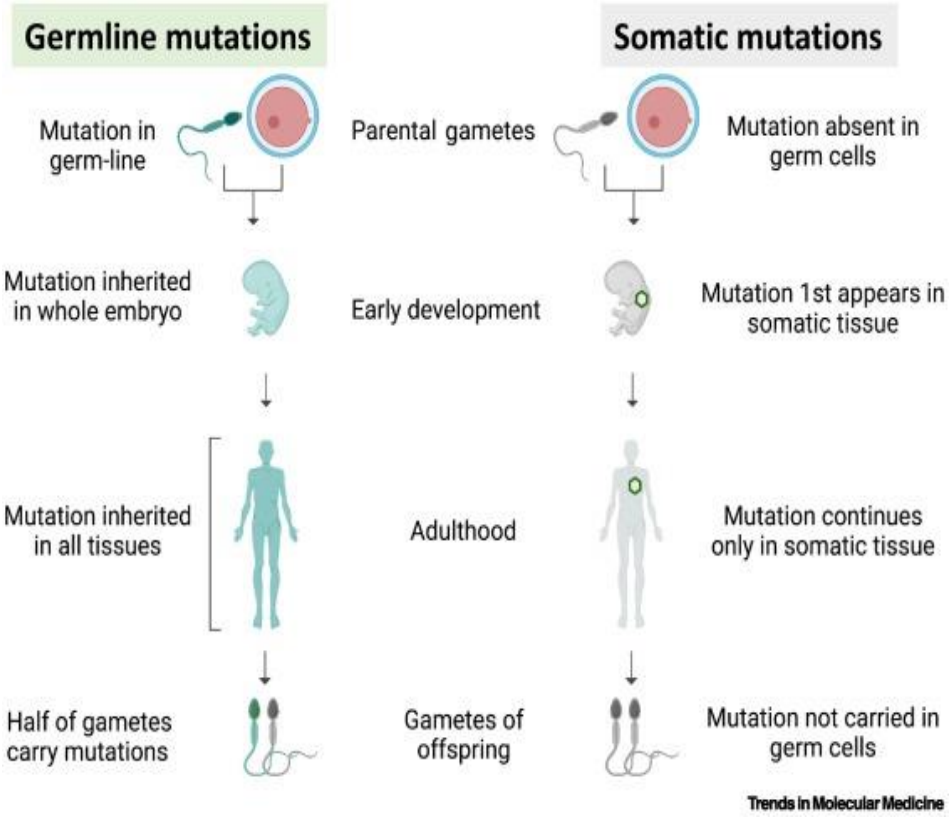

Diagram 9: Germline and Somatic Mutations

Somatic mutations are acquired after birth and can be found in part of the body cells, whereas germline mutations are hereditary and found in all body cells.

11.3 Target Sequencing for Diagnosis and Detection

Before becoming clinically evident, many tumours in humans develop for years, mostly inside the organ. This period offers a sizable window of opportunity for early diagnosis which may enable a circumscribed tumour to be surgically removed to cure the patient. This model also applies for canine cancer as with early discovery and appropriate surgical excision, clinical results are frequently favourable in cases of certain canine malignancies such as mast cell tumours and soft tissue sarcomas. [51]

Numerous genetic changes are present in every patient´s cancer, and no malignancies are the same even within the same cancer kind. There is no proven relationship between a certain genetic mutation and a specific tumour type. For instance, the canine BRAF V600E mutation is frequently seen in transitional cell carcinoma but is also prevalent in several canine cancer types, and similarly, the mutation is most frequently observed in human melanoma but can also be found in other malignancies. [52]

Different cancer types may respond differently to a given mutation, depending on the treatment implications. For instance, vemurafenib, a medication that targets the BARF gene, is more successful in treating melanomas in humans than in other cancer types. [53]

Uncertain diagnosis is a significant difficulty for veterinarians treating cancer patients. It is challenging to establish a clear treatment plan and prognosis in the absence of a conclusive diagnosis. A histopathology and/or a cytopathology assessment of the suspected tumour or mass is usually the first step towards a clear diagnosis, which is essential to establish the best course of therapy and prognosis. A pathologic diagnosis, however, may not always be attainable for several reasons, including insufficient differentiation of neoplastic cells, or lack of distinctive characteristics in the sample. Even with extra tests, a diagnosis can still be challenging which complicates the case management for clinicians by impeding the creation of a conclusive treatment plan and avoiding an accurate prognosis. Thus, more techniques are required to get beyond these diagnostic challenges.

Mutations of the tumours are essential to the characteristics of cancer that determine how the disease behaves clinically. Understanding the role of genomic alterations in cancer makes it possible to develop genomic biomarkers to help with the diagnosis, as well as implement precision medicine to target mutations and dysregulations of specific pathways in cancers in individual patients.

By Sequencing specific genes, it is possible to use small and cost-effective panels that concentrate on mutations with a high amount of biomarker evidence. This method is relatively low-cost and has high clinical actionability, which makes it valuable for veterinary oncology. Genomic testing can benefit especially with the use of a reporting framework that incorporates structured, comprehensive, and peer-reviewed genomic biomarker data, which are closely followed and regulated in the field of human cancer genomic diagnostics. A pathogenic mutation that is known to be enriched in, a particular tumour type in people or dogs is what makes up a diagnostic genomic biomarker. A brief, informative report for the veterinarian can include a description of each tumour's mutations, evidence statements indicating their function as biomarkers, and a qualitative assessment of the genomic results concerning the pathology findings and clinical history of the patient.

A recent study showed records of client-owned dogs with clinical genomic findings from results conducted for cancer or suspected malignancy, where the diagnosis was unclear and if the documented medical history or pathology report included any of the following terms: anaplastic, poorly differentiated, likely, potential, suspicious, suggestive, malignant neoplasm, round-cell tumour, atypical, and probable. If the pathologist offered at least one other differential diagnosis in addition to the diagnosis that did not contain a particular histotype (such as lymphoma, histiocytic sarcoma, or squamous cell carcinoma), the diagnosis was deemed equivocal. Regardless of whether a full pathology evaluation was carried out, dogs with strongly suspected malignancies for example dogs with pulmonary lesions with a history of distinct tumour histology, were included. Dogs with a clear diagnosis (i.e., a precise histotype diagnosis) were not included. After confirming that the samples met quality standards, the targeted genomic regions were enriched using a proprietary custom panel of hybridization-based capture probes, which allowed for the sequencing of the regions. This allowed for the evaluation of the presence of multiple mutation types, such as internal tandem duplications (ITDs), copy number variants (CNVs), and single-nucleotide variants (SNVs).

The study showed quantitative proof of the high value of genetic diagnostic assay that provides information that was previously unavailable and therapeutically relevant. Patients with ambiguous cancer diagnoses were included and 46% had already undergone at least one previous diagnostic test without receiving a clearer diagnosis. In the absence of a diagnosis, the clinician is forced to treat the patient empirically, which

makes prognosis estimation more difficult. In this case, more than half of patients (54%) were able to receive a clear diagnosis. Prognostic and/or therapeutic information was offered for patients for whom the diagnosis remained difficult.

It can be difficult to convert a list of genetic mutations—even with biomarker annotations—into clinically actionable information, thus for every sample, a qualitative analysis is regularly carried out and shared with the veterinarian. This procedure is comparable to that used in pathology reports when a clinical history and microscopic findings are combined by a qualified specialist. A brief synopsis and an interpretation of the mutations for clinical use are included in this investigation. A brief synopsis and an interpretation of the mutations for clinical use are included in this investigation. Because different mutations have varying degrees of diagnostic specificity, a qualitative analysis is especially relevant for patients with ambiguous diagnoses. This means that to support a diagnosis of cancer and/or a particular cancer type, it is necessary to evaluate an aggregate of mutations (i.e., the presence, absence, and combination of various mutations) within the sample's mutational milieu considering the tumour location, pathology reports, and clinical history. Consequently, the study particularly examined each case's qualitative analysis to see whether diagnostic clarity was attained. This research is particularly useful for veterinarians as genomics is used more and more in veterinary treatment, regardless of the diagnosis of cancer. [54]

12. Conclusion

The research aimed to target the development of canine tumour research and its diagnostic, therapeutic and prognostic factors using more targeted methods such as Next-Generation Sequencing. Targeted genome sequencing has been progressing and increasing its importance in the veterinary field due to the increase in fatalities linked to somatic mutations of cancer.

Exact genomic sequencing is used to detect mutations such as single nucleotide substitution, insertions, deletions, or duplications. Detection of unusual mutation modifications is increased by using Next-Generation sequencing as this method is capable of testing multiple genes at once. Additionally, it can be used for spotting innate cancer disorders from very early on as it is the most direct way to test hereditary tumours.

Furthermore, it is not only used for diagnostic purposes but also for therapeutic parameters, which can determine the patient's outcome. Formerly tumours were identified by their morphological appearance, however, it is possible to identify cancer by genetic variations with the help of Next-Generation Sequencing. In addition to that, one can target these genetic alterations with specific medication. Not only is it possible to use appropriate therapy, but also confirm if a drug is drug is ineffective, or even if the cancer has developed resistant mutations to the therapy. All in all, veterinarians have the option to distinguish if a drug is more adverse than beneficial for the patient as the drug will target the mutation of the cancer.

On the other hand, it might be challenging to comprehend the information related to canine tumours as the database is not as developed as in human oncology. The development regarding the detection, characterisation, and management in canine oncology is still growing and therefore only a few scientific advances have been progressing in the veterinary field of oncology. Nonetheless, since the publication of the canine genome in 2005, the progress has been increasing making it more efficient and comprehensible to understand the resistance, pathophysiology, breed disposition and biomarker identification.

Unfortunately, elderly dogs develop cancer which remains for the rest of their lifetime, and if there is microscopic or macroscopic spreading, treatment can no longer assist the animal, which poses a significant chance of casualty. It is of great importance that owners

are aware of the genetic predispositions which allow early testing for dogs and therefore prevent the risks of early somatic changes which not only cause mutation and the development of tumour formation but also the invasion of tissues and lymphatic pathways.

Overall Next-Generation Sequencing is a useful tool in canine tumour research for early diagnosis, as new prognostic tests and cancer therapy by using targeted sequencing. It benefits the outcome of an even more accurate diagnosis which leads to a more exact prognosis. In addition to that, Next-Generation Sequencing allows patient-specific tailored therapy by targeting the mutation of individualised and unique tumours. With target identification of cancer, more accuracy and efficiency can be approached not only for better treatment options but also to provide precise a diagnosis which will impact the course and the outcome of the patient. With its extremely high throughput, scalability and speed, Next-Generation Sequencing is a method that can sequence full genomes of multiple patients in a short time, making it possible to reduce the number of tests to find the causative mutation.

13. References

1. Applied Biological Materials INC. (2015). Next Generation Sequencing. https://info.abmgood.com/next-generation-sequencing-ngs-introduction. Accessed 31/10/2023.
2. Sanger F, Nicklen S, Coulson AR (1997) DNA sequencing with chain-termination inhibitors. 74:5363-5467 . Accessed 24/09/2023.
3. Barba, M., Czosnek, H., & Hadidi, A. (2013). Historical Perspective, Development and Applications of Next-Generation Sequencing in Plant Virology. *Viruses*, *6*(1), 106-136. https://doi.org/10.3390/v6010106. Accessed 24/09/2023.
4. ClevaLab (2023) Next Generation Sequencing – A Step-By-Step Guide to DNA Sequencing. In: YouTube. https://www.youtube.com/watch?v=WKAUtJQ69n8. Accessed 24/09/2023.
5. Sambrook J, Russell DW. (2006) Fragmentation of DNA by nebulization. CSH Protoc. 2006 Sep 1;2006(4):pdb.prot4539. doi: 10.1101/pdb.prot4539. PMID: 22485920. Accessed 15/10/2023.
6. Sambrook J, Russell DW. (2006) Fragmentation of DNA by sonication. *CSH Protoc*. 2006;2006(4):pdb.prot4538. Published 2006 Sep 1. doi:10.1101/pdb.prot4538. Accessed 15/10/2023.
7. Lynn Apone, Ph.D., Eileen Dimalanta, PhD, Fiona Stewart, PhD, New England Biolabs Inc. (2017), Improving Enzymatic DNA Fragmentation for Next Generation Sequencing Library Construction chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.neb.com/en/-/media/nebus/files/feature-articles/nebexpressions_feature_nebnext_fs_issueiii_2017.pdf?rev=d8e4239e932a4a5bb6b9b5f833f31143&hash=056ADED33DAA63D58F36CDB42D508ACA. Accessed 15/10/2023.
8. Biosite (2020) https://nordicbiosite.com/blog/size-selection-for-ngs. Accessed 24/10/2023.
9. Andrew Gane (2028), Size selection brings better data to NGS workflows. https://www.cytivalifesciences.com/en/us/news-center/better-data-via-size-selection-10001. Accessed 24/10/2023.
10. Ambardar, S., Gupta, R., Trakroo, D., Lal, R., & Vakhlu, J. (2016). High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology*, *56*(4), 394-404. https://doi.org/10.1007/s12088-016-0606-4. Accessed 25/10/2023.
11. Illumina (2023). Understanding the NGS workflow. https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html#:~:text=During%20the%20sequencing%20step%20of,copies%20of%20single%2Dstranded%20DNA. Accessed 25/10/2023.
12. Illumina (2016) Overview of Illumina Sequencing by Synthesis Workflow. In: https://www.youtube.com/watch?v=fCd6B5HRaZ8. Accessed 25/10/2023.
13. Quick Biochemistry Basics (2023). Emulsion PCR https://www.youtube.com/watch?v=qKouzbp1RWI. Accessed 25/10/2023.
14. Bruijns, B., Hoekema, T., Oomens, L., Tiggelaar, R., & Gardeniers, H. (2022). Performance of Spectrophotometric and Fluorometric DNA Quantification Methods. *Analytica*, *3*(3), 371-384. https://doi.org/10.3390/analytica3030025. Accessed 25/10/2023.

15. Nietsch, R., Haas, J., Lai, A., Oehler, D., Mester, S., Frese, K. S., Sedaghat-Hamedani, F., Kayvanpour, E., Keller, A., & Meder, B. (2016). The Role of Quality Control in Targeted Next-generation Sequencing Library Preparation. *Genomics, Proteomics & Bioinformatics*, *14*(4), 200-206. https://doi.org/10.1016/j.gpb.2016.04.007. Accessed 25/10/2023.
16. Adriana Gallego, Ph.D. (2023). Quality Control of Library Preps. https://goldbio.com/articles/article/Quality-Control-Library-Preps#:~:text=Quality%20control%20techniques%20use%20qPCR,possible%20contaminants%20or%20are%20degraded. Accessed 26/10/2023.
17. Illumina dye sequencing. (2023). In *Wikipedia*. https://en.wikipedia.org/wiki/Illumina_dye_sequencing. Accessed 26/10/2023.
18. Applied Biological Materials (2015). Next Generation Sequencing (NGS) – An Introduction. https://www.youtube.com/watch?v=jFCD8Q6qSTM. Accessed 26/10/2023.
19. Illumina. (2016). Overview of Illumina Sequencing by Synthesis Workflow. https://www.youtube.com/watch?v=fCd6B5HRaZ8. Accessed 27/10/2023.
20. Micheal Greenwood (2018). What is Pyrosequencing? https://www.news-medical.net/life-sciences/What-is-Pyrosequencing.aspx. Accessed 27/10/2023.
21. Sequencing by ligation. (2023). In *Wikipedia*. https://en.wikipedia.org/wiki/Sequencing_by_ligation. Accessed 28/10/2023.
22. Applied Biological Materials (2015). Next Generation Sequencing (NGS) – An Introduction. https://www.youtube.com/watch?v=jFCD8Q6qSTM. Accessed 28/10/2023.
23. Applied Biological Materials (2015). Next Generation Sequencing. https://info.abmgood.com/next-generation-sequencing-ngs-introduction. Accessed 28/10/2023.
24. Ion semiconductor sequencing. (2023). In *Wikipedia*. https://en.wikipedia.org/wiki/Ion_semiconductor_sequencing. Accessed 28/10/2023.
25. Christopher Nobles. Ph.D. (2018) Filtering Sequences. https://nuckit.readthedocs.io/en/latest/pages/filt.html. Accessed 28/10/2023.
26. Altuna Akalin (2020). Computational Genomics with R. https://compgenomr.github.io/book/filtering-and-trimming-reads.html. Accessed 28/10/2023.
27. Samuel Daniel Lup. Carla Navarro-Quiles. Jose Luis Micol (2023). Versatile mapping-by-sequencing with Easymap v.2. https://www.frontiersin.org/articles/10.3389/fpls.2023.1042913/full#:~:text=Mapping%2Dby%2Dsequencing%20combines%20Next,for%20mapping%2Dby%2Dsequencing. Accessed 20/10/2023.
28. Uddin, I., Woolston, A., Peacock, T., Joshi, K., Ismail, M., Ronel, T., Husovsky, C., & Chain, B. (2018). Quantitative analysis of the T cell receptor repertoire. *Methods in Enzymology*, *629*, 465-492. https://doi.org/10.1016/bs.mie.2019.05.054. Accessed 28/10/2023.
29. ClevaLab (2023) Next Generation Sequencing – A Step-By-Step Guide to DNA Sequencing. In: YouTube. https://www.youtube.com/watch?v=WKAUtJQ69n8. Accessed 28/10/2023.
30. Welch, J. S., Westervelt, P., Ding, L., Larson, D. E., Klco, J. M., Kulkarni, S., Wallis, J., Chen, K., Payton, J. E., Fulton, R. S., Veizer, J., Schmidt, H., Vickery, T. L., Heath, S., Watson, M. A.,

Tomasson, M. H., Link, D. C., Graubert, T. A., DiPersio, J. F., . . . Wilson, R. K. (2011). Use of whole genome sequencing to diagnose a cryptic fusion oncogene. *JAMA: The Journal of the American Medical Association*, *305*(15), 1577. https://doi.org/10.1001/jama.2011.497- Accessed 30/10/2023

31. Marchetti A, Del Grammastro M, Filice G, et al. Complex mutations & subpopulations of deletions at exon 19 of EGFR in NSCLC revealed by next-generation sequencing: potential clinical implications. *PLoS One*. 2012;7(7): e42164. doi: 10.1371/journal.pone.0042164. Accessed 31/10/2023.
32. Tran B, Dancey JE, Kamel-Reid S, et al. Cancer genomics: technology, discovery, and translation [published correction appears in J Clin Oncol. 2012 Apr 1;30(10):1149]. *J Clin Oncol*. 2012;30(6):647-660. doi:10.1200/JCO.2011.39.2316. Accessed 30/10/2023.
33. Walsh T, Lee MK, Casadei S, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proc Natl Acad Sci U S A*. 2010;107(28):12629-12633. doi:10.1073/pnas.1007983107. Accessed 30/10/2023.
34. Walsh T, Casadei S, Coats KH, et al. Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *JAMA*. 2006;295(12):1379-1388. doi:10.1001/jama.295.12.1379. Accessed 30/10/2023.
35. Mardis ER. Applying next-generation sequencing to pancreatic cancer treatment. *Nat Rev Gastroenterol Hepatol*. 2012;9(8):477-486. doi:10.1038/nrgastro.2012.126. Accessed 31/10/2023.
36. Genome Atlas Network, T. C. (2012). Comprehensive molecular portraits of human breast tumors. *Nature*, *490*(7418), 61. https://doi.org/10.1038/nature11412. Accessed 31/10/2023.
37. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999-2009. doi:10.1056/NEJMoa021967. Accessed 31/10/2023.
38. Desmedt, C., Voet, T., Sotiriou, C., & Campbell, P. J. (2012). Next generation sequencing in breast cancer: First take-home messages. *Current Opinion in Oncology*, *24*(6), 597. https://doi.org/10.1097/CCO.0b013e328359554e. Accessed 31/10/2023.
39. Namboodiri, A. M., & Pandey, J. P. (2011). Differential inhibition of trastuzumab- and cetuximab-induced cytotoxicity of cancer cells by immunoglobulin G1 expressing different GM allotypes. *Clinical and Experimental Immunology*, *166*(3), 361-365. https://doi.org/10.1111/j.1365-2249.2011.04477.x. Accessed 31/10/2023.
40. Lee WL, Cheng MH, Chao HT, Wang PH. The role of selective estrogen receptor modulators on breast cancer: from tamoxifen to raloxifene. *Taiwan J Obstet Gynecol*. 2008;47(1):24-31. doi:10.1016/S1028-4559(08)60051-0. Accessed 31/10/2023.
41. Honeyman JN, Simon EP, Robine N, et al. Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. *Science*. 2014;343(6174):1010-1014. doi:10.1126/science.1249484. Accessed 01/11/2023.

42. Nathanson DA, Gini B, Mottahedeh J, et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science*. 2014;343(6166):72-76. doi:10.1126/science.1241328. Accessed 01/11/2023.
43. Nathanson DA, Gini B, Mottahedeh J, et al. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science*. 2014;343(6166):72-76. doi:10.1126/science.1241328. Accessed 01/11/2023.
44. Pinho SS, Carvalho S, Cabral J, Reis CA, Gärtner F. Canine tumors: a spontaneous animal model of human carcinogenesis. *Transl Res*. 2012;159(3):165-172. doi:10.1016/j.trsl.2011.11.005. Accessed 01/11/2023.
45. Davis BW, Ostrander EA. Domestic dogs and cancer research: a breed-based genomics approach. *ILAR J*. 2014;55(1):59-68. doi:10.1093/ilar/ilu017. Accessed 01/11/2023.
46. Schiffman JD, Breen M. 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):20140231. doi:10.1098/rstb.2014.0231. Accessed 01/11/2023.
47. Fiala, C., & Diamandis, E. P. (2018). Utility of circulating tumor DNA in cancer diagnostics with emphasis on early detection. *BMC Medicine*, *16*. https://doi.org/10.1186/s12916-018-1157-9. Accessed 01/11/2023.
48. Del Monte U. Does the cell number 10(9) still really fit one gram of tumor tissue?. *Cell Cycle*. 2009;8(3):505-506. doi:10.4161/cc.8.3.7608. Accessed 01/11/2023.
49. Rivera P, Melin M, Biagi T, et al. Mammary tumor development in dogs is associated with BRCA1 and BRCA2. *Cancer Res*. 2009;69(22):8770-8774. doi: 10.1158/0008-5472.CAN-09-1725. Accessed 01/11/2023.
50. Forbes C, Fayter D, de Kock S, Quek RG. A systematic review of international guidelines and recommendations for the genetic screening, diagnosis, genetic counseling, and treatment of *BRCA*-mutated breast cancer. *Cancer Manag Res*. 2019; 11:2321-2337. Published 2019 Mar 22. doi:10.2147/CMAR.S189627. Accessed 01/11/2023.
51. Milovancev M, Tuohy JL, Townsend KL, Irvin VL. Influence of surgical margin completeness on risk of local tumour recurrence in canine cutaneous and subcutaneous soft tissue sarcoma: A systematic review and meta-analysis. *Vet Comp Oncol*. 2019;17(3):354-364. doi:10.1111/vco.12479. Accessed 01/11/2023.
52. Mochizuki H, Breen M. Comparative Aspects of BRAF Mutations in Canine Cancers. *Vet Sci*. 2015;2(3):231-245. Published 2015 Aug 24. doi:10.3390/vetsci2030231. Accessed 01/11/2023.
53. Sclafani F, Gullo G, Sheahan K, Crown J. BRAF mutations in melanoma and colorectal cancer: a single oncogenic mutation with different tumour phenotypes and clinical implications. *Crit Rev Oncol Hematol*. 2013;87(1):55-68. doi: 10.1016/j.critrevonc.2012.11.003. Accessed 01/11/2023.
54. Chon, E., Wang, G., Whitley, D., Sakthikumar, S., Warrier, M., Wong, S., Duran, N., Adkins, J., Boateng, M., Zhu, Z., Facista, S., Haworth, D., & Hendricks, W. (2023). Genomic tumor analysis provides clinical guidance for the management of diagnostically challenging cancers in dogs. *Journal of the American Veterinary Medical Association*, *261*(5), 668-677.

https://avmajournals.avma.org/view/journals/javma/261/5/javma.22.11.0489.xml. Accessed 04/11/2023.

Diagrams

1. **Library Preparation.** https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-022-01565-1. Accessed 29/10/2023.
2. **Bridge Amplification.** https://atdbio.com/nucleic-acids-book/Next-generation-sequencing. Accessed 28/10/2023.
3. **Emulsion PCR.** https://atdbio.com/nucleic-acids-book/Next-generation-sequencing#figure-emulsion-pcr. Accessed 28/10/2023.
4. **Pyrosequencing.** https://microbenotes.com/next-generation-sequencing-ngs/. Accessed 28/10/2023.
5. **Sequencing by Ligation.** https://atdbio.com/nucleic-acids-book/Next-generation-sequencing#figure-emulsion-pcr. Accessed 29/10/2023.
6. **Ion-Semiconductor Sequencing.** https://apollo-institute.org/ion-torrent-sequencing/. Accessed 29/10/2023.
7. **Molecular Subtypes of Breast Cancer.** https://www.mdpi.com/2072-6694/13/18/4685. Accessed 31/10/2023.
8. **Food and Drug Administration-approved drugs.** https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-015-0203-x/tables/2. Accessed 01/11/2023.
9. **Germline and Somatic Mutations.** https://www.cell.com/trends/molecular-medicine/fulltext/S1471-4914%2823%2900072-2. Accessed 01/11/2023.

I hereby confirm that I am familiar with the content of the thesis entitled

NEXT-GENERATION SEQUENCING AS A USEFUL TOOL IN CANINE TUMOUR RESEARCH

written by Nora Krafft

(student name) which I deem suitable for submission and defence.

Date: Budapest, 03 day 11 month 2023 year

Peter Vajdovich, Head of the department

..... Supervisor name and signature

University of Veterinary Medicine Budapest, Department of Clinical pathology and Oncology

..... Department

UNIVERSITY OF VETERINARY MEDICINE, BUDAPEST

founded in 1787, EU-accredited since 1995

INTERNATIONAL STUDY PROGRAMS

secretary, student@univet.hu

Thesis progress report for veterinary students

Name of student: Nora Krafft

Neptun code of the student: CNZ2RH

Name and title of the supervisor: Dr. Vajdovich Peter

Department: Department of Clinical Pathology and Oncology

Thesis title: Next Generation Sequencing as a useful tool in Canine Tumor Research

Consultation – 1st semester

	Timing			Topic / Remarks of the supervisor	Signature of the supervisor
	year	month	day		
1.	2022	09	04	Discussion about the Topic and Title	
2.	2023	02	22	General Information, Thesis Structure	
3.	2023	03	17	Clarification about Introduction and History about NGS	
4.					
5.					

Grade achieved at the end of the first semester: ……

Consultation – 2nd semester

	Timing			Topic / Remarks of the supervisor	Signature of the supervisor
	year	month	day		
1.	2023	05	20	Searching for articles	
2.	2023	09	20	Summarizing articles	
3.	2023	10	25	Discussion about thesis text writing	
4.					

5.					

Grade achieved at the end of the second semester: ……5……..

The thesis meets the requirements of the Study and Examination Rules of the University and the Guide to Thesis Writing.

I accept the thesis and found suitable to defence,

……
signature of the supervisor

ÁLLATORVOSTUDOMÁNYI EGYETEM Kórélettani és Onkológiai Tanszék 1787 BUDAPEST

Signature of the student: ……

Signature of the secretary of the department: ……

Date of handing the thesis in…17./November/2023

ÁLLATORVOSTUDOMÁNYI EGYETEM Kórélettani és Onkológiai Tanszék 1787 BUDAPEST