U.PORTO

MESTRADO EM ONCOLOGIA

ESPECIALIZAÇÃO EM ONCOLOGIA LABORATORIAL

Liquid Biopsies: Effects of Long-Time Storage

Guilherme Silva

M

2022

U.PORTO

Guilherme Silva

Liquid Biopsies: Effects of Long-Time Storage

M.ICBAS 2022

Liquid Biopsies: Effects of Long-Time Storage

Guilherme Silva

Guilherme Figueiredo Machado Silva

Liquid Biopsies: Effects of Long-Time Storage

Dissertação de Candidatura ao grau de **Mestre em Oncologia** –
Especialização em Oncologia Laboratorial submetida ao Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto

Orientadora: **Professora Doutora Carmen de Lurdes Fonseca Jerónimo**
Professora Catedrática Convidada
Departamento de Patologia e Imunologia Molecular
Instituto de Ciências Biomédicas Abel Salazar - Universidade do Porto
Investigadora Auxiliar, Coordenadora do Grupo de Epigenética e Biologia do Cancro,
Diretora do Centro de Investigação
Centro de Investigação
Instituto Português de Oncologia do Porto Francisco Gentil, E.P.E

Coorientadora: **Bianca Troncarelli Flores, PhD**
Investigadora Convidada do Grupo de Epigenética e Biologia do Cancro
Centro de Investigação
Instituto Português de Oncologia do Porto Francisco Gentil, E.P.E
Investigadora Auxiliar do Departamento de Genética Clínica
University Hospital of Southern Denmark - Vejle

This study was funded by a grant of the Research Centre of Portuguese Oncology Institute of Porto (Grant PI-27-CI-IPOP-27-2016)

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer à Professora Doutora Carmen Jerónimo por me ter proposto este tema, por ter aceite orientar e acompanhar-me neste percurso, por toda a sua paciência e, sobretudo, por me ter dado a honra de integrar o seu grupo de investigação.

Também tenho muito a agradecer à Bianca por ter aceite o convite de ser minha co-orientadora, tendo sido sempre incansável no auxílio que me prestou na realização desta dissertação, pela calma que sempre me transmitiu (mesmo nos momentos em que fui mais “difícil”), estando sempre lá para desligar o meu “complicómetro”.

Não posso também deixar de endereçar o meu mais sincero agradecimento à Carina, pela sua amizade, e por todo o seu auxílio na vertente prática desta dissertação, proporcionando-me experiência no laboratório, nomeadamente no que respeita ao processamento de amostras e à catalogação destas no biobanco.

Os membros do Centro de Investigação 3 também foram cruciais na minha dissertação, sendo de realçar o ambiente agradável e acolhedor que sempre proporcionaram.

Por último, mas não menos importante, tenho de realçar o impacto fundamental da minha família por todo o incentivo e apoio emocional que me providenciaram, e por terem sempre acreditado em mim. Sem eles é provável que teria desistido a meio.

Muito obrigado a todos! É graças a vocês que o meu objetivo foi cumprido!

RESUMO

O impacto do cancro na sociedade atual não deve ser subestimado. De acordo com o Global Cancer Observatory (GLOBOCAN), em 2020 a nível mundial, o cancro da mama teve o maior número de casos de todos os tipos de cancro para ambos os sexos em conjunto e também nas mulheres, registando-se 2.261.419 casos, para além de ser o quinto cancro mais mortífero em ambos os sexos e o mais mortífero nas mulheres, registando-se 684.996 mortes. Os testes de rastreio do cancro da mama podem levar à deteção precoce do cancro da mama, numa fase mais fácil de controlar, tratar e curar, o que permitirá resultados melhorados e um melhor prognóstico. Pese embora os amplos benefícios que o rastreio do cancro da mama acarreta, também pode trazer vários problemas e desafios, incluindo resultados falsos negativos ou positivos, sobrediagnóstico e cancros resultantes de radiação. Por esse motivo, cfDNA e CTCs têm sido explorados extensivamente para a identificação de genes que podem ser usados como biomarcadores para diagnósticos mais precoces e seleção de terapias em avanço, melhorando, desta forma, o prognóstico e a qualidade de vida dos pacientes. O biobanco é um arquivo de amostras biológicas que podem variar entre DNA/RNA, amostras de fluídos corporais (ex: sangue e urina) e amostras de tecido em condições ideais de preservação, para que possam ser mais tarde utilizados em investigação e em testes de diagnóstico. Devido à importância de estabelecer regras bem definidas para a criação e gestão dos biobancos para evitar o uso indevido de dados e uniformizar práticas nos biobancos, o objetivo primordial desta dissertação é quantificar e analisar a qualidade do DNA e RNA existente nas amostras de biópsias líquidas recolhidas de doentes de cancro da mama armazenadas no biobanco do IPO-Porto. Alguns dos benefícios de obter a avaliação da qualidade do material genético é a possibilidade de comparar diretamente amostras (por exemplo, antes e depois do envio, comparar a integridade do mesmo tecido entre os diferentes laboratórios, etc.) e assegurar a reprodutibilidade e confiança das experiências, também fazendo parte da rotina de várias experiências, tais como Next Generation Sequencing (NGS), microarray, e muitas outras. Por esse motivo, é de grande importância garantir a manutenção da integridade do material genético no biobanco e estabelecer controlo de qualidade internos, fulcrais para projetos de investigação.

ABSTRACT

The impact of cancer in current society must not be underestimated. According to the Global Cancer Observatory (GLOBOCAN), in 2020 worldwide, breast cancer had the highest number of cases of all cancers for both sexes combined as well as in women, with 2.261.419 of them, and was the fifth most deadly cancer for both sexes combined and the deadliest for women, with 684.996 deaths. Breast cancer screening tests can lead to the detection of breast cancer at an earlier stage, during which the disease is easier to control, treat and cure, which then can lead to improved outcomes and better prognosis. Even despite breast cancer screening having many benefits, it can also bring many problems and challenges, including false negative or positive results, over-diagnosis and radiation induced cancer. Because of that, cfDNA and CTCs have been extensively explored for identification of genes that can be used as biomarkers for early diagnosis and therapy selection in advance, improving the patient prognosis and quality of life. The biobank is an archive of biological samples that can range from DNA/RNA to bodily fluids (e.g. urine and blood samples) and tissue samples, in ideal conditions for preservation purposes, so that they can later be used in investigative research and diagnostic tests, ensuring reproducibility of experiments, also being part of the workflow of several experiments, like Next Generation Sequencing (NGS), microarray, and many others. For this reason, it is of main importance to ensure the maintenance of the genetic material integrity on the biobank and establish internal quality controls, since most research projects start there.

TABLE OF CONTENTS

FIGURE INDEX

TABLES INDEX

LIST OF ABBREVIATIONS

WHO - World Health Organization
GLOBOCAN - Global Cancer Observatory
HRT - Hormone Replacement Therapy
UK - United Kingdom
RR - Relative Risk
MRI - Magnetic Resonance Imaging
IDC - Invasive Ductal Carcinoma
ILC - Invasive Lobular Carcinoma
TNM - Tumour-Node-Metastasis
OR - Oestrogen Receptors
PR - Progesterone Receptors
IPO - Instituto Português de Oncologia
nt - nucleotides
bp - base-pairs

INTRODUCTION

BREAST CANCER

o EPIDEMIOLOGY

The impact of cancer in current society must not be underestimated. According to the World Health Organization (WHO), around 9.6 million people died due to cancer in 2018, which was about 1 of each 6 deaths that occurred in that year, making it the second leading cause of death worldwide. Additionally, the economic impact is also increasing, with the total economic cost for cancer treatment in 2010 being estimated around US$ 1.16 trillion. Given that 30 to 50% of total cancer cases could be prevented not only through the avoidance of risk factors but also through already existing prevention strategies, it is of the utmost importance to maintain the investigation towards more effective cancer treatment and screening models [1]. Hence, to facilitate the researchers' access to high quality samples, it became necessary the creation of biobanks to comply with that need.

According to the Global Cancer Observatory (GLOBOCAN), in 2020 worldwide, breast cancer had the highest number of cases of all cancers for both sexes combined as well as in women, with 2.261.419 of them, and was the fifth most deadly cancer for both sexes combined and the deadliest for women, with 684.996 deaths [2]. In Portugal, breast cancer is the second most incident for both sexes and the most for women, with 7.041 cases, and the fifth most deadly for both sexes and the most for women, with 1.864 deaths [3]. These numbers show that breast cancer is one of the most important cancers to study, which is why it is essential to use breast cancer biopsy samples for investigation and clinical purposes, and for that reason we will use breast cancer as the cancer model for our work.

o RISK FACTORS

There are certain factors that increase the risk of getting breast cancer. While less than 10% of breast cancer are linked to inherited mutations, most breast cancers are tied to one or more environmental, reproductive and lifestyle risk factors, some of which could be modified [4]. Knowing these risk factors can lead to an earlier diagnosis and offer more treatment options for breast cancer.

These factors are the following:

- **Aging** – In addition to gender, the increase in age is one of the main risk factors for breast cancer, and it is highly tied to an increase in breast cancer incidence. In

America, studies revealed that in 2016, around 99.3% and 71.2% of all deaths caused by breast cancer occurred in women with at least 40 and 60 years of age, respectively [5].

- **Family History** – Approximately 25% of all breast cancer cases are correlated with family history. A person can be predisposed to have breast cancer if their mother or sister also had it. This inherited increase of breast cancer risk is due in part to certain genetic mutations to breast cancer related genes like *BRCA1* and *BRCA2* that are passed down from the parents [5].
- **Breast Cancer Related Genes** – Many oncogenes and antioncogenes related to breast cancer have been discovered, and certain mutations in these genes can enact a crucial role for carcinogenesis and cancer progression. Genes like breast cancer associated gene 1 and 2 (*BRCA 1* and *BRCA 2*) - anti-oncogenes for breast cancer under normal functioning - for example are presented in their mutated form in 20 to 25% of hereditary breast cancers and 5 to 10% of all breast cancers. The overexpression of human epidermal growth factor receptor 2 (HER2), an important oncogene for breast cancer, can be detected in 20% of primary breast cancers and it can indicate poor clinical outcomes. The overexpression of epidermal growth factor receptor (*EGFR* or *HER1* for humans) can be present in 30% of cases of inflammatory breast cancer, as well as appearing in more than half of triple-negative breast cancer cases. The oncogene *C-MYC* and its overexpression plays a crucial role in breast cancer initiation and progression and is frequently observed in the invasive stage of breast carcinomas [5].
- **Reproductive Factors** - It has been proven that several reproductive factors like early menarche, late menopause, low parity, and late age at first pregnancy can lead to an increase in breast cancer risk. Several studies have shown that for every year of delay in menopause there is an increase of breast cancer risk of about 3 %, and for every year of delay in menarche or each additional birth it leads to a decrease in breast cancer risk by 5 or 10%, respectively [5].
- **Oestrogen** – Oestrogen is a key risk factor for breast cancer, whether the oestrogen's source is endogenous (naturally produced by the ovary in pre-menopausal woman) or exogenous (through the ingestion of oral contraceptives or partaking in hormone replacement therapy (HRT)). For example, a study in the UK presented a relative risk (RR) of 1.66 between current users of HRT and people who never had it [5].
- **Lifestyle** – Many lifestyle habits like drinking excessive amounts of alcohol and ingesting too much food with high amounts of fat can increase the risk of breast cancer. The consumption of alcohol can lead to an increased level of oestrogen-

related hormones in the blood, which triggers the oestrogen receptor pathways. There is also growing evidence that smoking, especially at an early age, leads to an elevated risk of breast cancer occurrence, even if the relationship between smoking and breast cancer risk is still controversial [5].

- **SCREENING AND DIAGNOSIS**

Breast cancer screening tests can lead to the detection of breast cancer at an earlier stage, during which the disease is easier to control, treat and cure, which then can lead to improved outcomes and better prognosis. It also leads to a decrease of breast cancer cases detected at later stages. For example, the implementation of widespread screening practices like mammography to all women aged 40 and older in the United States in the 1980's and 90's has resulted in a large increase of early-stage breast cancer diagnoses, a noticeable decrease of late-stage breast cancer diagnoses and a decrease in breast cancer mortality [6].

According to the European Commission Initiative on Breast Cancer, it is recommended for women aged between 50 and 69 to undergo mammography screening once every one or two years given that this age group is the one that brings the biggest benefits from screening in terms of mortality reduction. The age groups of 40-49 and 70-74 can also be recommended for regular screening, even if its benefits are not as evident as the 50-69 age group. In the case of women with familial history of breast cancer, whether they have or not *BRCA* genetic mutations, it is recommended an annual screening with magnetic resonance imaging (MRI) on the breast, alongside mammography [7].

Although breast cancer screening has many benefits, it can also bring many problems and detriments that health professionals need to take caution to and deal with. Examples of those problems include false negative or positive results, overdiagnosis and radiation induced cancer. A false negative result usually leads to breast cancer staying undetected, allowing it to grow without any intervention to stop that growth, potentially ending up with the breast cancer being diagnosed in its late stage. False positives, on the other hand, can lead to the unnecessary use of resources due to further testing and even breast biopsies and everything else that entails, like noticeable distress and anxiety from the patient. Another issue is the risk of overdiagnosis, which consists of diagnosis of breast cancer as being more severe or aggressive, like claiming the tumour to be invasive when it is not, or at a later stage. This can lead to a more aggressive treatment protocol, that could potentially cause disfiguring consequences to the patient, like scarring, cardiac toxicity, and lymphedema. Statistically, around 1 in 8 women diagnosed with screening mammograms

from the ages 50 to 75 years are over-diagnosed. Finally, cancer that could be induced by radiation from screening exams, while very rare (around 0,4 – 1,2 per 10000 women screened during their lifetimes), is still something to be considered [6].

The diagnosis of breast cancer is usually done through clinical examination alongside imaging and then confirmed by a pathological assessment. Clinical examination consists of bimanual palpation of the breasts and regional lymph nodes, in conjunction with distant metastases assessment. Bilateral mammography and ultrasound of the breast and regional nodes can be part of the imaging process. Usually, the pathological assessment is carried out through a core needle biopsy which is guided by ultrasound [7]. A detailed personal medical history (age at diagnosis of breast cancer, previous breast biopsies, use of radiation for treatment of other cancers, etc.) and family history pertaining to breast and ovarian cancer occurring in first degree relatives, as well as other cancers is also important for diagnosis [8].

- **HISTOLOGICAL SUBTYPES**

Most of all breast cancers (more than 95%) are adenocarcinomas [9]. In addition, there are various subtypes of breast cancer, both non-invasive and invasive type (figure 1), which denote the diverse sphere of characteristics, whether genetically or morphologically, of breast cancer. Examples of non-invasive breast cancer include ductal carcinoma *in situ* (the most common subtype of non-invasive breast cancer, where the tumour limits itself to the breast duct) and lobular carcinoma *in situ* (where the tumour cells grow and develop in the breast lobules, and it does not usually expand beyond the lobules). While non-invasive breast cancer may never proliferate beyond their affected regional tissue, there is always a possibility that they could progress into an invasive type, so it is still recommended to go through some type of treatment to control that tumour, even if preferably a breast-preserving one. On the other hand, invasive breast cancer occurs when the tumour cells start to spill out from the ducts or lobules and spread into the rest of the breast tissue in early stages, and into other organs of the body in late stages (metastatic breast cancer), usually the brain, bones, lungs and liver [8]. Examples of this include:

- **Invasive ductal carcinoma (IDC)** – Breast cancer that presents malignant spread in the breast ducts alongside stromal invasion. IDCs are a morphologically diverse group of tumours that differentiates themselves through cytoarchitectural features. While some tumours are distinctive enough in terms of structural and behavioural characteristics to be presented as special subtypes, most of them (around 75%) do not have sufficient features to be classified as a specific histological subtype, being

described as a “no special type” IDC. This generic type of IDC constitutes about 40% to 75% of all cases of invasive breast carcinoma, making this the most common type of breast cancer [9].

- **Invasive lobular carcinoma (ILC)** (tumour originates in the breast lobules) – Breast cancer that begins its growth and expansion in the breast lobules, invading into other breast tissues afterwards. ILC is the second most common type of breast cancer after IDC, being part of 5% to 15% of all breast cancer cases and usually affecting woman of an older age demographic compared to classic IDC [9].

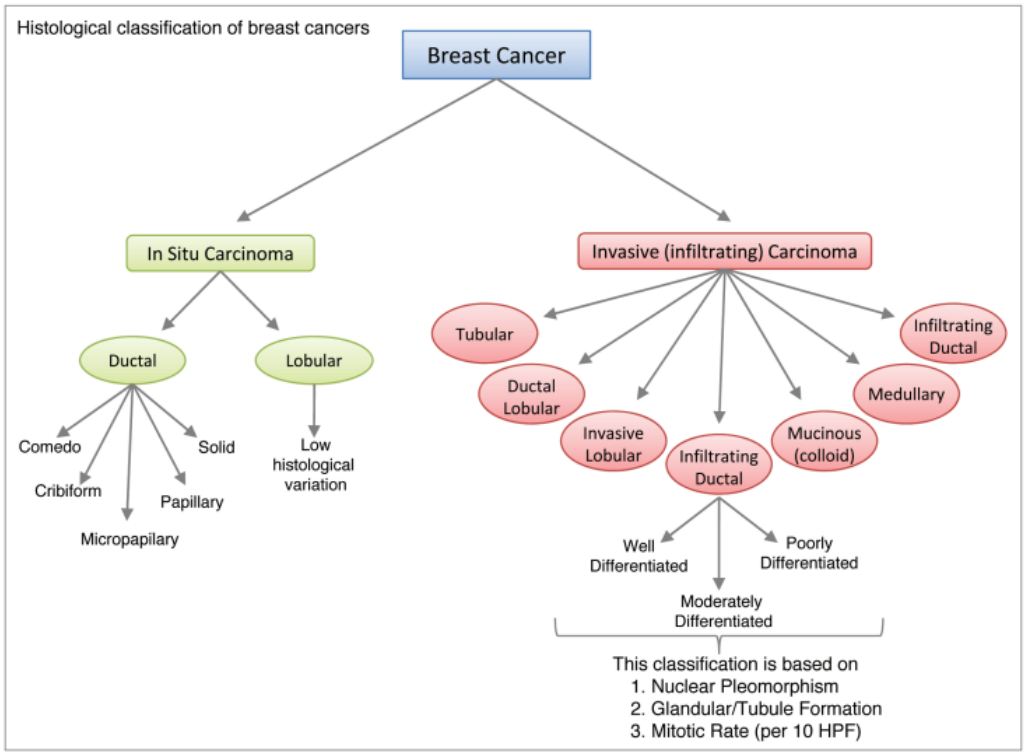

Figure 1 - Histological classification of breast cancer [10].

○ STAGING

The process of staging is necessary to determine the extent of the progression of the cancer, which helps to determine the prognosis and to establish the best possible treatment for each specific case. Currently, the most utilised and accepted system for breast cancer staging is the Tumour-Node-Metastasis (TNM) system, which allows for a quick, simple, and standardised way to communicate the staging information for all medical personnel. The TNM system works through the measuring of three parameters: the size of the primary tumour (T); the number of metastasized lymph nodes (N) and the presence or absence of distant metastasis (M) [11]. Depending on the result obtained from the

measurement of all three parameters, it can be assigned one out of five stages to the tumour that can go from 0 to IV. In breast cancer, each one of these categories describe the progress of the disease in the following way:

- **Stage 0** – Non-invasive carcinoma, the tumour stays restricted to its origin tissue, with no signs of invasion observed.
- **Stage I** – It englobes two categories (IA and IB). Stage IA describes a primary tumour no bigger than 2 cm with little to no metastasis found in lymph nodes, while stage IB describes a tumour with at least one lymph node metastasis bigger than 0.2 mm**.**
- **Stage II** – Also has two categories like stage I (IIA and IIB). Stage IIA describes a tumour with lymph node metastasis, axillary and/or sentinel, that can be smaller or larger than 2 cm but cannot be bigger than 5 cm. Stage IIB describes a tumour with no lymph node metastasis that is usually bigger than 5 cm.
- **Stage III** – It has three categories (IIIA, IIIB and IIIC). Stage IIIA describes a tumour with 4 to 9 axillary and/or sentinel lymph nodes metastasized. Stage IIIB describes a tumour of any size that causes swelling or ulcers on the breast skin (inflammatory breast cancer) while also having up to 9 lymph nodes metastasized. Finally, stage IIIC describes a tumour with at least 10 lymph nodes metastasized, including nodes above and below the clavicle.
- **Stage IV** – The cancer has spread to other organs, forming distant metastasis mainly in the lungs, bones, brain, and liver [8].

o PROGNOSIS FACTORS AND PREDICTIVE BIOMARKERS

The most important biomarkers that can be searched through liquid biopsies for breast cancer patients are oestrogen receptors (OR), progesterone receptors (PR) and HER2 gene levels. OR is not only a great prognostic and predictive biomarker for therapy response but can be also used mainly as a predictive biomarker for endocrine therapy. Given that oestrogen can promote cancer cell growth through its association with oncogenes like MYC and cyclin D, and that OR allows oestrogen to apply its effects, it is presumed that OR levels can predict the response of breast cancer to antiestrogenic therapy. PR can be induced by oestrogen, and its presence can be used as a marker for a functional OR. Moreover, the interaction between progesterone and PR could induce a change in the OR chromatin binding site, with that change usually promoting genes involved in cell proliferation and silencing genes associated with differentiation, apoptosis, and cell

cycle arrest. High levels of HER2 alongside activation of signalling pathways like PI3K/AKT, MAPK and cell membrane deformity can promote metastasis, invasion, and proliferation of cancer cells. The analysis of all the three biomarkers in conjunction is highly recommended by most health experts for newly diagnosed breast cancer patients [12].

cfDNA and CTCs have been extensively used for identification of genes that can be used as biomarkers. Maltoni et al. analysed the role in prognosis of cfDNA quantity and integrity of frequently pathogenic genes in breast cancer, like HER2, MYC and PI3KCA. From 2002 to 2010, samples from 79 breast cancer patients before surgery and 10 healthy donors were collected. The researchers observed that breast cancer patients showed significantly higher amounts of cfDNA when compared to healthy controls who showed integrity and lower apoptosis events using real-time PCR [13].

- **TREATMENT**

The purpose of treatment can diverge depending on if you are trying to treat nonmetastatic or metastatic breast cancer. In the case of the former, the treatment is mostly curative in nature, in which it attempts to eliminate the primary tumour from the breast and regional lymph nodes and prevent the cancer from proliferating and metastasising in other organs. For the latter, the treatment used has as its objective to prolong patients' life, as well as to improve life quality and to alleviate symptoms caused by the disease, being palliative in nature [14].

Treatments for nonmetastatic breast cancer may require surgery to remove axillary lymph nodes or even for sampling purposes (biopsies), alongside postoperative radiation. Surgery can be breast-conserving, like lumpectomy (which removes only breast tissue that contains the tumour) or remove the entire breast (like mastectomy). The choice between one of the two relies on the extent of the cancer progression and the patient's preference, usually related to cosmetic reasons [8]. Therapy can be applied before surgery (neoadjuvant), after surgery (adjuvant), or even both, depending on what is adequate to apply. The targeted treatment protocol will also depend on the molecular subtype of the cancer, given that certain types of treatments will only prove to be efficient on the presence or absence of hormonal receptors (HR) or HER2 in at least a part of all breast cancer cells [14].

For breast cancer that are positive for HR but negative for HER2, the targeted therapy most used is endocrine therapy with the intent to neutralise tumour growth promoted by oestrogen. Endocrine therapy usually consists of daily intake of oral anti-

oestrogen medication, typically tamoxifen, for 5 years. This treatment has been proven to reduce by around 50% the recurrence rate of breast cancer in the first 5 years after diagnosis when compared with no endocrine therapy. Some patients may be treated with chemotherapy alongside endocrine therapy if considered adequate [14].

In relation to breast cancer of the HER2+ subtype, immunotherapy like the usage of monoclonal antibodies can be an effective targeted treatment against it. Trastuzumab is usually one of the main monoclonal antibodies used for breast cancer treatment due to it targeting the extracellular domain of HER2, which stimulates a response from the immune system that leads to the destruction of any cancer cell that possesses the antigen targeted by trastuzumab previously mentioned. Antibodies like pertuzumab and neratinib may also prove to be effective in breast cancer treatment, especially with high-risk HER2+ breast cancer. Chemotherapy can be used in conjunction with this therapy [14].

For triple-negative breast cancer (HR- and HER2-), chemotherapy is the only treatment (besides surgery and radiation therapy) that has proven to be efficient for this subtype, given that treatments like endocrine therapy and monoclonal antibody therapy are ineffective against it due to the absence of HR and HER2 in the cancer cells. A study concluded that a chemotherapy regimen consisting of a high dose of anthracycline applied in early breast cancer patients reduced 10-year breast cancer mortality by approximately a third, with most of that reduction occurring in the first 5 years after diagnosis [14].

BIOBANKS

The biobank are archives of biological samples that can range from DNA/RNA to blood and urine samples and even tissue samples, in ideal conditions for preservation purposes, so that they can later be used in investigative research and diagnostic tests. The biobank allows researchers to have an easy and standardised access to biological samples as well as all the clinical information connected to them, like the type of sample, its characteristics, and relevant sociodemographic features about the samples' donor [15]. The samples collected for the biobank go through a cataloguing process that allows for its quick identification when looking for adequate samples for research. Importantly, all this information is protected with different levels of passwords.

Given the importance of establishing well-defined rules for the creation and management of biobanks to avoid data misuse and uniformize practice in the biobanks, every country defines their own laws, although there are directives defined by the European Union that all member states must comply. In Portugal, the law that regulates that practice and establishes the ethical boundaries related to it is published in the *"Diário da República"*,

under the Law n.º 12/2005 of the January 26th, updated by the Law n.º 26/2016 of the August 22nd (transposing the Directive 2003/4/CE and 2003/98/CE of the European Parliament and of the Council of the European Union, of 28th January 2003 and 17th November 2003, respectively). Despite this law englobing the countless aspects surrounding the procedure related with the biobanks, in this dissertation we intend to refer only to the main points pertinent for this work. These include:

- The requirement of a written informed consent from the sample's donor, not only for the collection of the biological material but also for its storage in the biobank; given that the biological material is owned by its donor, the consent can be removed at any time by the sample's owner, in which case, the sample must be destroyed.
- The use of samples for medical or scientific purposes demands its authorization from the people involved or legally responsible for them; any sample stored in the biobank must not be identified, its access must be controlled, and the number of people authorised to do it limited, as well as the safety of the samples must be assured in order to respect the privacy and confidentiality of the donor, in accordance with the Data Protection Law;
- Any change in the sample's status or in the investigation project or the biobank must be reported to the material's owner; just like in the case of the biological material, any personal information obtained during the collection and the storage process is also owned by its donor, and its utilisation is only permitted for medical and scientific purposes; in addiction, the personal information is also under the same scrutiny as the biological material in respect to its confidentiality and security.

Biobanks can be generalised or specialised. Generalised biobanks collect samples of all types and from all groups of people, either with pathologies or healthy. The samples tend to come from completely anonymous sources allowing the use of big quantities of biological data [15]. On the other hand, specialised biobanks focus their collection on a specific type of biological material or even a certain group of patients. Examples of specialised biobanks can range from stem cells banks or biobanks that collect samples from patients with heart disease or even cancer (tumour banks) [15]. It is specifically about the latter one that this work is mostly focused.

A tumour bank is a more specific type of biobank that stores mainly biological material collected from cancer patients. The task of sample collection and cataloguing is usually left to the pathologist, which makes sure that the tumour bank functions properly [16].

Tumour banks' main objective is to facilitate and accelerate cancer research through making possible the access to an abundant quantity of samples preserved in its highest quality possible today. The tumour bank also allows for a bigger proximity between the researchers and the clinical team thanks to the clinical annotation cataloguing process previously described, as well as it provides jobs to various technicians and pathologists and develops researchers, while promoting high quality scientific projects. These projects can then lead to an easier discovery of new therapies and/or medicine [16].

IPO-Porto's Biobank was founded in September 2012, and it can be found in the Department of Pathology [17]. It stores all types of tumoral and non-malignant tissues, mostly solid and liquid biopsies collected from cancer patients, although the latter type of samples is the one most pertinent for our work.

○ LIQUID BIOPSIES

According to the National Cancer Institute (USA), liquid biopsies are samples of biological fluids that are collected to search for tumour biomarkers, which can be any molecule or cell produced by the tumour that is then released in biological liquids like blood. They can be used for early-stage cancer diagnosis, treatment planning and relapse detection [18]. Liquid biopsies not only allow for a non-invasive method of cancer stage assessment and treatment response, but its non-invasive property enables continuous and constant real-time follow ups to determine tumour progression and evolution, something that is unfeasible with tissue biopsies. They also allow for the detection of a bigger range of biomarkers than a tissue biopsy is usually able to trace, given that due to spatial tumour heterogeneity, different parts of the tumour could express different biomarkers, and a tissue biopsy on a single fragment of the tumour will only detect some of them. Liquid biopsies don't have that problem because all the biomarkers expressed by the tumour in its entirety are expelled into the bloodstream or urine, which will then be detected during liquid biopsy analysis [19].

All the biomarkers that can be found in liquid biopsies can be categorised in these different subgroups, namely circulating tumour cells, "tumour-educated platelets", exosomes and cell-free circulating nucleic acids (which includes circulating tumour DNA (ctDNA), mRNA, miRNA, and long non-coding RNA) [19]. From these, the biomarker subgroups that will be most relevant for our work, especially for the plasma biopsy analysis, will be ctDNA and RNA. ctDNA is the fraction of all cell-free circulating DNA that was directly released by the tumour. Therefore, the tumoral origin of this biomarker means that its

presence in the biopsy could indicate the existence of a tumour in an earlier stage than through histological analysis. Several studies also corroborate with that notion. It has been also noticed that there is a correlation between the quantity of ctDNA and tumour burden, since higher concentrations of this biomarker are detected in advanced cancers compared with localised ones. The quantitative measurement of ctDNA and its variation, especially after curative treatment, can be a strong predictor not only to the tumour's response to that treatment, but also to a potential cancer recurrence if applied in patients with localised tumours [19]. RNA, specifically cell-free RNA (cfRNA), consists of detached fragments of mRNA and miRNA that were released by the tumour into the blood. The abundance and composition of miRNA in specific have been shown to be altered in cancer patients when compared to healthy individuals, and they can also be used as a therapeutic target for cancer given the significance of their role in regulating the expression of the genes involved in angiogenesis during cancer metastasis [13]. MiRNAs can also be considered to have an important role in tumour growth and treatment resistance [19]. In the case of mRNAs, because of the role they have in intracellular protein translation, they can reflect the status of the intracellular process and could therefore be good biomarkers for cancer diagnosis and therapeutic monitoring [20]. They are also in higher concentrations in the blood than ctDNA, especially in early-stage cancer patients, which can be useful for diagnostic and preventive purposes [13]. cfRNAs might prove to be valuable in determining the gene expression profile of the tumour, which may not only indicate many somatic mutations at the DNA level, but also any epigenetic changes and other alterations in cellular pathways [13]. All the points mentioned beforehand make ctDNA and cfRNA the two most important groups of biomarkers to measure in our work.

PRESERVATION AND LIQUID BIOPSIES QUALITY AND STORAGE

Circulating tumour DNA (ctDNA) can be used to follow up genetic mutations, rearrangements, amplifications, and copy number variations in patients during the treatment. In healthy donors, ccfDNA usually is found in low concentration in plasma with a median of about 5 ng/ml but differs between individuals. In general, increased ccfDNA content in blood can be seen as an indicator of abnormally high cell death connected to different pathological conditions. In cancer patients, the concentration of ctDNA can vary considerably, depending on tumour size, stage, location, and other factors, with a proportion of ctDNA between 0.01 and 90%. Majority of reports in the subject describe that ctDNA is

more fragmented compared to ccfDNA from normal tissues. To correctly collect cfDNA, two consecutive centrifugation steps to generate plasma are applied in many if not most laboratories, a procedure initially cited by Chiu et al. [21].

After the plasma separation, it must be stored frozen at 20 °C or at 80 °C and repeated freeze-thaw cycles should be avoided. The chosen extraction method can have a profound impact on DNA yield and concentration, so it is crucial to choose an adequate isolation methodology. For genomic quantification, qPCR-based methods are usually preferred over spectrophotometry because impurities might interfere with the spectrophotometric measurements and their reliability usually declines in the lower DNA concentration range [21].

Quality control consists of specific tests defined by the quality assessment program to be performed to monitor procurement, processing, preservation and storage, specimen quality, and test accuracy. These tests may include performance evaluations, testing and controls used to determine the accuracy and reliability of the biobank's equipment, to maintain and guarantee that the supplies, reagents, equipment, and facilities are working properly. QA requires a systematic monitoring and evaluation of all aspects of the biobank's procedures; it covers the way in which the biobank is operated since the sample collection as well as the quality of the samples and data held [22].

AIMS

In order to uphold an adequately functional biobank, it will be necessary to conduct a rigorous quality control of the liquid biopsy samples that are stored inside them to make sure that all of the samples satisfy all the quality parameters necessary for its utilisation on investigation projects. A crucial step of this control, and the main purpose of our work, will be to test and determine the maximum preservation time in which the samples are able to be stored within the minimally acceptable quality, so that we will be able to determine which samples are still good enough to be utilised and which ones need to be disposed of due to its quality not being up to par with the quality parameters established. Beyond that, our project aims to assess the effect of liquid biopsies long term storage in nucleic acids quantity and integrity.

MATERIAL AND METHODS

SAMPLE PROCESSING

Before discussing sample selection, first we will need to address the sample processing established in IPOPorto's biobank. After we collect the liquid biopsies from cancer patients, these are then processed before being stored, separating them into urine cells (pellet) and cell-free urine, in the case of urine, and blood plasma and leukocytes, in the case of the blood samples. In blood processing, all samples must be centrifuged right after their collection (up to four hours) and then have the blood plasma (supernatant) removed into separate eppendorfs, so that they can be stored in the tumour bank. Then, the pellet must go through cycles of suspension in AKE and PBS and centrifugation (at 4°C) until only the leukocytes remain, in which the cells must be suspended in PBS one last time, placed in a new 1.5 mL tube, centrifuged and the supernatant discarded, storing the pellet in the biobank. In urine processing, the samples are centrifuged (at 4 °C) and the supernatant is poured into a new 4.5 mL cryopure tube and then stored in the biobank. The pellet is then suspended in PBS and centrifuged, with this process performed twice: once in the falcon tube and once in the 1.5 mL tube, with the supernatant being discarded each time. The urine cells pellet can then be stored in the biobank.

SELECTION CRITERIA

For this project, we established a database with the relevant information available about the patients that provided informed consent for biobanking (CES IPOPFG-EPE 019/08), whose liquid biopsies were used for this dissertation. The samples are randomised, and they are picked in accordance with the following exclusion and inclusion parameters. Those are:

1. The samples were selected from female patients that had invasive breast cancer from the time of collection of the liquid biopsies.
2. The samples were collected around the time of diagnosis, so follow-up samples were excluded.
3. Ten samples for each type of liquid biopsy were used in the study (blood plasma, urine pellets and urine supernatant) per year, from 2016 to 2020.
4. The patients included in the database provided all three types of liquid biopsy available in the biobank to facilitate data processing for the results.

RNA/DNA ISOLATION

After the establishment of the database, we used the chosen samples to extract DNA and RNA from them, by using magLEAD 12gC extraction device, and followed the manufacturer's instructions for the Mag DEA Dx SV 400 protocol in the case of DNA extraction and the Mag DEA Dx SV RNA protocol in the case of RNA extraction.

Following the extraction of the genetic material from the samples, we proceeded to measure the quality and concentration of the DNA and RNA fragment strands using the TapeStation software and various ScreenTape assay kits.

Figure 2 - MagLEAD 12gC extraction device (PSS Bio System Network, produced in more than ten countries, including Japan, the United States and Europe [23])

RNA/DNA QUALITY AND CONCENTRATION EVALUATION

The Tapestation 4200 is an instrument that quantifies and analyses nucleic acid strands by size using electrophoresis, a technique that separates molecules like DNA and RNA by applying an electric charge that forces the DNA and RNA fragments to move through a matrix. Smaller fragments move faster and advance further than bigger ones due to being easier for small fragments to transverse in the matrix. It detects fluorescently marked double-strand DNA, which can include genomic and cell-free DNA, and fluorescently marked RNA, whether it is eukaryotic or prokaryotic [24]. Besides offering sample loading, separation, and imaging, it also provides various DNA and RNA assays that are adequate for a vast size and concentration range for the purpose of analysing fragmented DNA/RNA, as well as PCR products and DNA libraries [25]. The 4200

TapeStation can evaluate up to 96 samples in one session, obtaining results within 1 to 2 minutes per sample, while only needing not more than 2 µL for each sample [26]. The visualisation of the results obtained from the TapeStation 4200 was performed by the TapeStation Analysis Software in a computer.

Figure 3 - TapeStation 4200 and laptop with TapeStation Analysis Software (Agilent Research Laboratories, California, USA)

For the TapeStation, the matrix in which the electrophoresis was carried out was the ScreenTape, to streamline and accelerate the process. The ScreenTape contains buffer chambers (contains optimised buffers for the effective separation of nucleic acid fragments), electrodes (creates an electric current that goes across the ScreenTape, making the need for any other electrophoresis equipment unnecessary) and gel (designed specifically to separate nucleic acids) [24].

Figure 4 - ScreenTape devices for DNA and RNA assays

We used three different types of ScreenTape assay for our DNA and RNA quality assessment: the High Sensitivity D1000 ScreenTape, the Genomic DNA ScreenTape and the RNA ScreenTape. The first is designed for DNA fragments within the 35-1000 bp size range [27] and was used for the plasma and urine supernatant DNA samples. The second

one was used to assess the integrity of genomic DNA samples, being more adequate for DNA fragments from 200 to around 60000 bp [28], was used for the urine pellet DNA samples. And the final one was developed for the integrity assessment of eukaryotic and prokaryotic total RNA [29] and it was used for the RNA samples.

In order to prepare a DNA assay with High Sensitivity D1000 Screentape, the reagents were at room temperature for 30 minutes, and then the mix vortexed before use. After that, the sample was prepared by mixing 2 µL of High Sensitivity D1000 Sample Buffer with 2 µL of DNA sample (or High Sensitivity D1000 Ladder, in the case of the ladder), then spined down and vortexed at 2000 rpm for 1 minute, spinning down again to position the sample at the bottom of the tube [27]. For Genomic DNA ScreenTape assays, the process was basically the same, except the preparation of the sample: 10 µL of Genomic DNA Sample Buffer were mixed with 1 µL of DNA sample, and for the ladder we used 10 µL of Buffer and 1 µL of Genomic DNA Ladder if we insert 1 or 2 ScreenTape devices, or 20 µL of Buffer and 2 µL of Ladder in the case of more than 2 ScreenTapes [25]. For the RNA ScreenTape assays, the process was the same as the former two, except that the preparation of the samples involved mixing 5 µL of RNA Sample Buffer with 1 µL of RNA sample (or RNA Ladder, in the case of the ladder) [29].

PARAMETERS EVALUATED ON THE TAPESTATION ON SAMPLES' QUALITY

Several parameters were used to evaluate the quality of the genetic material found in the samples, all of which are limited on the region of fragment sizes utilized to analyse those same fragments of DNA/RNA. The first one of them is the average size of the fragments found inside the region. The next one is the concentration of all genomic material with the fragment sizes found inside the region. And finally, the percentage of all genomic material of the sample that is in that fragment region. Medians of all average fragment sizes and concentrations of all samples of a given year and type were calculated. We decided to use medians instead of averages to offset any outliers found in the results.

WHY THAT SPECIFIC RANGES WERE CHOSEN TO EVALUATE IN EACH TYPE OF SAMPLE

In the RNA ScreenTape assay, for the plasma and supernatant samples the analysis range of 50 to 1000 nucleotides (nt) was chosen, while for the pellet samples the range was

extended from 50 to 10000 nt. For the DNA plasma and supernatant samples, given that we used High Sensitivity ScreenTape, the analysis range was set from 100 to 900 base-pairs (bp), while the pellet samples were assessed with Genomic DNA ScreenTape, being the region adjusted between 200 to 55000 bp. This was because the pellet samples have much more genetic material and therefore a higher amount of big DNA/RNA fragments. Additionally, the pellet samples' genetic material derives from leukocytes while plasma and supernatant samples consists of mainly cell-free DNA/RNA. So, the pellet samples are more preserved, needing a wider region than the other ones to reflect the increased concentration and fragment size in relation to other samples.

STATISTICAL ANALYSIS

GraphPad Prism 6.0 (GraphPad Software Inc., USA) was used to perform the statistical analysis of the results. Non-parametric Mann-Whitney U test was used to compare two groups. Additionally, when three or more groups were compared, the non-parametric Kruskal-Wallis's test was used, followed by Mann-Whitney U test for pairwise comparisons and Bonferroni´s correction, when applicable. The p-value was considered statistically significant at values lower than 0.05. Moreover, significance is shown versus the control group and the values are represented as the following: * - $p<0.05$, ** - $p<0.01$, *** - $p<0.001$ and **** - $p<0.0001$.

RESULTS

A total of 150 samples (50 blood plasma, 50 urine pellets and 50 urine supernatants), obtained from 50 breast cancer patients between 2016 and 2020 were used in this study. The total cohort of 50 patients represents most women with Invasive Carcinoma and a median age of 57.5 years, without Lymphovascular invasion, a great majority positive for Oestrogen and Progesterone hormones and low stage (Table 1).

Table 1 - Clinicopathological features of the patients that provided the liquid biopsies.

Clinicopathological features	Total Cohort
Patients (n)	50
Age median (range)	57.5 (34 – 84)
Histological type (%)	
Invasive Carcinoma, no special type (NST)	36 (72)
Invasive Lobular Carcinoma	4 (8)
Other invasive carcinoma subtypes	10 (20)
Lymphovascular invasion (%)	
No	32 (64)
Yes	18 (36)
Not determined	0
Grade (%)	
G1 & G2	20 (40)
G3	28 (56)
Not determined	2 (4)
Oestrogen Receptor Status (%)	
Positive	43 (86)
Negative	7 (14)
Progesterone Receptor Status (%)	
Positive	38 (76)
Negative	12 (24)
Primary tumour (T) (%)	
T1 & T2	44 (88)
T3 & T4	6 (12)
Regional lymph node (N) (%)	
N0	32 (64)
N+	18 (36)
Stage (%)	
I/II	43 (86)
III/IV	7 (14)

The results obtained after DNA and RNA quantity and quality assessment for the biobank samples were elucidative of their conservation state after several years of storage (Table 2-5).

Overall, DNA and RNA measurements denote the presence of two groups, which are distinguished by a clear difference in fragment size and concentration. One group is represented mostly by plasma and supernatant samples (lower fragment size and concentration), and another group in its majority composed by pellet samples (higher fragment size and concentration). Within these main groups, we could identify some particularities.

Within the same sample type, it is possible to identify some variation, and quantification range differed in certain years. These specific points can be explained by routine procedures on the processing of the samples. For example, the year 2018, where we can observe in a general way, a lower minimum concentration in all groups. Generally, in 2017 the minimum quantification was higher than in the other tested years.

Contrarily, no differences were apparent regarding the fragment size over the years, and thus in the genomic material quality.

Table 2 - Concentration and size ranges for each sample type and year in the small DNA fragments' group.

DNA (Small fragments)						
Years	Concentration Range (min-max)			Size Range (min-max)		
	Plasma (pg/µL)	Supernatant (pg/µL)	Pellet (ng/µL)	Plasma (bp)	Supernatant (bp)	Pellet (bp)
2016	12.6 – 45.4	7.75 – 24.3	0.712 – 2.87	46 – 51	45 – 53	589 – 996
2017	13.8 – 37.3	6.54 – 19.7	0.625 – 2.31	48 – 52	45 – 50	586 – 972
2018	8.66 – 29.7	4.88 – 19.2	0.629 – 2.54	47 – 52	45 – 53	605 – 940
2019	11.4 – 24.4	9.59 – 19.5	0.585 – 2.16	46 – 50	45 – 51	611 – 816
2020	11.5 – 26.2	6.77 – 17.1	0.737 – 1.55	46 – 51	45 – 50	573 – 814

Table 3 - Concentration and size ranges for each sample type and year in the large DNA fragments' group.

DNA (Large fragments)						
Years	Concentration Range (min-max)			Size Range (min-max)		
	Plasma (pg/µL)	Supernatant (pg/µL)	Pellet (ng/µL)	Plasma (bp)	Supernatant (bp)	Pellet (bp)
2016	0.271 – 5.64	0.83 – 6.65	0.32 – 7.77	337 – 427	278 – 433	11938 – 24430
2017	4.14 – 9.72	1.49 – 3.04	0.218 – 3.61	329 – 441	332 – 454	8921 – 25261
2018	3.08 – 10.1	1.19 – 17	0.184 – 40.8	365 – 454	274 – 432	9955 – 34085

2019	3.6 – 9.04	1.06 – 4.73	0.183 – 8.91	384 – 471	312 – 413	9696 – 32031
2020	1.68 – 10	1.28 – 27.1	0.248 – 4.11	366 – 447	294 – 454	13495 – 31098

Table 4 - Concentration and size ranges for each sample type and year in small RNA fragments' group.

RNA (Small fragments)						
Years	Concentration Range (min-max)			Size Range (min-max)		
	Plasma (pg/µL)	Supernatant (pg/µL)	Pellet (pg/µL)	Plasma (nt)	Supernatant (nt)	Pellet (nt)
2016	11.7 – 15.8	10.3 – 15.9	34.4 – 407	64 – 66	62 – 65	238 – 388
2017	9.12 – 18.3	9.7 – 22	83.7 – 1800	61 – 64	61 – 66	221 – 444
2018	6.94 – 12.8	11.8 – 27.1	28.3 – 1300	57 – 66	63 – 68	212 – 440
2019	10.4 – 19.3	11 – 21.5	22.3 – 511	62 – 64	61 – 65	222 – 429
2020	9.02 – 16.9	15.9 – 20.8	36.7 – 845	62 – 67	60 – 64	275 – 422

Table 5 - Concentration and size ranges for each sample type and year in the large RNA fragments' group.

RNA (Large fragments)						
Years	Concentration Range (min-max)			Size Range (min-max)		
	Plasma (pg/µL)	Supernatant (pg/µL)	Pellet (pg/µL)	Plasma (nt)	Supernatant (nt)	Pellet (nt)
2016	5.21 – 15.6	3.43 – 19.2	2.32 – 16.1	508 – 582	412 – 569	2095 – 7925
2017	7.37 – 15.9	1.62 – 31.2	3.51 – 85	516 – 576	428 – 523	3319 – 6967
2018	0.156 – 10.9	2.62 – 13.6	2.08 – 87.3	431 – 661	376 – 529	4136 – 8338
2019	1.72 – 9.53	1.36 – 18.1	1.18 – 88.4	449 – 522	449 – 536	4723 – 7580
2020	2.58 – 15.7	4.27 – 14.3	3.48 – 45.1	368 – 589	433 – 629	3861 – 8611

DNA Analysis

The PCA graph for the small DNA fragments (Figure 5) denotes the presence of two large compact groups, that present different fragment sizes. The group on the left mostly refers to plasma and supernatant samples (lower fragment size), whereas the right group are in its majority pellet samples (higher fragment size). The higher fragment size for the pellet samples is mainly due to a higher amount of genetic material relative to the plasma and supernatant samples. The pellet sample group also demonstrated a higher variation in the concentration than the other group.

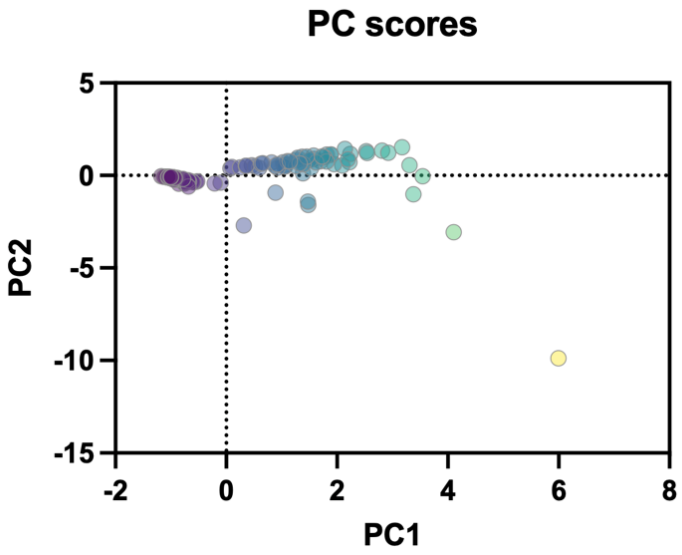

Figure 5 - PCA graph detailing the spread of all samples along a concentration and fragment size axis for the small DNA fragments.

Regarding large DNA fragments (Figure 6), two distinct groups were apparent, that also differed by their fragment size, with the left group being composed by plasma and supernatant samples (lower fragment size) and the right group by pellet samples (higher fragment size). In this graph, the left group is highly compact, showing little variation between all plasma and supernatant samples in terms of genetic material concentration and percentage, as well as fragment size. The right group is spread-like an almost horizontal line, signifying a large variation in fragment size but a very small variation in concentration between pellet samples.

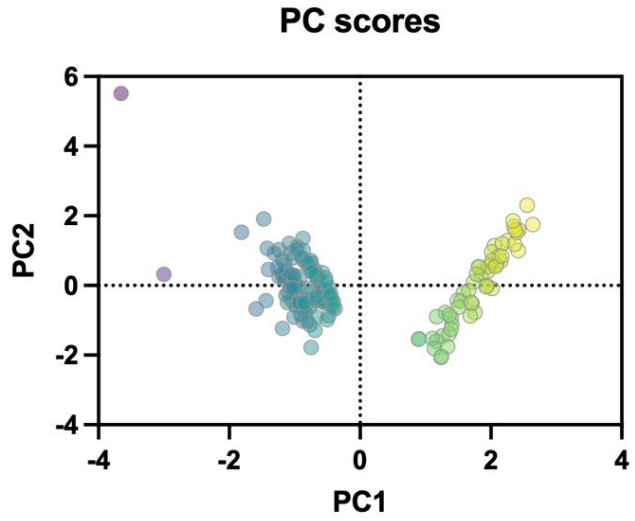

Figure 6 - PCA graph detailing the spread of all samples along a concentration and fragment size axis for the large DNA fragments.

For the plasma samples, the concentration of the small fragments of DNA (analysis region between 35 and 100 bp for plasma and supernatant DNA samples) (Figure 7) was mostly consistent throughout the years, with no statistically significant differences among the tested years. In general, there was a slight decline in concentration between 2016 (23.35 pg/μL) and 2018 (14.25 pg/μL), followed by a slight increase from 2018 to 2019 (18.15 pg/μL), while a decrease was observed between 2019 and 2020 (16.2 pg/μL). The average size of the small DNA fragments was also consistent throughout the years, with the averages being the same in 2016 and 2017 (50.5 bp), being the shortest observed in 2019 (47 bp). Significant differences were detected for small fragments quantity between the 2017 and 2019 groups ($p < 0.01$).

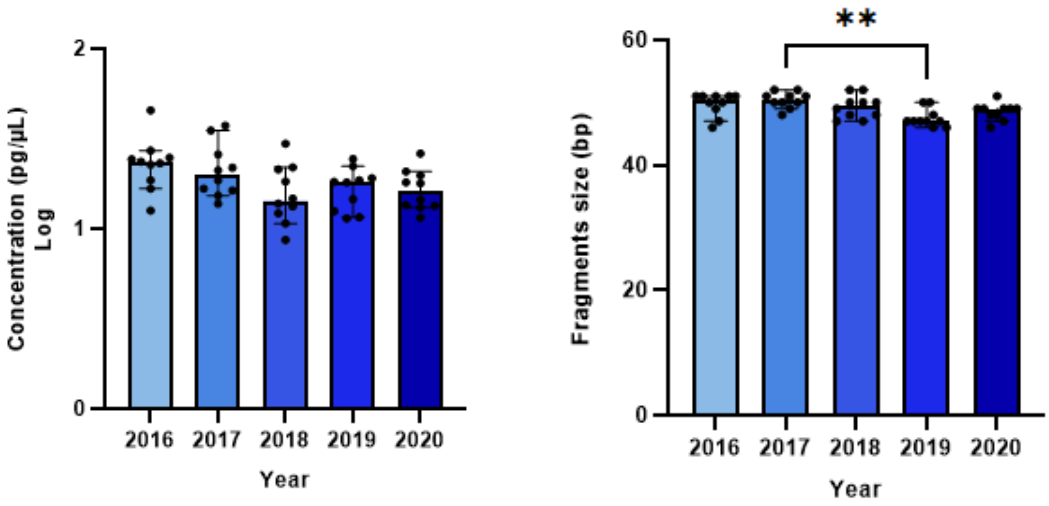

Figure 7 - Bar charts detailing the mean average concentration (left) and fragment size (right) for small DNA fragments in plasma samples categorised by year, with 95% confidence intervals and statistically significant differences between years.

For large DNA fragments in plasma samples (analysis region between 100 and 900 bp for plasma and supernatant DNA samples) (Figure 8), a noticeably increased concentration was found in 2017 samples (6.66 pg/μL), although statistically different concentrations were observed between 2017 and 2016 (4.28 pg/μL) ($p < 0.05$), and between samples collected in 2017 and 2020 (3.34 pg/μL) ($p < 0.05$). Fragment's sizes were consistent throughout the years, with no significant differences being registered.

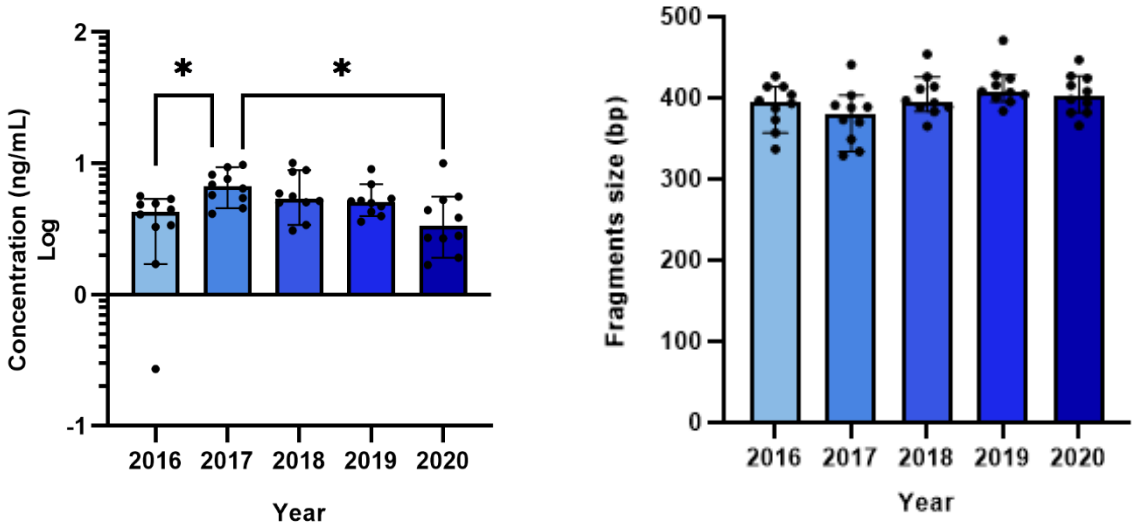

Figure 8 - Bar charts detailing the mean average concentration (left) and fragment size (right) for large DNA fragments in plasma samples categorised by year, with 95% confidence intervals and statistically significant differences between year groups included in the charts.

In supernatant samples, the small DNA fragments (Figure 9) maintained approximately the same average concentration levels among the tested years, only a small decrease was found in the samples collected in 2017 (11.65 pg/µL) and 2018 (10.55 pg/µL) in comparison to 2016 (16.55 pg/µL), and from 2019 (13.45 pg/µL) to 2020 (9.80 pg/µL). All changes were not statistically significant. The same can be said about the average fragment size, only varying between 47 and 48 bp throughout the years.

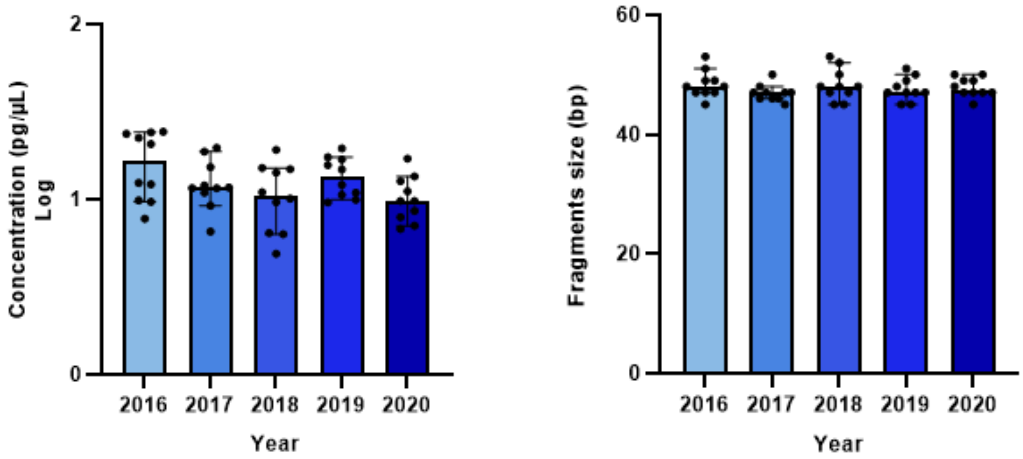

Figure 9 - Bar charts detailing the mean average concentration (left) and fragment size (right) for small DNA fragments in supernatant samples categorised by year, with 95% confidence intervals included in the charts.

In supernatant samples, the large DNA fragments (Figure 10) average concentration was similar in all years, having a slight but not significant increase in 2018 (3.06 pg/μL). That can also be observed for the average fragment size.

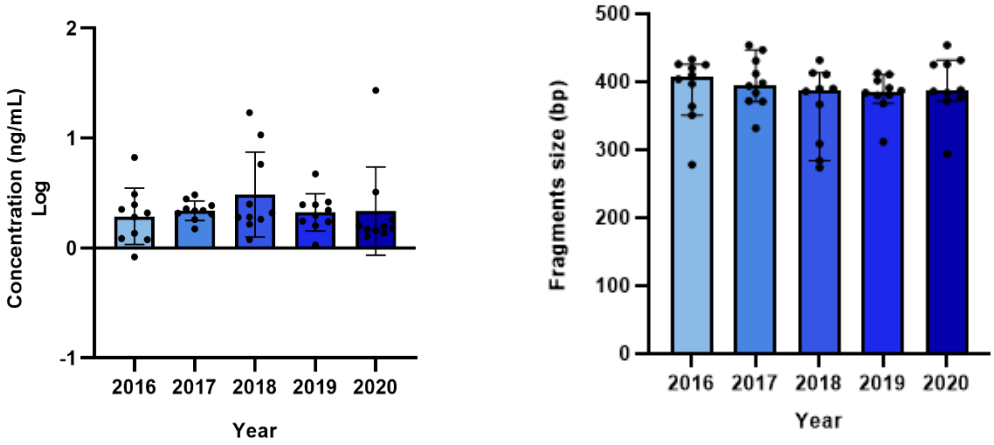

Figure 10 - Bar charts detailing the mean average concentration (left) and fragment size (right) for large DNA fragments in supernatant samples categorised by year, with 95% confidence intervals included in the charts.

In pellet samples, the small DNA fragments (analysis region between 200 and 2000 bp for pellet DNA samples) (Figure 11) concentration was rather small (1.05 ng/μL – 1.29 ng/μL), being similar independently of the year of collection. In the same line, average fragment size of around 700 bp was consistent in all tested samples.

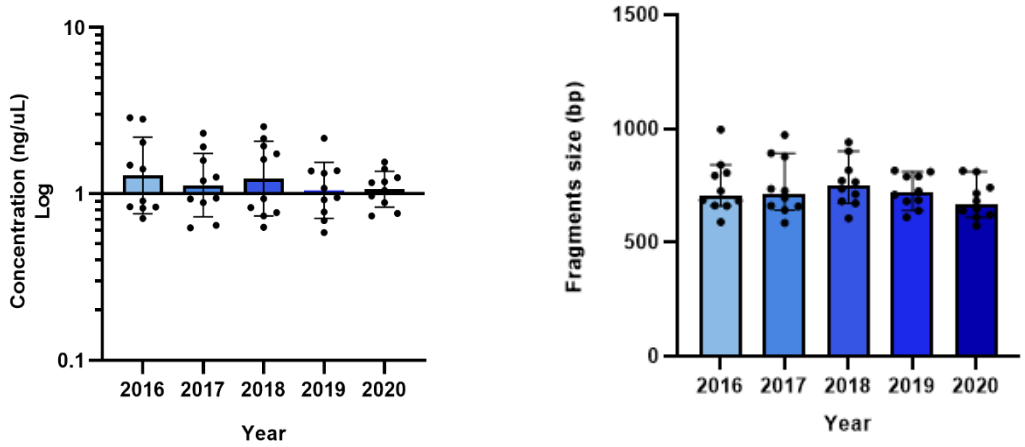

Figure 11 - Bar charts detailing the mean average concentration (left) and fragment size (right) for small DNA fragments in pellet samples categorised by year, with 95% confidence intervals included in the charts.

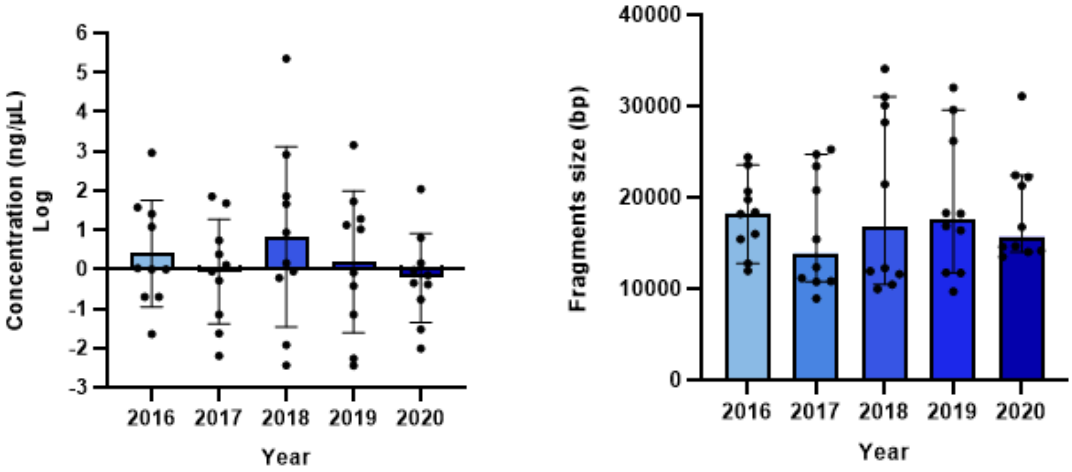

Figure 12 - Bar charts detailing the mean average concentration (left) and fragment size (right) for large DNA fragments in pellet samples categorised in year groups, with 95% confidence intervals included in the charts.

Concerning the large DNA fragments in pellet samples (analysis region between 2000 and 60000 bp for pellet DNA samples) (Figure 12), similar low concentration levels were observed, varying between 0.86 ng/µL - 1.77 ng/µL. A similar pattern was observed through the years for the average fragment size (13889 bp – 18294 bp). None of the variations were statistically significant.

RNA analysis

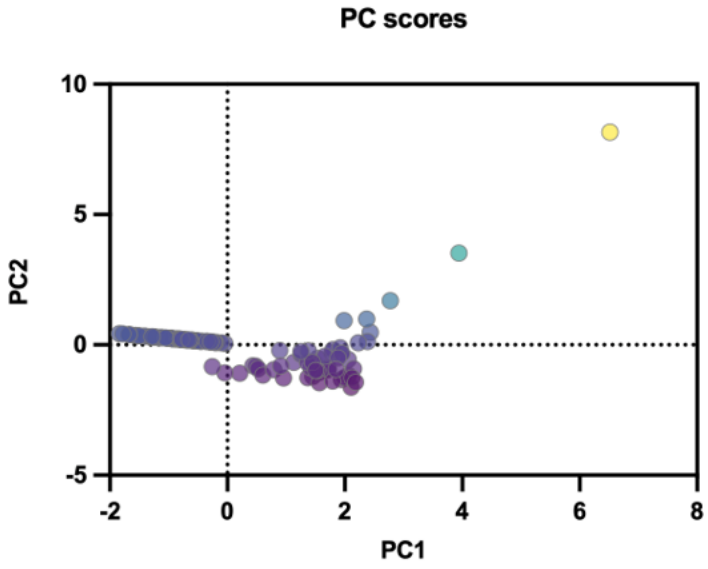

Figure 13 - PCA graph detailing the spread of all samples along a concentration and fragment size axis for the small RNA fragments.

In the PCA graph (Figure 13) regarding the small RNA fragments, two distinct groups were obtained, the group on the left includes plasma and supernatant samples (lower fragment size) and the group on the right refers pellet samples (higher fragment size). For the left group, the dots spread in an almost horizontal line, shows the moderate fragment size variation between plasma and supernatant, but almost no variation was observed in terms of concentration. The right group samples are diffusely distributed comparing to the left group but still is moderately compact, excluding the outliers. Thus, in general, a middling variation amplitude was found in pellet samples, both for concentration and fragment size. All outliers presented a much higher concentration and fragment size than the others.

The PCA graph for the large RNA fragments (Figure 14) has two distinct groups, where once again the left group is comprised by plasma and supernatant samples (lower fragment size) and the right size contains the pellet samples (higher fragment size). The left group is spread in a declining diagonal line, suggesting that RNA higher average fragment size, inversely relates with concentration. The dots in the right group are scattered out in a wide area, translating a large variation not only for concentration but also for fragment size in pellet samples.

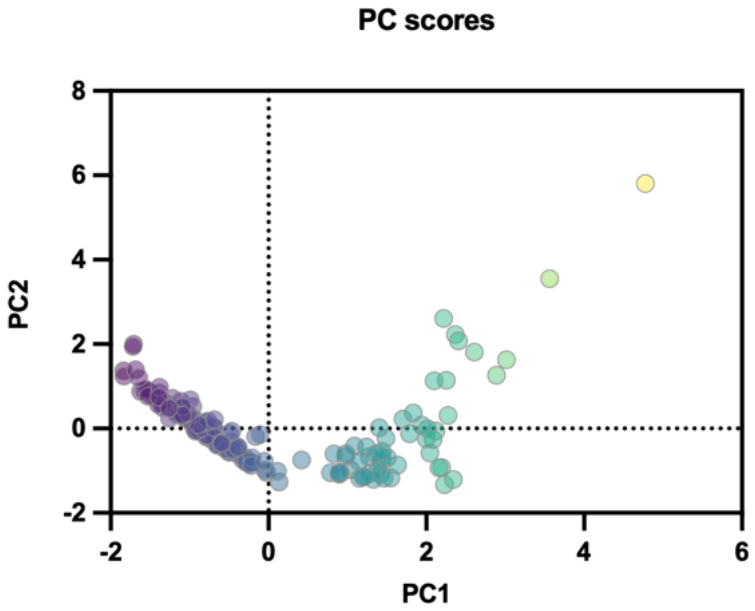

Figure 14 - PCA graph detailing the spread of all samples along a concentration and fragment size axis for the large RNA fragments.

The average concentration of small RNA fragments detected in plasma samples (analysis region between 40 and 100 nt for plasma and supernatant RNA samples) (Figure 15) were significantly lower in 2018 (12.1 pg/μL), comparing with 2016 (14.95 pg/μL) ($p <$

0.01), 2017 (14 pg/µL) ($p < 0.05$) and 2019 (14.45 pg/µL) ($p < 0.05$). A similar trend was observed concerning fragment size, but significant differences were only observed between the samples collected in 2016 (65 nt) and 2018 (62 nt), being smaller in the latter ($p < 0.001$).

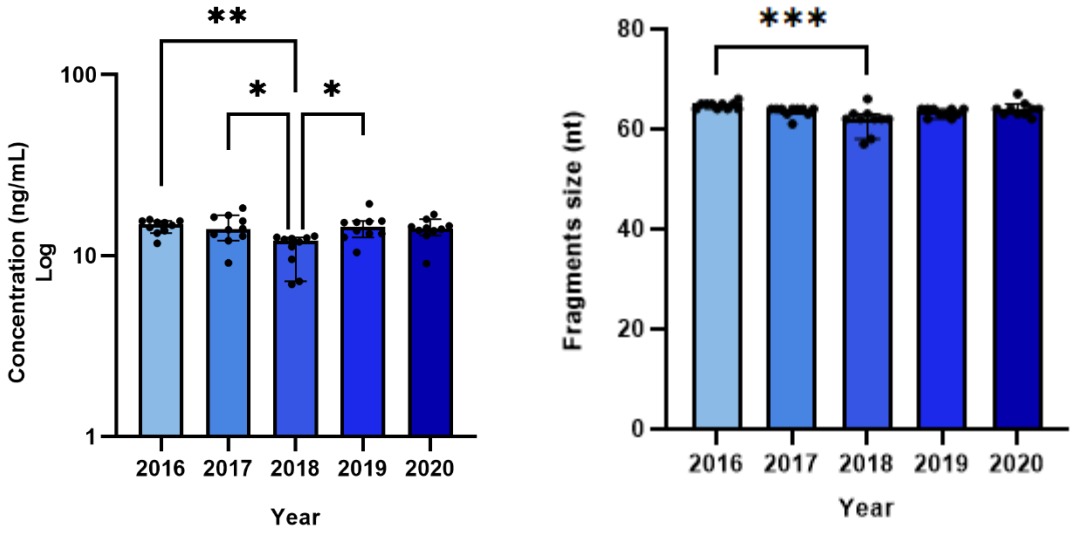

Figure 15 - Bar charts detailing the mean average concentration (left) and fragment size (right) for small RNA fragments in plasma samples categorised by year, with 95% confidence intervals and statistically significant differences annotated.

In the case of the large RNA fragments found in plasma samples (analysis region between 100 and 1000 nt for plasma and supernatant RNA samples) (Figure 16), significant differences were found between samples collected in 2016 (10.5 pg/µL) and 2018 (5.25 pg/µL) ($p < 0.05$), 2017 (11.85 pg/µL) and 2018 ($p < 0.01$), and 2017 and 2020 (5.08 pg/µL) ($p < 0.05$). The average fragment sizes were similar in samples collected from 2016 (551 nt) to 2018 (545 nt), suffering a sizeable decrease in 2019 (472 nt), that was maintained in 2020 (484 nt). Nonetheless, significant changes were observed for average fragment sizes, between 2016 and 2019 ($p < 0.01$), 2016 and 2020 ($p < 0.05$), 2017 (548 nt) and 2019 ($p < 0.01$), 2017 and 2020 ($p < 0.05$), and 2018 and 2019 ($p < 0.05$).

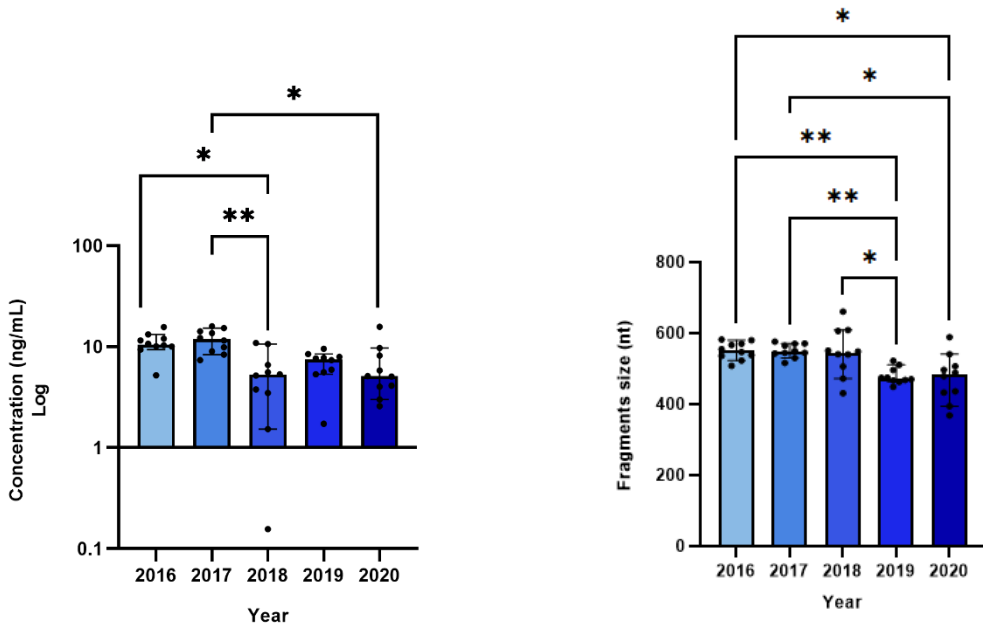

Figure 16 - Bar charts detailing the mean average concentration (left) and fragment size (right) for large RNA fragments in plasma samples categorised in year groups, with 95% confidence intervals and statistically significant differences between year groups included in the charts.

For the supernatant samples, the average concentration of small RNA fragments (Figure 17) remained very consistent in almost all years except in samples collected in 2020 (18.45 pg/μL), that presented increased amounts, being significantly higher than in samples collected in 2016 (14.2 pg/μL) ($p < 0.01$). As for the average fragment sizes, a consistent increase was observed throughout the years, excepting those of year 2020 (62 nt), that displayed a significant decrease. Nonetheless, statistical differences were only observed between the small RNA fragments extracted from samples collected in 2018 (64 nt) and 2020 ($p < 0.01$).

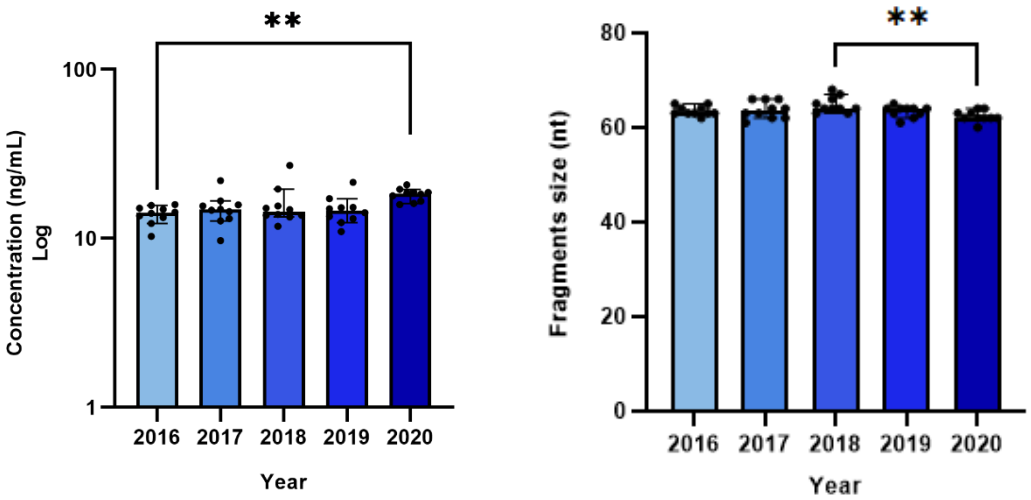

Figure 17 - Bar charts detailing the mean average concentration (left) and fragment size (right) for small RNA fragments in supernatant samples categorised in year groups, with 95% confidence intervals and statistically significant differences between year groups included in the charts.

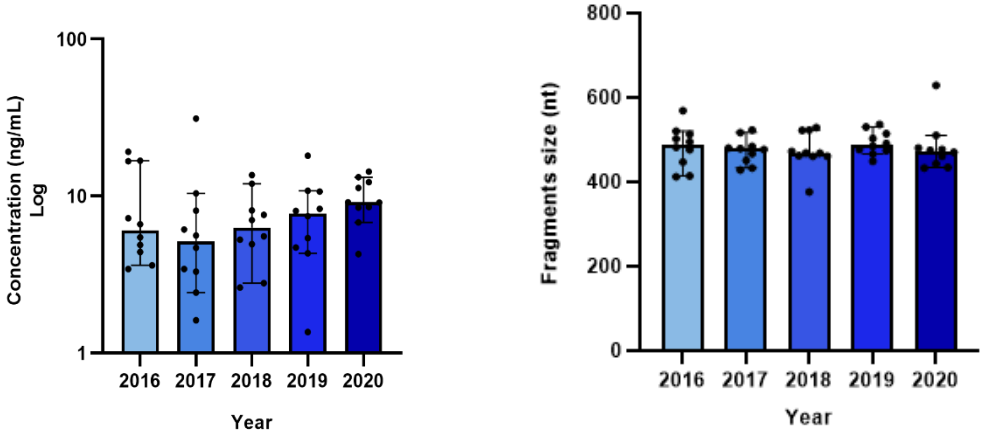

Figure 18 - Bar charts detailing the mean average concentration (left) and fragment size (right) for large RNA fragments in supernatant samples categorised in year groups, with 95% confidence intervals included in the charts.

The average concentration of large RNA fragments extracted in supernatant samples (Figure 18) were slightly decreased in samples collected in 2016 (6.04 pg/μL) to 2017 (5.15 pg/μL) than in other years, although not significantly. The average fragment sizes were rather consistent among the tested years.

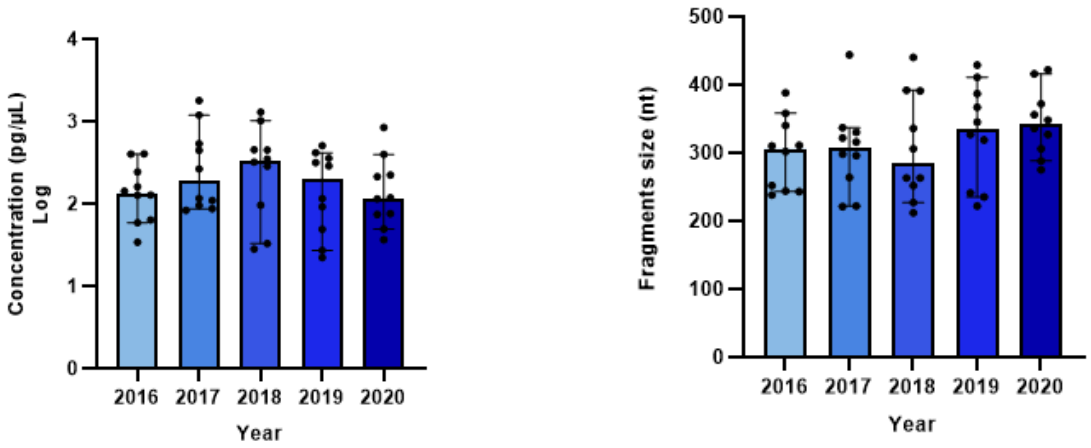

Figure 19 - Bar charts detailing the mean average concentration (left) and fragment size (right) for small RNA fragments in pellet samples categorised in year groups, with 95% confidence intervals included in the charts.

The average concentration of small RNA fragments (analysis region between 40 and 1000 nt for pellet RNA samples) was slightly higher in pellet samples (Figure 19) collected in 2018 (337 pg/µL), that in other years, though no significant changes were detected. The average fragment size also maintained consistency for all years, only registering a very modest decrease in 2018 (285 nt) compared to the other years.

Concerning the average concentration of large RNA fragments in pellet samples (analysis region between 1000 and 10000 nt for pellet RNA samples) (Figure 20), apparent increased values were observed in 2017 (20.85 pg/µL) and 2018 (35.8 pg/µL), though without significative differences. The average fragment size was about 5000 nt in all the tested years, except in 2018, whose fragment size slightly increased close to 7000 nt.

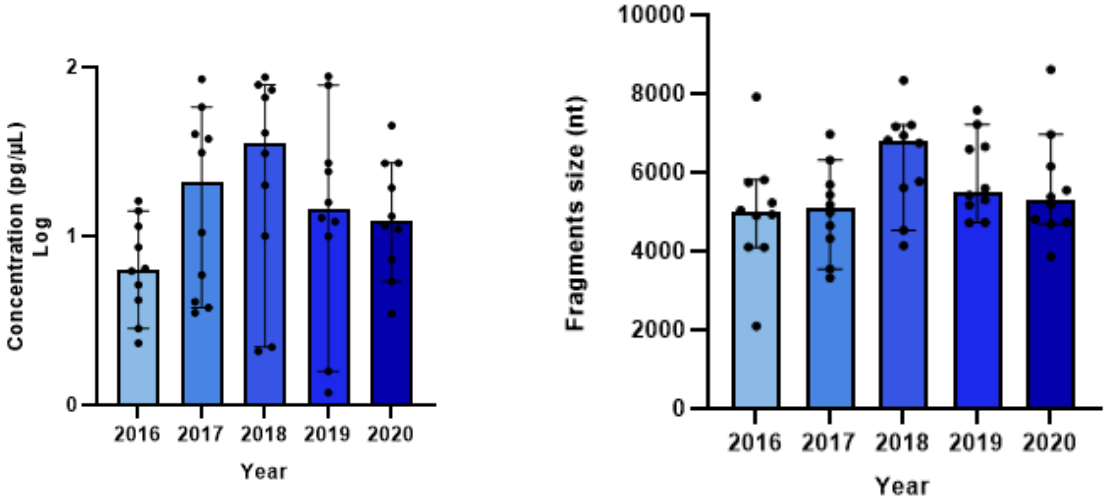

Figure 20 - Bar charts detailing the mean average concentration (left) and fragment size (right) for large RNA fragments in pellet samples categorised in year groups, with 95% confidence intervals included in the charts.

DISCUSSION

The aim of this project was to analyse cfDNA and cfRNA samples from the IPO-Porto biobank in terms of its quality and quantity; comparing samples from different years and determining the effects of long-term biobank storage on the genetic material found in breast cancer liquid biopsies, to ultimately establish quality criteria for biobank guidelines.

For small DNA fragments in plasma samples, the concentration range reveals a consistency between all the years, which is also reinforced by the concentration graph that shows no significant changes in any of the years. When compared with the larger DNA fragments, its concentration range has higher minimums in all years and higher maximums. The same happens for supernatant and pellet samples. A possible explanation for why the average concentration of small DNA and RNA fragments is higher than the average concentration of large DNA and RNA fragments can be due to the amount of time the samples spent outside the cryogenic freezer during the processing procedure.

Conversely, variability occurs for large DNA fragments and RNA (both small and large) isolated from all the sample types.

For 2016, significant changes in nucleotide material concentration with other years can be noticed, especially in plasma samples for large DNA fragments from 2017; small RNA fragments and large RNA fragments from 2018. Also, the same difference is observed in the supernatant samples for small RNA fragments from 2020. In the case for 2017, these statistically significative differences with other sample year groups can be seen in plasma samples for large DNA fragments from 2016 and 2020, small RNA fragments from 2018 and large RNA fragments from 2018 and 2020. The substantial changes in concentration observed with 2018 can be found in plasma samples for small RNA fragments from 2016, 2017 and 2019 and large RNA fragments from 2016 and 2017. The only significant difference of concentration that was detected with 2019 was for the small RNA fragments in plasma samples from 2018. Finally, the statistically substantial changes with 2020 are seen in plasma samples for large DNA and RNA fragments from 2017, and supernatant samples for small RNA fragments from 2016. These results show that almost all samples maintained a high level of quality and concentration that suffered a minimal amount of degradation from 2016 to 2020 while being stored in the IPO-Porto tumour bank, with its quality and reproducibility still being high enough to be utilised in biomedical investigation. This means that liquid biopsies can be stored in the biobank for more than 5 years before being discarded. The results obtained were corroborated by another dissertation from *Rodrigues, F.* [18], where all groups of DNA samples extracted from whole blood categorized by storage time (<1 year, 2 years and >3 years) had no statistically significant differences between themselves in regards to concentration. However, in group T1 (DNA

extracted from previously stored whole blood) a statistically significant difference in the 260/280 ratio can be observed between the <1 year group and the other two groups. That difference is however not important given that all ratio values are above 1.8, which means that the DNA is considered free of proteins and other contaminants, and therefore within the acceptable parameters of quality and purity.

Some explanations could be given to justify some differences in the quality of the samples throughout the years. For example, instead of having only one proficient technician processing the samples, the samples were processed by several persons with different experiences in processing. Moreover, the cryogenic freezer being opened unnecessarily and/or excessively leads to the deterioration of genetic material due to needless freezing and thawing cycles that could be avoided with effective processing and managing procedures that minimises those cycles.

The effect of freezing and thawing cycles on the genetic material was better explored by Shao *et al.* (2012) [30], where genomic DNA samples extracted through QIAmp Blood Maxi Kit, Gentra Puregene Blood Kit or manual phenol/chloroform method with ethanol precipitation revealed significant degradation of the genetic material with the increase of the number of freeze/thaw cycles, with the degradation being more noticeable the larger the DNA chain is. The average size of DNA after 18 freeze/thaw cycles would be around 25 kb, with the method of extraction and starting DNA size having no influence on that average. DNA in higher concentrations were generally more resistant to degradation than DNA in lower concentrations.

Genomic analysis as described above, can be influenced by many aspects. One of them is the technique used to isolate the DNA or the RNA. Different isolation kits can result in several different outcomes regarding genetic material quality and quantity. For example, our work used the magLEAD 12gC extraction device, an automatized process that is generally more precise than manual extraction techniques due to the absence of human error in the realisation of the process.

Furthermore, the method of quantification and quality assessment may influence the results obtained, as the precision and accuracy may vary depending on the procedure used. For example, the TapeStation platform is a more precise way of measuring the quality and quantity of genetic material than the gel electrophoresis of the paper cited as [30]. Other methods like the NanoDrop spectrophotometer measure not only the quantity of genetic material but also its purity through the absorbance of all molecules contained in the sample that absorb around 230, 280 and 260 nm. The spectrophotometer then calculates the ratio

between absorbance at 260 nm (most nucleotides absorb at 260 nm) and absorbance at 230 or 280 nm (many contaminants like protein or phenol absorb at either 230 or 280 nm) in order to obtain a numerical value that determines the purity grade of DNA or RNA [31].

The RNA integrity and quality (IQ) assay utilises three standards consisting of: a small or degraded RNA; a large or intact RNA and a blank. Samples are analysed using a dye, and two emission signals combined using an algorithm to yield a quality score representative of the ratio between small and large RNAs in the sample [32]. Through that, a RIN number is generated. The RIN algorithm is based on features that contribute information about the RNA integrity [33].

Some of the benefits of obtaining the RIN are the assessment of the integrity of RNA, the possibility of directly comparing RNA samples (for example, before and after shipment, comparing integrity of same tissue across different labs, etc.) and ensuring repeatability and trustability of experiments, also being part of the workflow of several experiments, like Next Generation Sequencing (NGS), microarray, and many others [34]. For this reason, it is of main importance to ensure the maintenance of the genetic material integrity on the biobank and establish internal quality controls, since most research projects start there.

CONCLUSIONS & FUTURE PERSPECTIVES

Nowadays, liquid biopsies have shown many promising results for cancer patients' monitoring, prognosis, and treatment selection. However, using ctDNA for cancer screening and early detection routinely remains to be solved. The biggest challenge is the low concentration of genetic material in the blood, as well as the respective quality. Such quality and integrity must be evident and standardised before liquid biopsy can replace the current use of mammography to detect early disease, although highly sensitive assays can detect very small amounts of genetic material nowadays, and the main cause of quality and integrity loss is inadequate storage manners.

In our hands, small DNAs and RNAs obtained from 3 different types of clinical material does not seem to be influenced by the tested storage time. The same can be said about the large DNA and RNA fragments.

More work is needed to further improve the technologies in isolating and analysing tumour-derived materials in the blood. It is still important to investigate the genetic material isolation methods, storage, quantification and quality evaluation methods, for the next steps of liquid biopsy in the clinic. Our work remains a small but important part of a bigger scenario for translational research.

Liquid biopsies in breast cancer have yielded promising results and the outlook remains optimistic. With time, it is certainly possible that liquid biopsies may play an even greater role in the breast cancer clinic.

REFERENCES

1. *Cancer : Key Facts*. [cited 2020 December]; Available from: https://www.who.int/news-room/fact-sheets/detail/cancer.
2. *Breast Globocan 2020*. [cited 2020 December]; Available from: https://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf.
3. *Portugal Globocan 2020*. [cited 2020 December]; Available from: https://gco.iarc.fr/today/data/factsheets/populations/620-portugal-fact-sheets.pdf
4. Rojas, K. and A. Stuckey, *Breast Cancer Epidemiology and Risk Factors.* Clin Obstet Gynecol, 2016. **59**(4): p. 651-672.
5. Sun, Y.S., et al., *Risk Factors and Preventions of Breast Cancer.* Int J Biol Sci, 2017. **13**(11): p. 1387-1397.
6. Winters, S., et al., *Breast Cancer Epidemiology, Prevention, and Screening.* Prog Mol Biol Transl Sci, 2017. **151**: p. 1-32.
7. Cardoso, F., et al., *Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-updagger.* Ann Oncol, 2019. **30**(8): p. 1194-1220.
8. Akram, M., et al., *Awareness and current knowledge of breast cancer.* Biol Res, 2017. **50**(1): p. 33.
9. Makki, J., *Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance.* Clin Med Insights Pathol, 2015. **8**: p. 23-31.
10. Malhotra, G.K., et al., *Histological, molecular and functional subtypes of breast cancers.* Cancer Biol Ther, 2010. **10**(10): p. 955-60.
11. Cserni, G., et al., *The new TNM-based staging of breast cancer.* Virchows Arch, 2018. **472**(5): p. 697-703.
12. Barzaman, K., et al., *Breast cancer: Biology, biomarkers, and treatments.* Int Immunopharmacol, 2020. **84**: p. 106535.
13. Alimirzaie, S., M. Bagherzadeh, and M.R. Akbari, *Liquid biopsy in breast cancer: A comprehensive review.* Clinical genetics, 2019. **95**(6): p. 643-660.
14. Waks, A.G. and E.P. Winer, *Breast Cancer Treatment: A Review.* JAMA, 2019. **321**(3): p. 288-300.
15. Riso, B., *O estatuto social do biobanco de pesquisa clínica.* 2016.
16. Carvalho, L., et al., *Banco de Tumores: imperativo na medicina.* 2007.
17. *Serviço de Anatomia Patológica IPO Porto*. [cited 2021 February]; Available from: https://ipoporto.pt/eu-doente/servicos/servico-de-anatomia-patologica/.
18. Rodrigues, F., *Avaliação da qualidade das amostras de biópsias líquidas de pacientes com tumores urológicos, colhidas e armazenadas no Biobanco-iMM, CAML*, in *Faculdade de Medicina da Universidade de Lisboa*. 2019, Universidade de Lisboa: Universidade de Lisboa p. 59.
19. Poulet, G., J. Massias, and V. Taly, *Liquid biopsy: general concepts.* Acta cytologica, 2019. **63**(6): p. 449-455.
20. Junqueira-Neto, S., et al., *Liquid biopsy beyond circulating tumor cells and cell-free DNA.* Acta cytologica, 2019. **63**(6): p. 479-488.
21. Grölz, D., et al., *Liquid biopsy preservation solutions for standardized pre-analytical workflows—venous whole blood and plasma.* Current pathobiology reports, 2018. **6**(4): p. 275-286.
22. Carter, A. and F. Betsou, *Quality assurance in cancer biobanking.* Biopreservation and biobanking, 2011. **9**(2): p. 157-163.
23. *magLEAD*. [cited 2021 September]; Available from: https://pss.co.jp/english/product/magtration/lead6-12gc.html.
24. *4200 TapeStation System Manual*. 02/2021 ed.: Agilent Technologies, Inc.

25. *Comparison of DNA Assays Using the 4200 TapeStation System and 2100 Bioanalyzer System*. [cited 2021 April]; Available from: https://www.chem-agilent.com/pdf/5991-9093EN.PDF.
26. *Complete Success begins with Sample Quality Control. TapeStation*. [cited 2021 April]; Available from: https://www.agilent.com/cs/library/brochures/brochure-dna-rna-sample-quality-control-tapestation-5994-0060EN-agilent.pdf.
27. *Agilent High Sensitivity D1000 ScreenTape System Quick Guide*. [cited 2021 April]; Available from: https://www.agilent.com/cs/library/usermanuals/Public/ScreenTape_HSD1000_QG.pdf
28. *Genomic DNA ScreenTape Assay for TapeStation Systems Quick Guide*. [cited 2021 April]; Available from: https://www.agilent.com/cs/library/usermanuals/public/gDNA_QuickGuide.pdf.
29. *Agilent RNA ScreenTape Quick Guide for TapeStation Systems*. [cited 2021 April]; Available from: https://www.tuni.fi/sites/default/files/media/file/tapestation-4150_rna_quickguide_0.pdf
30. Shao, W., S. Khin, and W.C. Kopp, *Characterization of effect of repeated freeze and thaw cycles on stability of genomic DNA using pulsed field gel electrophoresis.* Biopreserv Biobank, 2012. **10**(1): p. 4-11.
31. *T042-TECHNICAL BULLETIN NanoDrop Spectrophotometers*. [cited 2022 March]; Available from: dna.uga.edu/wp-content/uploads/sites/51/2019/02/Note-on-the-260_280-and-260_230-Ratios.pdf.
32. *Qubit RNA IQ Assay: a fast and easy fluorometric RNA quality assessment* [cited 2022 March]; Available from: assets.thermofisher.com/TFS-Assets/BID/Application-Notes/qubit-rna-iq-assay-fluorometric-rna-quality-assessment-app-note.pdf.
33. Mueller, O., S. Lightfoot, and A. Schroeder, *RNA integrity number (RIN)–standardization of RNA quality control.* Agilent application note, publication, 2004. **1**: p. 1-8.
34. *The RNA Integrity Number (RIN)*. [cited 2022 March]; Available from: https://www.gene-quantification.de/RIN-2.pdf.