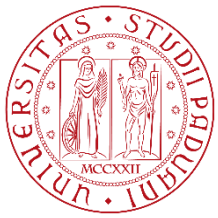

UNIVERSITÀ DEGLI STUDI DI PADOVA

CORSO DI LAUREA MAGISTRALE IN MEDICINA E CHIRURGIA

Dipartimento di Neuroscienze - DNS
Direttore: Prof. Raffaele De Caro

U.O.C. Chirurgia Plastica
Direttore: Prof. Franco Bassetto

TESI DI LAUREA

SINGLE CELL TRANSCRIPTOMIC ANALYSES OF HUMAN MAMMARY GLAND AND BREAST TUMORS

Relatore: Prof. Franco Bassetto

Correlatore: Prof. Vincenzo Vindigni

Correlatore: Prof. Stefano Piccolo

Laureanda: Caterina Ottaviano

Anno Accademico

2021-2022

UNIVERSITÀ DEGLI STUDI DI PADOVA

CORSO DI LAUREA MAGISTRALE IN MEDICINA E CHIRURGIA

Dipartimento di Neuroscienze - DNS
Direttore: Prof. Raffaele De Caro

U.O.C. Chirurgia Plastica
Direttore: Prof. Franco Bassetto

TESI DI LAUREA

SINGLE CELL TRANSCRIPTOMIC ANALYSES OF HUMAN MAMMARY GLAND AND BREAST TUMORS

Relatore: Prof. Franco Bassetto

Correlatore: Prof. Vincenzo Vindigni

Correlatore: Prof. Stefano Piccolo

Laureanda: Caterina Ottaviano

Anno Accademico

2021-

INDEX

ABSTRACT

Background

Breast cancer is the most frequently diagnosed cause of death from cancer in women world-wide, and the second cause of death from cancer in women in developed countries. Besides a significant family history of breast or ovarian cancer, the risk of breast cancer increases with age. This correlation with age could be linked to the continuous and progressive endocrine proliferative stimulus that the mammary epithelium undergoes over the years, together with the progressive damage to DNA and accumulation of epigenetic changes that modifies the balance in the expression of oncogenes and tumor suppressor genes.

Purpose of the study

The purpose of our study is to collect normal samples of human mammary gland that will be used as healthy controls for pathological tissue specimens from breast primary tumors and metastases. Our study is part of a larger study, whose aim is to investigate breast cancer heterogeneity at a single-cell level, through the analyses of the transcriptome of the different cells that make up the human mammary gland involved in breast tumorigenesis. Through transcriptional characterization of cells, we aim to reach a more comprehensive evaluation of mammary cell diversity, that could serve as reference point to pinpoint cell states specifically associated to tumorigenesis and, in particular, to acquisition of metastatic proclivity. To do this, all the samples will be analyzed through two different sequencing strategies: single cell RNA sequencing of both epithelial, stromal and microenvironmental cells of the breast tissue, and RNA sequencing of a bulk population of epithelial breast cells.

Materials and methods

Samples were obtained from five consenting healthy patients (between 47 and 63 years old), during reduction mammoplasties, and they were processed through mechanical and enzymatic tissue dissociation to reach single cell suspensions.

Three of the human mammary gland samples were destined to single-cell RNA sequencing analyses, following Chromium Next GEM protocol. Cells were loaded into a microfluidic chip that enables to encapsulate each cell into reaction vesicles where reactions occur. After reverse transcription inside each reaction droplet, all the cDNAs were tagged with various oligos specific for each different cell (10x Barcodes), and for each distinct

transcript within the cell (UMIs). Thanks to this barcoding, during data analysis, after sequencing (by Illumina sequencer), each transcript can be mapped back to the origin cell, and all transcripts from a single cell can be quantified.

On the other hand, three samples were assigned to RNA analyses in bulk of epithelial cell population, that were previously selected through fluorescence-mediated cell sorting (FACSAria flow cytometer), to obtain a higher resolution of the sequencing data.

Results

RNA sequencing and data analyses of our healthy samples is still in progress.

Single-cell RNA sequencing technique allows to detect and sequence the higher expressed genes of the processed cells at a single-cell level.

Sequencing of bulk populations will offer a deeper analysis of the gene expression of our samples, with a higher number of detected genes (not only those more highly expressed) but limited to sorted epithelial cells.

The resulting output will be analyzed and used as a healthy control for pathological samples from breast primary tumor and brain metastasis.

Here we also present scRNA sequencing analyses of a primary breast tumor and a breast cancer brain metastasis. Through this we reconstructed the complexity of different cell populations in primary and metastatic cancer and, by integrating CNV and transcriptomic analyses, we characterized primary tumor and metastatic cells at the molecular level.

Conclusions

At the moment, this study is investigating the gene expression of normal samples that will be used as the baseline from which pathological tissues can be explored. Certainly, further studies will be continued to make useful what has been collected up until now, with the purpose of trying to understand the molecular mechanisms and the various pathways that guide breast primary tumors development and spreading within the human body.

1.0 NORMAL AND MALIGNANT ASPECTS OF THE HUMAN MAMMARY GLAND

1.1 ANATOMY OF THE HUMAN MAMMARY GLAND [1]

In young adult females, each breast is a rounded eminence largely lying within the superficial fascia anterior to the upper thorax. Breast shape and conformation depend on genetic, racial, and dietary factors and on the age, parity, and menopausal status of the individual. The trunk superficial fascial system splits to enclose the breast to the pectoralis fascia, part of the deep fascial system. The inframammary crease is a zone of adherence of the superficial fascial system to the underlying chest wall at the inferior crescent of the breast. The breast lies on the deep pectoral fascia. Between the breast and the deep fascia, the loose connective tissue in the “submammary space” allows the breast some degree of movement on the deep pectoral fascia.

The nipple projects from the center of the breast anteriorly. The skin covering the nipple and the surrounding areola contains numerous sweat and sebaceous glands that open directly on the skin surface. The skin of the nipple and areola is rich in melanocytes and is therefore darker than the skin covering the remainder of the breast.

The breasts are made up by lobes that contain a network of glandular tissue consisting of branching ducts and terminal secretory lobules within a connective tissue stroma. The terminal duct lobular unit is the functional milk secretory component of the breast; pathologically, it gives rise to primary malignant lesions within the breast.

Although the lobes are usually described as discrete territories, they interlace in three dimensions and merge at their edges; they cannot be distinguished during surgery.

The connective tissue stroma that surrounds the lobules is dense and fibrocollagenous, whereas intralobular connective tissue has a loose texture that allows the rapid expansion of secretory tissue during pregnancy.

Fibrous strands or sheets consisting of condensations of connective tissue between the layer of deep fascia that covers the muscles of the anterior chest wall and the dermis. These suspensory ligaments (of Astley Cooper) are often well developed in the upper part of the breast and support the breast tissue, helping to maintain its non-ptotic form. In the normal breast, fibrous tissue also surrounds the glandular components and extends to the skin and nipple, assisting the mechanical coherence of the gland. The interlobar stroma contains variable amounts of adipose tissue, which is responsible for much of the increase in breast size during puberty.

1.1.1 VASCULAR SUPPLY

Breasts are supplied by different artery branches. The axillary artery supplies blood thanks the superior thoracic artery, the pectoral branches of the thoraco-acromial artery, the lateral thoracic artery, and the subscapular artery. The internal thoracic artery provides perforating branches to the anteromedial part of the breast. While the second to fourth anterior intercostal arteries provide perforating branches more laterally in the anterior thorax side.

Blood is drained by the circular venous plexus that surround the areola and by veins that go along with the corresponding arteries, from the glandular tissue of the breast into the axillary, internal thoracic and intercostal veins.

1.1.2 LYMPHATIC DRAINAGE

The lymphatic flow of the breast plays a fundamental role in metastatic dissemination, which occurs principally by lymphatic routes. The main lymphatic drainage of the breast derives from the dermal network. The direction of lymphatic flow within the breast goes in parallel with the major venous tributaries and enters the regional lymph nodes via the extensive periductal and perilobular network of lymphatic channels. The majority of these lymphatics drain into the axillary group of regional lymph nodes directly or through the retro-areolar lymphatic plexus. Some lymphatic channels drain in the subclavicular lymph nodes following the vascular channels, after penetrating pectoralis and joining lymphatics who drain the deeper parenchymal tissues.

Lymphatic from the left breast finally drain into the thoracic duct and then, into the left subclavian vein. Lymphatic from the right drain into the right subclavian vein, near its junction with the internal jugular vein.

Part of the medial side of the right breast drain into the internal thoracic group of lymph nodes. This chain may drain downward to the groin toward the superior and inferior epigastric lymphatic routes. The connection of lymphatics across the midline may provide access of lymphatic flow to the opposite axilla.

Axillary nodes drain more than 75% of the lymph from the breast. There are 20-40 nodes grouped as pectoral (anterior), subscapular (posterior), central and apical. As regards surgical classification, nodes are described in relation to pectoralis minor:

- Level 1: the low nodes, those who are localized below the pectoralis minor.
- Level 2: the middle group, who have as component lymph nodes behind the muscle.
- Level 3: the upper or apical nodes, localized between the upper border of pectoralis minor and the low border of the clavicle.

Rotter's nodes (interpectoral group) are one or two other nodes localized between pectoralis minor and major.

The presence of metastatic lymph nodes has strong prognostic significance and might influence therapeutic decisions. For this reason, axillary lymph node dissection may be performed in women with breast cancer. However, this dissection can lead to chronic postoperative complications, such as pain, seroma formation, reduced arm mobility, altered sensations and lymphoedema.

Lymphatic mapping with sentinel lymph node biopsy has become an important technique in the staging of women with early breast cancer. A radiolabeled colloid is injected into either the subareolar tissue of the index quadrant of the breast or the peritumoral and intradermal tissue lying on the top of the primary tumor. At the time of surgery, a vital blue dye is injected after general anesthesia is achieved. The combination of radioisotope and dye provide the most accurate means of localizing the sentinel node. It represents the first draining node of the axilla, and it must be surgical removed for histopathological analyses, to detect the presence of metastases.

1.1.3 INNERVATION

The breast is innervated by anterior and lateral branches of the fourth and the sixth intercostal nerves. Those nerves carry sensory and sympathetic efferent fibers. Nipple is innerved by an extensive plexus formed by the anterior branch of the lateral cutaneous branch of T4; its sensory fibers terminate at the epithelium level as free endings, Meissner corpuscles and Merkel disc endings.

Secretory activities of the gland are regulated by ovarian and hypophysial hormones.

1.1.4 MICROSTRUCTURE

The mammary gland is a complex tissue comprised of an epithelial parenchyma enclosed in an array of stromal cells that regulate its proliferation, differentiation and survival [2]. The epithelial cells elaborate the ductal network, whereas a variety of stromal cells or connective tissues with extracellular matrix (ECM) protein supports the mammary glands. The major components of stromal connective tissues are:

- Adipocytes, which constitute the mammary fat pad and retain the inserted ductal network.
- Fibroblasts which support the hematopoietic system.
- Vascular endothelial cells which support the blood vessels.
- Various innate immune cells (both macrophages and mast cells).

- Nerves.

The fatty stroma is the supportive network for the epithelium bi-layered structure and provides nutrients, blood supply, and immune defenses besides the physical structure to the gland. Each specific stromal cell secretes various signals for specific aspects of the development and function of the epithelium [3].

The microstructure of the breast tissue varies with age, time in menstrual cycle, pregnancy, and lactation. As regards the mature, resting breast, ducts are lined by columnar epithelium, which is two cells thick in the larger ducts, while only one cell thick (columnar or cuboidal cells) in the smaller ones.

There are two main types of epithelial cells in the mammary gland, namely, luminal and basal. The luminal epithelium forms the inner layer of the ducts as a lactiferous duct.

In contrast, the basal epithelium consists of myoepithelial cells that form the outer layer of mature mammary ducts. It also harbors stem and progenitor cells, which form both luminal and myoepithelial cells/layer [3].

These two types of epithelial cells can be distinguished through immunophenotypic characterization. Cytokeratins 8 and 18 are expressed in the luminal layer, whereas cytokeratins 5/14 and the transcription factor p63 characterize the basal epithelial layer [4].

1.1.5 POSTMENOPAUSAL BREAST

After the menopause, progressive atrophy of lobules and ducts occurs, and a fatty replacement of the glandular breast tissue takes place. Only few ducts may remain. The stroma becomes much less cellular and collagenous fibers declines. The amount of adipose tissue varies widely between individuals.

1.2 PHYSIOLOGY OF THE BREAST

The breasts begin to develop during puberty, thanks to the estrogen stimulation of the monthly female sexual cycle. Estrogens promote growth of the breasts' mammary glands and the deposition of fat to give breast mass. A greater growth occurs during the high-estrogen state of pregnancy, during which the glandular tissue become completely developed for milk production. All through pregnancy the ductal system of the breast grows and branches, the stroma increases in quantity, and large volume of fat is laid down in the stroma. Other hormones important for ductal system growth are growth hormone, prolactin, adrenal glucocorticoids, and insulin.

Progesterone plays the pivotal role for the final development of breasts into milk-secreting organs. Progesterone, acting synergically with the other hormones mentioned above, promotes additional growth of breast lobules, with budding of alveoli and development of secretory characteristics in the cells of the alveoli [5].

1.3 BREAST CANCER

Breast cancer is the most frequently diagnosed cause of death from cancer in women world-wide, the second leading cause of death from cancer in women in developed countries, and the leading cause of death from cancer in low and middle-income countries, where a high proportion of women presents with advanced disease, which leads to a poor prognosis. Established risk factors for breast cancer include age, family, or personal history of breast cancer or of precancerous lesions, reproductive factors, hormone treatment, alcohol consumption, obesity (for postmenopausal breast cancer only), exposure to ionizing radiation, and genetic predisposition. Screening for breast cancer aims to reduce mortality from this cancer, as well as the morbidity associated with advanced stages of the disease, through early detection in asymptomatic women. The key to achieving the greatest potential effects from this screening is providing early access to effective diagnostic and treatment services. Comprehensive quality assurance is essential to maintaining an appropriate balance between benefits and harms [6].

1.3.1 EPIDEMIOLOGY

Excluding skin cancer, breast cancer is the most frequently diagnosed cancer in women, a population where about one out of three malignant tumors (30%) is breast cancer. Considering the frequencies in the different age groups, breast cancer represents the most frequently diagnosed tumor among women in the 0-49 age group (41%), in the 50-69 years age group (35%), and in the older 70+ years age group (22%). The trend of incidence rate for breast cancer in Italy appears to be slightly increasing (+0.3% per year). The differences between macro-areas observed between 2010 and 2014, which confirm a higher incidence in Northern Italy (162.9 cases/100.000 women) than in Central Italy (141.5 cases/100.000 women) and in Southern Italy & Islands (127.1 cases/100.000 women), result from the combination of the various factors involved, from the varying implementation of screening mammography to the uneven distribution of the risk factors) [8].

In 2020 are estimated 12.500 deaths [9]. The breast cancer mortality rate is about 13,2/100.000, although there is a continuing downward trend in mortality (-6% from 2015

to 2020 [10]) attributable to the wider availability of early detection programs and therefore early diagnosis, and to therapeutic progress.

The 5-year survival rate of women with breast cancer in Italy is 87%. This value does not differ much across different age groups: 91.4% in women aged 15-44, 93.9% among women aged 45-54, 93.2% among women aged 55-64, and 91.7% in 65-74 aged women, while it is slightly lower among elderly women (76.1%) [9].

On the whole there are 834.200 women in Italy who have been diagnosed with breast cancer, equal to 43% of all women living with a prior cancer diagnosis and 23% of all prevalent cases (men and women) [10].

1.3.2 RISK FACTORS AND PREVENTION

The risk of developing breast cancer increases with age, with the probability of developing breast cancer being 2.3% up to 49 years (1 in 43 women), 5.4% in the 50-69 age group (1 in 18 women), and 4.5% in the 70-84 age group (1 in 22). This correlation with age could be linked to the continuous and progressive endocrine proliferative stimulus that the mammary epithelium undergoes over the years, together with the progressive damage to DNA and accumulation of epigenetic changes that modifies the balance in the expression of oncogenes and tumor suppressor genes. The incidence curve increases exponentially until the age of menopause (around 50-55 years), and then slows down reaching a plateau after menopause, to subsequently increase again after 60 years of age. This specific trend is linked both to the endocrinological history of a woman, and the availability and coverage of mammography screening programs.

Other risk factors that have been identified are:

- Reproductive factors: a long fertile period, with early menarche and late menopause, and therefore a longer exposure of the glandular epithelium to the proliferative stimuli of ovarian estrogens; nullity, first full-term pregnancy after 30, no breastfeeding.
- Hormonal factors: increased risk in women taking hormone replacement therapy during menopause, especially if based on synthetic estrogen-progestins with androgenic activity; increased risk in women taking oral contraceptives.
- Dietary and metabolic factors: high consumption of alcohol and animal fats and low consumption of vegetable fibers seem to be associated with an increased risk of breast cancer. Diet and behavior leading to obesity and metabolic syndrome are also increasingly important. Obesity is a recognized risk factor, probably linked to the excess of fat tissue that in postmenopausal women is the main source of synthesis of

circulating estrogen, resulting in excessive hormone stimulation of the mammary gland. Metabolic syndrome also increased the risk of breast cancer: it is hypothesized that subjects with metabolic syndrome show a resistance to insulin, which act on the membrane receptor of insulin-like growth factor 1 (IGF-1R), activating the intracellular signal pathways essential for neoplastic growth. This syndrome is based on genetic predisposition, but its development is clearly favored by sedentary lifestyles and high-calorie diets rich in fats and simple carbohydrates. Consequently, by acting on these modifiable risk factors through regular daily physical activity combined with a balanced diet the risk of developing breast cancer could be reduced by improving the metabolic hormonal balance of the woman.

- Prior radiotherapy (of the chest, and especially before 30 years of age) and prior breast dysplasia or neoplasm.
- Familiarity and heredity: although most breast cancers are sporadic forms, 5-7% are linked to hereditary factors, ¼ of which are determined by the mutation of two genes: BRCA-1 and BRCA-2. In women with BRCA-1 mutations the life-time risk of breast cancer is 65%, while in women with BRCA-2 mutations it is 40%. Other hereditary factors are represented by: mutations of the ATM gene (Ataxia Telangiectasia Mutated) or CHEK2 gene (checkpoint kinase 2 gene), whose protein plays a critical role in the DNA damage checkpoint; mutation of the PALB2 gene, that encodes a protein involved in genome maintenance (double strand break repair); Li-Fraumeni Syndrome (p53 mutation); Cowden Syndrome (PTEN gene mutation); Peutz-Jeghers syndrome.

1.3.3 HISTOLOGIC CLASSIFICATION [6][10][11]

The 2019 World Health Organization's new classification of breast tumors provides for different histological types of breast cancer (Table I).

This classification has got some differences from the previous classification (WHO 2012).

1) The prognostic value of TILs (Tumor Infiltrating Lymphocytes) is now recognized.
2) As regards the histologic grade is recommended to report mitotic number per mm^2.
3) The breast carcinoma with medullary features is considered as NST carcinoma rich in lymphocytes, and not yet as a separate entity.
4) The group of neuroendocrine neoplasms includes neuroendocrine tumor (NET) and neuroendocrine carcinoma (NEC). NET is a lesion with morphological features that are typical of a neuroendocrine proliferation: nest or trabecular growth, diffuse neuroendocrine markers expression, low grade (G1, G2). These lesions are usually

characterized by the estrogen receptor expression, low Ki67, and HER2-. Their management does not differ from that of the other carcinomas. NEC is a poorly differentiated neoplastic lesion (G3). It is possible to differ the small cell neuroendocrine carcinoma from the large cell neuroendocrine carcinoma. A lesion is classified as “NST with neuroendocrine differentiation” when it does not present the morphological characteristic of the previous histological types.

5) Two new histological types are now included in WHO classification: mucinous cystoadenocarcinoma and tall cell carcinoma with reversed polarity. Even if there are only few cases of these histologic types of breast cancer, they are both triple negative tumors with a good prognosis.
6) The invasive micropapillary carcinoma, that is a variant of the mucinous carcinoma, should appear at a young age and it presents more frequent neoplastic lymphovascular embolization and metastatic lymph nodes, when it is compared with classic mucinous carcinoma.

Table I: Breast cancer histologic classification – WHO 2019.

HISTOLOGIC CLASSIFICATION OF BREAST CANCER – WHO 2019
EPITHELIAL TUMORS OF THE BREAST Invasive breast carcinoma of no special type/Invasive ductal carcinoma Invasive lobular carcinoma Tubular carcinoma Cribriform carcinoma Mucinous carcinoma Mucinous cystoadenocarcinoma Invasive micropapillary carcinoma Adenocarcinoma with apocrine differentiation Metaplastic carcinoma

Rare tumors salivary glands-like:

Acinic cell carcinoma

Adenoid cystic carcinoma

Secretory carcinoma

Mucoepidermoid carcinoma

Polymorphous adenocarcinoma

Tall cell carcinoma with reversed polarity

Neuroendocrine neoplasms:

Neuroendocrine tumor (NET)

Neuroendocrine carcinoma (NEC)

Benign epithelial proliferations and precursors:

Usual ductal hyperplasia

Columnar cells lesion

Atypical ductal hyperplasia

Benign adenosis and sclerosant lesions:

Sclerosant adenosis

Adenoma with apocrine differentiation

Microglandular adenosis

Radial scar/complex sclerosing lesion

Adenomas:

Tubular adenoma

Lactating adenoma

Ductal adenoma

Epithelial-myoepithelial tumors:

Pleomorphic adenoma

Adenomyoepithelioma

Adenomyoepithelioma with carcinoma

Epithelial-myoepithelial carcinoma

Papillary neoplasms: Intraductal papilloma Ductal carcinoma in situ Incapsulated papillary carcinoma Incapsulated papillary carcinoma with invasion Solid-papillary carcinoma in situ Solid-papillary carcinoma with invasion Intraductal papillary adenocarcinoma with invasion Non-invasive lobular neoplasia: Atypical lobular hyperplasia Lobular carcinoma in situ Classic lobular carcinoma in situ Florid lobular carcinoma in situ Pleomorphic lobular carcinoma in situ Ductal carcinoma in situ (DCIS): Intraductal carcinoma, non-invasive, NAS, DCIS of low nuclear grade DCIS of intermediate nuclear grade DCIS of high nuclear grade
FIBROEPITHELIAL TUMORS AND BREAST HAMARTOMAS Hamartoma Fibroadenoma Phyllodes tumors
NIPPLE TUMORS Nipple adenoma Syringomatous tumor Paget disease
MESENCHIMAL TUMORS OF THE BREAST Vascular tumors

Fibroblastic and myofibroblastic tumors Peripheral nerves tumors Smooth muscle tumor Adipose tissue tumors Other mesenchymal tumors and tumor-like lesions
LYMPHOEMATOPOIETIC TUMORS OF THE BREAST Lymphomas
MALE BREAST CANCER Invasive carcinoma Carcinoma in situ
METASTATIC CANCER
GENETIC SYNDROMES ASSOCIATED WITH BREAST CANCER

1.3.4 MOLECULAR CLASSIFICATION [10]

Breast cancer is an heterogenous disease and patients with breast cancer characterized by similar clinicopathological features should have a different clinical course.

It is possible to identify four invasive breast cancer subtypes through gene expression analyses:

- Luminal A: breast cancer with high hormonal receptors expression.
- Luminal B: breast cancer with hormonal receptor expression, high recurrence risk, higher proliferative rate, and high proliferation genes expression.
- HER2-enriched: breast cancer characterized by HER2 expression.
- Basal like: breast cancer without hormonal receptor and HER2 expression, and with an elevated basal cytokeratin expression (CK5/6, CK14).

As regards prognosis, Luminal A subtype has got a favorable prognosis, better than that of Luminal B subtype, while HER2 positive and Basal like breast cancer are characterized by the worse prognosis.

Thanks to further pathological and molecular knowledge, other breast cancer subtypes are now identified. One of them is the “claudin low” subtype, which has not hormonal receptor and HER2 expression, and it is defined by stem cell markers expression, low claudin expression (that are proteins involved into cell-to-cell junction) and the presence of lymphocytic infiltrate, which is correlated with cancer growth. Its prognosis is bad.

Other breast cancer subtypes with various molecular biology and clinical course are basal like 1 and 2 (BL1 and BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem-like (MSL) and luminal androgen receptor (LAR). Two of them, immunomodulatory and mesenchymal stem-like subtypes, are defined by lymphocytic/inflammatory infiltrate and not by different gene expression of tumor cells.

Hormonal receptors, Ki67 and HER2 immunohistochemical determination allows to identify the four phenotypical groups of breast cancer. Such determined groups present a relative parallel with the four different gene expression profiles. The immunophenotypically groups with a clinical and a therapeutical significance, also as regards the adjuvant therapy, are:

- Luminal A: hormonal receptors positive (positivity rate >20%), HER2 negative, low proliferative rate (Ki67 ≤14%, cut-off 20%).
- Luminal B/HER2 negative: hormonal receptors positive, HER2 negative, high proliferative rate.
- Luminal B/HER2 positive: hormonal receptors positive, HER2 overexpressed (IHC score 3+) or amplified (FISH or other ISH methods), any proliferative rate.
- HER2 positive (non-luminal): HER2 overexpressed (IHC score 3+) or amplified (FISH or other ISH methods) and both of hormonal receptor negative.
- Triple negative: absence of hormonal receptors expression and HER2 negative. The parallelism between the “triple negative” phenotype, identified by IHC, and the “basal like” subtype, defined on the basis of gene expression analyses, exists only in the 80% of cases. Some particular histotypes are included into the “triple negative” subtype, such as the typical medullary and the adenoid-cystic, with a low recurrence risk.

Nowadays, some tests for genetic profiles analyses (Prosigma®, Mamaprint®, Oncotype DX®), that consider a restricted number of genes, can be performed in some countries to assess the breast cancer genetic profile. In combination with the other clinic-pathological,

radiological, and molecular indexes, and patient's comorbidity and preferences, molecular tests allow to get the so-called personalized medicine.

These tests are suggested to select patients with hormonal receptors positive and HER2 negative breast cancer, that can derive benefit form an adjuvant chemotherapeutical treatment added to endocrine therapy. Vice versa they are not recommended in HER2 positive or triple negative breast cancer assessment.

Molecular tests' role is to be pro or against adjuvant chemotherapy. They don't play any role in patients that cannot receive any adjuvant chemotherapy, because of comorbidities or tumor risk or stage, or in patients who need chemotherapy (stage III). On the contrary, they are useful in breast cancers that are 1-3 cm in the greatest diameter, with no or 1-3 metastatic lymph nodes, and intermediate proliferative rate. In any case molecular tests aren't the only factor that is necessary to considerate to decide about chemotherapeutic treatment.

According to the Italian decree that was published in the G.U. on 07/07/2021, genomic tests are allowed for uncertain cases of patients with early-stage breast cancer (stage I-IIIA) ER+/HER2- (identified by clinical, histopathological, and radiological stratification), when further definitions of possible benefits of an adjuvant post-operative chemotherapy added to hormonal therapy are necessary [12]. On the contrary, genomic tests aren't recommended for patients with early-stage breast cancer (stage I-IIIA) ER+/HER2-, who are identified as patients with a low risk of recurrence, or patients with high risk of recurrence.

- Breast cancer with a low risk of recurrence is defined by five features: G1, T1 (a-b), Ki67 <20%, ER>80%, N0. In this case the recommended treatment is hormonal therapy without adjuvant post-operative chemotherapy.
- Breast cancer with a high risk of recurrence is defined by at least four of these features: G3, T3-T4, Ki67>30%, ER<30%, N+ (no indications for testing if >3 positive lymph nodes). In this case the recommended treatment is hormonal therapy combined with adjuvant post-operative chemotherapy.

The genomic test is performed on the breast cancer sample obtained from the surgical resection, after formalin fixation and paraffine inclusion [12].

1.3.5 BREAST STAGING - TNM CLASSIFICATION AND STAGING WORKFLOW

The eighth edition of the TNM (tumor, regional nodes, distant metastases) staging system adds biological markers to modify anatomic staging.

There are five stages of breast cancer, each one is determined by the combination of three parameters:

- T: tumor size (cm)
- N: nodes with cancer (number of nodes)
- M: distant metastases (0= no spread; 1= it has spread)

The different stages are [10]:

Stage 0: noninvasive cancer (Tis, N0, M0). The disease is presented only in the ducts and lobules of the breast. There isn't any spread in the surrounding tissue.

Stage I: invasive cancer. The lesion also involves the normal breast tissue. There are two types. 1) Stage IA (T1, N0, M0): the tumor is ≤2cm. It has not spread to the lymph nodes. 2) Stage IB (T0/T1, N1 mi, M0): the tumor is in the breast, and it is ≤2cm; or there is no tumor in the breast tissue, and it is in the lymph nodes of the breast.

Stage II: invasive cancer. 1) Stage IIA (T0/T1, N1, M0; T2, N0, M0): tumor may not be found in the breast, but to 1 to 3 lymph nodes are involved because of cancer cells spreading; or tumor of 2-5 cm is detected in the breast with or without spread to the axillary lymph nodes. 2) Stage IIB (T2, N1, M0; T3, N0, M0): tumor is 2-5cm, and it has spread to 1-3 axillary lymph nodes; or the tumor is >5 cm but there isn't any axillary lymph node spreading.

Stage III: invasive breast cancer. 1) Stage IIIA (T0/T1/T2, N2, M0; T3, N1/N2, M0): tumor is in the breast (any size), or no tumor is found in the breast but is in the lymph nodes. More than 4 lymph nodes are involved in the breast or axilla. It has not spread to other parts of the body. 2) Stage IIIB (T4, N0/N1/N2, M0): tumor may be any size and the chest wall is involved by the disease. It may cause swelling of the breast and may be in up to 9 lymph nodes. Inflammatory breast cancer is considered Stage IIIB. 3) Stage IIIC (any T, N3, M0): there may be no sign of cancer in the breast, or a tumor may be any size and may have spread to the chest wall or breast skin. 10 or more axillary lymph nodes are involved, or nodes above or below the collarbone or breastbone.

Stage IV (any T, any N, M1): metastatic breast cancer. The tumor can be any size and the disease has spread to other organs and tissues (bones, lungs, brain, liver, distant lymph nodes, or chest wall).

1.3.5.1 CLINICAL CLASSIFICATION (AJCC 2017 CLASSIFICATION)

Primary tumor (T):

Tx: primary tumor cannot be assessed

T0: no evidence of primary tumor

Tis: carcinoma in situ. Tis (DCIS): ductal carcinoma in situ; Tis (Paget): Paget disease of the nipple not associated with invasive and/or carcinoma in situ in the underlying parenchyma of the breast.

T1: tumor up to 20mm in the greatest diameter. T1mi: microinvasion≤1mm; T1a: 1-4.9mm; T1b: 5-9.9mm; T1c: 10-19.9mm.

T2: tumor >20mm but <50mm in the greatest diameter.

T3: tumor >50mm in the greatest diameter.

T4: tumor of any size with direct extension to the chest wall and/or to the skin (ulceration of the skin or nodules). T4a: extension to the chest wall (excluding pectoral muscle involvement); T4b: ulceration and/or ipsilateral macroscopic satellite nodules and/or edema (including peau d'orange) of the skin that doesn't meet the criteria for inflammatory carcinoma; T4c: both T4a and T4b; T4d: inflammatory carcinoma, which is primarily a clinical diagnosis. Inflammatory carcinoma is the most malignant form of breast cancer, even if it represents less than 3% of all cases. Clinical findings consist of a rapidly growing, sometimes painful mass that involves and enlarges the breast. It is characterized by typical skin changes involving one third or more of the breast skin. Skin changes may be due to carcinomatous invasion of the subdermal lymphatics, with resulting lymphoedema and hyperemia [13].

Regional lymph nodes (N):

Nx: lymph nodes involvement cannot be assessed.

N0: no regional lymph nodes metastases (instrumental and clinical examination).

N1: metastases to movable ipsilateral axillary lymph nodes (level I-II).

N2: metastases to movable ipsilateral axillary lymph nodes (level I-II) that are clinically fixed or matted; or in clinically detectable ipsilateral internal mammary nodes in the absence of clinically evident metastases in axillary nodes. 1) N2a: metastases in ipsilateral

axillary nodes fixed to other structures; 2) N2b: metastases only ipsilateral internal mammary nodes and no metastasis in ipsilateral axillary nodes.

N3: metastases in one or more ipsilateral subclavicular lymph nodes (level III axillary) with or without level I, II axillary nodes involvement; or ipsilateral internal mammary nodes with level I, II metastases; or metastases in one or more ipsilateral supraclavicular nodes with or without axillary or internal mammary lymph nodes involvement. 1) N3a: ipsilateral subclavicular lymph nodes involvement; 2) N3b: internal mammary and axillary lymph nodes involvement; 3) N3c: supraclavicular lymph nodes involvement.

Distant metastasis (M):

Mx: distant metastases cannot be verified.

M0: no clinical or radiological evidence of distant metastases.

cM0(i+): no clinical or radiological evidence of distant metastases, but deposits of tumor cells detected by molecular biology or microscopically in blood, bone marrow or other tissue other than regional lymph nodes, not exceeding 0.2mm in absence of signs or symptoms of metastases.

M1: distant metastases detected by typical clinical and radiological investigations and/or histologically proven metastases greater than 0.2mm (pM).

1.3.5.2 PATHOLOGICAL CLASSIFICATION (AJCC 2017 CLASSIFICATION)

Primary tumor (pT): the pathological classification of the primary tumor corresponds to the clinical classification.

Regional lymph nodes (pN)

pNx: regional lymph nodes cannot be assessed.

pN0: no regional lymph node metastases identified or isolated tumor cells (ITCs) only. Isolated tumor cells are defined as small aggregates of cells no larger than 0.2mm or single tumor cells or a small grouping of cells with less than 200 cells in a single histological section. These cells can be detected by traditional histological methods or by immunohistochemical methods. Nodes containing only ICTs are excluded from the total count of positive lymph nodes for N classification purposes, but should be included in the total number of nodes tested.

- pN0 (i-): no regional lymph nodes metastases by histology (EE staining), negative by immunohistochemistry.

- pN0 (i+): presence of ITCs in regional lymph nodes not exceeding 0.2mm (detected by EE staining or immunohistochemistry).
- pN0 (mol): no regional lymph nodes metastases by histology, RT-PCR (real-time polymerase chain reaction) negative.
- pN0 (mol+): RT-PCR positive but no regional lymph nodes metastases by histology or immunohistochemistry; no ITCs are detected.

pN1: micrometastases or metastases in 1-3 ipsilateral axillary lymph nodes; and/or micrometastases or macrometastases in ipsilateral internal mammary nodes detected by sentinel node biopsy, but not clinically detectable.

- pN1mi: micrometastases (aggregate of contiguous tumor cells larger than 0.2mm and/or more than 200 cells, but no larger than 2mm).
- pN1a: metastases in 1-3 ipsilateral axillary lymph nodes, including at least one metastasis larger than 2mm in greatest dimension.
- pN1b: metastases in internal mammary lymph nodes, excluding ITCs.
- pN1c: combination of pN1a and pN1b.

pN2: metastases in 4-9 ipsilateral axillary lymph nodes; and/or ipsilateral internal mammary lymph nodes at instrumental examinations with no axillary lymph nodes metastases.

- pN2a: metastases in 4-9 axillary lymph nodes, including at least one of them larger than 2mm in greatest dimension.
- pN2b: clinically detectable metastases in internal mammary lymph nodes, with or without histological confirmation, without axillary lymph nodes metastases.

pN3: metastases in 10 or more ipsilateral axillary lymph nodes; or in ipsilateral subclavicular (level III axillary) lymph nodes; or metastases in ipsilateral internal mammary lymph nodes confirmed by imaging with metastases in one or more level I-II axillary lymph nodes; or metastases in more than 3 axillary lymph nodes and internal mammary lymph nodes, with microscopic or macroscopic metastases shown by sentinel lymph node biopsy, but not clinically detectable; or metastases in ipsilateral supraclavicular lymph nodes.

- pN3a: metastases in 10 or more ipsilateral axillary lymph nodes (at least one larger than 2 mm in greatest dimension); or metastases in subclavicular lymph nodes (level III axillary).
- pN3b: pN1a or pN2a in the presence of cN2b (ipsilateral internal mammary lymph nodes positive by imaging), or pN2a in the presence of pN1b.
- pN3c: metastases in ipsilateral supraclavicular lymph nodes.

The eighth edition of AJCC classification recommends considering information about prognostic factors such as histological grading, hormonal receptor status (ER and PgR) and HER2 status, because they can influence staging in the prognostic version of staging itself. All these factors are now included in the Prognostic Classification, making the staging system more precise, flexible, and personalized.

1.3.6 PROGNOSTIC AND PREDICTIVE FACTORS

Prognostic factors are related to patient's prognosis (survival), while predictive factors to the probability of the efficacy of an antitumoral treatment.

Some important prognostic factors are:

- Tumor size
- Axillary lymph nodes involvement. The presence of ITCs or micrometastasis seems not to be significant.
- Histological grading: a high histological grade (G3) is considered an unfavorable prognostic factor. An intermediate histological grade could be reclassified as G3 or molecular G1 analyzing its gene profile.
- Proliferation (Ki67): proliferative activity is measured by the Ki67 labeling index (percentage of tumor cell nuclei that are stained with Mib1 protein antibody encoded by the KI67 gene) is now a recognized prognostic factor.
- Histological type: tubular, cribriform, adenoid cystic and histotypes have a favorable prognosis. As regards lobular carcinoma there are different histological variants, and the classic variant has the better prognosis. Lobular carcinoma is typically multifocal and presents distinctive metastatic sites (gastrointestinal tract, ovary, serous membranes).
- Vascular invasion
- HER2 status: it is an established prognostic factor and predictive factor for HER2 drugs response, and for hormone therapy resistance. HER2 status can be assessed using immunohistochemistry (IHC), which determines the expression of HER2 receptor, and in situ hybridization (ISH), which measures gene amplification. Tumor is defined HER2 positive if a 3+ score emerged from immunohistochemistry, or if ISH method highlights gene amplification. If immunohistochemistry score is 2+, it is important to investigate gene amplification. Recently, new anti-HER2 strategies, such as Trastuzumab Deruxtecan and Trazstuzumab Duocarmazine, have been identified as promising

therapies for low-HER2 breast cancer. Low-HER2 breast cancer is defined as: 2+ IHC score, without HER2 gene amplification, or 1+ IHC score (ISH not performed or negative).

- Hormonal receptors status: it is important to determine both estrogen and progestin receptor status and report the percentage of positive cells. Hormonal receptor status is defined positive with at least 1% positive invasive tumor nuclei in the sample on immunohistochemical testing [14]. A relationship between receptors' level and response rate to hormonal treatments exists in both metastatic disease and in adjuvant and neoadjuvant setting. Tumors with high levels of receptors are the most likely to benefit from hormonal treatment. In any case, other factors may influence the hormone responsiveness of breast cancer, such as HER2 status, histological grade and Ki67. It is possible to distinguish ER-low tumors (ER levels between 1% and 9%), that are <2% of all ER-positive tumors, and ER positive tumors, whose expression levels are >10%. ER-low tumors have a better prognosis than the ER-positive tumors, they often present a basal-like genomic, and have a neoadjuvant therapy response like that of triple negative tumors.
- Multifocality, the presence of several tumoral foci divided by normal parenchyma. "Satellite nodes" of the primary node are defined as lesions located less than 5mm from it and separated by normal parenchyma. It is important to sign the number of invasion foci on the diagnostic report. Multifocality has got a substantial impact on lymph nodes metastasis, increased local recurrence, and increased risk of cancer-related death. "Diffuse carcinoma" is a new term, that has been coined to indicate a tumor with a lobular growth pattern and with the involvement of one or more quadrants. These tumors are often not easy to identify in radiology and ultrasound scans, and, together with multifocal growth, they carry a high risk of disease-related death (4.14, for multifocal growth, and 2.75, for diffuse growth, times higher), regardless of the tumor immunophenotype.
- Intratumoral lymphocytes (TILs): breast carcinomas with a high intratumoral stromal lymphocytic infiltrate have a better prognosis than those with lymphocytic depletion. The highest degree of TILs is shown by triple negative and HER2-positive breast cancers.
- Age (worse prognosis if <35).
- Gene expression profiles (Luminal A, Luminal B, HER2-enriched, Basal-Like).

- Genomic classificators.

Predictive factors of target therapy response are:

- HER2 and hormonal receptor status
- PD-L1: percentage of tumor size with PD-L1 positive immune cells infiltrating tumor ≥1%.
- PIK3CA mutation
- BRCA1 and BRCA2 germline mutation.

1.3.7 TREATMENT

Most patients with early breast cancer can be treated and cured. Treatment with a curative intent is recommended for clinical stage I, II, and III disease. A multimodality strategy should be undertaken for locally advanced (T3, T4) breast cancer and even for inflammatory tumors. Palliation becomes the preferred strategy for metastatic disease and for unresectable local cancers [15].

DCIS TREATMENT

DCIS represents 25% of all treated breast cancers. The most common surgical treatment for DCIS is the conservative surgery with a wide excision with clean margins. Post-operative radiotherapy is recommended for patients with intermediate/high grade DCIS. Sometimes mastectomy is preferred

Systemic therapy with tamoxifen can be considered in patients with ER+ DCIS, after conservative surgery and radiotherapy.

Lobular in situ neoplasia/lobular in situ carcinoma is a risk factor and a non-obligated precursor for the breast cancer development (the relative risk is 9-10 times higher than that of the general population). Surgical treatment is not recommended for the lobular neoplasia. After the diagnosis of lobular neoplasia, the possible options are:

1) Surveillance: clinical examination each 6-12 months and annual mammography.
2) Chemoprevention.
3) Prophylactic bilateral mastectomy, only after multidisciplinary evaluation.

As regards pleomorphic LN, the surgical excision of the lesion is recommended in this case, because of its high risk of upgrade to invasive cancer at the definitive histology (20-36% of cases).

TREATMENT OF OPERABLE INVASIVE CARCINOMA

A. Locoregional treatments

Conservative surgery associated with breast radiotherapy (whole breast irradiation) is the first choice for invasive breast cancer stage I-II, and only selected more advanced lesions. Mastectomy is allowed when conservative surgery is not practicable.

Surgical treatment should be performed within 4-6 weeks after the diagnostic biopsy or after taking charge the patient.

After the mastectomy, breast reconstruction is desirable as it improve patients' quality of life, it isn't associated with higher loco-regional recurrence, and it doesn't interfere with their diagnosis. Breast reconstruction can use autologous tissues or prosthetic implants.

Skin sparing mastectomy and nipple sparing mastectomy are considered appropriate when reconstruction is immediate and the esthetic and psychological advantages are manifest. The nipple sparing approach is considered acceptable only when resection margins near to the nipple aren't involved.

Sentinel lymph node biopsy is the therapeutical standard for patients with breast cancer at stage I-II and clinically non metastatic lymph nodes or clinically suspicious lymph nodes but with negative cytology. It is performed also in multifocal breast cancer and in case of previous breast/axilla surgery.

Axillary dissection with removal of at least 10 lymph nodes is recommended:

- In presence of clinically pathological axillary lymph nodes confirmed by preoperative cytological and micro-histological testing.
- In presence of positive sentinel lymph node, with macro-metastasis at the histological examination, in selected cases after multidisciplinary discussion.
- If the sentinel lymph node couldn't be detected.
- In case of T4 breast lesions and inflammatory breast cancer.

Axillary dissection usually includes level I and II of the lymph node axis. Level III dissection is performed only in case of level II or III macroscopic pathological involvement.

Selective dissection is an alternative to the traditional axillary dissection. It can be used for N1 breast cancers, and it reduces post-surgical lymphoedema incidence.

As regards radiotherapy, whole breast irradiation is considered in most of the patients with breast cancer after conservative surgery because it reduces the local recurrence risk and the mortality rate.

As regards timing, radiotherapy should be started after healing of the surgical wound and within 8-20 weeks. In case of adjuvant chemotherapy, radiotherapy should be started within 4-6 weeks after the end of the chemotherapy.
In patients with locally advanced breast cancer (T3-4, N0/N1; any T, N2-3) radiotherapy can be performed also when surgical treatment is not possible.

B. Adjuvant systemic treatments

Adjuvant systemic treatment after the surgical approach has an important role. In fact, it can reduce the risk of recurrence and death.
It includes polychemotherapy, hormonal therapy and biological therapy (trastuzumab).

a) Adjuvant hormonal therapy

It reduces recurrence and death risk in patients with hormone-responsive breast cancer, that is defined by ER and/or PgR immunohistochemical expression at least in 1% of the tumor cells. The specific choice about hormonal treatment is guided by the menopausal status and the recurrence risk of the patient. It is suggested to assess the serial determination of FSH/estradiol to determine the menopausal status of the patient. The standard duration of this therapy is at least five years.
Women in a premenopausal status can receive a hormonal adjuvant therapy based on tamoxifen 20mg/os/die +/- ovarian suppression. Ovarian suppression is added in function of patient's recurrence risk and desires. It is possible to add an aromatase inhibitor in premenopausal patients with high risk of recurrence and treated with tamoxifen + ovarian suppression. Ovarian suppression can be reached through bilateral ovariectomy, ovarian irradiation, or LHRH analogues (intramuscular administration for 28 days). Women in post-menopausal status can receive a hormonal adjuvant therapy based on aromatase inhibitors.

b) Adjuvant chemotherapy

It should be started after postoperative course is ended, within 4-8 weeks after surgery and no later than 90 days. The optimal duration of this therapy is from 4 to 8 cycles.
Polychemotherapy strategies should be preferred rather than monochemotherapy.
First generation regimens are based on the combination of cyclophosphamide, methotrexate, and fluorouracil (CMF). They are not so used today.
Second generation regimens contain anthracyclines, and on average they are more effective than CMF-like regimens. These regimens are burdened by increased acute and delayed toxicity, that includes the development of congestive heart failure and acute myeloid leukemia.

Third generation regimens include regimens containing anthracyclines and taxanes administered sequentially or in combination.

c) Adjuvant anti-HER2 therapy

The monoclonal antibody trastuzumab can be added to chemotherapy in patient with HER2 + operated breast cancer, who are candidates for adjuvant therapy. Trastuzumab is a recombinant monoclonal antibody for the extracellular domain of HER2 and it should be administered for one year. An increased risk of cardiotoxicity is associated with trastuzumab administration.

Thanks to immunohistochemical evaluation different subtypes of breast cancer can be identified. Adjuvant systemic therapy differs from one subtype to another.

1) Luminal A: low grade, good prognosis, high hormonal therapy sensitivity, and lower chemotherapy sensitivity. It can be treated with hormonal adjuvant therapy, with chemotherapy in selected cases.
2) Luminal B (HER2 -): more aggressive phenotype than Luminal A subtype. Luminal B tumors are often high-grade tumors with a worse prognosis. They can be treated with chemotherapy added to hormonal treatment.
3) Luminal B (HER2 +): the key therapy is chemotherapy+trastuzumab added to hormonal therapy.
4) HER2 + (non-luminal): the key therapy is chemotherapy+trastuzumab. It is usually adopted for tumor >1cm or with positive lymph nodes.
5) Triple negative: the key therapy is chemotherapy. It is recommended for tumors >1cm or with positive lymph nodes. Adjuvant chemotherapy can be considered for pT1a pN0/pN1 G3 or high Ki67 tumors.

NEOADJUVANT SYSTEMIC THERAPY

Neoadjuvant systemic therapy refers to the breast cancer systemic treatment administered before the potentially radical surgical procedure. Typically, this treatment is based on chemotherapy, possibly associated with biological drugs, although there is increasing interest in expanding the role of neoadjuvant endocrine therapy in some patients with endocrine-sensitive disease.

Before systemic neoadjuvant treatment starting, histopathologic diagnosis, with HER2 and hormone receptor status is recommended.

Neoadjuvant treatment strategies differ from one to another tumor subtype.

- HR+/HER2-: chemotherapy, or hormone therapy in selected cases.

- HER2+: chemotherapy associated with anti-HER2 agents.
- Triple negative: chemotherapy.

As regards endocrine therapy in neoadjuvant settings, tamoxifen is suggested for elderly patients with breast cancer. Clinical response can be reached in more than 30% of patients treated with tamoxifen-based neoadjuvant therapy. This therapy should be administered at least for 3-4 months.

Periodic revaluations should be performed in neoadjuvant-treated patients, during the period of treatment, to evaluate tumor response and to make sure that the tumor does not progress. For patients with operable tumors, if tumor progression is highlighted, it is recommended to perform surgical removal of the lesion. Patients with inoperable tumors should be tried a new chemotherapy strategy, with non-cross resistant drugs, with the intention of allowing successive breast surgery and/or radiation therapy.
MRI is sometimes used at the end of a neoadjuvant treatment for the definitive response evaluation and for the “exceptional responders” identification, who are associated with pCR in a high percentage of cases.

As regards breast surgery after neoadjuvant treatment, it is often a conservative surgery, and it needs a particular expertise and a multidisciplinary collaboration between radiologists, pathologists, oncologists, radiotherapists, and plastic surgeons. Breast surgery is usually performed within 3-8 weeks after the end of neoadjuvant treatment.

TREATMENT OF THE METASTATIC DISEASE

Most of the metastatic breast disease are identified during the post-treatment follow-up of a localized disease, and only 5% of metastatic disease is diagnosed de novo. After the diagnosis of metastatic breast cancer, it is appropriate to perform a complete restaging: clinical examination, laboratory tests, whole body imaging exams (TC, PET, bone scintigraphy), district radiological exams (TC, RM), metastatic lesion biopsy.
Biopsy of the metastatic mass should be considered because it may help to confirm the diagnosis of metastatic breast cancer, and it may distinguish non-malignant lesions or a new primary tumor. It also may show a change in the receptor status of metastases compared to primary tumors, providing useful information for the selection of the better systemic treatment.
Metastatic disease can be distinguished into two situations: with low short-term mortality risk (indolent disease) or with intermediate/high risk (aggressive disease).

Indolent disease is defined by a long disease-free interval (>24 months from the end of the adjuvant therapy), metastatic involvement of bones and/or soft tissues, a limited number of metastatic lesions (such as lung metastasis of small volume and limited number, or a limited hepatic involvement, <30%).

Aggressive disease usually occurs with visceral crisis, or a high number of metastases in multiple organs, or with a short disease-free interval (metastasis occurs during adjuvant treatment, or within 3 months from its end).

"Visceral crisis" is a state characterized by severe disfunction of an organ, with the possibility of rapid evolution and risk of imminent death, which requires treatments with rapid effect. Some typical conditions are diffuse pulmonary lymphangiomatosis, liver or respiratory failure, or neoplastic meningosis.

Treatment of metastatic breast cancer proposes to prolong survival, reduce, or delay symptoms appearance, and improve patients' quality of life. The therapeutical choice is based, primarily, on the biological features of the disease (hormone receptors and HER2 status), then on all the factors cited above.

Although the goal of treatment in metastatic disease is to control the disease, only a limited number of patients, those with oligometastatic disease, can reach long survival or recovery.

Bones are the first metastatic site in 20-30% of patients with breast cancer. More than 80% of patients who die from metastatic breast cancer have bone lesions. Treatment with bisphosphonates and denosumab should be considered if lytic or mixed bone metastases are diagnosed and require treatment in order to reduce the risk of adverse skeletal events or in case of hypercalcemia.

In case of diffuse bone lesions, radiometabolic therapy with Strontium-89(^{89}Sr)^{102} plays a possible role in pain palliation, nevertheless frequent severe side effects such as leukopenia and thrombocytopenia.

Surgery is indicated in selected cases of metastatic breast cancer. In selected cases, such as ER+/PR+ or HER2+ tumors stage IV de novo, adding surgery to systemic treatment can be advantageous. Surgery can be also indicated in case of:

- Vertebral metastases with spinal compression.
- Single visceral metastasis (liver, lung, brain).
- Pathological fractures.
- Pleural or pericardial effusions.

Radiation therapy plays an important role in the palliative care and in oncological emergencies treatment (for example spinal cord compression) of metastatic disease. Its purpose is to improve patient's symptoms and quality of life. It is frequently used in association with systemic therapies.

1.3.13 GENETIC COUNSELING

About 18% of all breast cancers are due to familiarity, while 13% are due to hereditary predisposition. BRCA1 and BRCA2 are the two main genes involved in breast cancer hereditary predisposition. An increased risk of developing breast and/or ovarian cancer is related to mutation of these two genes. BRCA2 gene's mutations also determined an increased risk of male breast cancer. Mutations of other genes are also responsible for an increased risk of breast cancer developing.

It is considered appropriate to suggest oncogenetic counseling to a person who meets at least one of these criteria:

Personal history of:

a) Known pathogenetic variant in a predisposing gene in a family member.
b) Male with breast cancer.
c) Woman with breast and ovarian cancer.
d) Woman with breast cancer <36 years.
e) Woman with triple negative breast cancer <60 years.
f) Woman with high-grade serous ovarian cancer at any age.

Personal history of breast cancer <50 and at least one first degree relative with:

a) Breast cancer <50 years.
b) Non-mucinous or borderline ovarian carcinoma at any age.
c) Bilateral breast cancer.
d) Male breast cancer.
e) Pancreatic cancer.
f) Prostate cancer.

Personal history of breast cancer >50 and breast, ovarian, pancreatic cancer familiarity in 2 or more relatives who have a first-degree relation with each other (including one who has first-degree relation with her).

Genetic test must first be carried out on a family member who has already developed the disease (index case), after appropriate genetic counseling. Only when test allows to obtain an informative result, it can be extended to other family members, from the age of 18.

The only approach that has proven effective in significantly reducing the risk of developing breast cancer is prophylactic surgery, in particular bilateral prophylactic mastectomy.

Even if the risk of developing ovarian cancer in BRCA-mutated woman is lower than the risk of developing breast cancer, bilateral prophylactic adnexectomy should be considered because of the lack of reliable methods for early diagnosis and the poor prognosis of ovarian cancer diagnosed at an advanced stage. Prophylactic salpingo-oophorectomy should be offered as a risk reduction option to all women with mutations of BRCA1 and BRCA2 genes from the age of 35-40 years, after satisfying their wish to have children.

2.0 SINGLE CELL TRANSCRIPTOMIC ANALYSES

2.1 PERSPECTIVES OF SINGLE CELL TRANSCRIPTOME ANALYSES

Breast cancer, and cancer in general, is associated with aberrations in gene expression related to both cell-intrinsic and extrinsic factors. Pivotal cancer features, such as tumor growth, metastatic involvement, and treatment resistance, that are guided by subsets of tumor cells, are governed by intra-tumor heterogeneity. Three are the determinants, that drives intra-tumor heterogeneity: a) genetic heterogeneity, due to subclonal mutations that appear during tumor evolution; b) epigenetic modulation and differentiation programs; c) environmental cues, or rather extrinsic and spatial factors.

Transcriptome study allow to investigate the integration of genetic, epigenetic and environmental stimuli of the cells. As regards cancer transcriptome analyses, it can show cancer cell fitness, behavior, and response to treatments.

Single-cell RNA sequencing (scRNA-seq) techniques enables gene expression analyses at a single cell level, investigating the transcriptional output of cells in both normal and tumoral tissue samples. Thanks to the identification of preferentially expressed genes, gene expression study permits to identify not only different “cell types” (malignant cells, epithelial cells, fibroblasts, etc.), but also various “cell states” (progression along the cell cycle, different metabolic states), or rather the diversity within each of the clusters defined as “cell types”.

The key of single-cell analyses is that it allows to know intra-tumor heterogeneity and intra-tumor expression programs, that can be found within numerous patients of a particular cancer type, and not in other cancer types. This characteristic reflects pivotal properties of the corresponding cells from which cancer developed. The expression of different associated genes, related to intra-tumor heterogeneity, may impact on important aspects of tumor progression, on cancer cells properties and vulnerabilities: drug resistance, metastasis, self-renewal [16].

Tumor progression is an evolutionary process. One of the purposes of single-cell transcriptomic analyses is to understand cancer initiation and evolution, thanks to the characterization of the different cell subpopulations. It is known that only few cells are sufficient for the emerging of a new phenotype, such as cancer stem cells, drug-resistant cells, and metastatic cells, through selection. Uncovering the relationships between various subpopulations (presence and frequency of them) and cancer clinical features, such as survival, metastasis, response to therapies, allows to deepen the knowledge on the oncogenetic process.

To study breast cancer disease is necessary to compare it with normal breast tissue. Normal cells from human mammary gland are required as healthy controls for malignant cells. Comparisons between malignant cells and normal cells not only show similarities, but also provide a more detailed vision of the differences between tumor cells and their supposed cells of origin, uncovering new aspects of cancer evolution.

Single-cell RNA analyses also permits the characterization of tumor ecosystems, and it allows to lay the basis for the identification of unique clinical phenotypes. It shed light on tumor classification previously based only on bulk measurements, disclosing unanticipated patterns of intra-tumor heterogeneity, whose understanding might offer new potential vulnerabilities that could be used as therapeutic targets.

2.2 SINGLE-CELL RNA SEQUENCING TECHNIQUES

Single-cell gene expression techniques enables gene expression analyses at a single cell level. There are different methods to study the transcriptome at the level of individual cells, one of them is the Chromium 10X technology, that uses a microfluidic droplet chemistry with encapsulation of cells and barcoding beads [17].

Single-cell RNA sequencing workflow comprises various passages:

a) GEM generation and barcoding
b) Post-GEM-RT cleanup and cDNA amplification
c) 3’ gene expression library construction
d) Sequencing: Illumina Sequencing by Synthesis

The key point of all droplet-based scRNA sequencing systems, such as Chromium 10X, is the capability to create individual water-in-oil reaction chambers (or GEMs) within each of which the system introduces both a cell, a bead, and the molecular reagents to create barcoded cDNA in a droplet (Figure 1). The reactions that are needed to obtain barcoded cDNA occurs inside each droplet, that performs the function of a reaction tube. Droplets are created by the introduction of two immiscible fluids into a microfluidic system resulting in a dispersed phase, typically of multiple isolated droplets contained within a continuous phase [16].

During the workflow, Gel Beads in-emulsion (GEMs) are generated into a microfluidic chip (Chromium Next GEM chip). Gel Beads in-emulsion or GEMs are the reaction vesicles formed combining single cells, Master Mix (containing RT Reagent B, Template Switch Oligo, Reducing Agent B, RT Enzyme C), Gel Beads containing primers and partitioning oil (Figure 2).

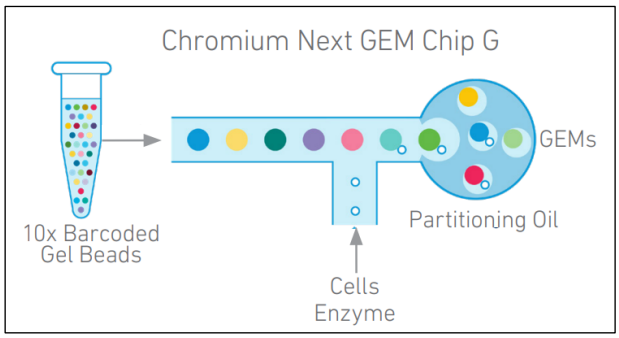

Figure 1: Schematic picture of GEM generation.
Modified from: Chromium Next GEM Single Cell 3' Reagent Kits – User guide, 10x Genomics.

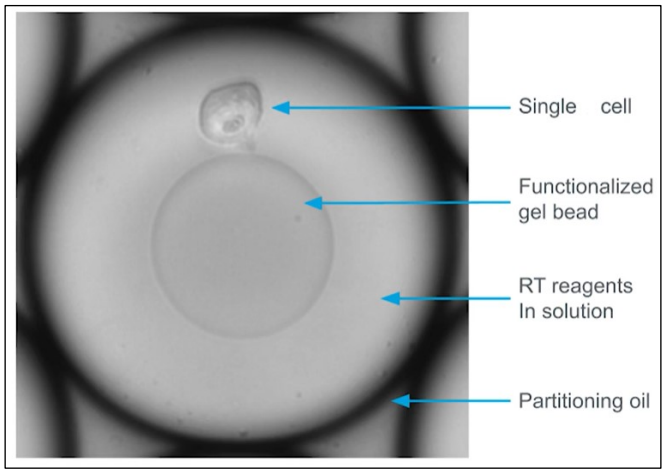

Figure 2: Microscopic picture of an individual GEM (Gel Beads in Emulsion). The bead is in the middle and the single captured cell is seen inside the GEM, that is surrounded by partitioning oil.
Modified from: Chromium Single Cell Expression v3.1 (Single Index), Chapter 2 – Biochemistry & Assay Scheme, 10x Genomics.

Each Next GEM Single Cell 3' Gel Bead contains millions of copies of oligonucleotides, each of which has the same 10xBarcode, but a different Unique Molecular Identifier (UMI) for every individual oligo (Figure 3). The presence of these oligos inside each GEM allow to "give the same name" to all the transcripts coming from the cell contained inside that reaction vesicle, thanks to the 10x Barcode, and to mark each mRNA, or better each cDNA after reverse transcription, of the same cell with a different nt-code (the UMI).

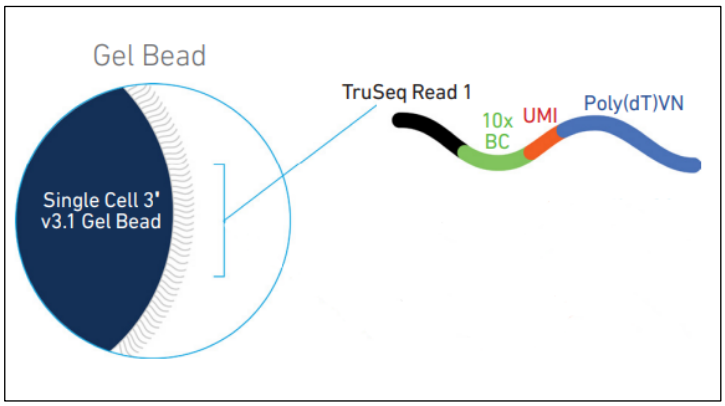

Figure 3: Schematic picture of Next GEM Single Cell 3' Gel Beads oligonucleotides.
Modified from: Chromium Next GEM Single Cell 3' Reagent Kits – User guide, 10x Genomics.

The components of these mRNA capture oligos are:

- Poly(dT)VN sequence: 30 nt Poly(dT) sequence, which enables the specific capture of polyadenylated mRNA molecules and avoids ribosomal RNAs capture. It consists of a stretch of deoxythymidines that anneal to poly(A) tails of eukaryotic mRNAs, which make up only 1–5% of total RNA. These primers are the optimal choice for constructing cDNA libraries from eukaryotic mRNAs, full- length cDNA cloning, and 3' rapid amplification of cDNA ends (3' RACE). Because of their specificity for poly(A) tails, oligo(dT) primers are not suitable for other RNA types, such as degraded RNA. The direct capture of poly-adenylated mRNA inside individual GEMs is mediated by the interaction between the Poly(dT)VN in the Gel Bead oligo and the poly-adenylated 3' end of the mRNA transcript. The VN bases at the end of Poly(dT) facilitate reverse trascriptase enzyme annealing, close to the 3'. 'N' is any base, 'V' is either adenine, cytosine, or guanine.

- 10x Barcode (cellular barcode): a 16 nt 10x Barcode which identifies the particular GEM, and thus will univocally identify each cell and all its transcripts. There are approximately 3.6 million defined 10x barcoded sequences. Each gel bead contains identically barcoded RT oligonucleotides; as a result, at the end of RT, all cDNAs from a single cell will have the same barcode, allowing the sequencing reads to be mapped back to their original single cells of origin.

- UMI: 12 nt Unique Molecular Identifier (UMI), to unequivocally label a single mRNA molecule. Unique molecular identifiers (UMIs) are a type of molecular barcoding, in which the RNA molecules to be amplified are tagged with random oligonucleotides.

The number of different tags is designed to significantly exceed the number of copies of each transcript species to be amplified, resulting in uniquely tagged molecules, allowing control for quantitative biases introduced by the amplification, providing error correction, and increasing accuracy during sequencing. The random sequence composition of the UMIs assures that every fragment-UMI combination is unique in the library. After PCR enrichment, without UMIs, distinguish if multiple copies of a fragment are caused by PCR clones or if they are real biological duplicated is not possible. By using UMIs, PCR clones can be found by searching for same fragment-UMI combinations, which can only be explained by PCR clones. The process of discarding PCR clones, ending up with one single representative, is known as de-duplication. If the de-duplication is performed after the mapping step, it is possible to use the multiple copies of PCR clones to enrich the quality of the read for the original fragment. In fact, the multiple copies can be used to correct for sequencing errors.

- TruSeq Read 1 (read 1 sequencing primer): a 22 nt Partial Illumina TruSeq Read 1 sequence. This sequence is functional to bind the oligonucleotide to the bead, and - in the following steps - for library construction and sequencing.

With this system, each mRNA molecule is coupled with a barcode that identifies its cell of origin, and a transcript-specific UMI. After sequencing, transcripts can be traced back to their original cell through the 10x Barcode and quantified by counting how many different UMIs were associated with each cell.

Thanks to the Chromium instrument it was possible to capture about 10.000 cells from each sample that has been loaded. Through the Chromium Next GEM technology, we expect to obtain above 2.500-3.000 genes per cell.

After barcode incorporation and molecular amplification for sequencing, the following step is the 3’ gene expression library construction. The purpose of preparing the library is to make the extracted DNA compatible with binding to the flow cell for sequencing through series of steps, at the end of which the final construct consists of a molecule that contains various sequences (Figure 4):

- DNA insert, that is the fragment of double stranded DNA from the original sample.
- The previous added tags, or rather a 10x Barcode, a UMI, and the Poly(dT) sequence.
- P5, P7, and two binding sites, that are required for paired-end sequencing on an Illumina sequencer. The two unique P5 and P7 adapters are on both ends of the

DNA insert and contain the binding regions for clustering. P5 and the P7 are complementary to the flow cell oligo lawn, and these facilitate library hybridization to the flow cell. The binding sites for sequencing include the Read 1 sequencing binding site for the forward read and the Read 2 sequencing binding site for the reverse read.

- Sample Indexes allow for multiplexing, which is the ability to combine multiple samples in a single sequencing run. Indexes are short sequences of DNA, normally about 6-10 bp that barcode or uniquely identify samples within a run. The presence of only one index signifies a single index library.

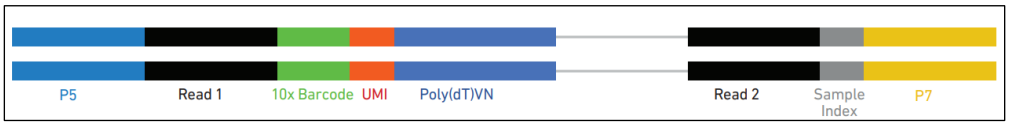

Figure 4: Scheme of the Single Cell 3' Gene Expression Library, final cDNA construct.
Modified from: Chromium Next GEM Single Cell 3' Reagent Kits – User guide, 10x Genomics.

After library construction, sequencing by Illumina sequencer can be performed and, then scRNA-seq data can be interrogated, and interpreted.

3.0 PURPOSE OF THE STUDY

Breast cancer development is an evolutionary process that starts from mammary epithelial cells and leads to primary tumor appearance and metastatic spreading.

Human tumors, and breast tumors, are complex ecosystems composed of different cells, including malignant, immune, and stromal cells, that act together determining the histopathological features of the tumor and the clinical outcome.

In this contest, transcriptomic profiling of biological samples has revolutionized our knowledges of cancer.

Our study is part of a larger study, whose aim is to investigate breast cancer heterogeneity at a single-cell level, through the analyses of the transcriptome of the different cells that make up the human mammary gland involved in breast tumorigenesis.

This study encompasses the analysis of the breast primary tumor, as well as metastatic breast disease, and, in particular, breast cancer brain metastases.

To explore the neoplastic context, it is mandatory to compare the results coming from primary tumor and metastatic samples with those obtained from healthy tissue specimens of human breast. For this reason, the purpose of this thesis is to collect and analyze human mammary gland samples that will be used as healthy controls for pathological tissue specimens.

These samples were obtained during reduction mammoplasties of healthy patients, and they were processed through mechanical and enzymatic tissue dissociation to reach a single cell suspension. This was the key point of the tissue digestion: to get a suspension composed of at least one million individual viable cells per milliliter, minimizing dead cells, debris, erythrocytes, and cell clumps.

Human mammary gland samples were then analyzed in two different ways, single-cell RNA sequencing, or bulk RNA sequencing of selected cell populations.

These two different types of transcriptome analysis offer distinct points of view of the same landscape: the human mammary gland.

sc-RNAseq analysis will allow to investigate intra-tissue heterogeneity at a single-cell level, albeit at a low resolution. Instead, bulk RNA sequencing and analysis permits to obtain an average profile of the gene expression of the sample that has been investigated, but with high resolution.

4.0 MATERIALS AND METHODS

4.1 HUMAN MAMMARY GLAND SAMPLES

Five human mammary glands samples from reduction mammoplasties were obtained from consenting patients. Patients' recruitment was allowed only after negative breast screening. Each sample was harvested in HBSS 2X P/S and kept cold until processing. Before single-cell preparation each sample was weighed, and a tissue section was obtained for microscopic examination.

4.2 TISSUE DISSOCIATION

Single-cell suspensions from reduction mammoplasties were prepared from fresh surgical specimens. Mammary glands were washed with PBS 2X P/S to drain as much blood as possible. For preparation of single-cell suspensions, breast samples were minced to a homogeneous slurry (into $1mm^3$ pieces) using scalpel blades and then digested with 50-60ml of Tissue Dissociation Medium. It contains: 600U/ml Collagenase (Life Technologies 177100-017) and 400U/ml Hyaluronidase (Sigma H3506) in Advanced DMEM/F12 supplemented with HEPES and GlutaMAX. 15cm cell culture dishes were left in incubator (37°C, 5% CO_2) for 14 hours.

After 14 hours the content of the dish was transferred in 50ml falcon tubes and then vigorously shaken manually. After decantation (in incubator in vertical position for 20/30 minutes) it could be possible to find this situation:

- Adipose tissue on the top of the falcon tube.
- Tissue dissociation medium, containing blood and debris, in the middle.
- Pellet of fragmented mammary ductal tree, enriched in epithelial cells, at the bottom of the falcon tube.

Adipose tissue and dissociation medium were trashed. After three washes with HBSS 2X P/S, each one followed by centrifugation at 0.3 RCF for 3 minutes, removal of the supernatant and changing the 50ml falcon tube to get rid of adipose tissue, the pellet enriched in epithelial cells was digested with Dispase 5mg/ml (10ml total) and 50µl DNaseI, at 37°C for 10 minutes.

After pipetting through a 5ml serological pipette, 0.25%v trypsin-EDTA (4ml) and 30µl DNaseI were added on top of Dispase and digestion incubated at 37°C for another 10 minutes.

After pipetting through a 5ml serological pipette with a p1000 tip attached to disaggregate cell clumps a single cell suspension was reached and digestion was stopped by adding FBS 150µl. BSA (bovine serum albumin), a major component of fetal bovine serum, acts as surfactant avoiding cell to cell adhesion in a single-cell suspension. It is added to minimize cell losses and aggregation.

Red blood cells were burst by incubating in 4ml haemolytic solution (0,64% NH_4Cl) at 4°C for 3 minutes. Haemolysis was stopped with 8ml of HBSS P/S or better with 1% BSA.

Cells were filtered using a 40 µl cell strainer and transferred into a 15ml tube (pre-washed with HBSS/PS 1% BSA). Filter cell suspensions with an appropriate cell strainer is helpful for removing large clumps and debris. The strainer should have a pore size that is larger than the maximum cell diameter in the sample, but small enough to catch larger clumps [18].

4.3 CELL COUNTING

Countess™ 3 Automated Cell Counter was used to count cells.

Sample was prepared by adding 10 µL of cell suspension to 10 µL of 0.4% trypan blue stain (for brightfield only). The sample mixture was mixed well by pipetting it up and down a few times. Then, 10 µL of the sample was pipetted into the half-moon shaped sample loading area. The sample was loaded into the chamber through capillary action and let in the chamber for 30 seconds.

The instrument captures the image and displays the results (total concentration, percentage, and concentration of live and dead cells).

The results screen allowed to identify the objects (i.e., cells) counted in each channel and included in the count results for further review. After reviewing the marked objects, it was possible to adjust the threshold for size, brightness, and/or circularity.

4.4 CYTOFLUORIMETRY AND FLUORESCENCE-ACTIVATED CELL SORTING (FACS)

Single cell pellet was resuspended in 500 µl of medium without FBS and with 0.1% BSA. 150 µl of the sample is used for scRNA sequencing. An aliquot of 50 µl remains unlabeled (control sample) and acts as a FACS baseline, to assess the sample autofluorescence. The leftover part (300 µl) was marked with antibodies to assess epithelial cell marker expression on normal tissue:

- 5 µl CD31-APC/Cy7
- 5 µl CD45-ECD
- 12.5 µl EpCAM FITC
- 10 µl CD49f-PE-Cy5

After incubation with antibodies on ice for 30 minutes, sample was diluted with HBSS 2X P/S to wash antibodies. After spinning for 5 minutes at 0.4 RCF, the supernatant was removed, and pellet was resuspended with 500 µl of sorting solution (PBS 1X Ca^{2+}/Mg^{2+} free, 1mM EDTA, 25mM HEPES pH 7.0, 0.1% BSA) and put into a flow-cytometry tube. Flow-cytometry data were collected and analyzed using BD FACS Aria.

4.5 10x GENOMICS CHROMIUM LIBRARY CONSTRUCTION

A 10x Genomics Chromium machine was used for single cell capture and cDNA preparation, according to the Single Cell 3' Protocol. Thanks to Next GEM technology partitioning of the inputs into a microfluidic chip was performed with a high resolution.

Silane magnetic beads were used to purify the obtained cDNA from the GEM reaction mixture. Barcoded cDNA was amplified via PCR, using primers complementary to the common partial Read 1 and Template Switch Oligo (TSO).

10x Genomics Chromium library was prepared through: a) enzymatic fragmentation, and size selection, that were used to optimize cDNA amplicon size; b) end repair, A-tailing, and ligation of adaptors (partial Read 2); c) additional cleanup and priming, followed by Sample Index PCR reaction to add the unique Sample Index and the adaptors needed for sequencing (P5 and P7). Sequencing was performed on an Illumina sequencer.

5.0 RESULTS

5.1 SAMPLE PROCESSING AND CELL COUNTING

Human mammary gland samples were obtained from five healthy women undergoing reduction mammoplasty at the Azienda Ospedaliera di Padova's Plastic Surgery Unit, directed by Prof. F. Bassetto. These patients were selected according to the following criteria, delineated by CancerORG protocol approved by the ethics committee of Azienda Ospedaliera di Padova Hospital: the serology for hepatitis B virus (HBV DNA and HBsAg), hepatitis C virus (HCV RNA and anti-HCV) and HIV virus (HIV RNA and anti-HIV) had to be negative, no mental disorders had to be observed and patient's age had to be over 18 years. To be eligible for the study, we further selected only patients that were over 45 years of age, to assure better age matching to tumor and metastatic samples. Moreover, a special informed consent had to be given to each patient, in order to explain the aim of the study and its characteristics, and every sample was anonymized through a coded numbering (Table II). Only one of the five samples collected was excluded from the study, due to an excess of red blood cells that contaminated the single cell suspension during dissociation.

Table II: Patient's age and type of analysis that was performed for each sample are shown.

Patient	Age (years)	Analysis
1	62	sc-RNA seq
2	58	-
3	63	sc-RNA seq bulk-RNA seq
4	54	sc-RNA seq
5	49	bulk-RNA seq

Mammoplasties derived samples were transferred to the lab, where they were subjected to mechanical and enzymatic tissue dissociation: extracellular matrix was digested, cell-cell junctions were disrupted, and each surgical sample was reduced into a single-cell suspension (see Methods section 4.2).

To obtain high quality data, one of the critical points of sample processing was to minimize the presence of cellular aggregates, dead cells, non-cellular nucleic acids, and potential

inhibitors of reverse transcription. Best practices for handling samples included biohazard safety regulations, sterile technique, nuclease-free agents, and minimizing of cell damage. Ideally, input cell suspension should contain more than 90% of viable cells. Non-viable and dying cells generally contain less total RNA and generally in a more fragmented state, which may not be efficiently captured by 10x Genomics Single Cell Solutions, and their presence in the input suspension may therefore decrease the apparent efficiency of cell partitioning and recovery.

Counting cells was an important passage to assess the quality of tissue dissociation and to decide the volumes of cell suspension in water to be added to the RT Master Mix, which is one of the first step of 10x Genomics Chromium library construction. For this, all dissociated tissues were subject to cell counting by Countess™ 3 Automated Cell Counter (see Methods section 4.3) to assess the total number of cells and viability percentages (Table III).

Table III: Cell counts, and live cell percentages obtained for each dissociated sample are shown. Two independent counts were performed for each sample.

Patient	Cell counting (live cells/ml)	Live cells (%)
1	$1.33*10^6$	83
	$1.20*10^6$	79
2	$1.04*10^6$	79
	$1.12*10^6$	79
3	$1.08*10^6$	89
	$1.09*10^6$	93

During sample processing few microliters of single-cell suspension were checked with the microscope to confirm complete single-cell dissociation and to verify the accurate representation of a heterogeneous sample, that should contain cells of variable sizes.

5.2 CYTOFLUORIMETRY

Before single cell capture for subsequent scRNA-seq analyses, we verified that each dissociated sample falls within the normal parameters of breast cellular heterogeneity, and that the previous mechanical and enzymatic dissociation, or patient specific variables, did not cause any major variation from normality. For this, cytofluorimetry assessment was performed to identify the four cellular subpopulations that compose the human mammary gland (luminal progenitor, mature luminal, basal, and stromal cells) and to check if these were normally represented in each sample [19] [20] [21].

Flow cytometry provides quantitative characterization of the sample with regard to viability, cell debris, and cell type composition using fluorescent-labelled monoclonal antibodies.

Cytofluorimeter output was analyzed in two-parameter dot plots, displaying two measurement parameters, one on the x-axis and one on the y-axis.
During cytofluorimeter output analysis, the cells of interest were selectively shown, through the so called “gating strategy”. Gating is an important procedure in flow cytometric data analysis used to selectively visualize the cells of interest and eliminate results from unwanted particles, such as dead cells and debris, or unlabeled cells. In a tissue sample containing a mixed population of cells, gating method allows to define the characteristics of particles for further analysis and to restrict analysis to only a subpopulation of cells.
In each plot, a quadrant marker divides the two-parameter plot into four sections to discriminate populations as negative, single positive or double positive.

First, we analyzed cells through two different parameters:

- FSC, or forward scattered channel, on the x-axis. It collects the forward scattered light signal, that is used to determine the particle size. Bigger particles will usually produce more forward scattered light than smaller ones, and larger cells will have a stronger forward scatter signal.
- SSC, or side scattered channel, on the y-axis. It collects the side scattered light signal, that usually provides information about the granularity and complexity of the cells.

Through gating, it was possible to select live cells, and to exclude dead cells and debris.

The next two parameters that were analyzed only in the selected population of cells were two markers, CD45 and CD31 (Figure 5).
CD45 antigen (leukocyte common antigen or Ly-5) is a unique and ubiquitous transmembrane glycoprotein with a molecular mass of about 200 kDa, also known as PTPRC (protein tyrosine phosphatase, receptor-type C). It is expressed on all nucleated hematopoietic cells [22]. For this reason, CD45 is used as a marker of all hematopoietic cells (blood cells), except for mature erythrocytes (red blood cells) and platelets.
CD31, also known as PECAM-1 (platelet endothelial cell adhesion molecule), is a 130-kDa transmembrane glycoprotein expressed by endothelial cells, platelets, macrophages and Kupffer cells, granulocytes, lymphocytes (B/T/NK cells), megakaryocytes, osteoclasts,

neutrophils, and certain tumors. CD31 is used as a marker to demonstrate the presence of endothelial cells [23].

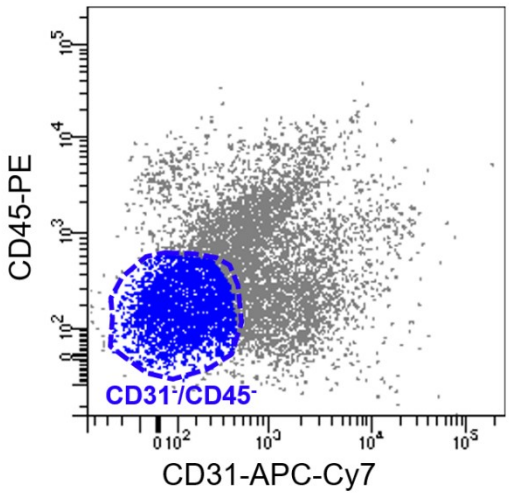

Figure 5: Representative cytofluorimetric dot-plot showing the gating strategy employed to exclude hematopoietic-lineage and endothelial cells. The double negative CD45-/CD31- population was gated (blue) for further analyses.

Only $CD45^-/CD31^-$ cells were gated (the blue cloud in figure 5), containing breast parenchymal cells (epithelial and stromal cells). Otherwise, $CD45^+$ and $CD31^+$ cells (lymphohematopoietic and endothelial cells) were not considered.

Finally, $CD45^-/CD31^-$ cells were studied for CD49f and EpCAM markers.

CD49f, or integrin α6,is among the proteins that have been identified in stem cell populations and somatic cells like keratinocytes, platelets, epithelial cells, and basal cells of the cornea. Integrin α6 has been identified as a biomarker of mammary stem cells (MaSCs), which are cells that display self-renewal properties and are able to regenerate new mammary tissue in vivo [24].

EpCAM/CD326, or epithelial cell adhesion and activating molecule, is a pan-epithelial differentiation antigen. It is a transmembrane glycoprotein that mediates calcium-independent homophilic adhesions on the basolateral surface of most epithelial cells. EpCAM is expressed on the basolateral membrane of all simple (especially glandular), pseudo-stratified, and transitional epithelia. EPCAM marks luminal mammary epithelial cells [7]. It is also frequently upregulated in carcinomas, but it is not expressed in cancers of non-epithelial origin.

Flow cytometry based on CD49f and EpCAM antibody direct staining separates breast tissue cells into stromal ($CD49f^-$ and $EpCAM^-$) and epithelial cells.

As shown in figure 6, the final plot shows four different cloudlike shapes, each-one representing a different cellular subpopulation.

- Mature luminal or LM cells ($CD49f^-/EpCAM^+$ subpopulation) are indicated by the upper-left cloud of the plot (orange).
- Luminal progenitor or LP cells ($CD49f^+/EpCAM^+$subpopulation) are shown in the upper-right cloud of the plot (light green).
- Basal cells ($CD49f^+/EpCAM^{lo/-}$) are represented in the lower-left cloud (purple).
- Stromal cells ($CD49f^-/EpCAM^-$) are the remaining cells localized in the lower-right quadrant (dark green).

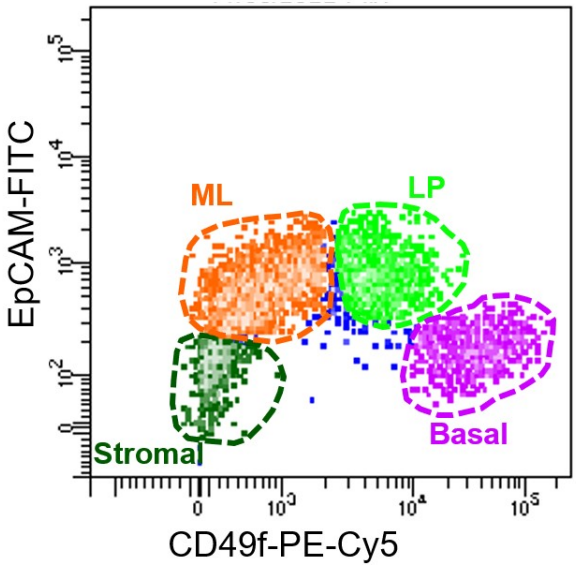

Figure 6: Representative cytofluorimetric dot-plot showing the gating strategy employed to evaluate the relative representation of the four cell populations composing the mammary gland parenchyma. Cells were gated as: mature luminal (LM; CD49f-/EpCAM+); luminal progenitor (LP; CD49f+/EpCAM+); basal ($CD49f+/EpCAM^{lo/-}$); stromal (CD49f-/EpCAM-).

To check tissue sample subpopulations is a critical step of this workflow, because it allows to confirm that the sample falls within the normal parameters of breast cellular heterogeneity.

Table IV shows the percentages of the different cellular populations that were identified thanks to flow cytometry. Each major cell type was present in each sample, and the percentages obtained were similar between samples and close to those expected [20].

Table IV: Percentages of the breast parenchyma cell subpopulations, identified through cytofluorimetry, are shown. The upper box shows the percentages of cells expressing specific markers. The box below shows the cellular percentages of the different clusters - stromal, basal, luminal progenitor (LP), mature luminal (ML) - calculated on all live cells.

Patient	CD31+ (%)	CD45+ (%)	EpCAM+ (%)	CD49f+ (%)	CD31-/CD45- (%)
1	1.8	6.9	88.9	31.3	90
2	5.1	8.7	48.9	25.5	56.5
3	2	33.3	31.6	30.6	53.3

Patient	Stromal (%)	Basal (%)	LP (%)	ML (%)
1	3.8	4.2	30.2	58.9
2	26.9	11.9	21	33.2
3	51.7	8.7	20.3	11.8

5.3 SINGLE-CELL LIBRARY CONSTRUCTION AND QUALITY CONTROLS

5.3.1 GEM GENERATION

Three of all samples were destined to single-cell RNAseq analyses, using 10x Genomic Chromium technology. The adequate volume of cell suspension (containing 16.500 cells) was loaded on the microfluidic chip (Chromium Next GEM chip), and it was placed in the Chromium Controller where Gel Beads in-emulsion (GEMs) were generated. Ideally, each GEM vesicle should contain both a single cell and a single bead, but in reality, drops may also contain: 1) no bead or no cells, 2) one or more cells, 3) one or more beads, 4) one or more cells and one or more beads [25]. The proportion of these different possibilities depends on the concentration of beads and cells. Consequently, to achieve optimal single cell resolution, cells were loaded at a limiting dilution, such that the majority (~90-99%) of generated GEMs contain no cell, while the remainder largely contain a single cell [17].

5.3.2 cDNA SYNTHESIS AND QUNTIFICATION

After cell capture, the Chromium Next GEM Single Cell 3' protocol was followed to lyse cells and obtain complementary DNA (cDNA). Briefly, immediately following GEM generation, within each GEM reaction vesicle, a single cell is lysed, the Gel Bead is dissolved, and primers are released. Primers are mixed with the cell lysate and the Master

Mix containing reverse transcription reagents. At this point reverse transcription of polyadenylated mRNA occurs and poly(dT)-primed mRNA is converted to complementary DNA (cDNA) by the reverse transcriptase.

During GEM incubation the reverse transcriptase enzyme uses the mRNA template to generate full length cDNA (Figure 7). First-strand synthesis incorporates the cell 10x Barcode and the UMI into the same cDNA molecule. When the enzyme reaches the 5' end of the transcript, it adds three untemplated cytosine bases to the cDNA strain. A Template Switch Oligo ending in three guanine bases anneals to these bases enabling the RT enzyme to incorporate the reverse complement of this oligo sequence at the end of this first strand of cDNA. The TSO (template switch oligo) is an oligo that hybridizes to untemplated C nucleotides added by the reverse transcriptase during reverse transcription. The TSO adds a common sequence to full-length cDNA that is used for downstream cDNA amplification.

Incubation of the GEMs produces barcoded, full-length cDNA from poly-adenylated mRNA. These strands cDNA contain the partial Read 1 sequence, the 10x Barcode, UMI, Poly(dT)VN, the insert and the TSO sequence.

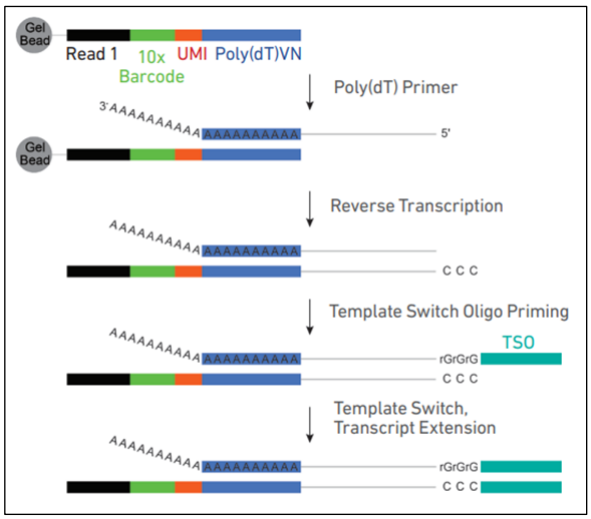

Figure 7: Scheme showing the cDNA synthesis reactions that occurs inside each GEM.
Modified from: Chromium Next GEM Single Cell 3' Reagent Kits – User guide, 10x Genomics.

After incubation, GEMs were broken simply by the addition of Recovery Agent to release the barcoded cDNA fragments into a single bulk aqueous pool, and pooled fractions were recovered (Figure 8). Silane magnetic beads were used to purify the first-strand cDNA from the post GEM-RT reaction mixture, which includes leftover biochemical reagents and primers.

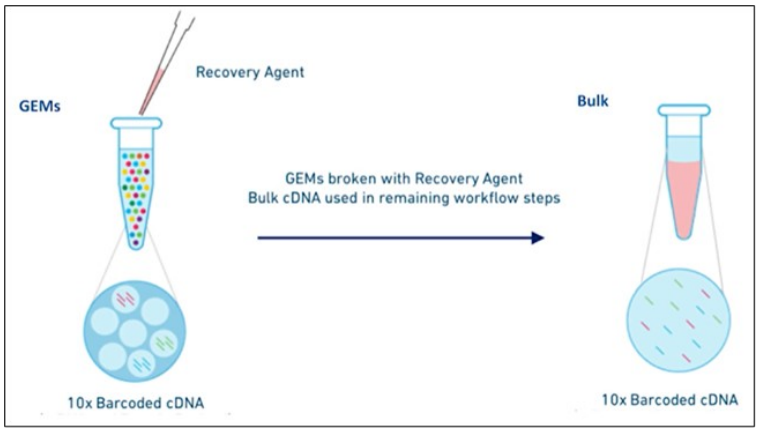

Figure 8: Schematic picture of the recovery phase: GEMs are broken and barcoded cDNAs are pooled. Modified from: Chromium Single Cell Expression v3.1 (Single Index), Chapter 3 – Biochemistry & Assay Scheme, 10x Genomics.

Barcoded, full-length cDNA was amplified by PCR, using primers complementary to the common partial Read 1 and TSO sequences present on each end of the cDNA molecule to generate sufficient mass for library construction.

The optimal number of cycles is a trade-off between generating sufficient final mass for library construction and minimizing PCR amplification artifacts, and it is selected based on the number of target cells to capture [17]. Considering the number of captured cells, we performed 12 cycles of cDNA amplification.

Product quality and yield were assessed through cDNA quantification, using a BioAnalizer (Agilent) with a High Sensitivity DNA kit (Figure 9).

cDNA quality assessment and quantification were performed for two main reasons:

a) To assess the quality of the single-cell capture: the right variability of the fragments is revealed by a rising graph, that is shifted to the upper marker, and characterized by various peaks. A graph with such characteristics means a high-quality capture. As shown in figure 9, representative cDNAs displayed the expected fragment size distribution (in the range from 500 to 5.000 bp).
b) To decide the number of PCR cycles in the following library preparation steps: the lower amount of cDNA, the higher number of PCR cycles, and vice versa. cDNA concentrations of our samples were those that are shown in Table V.

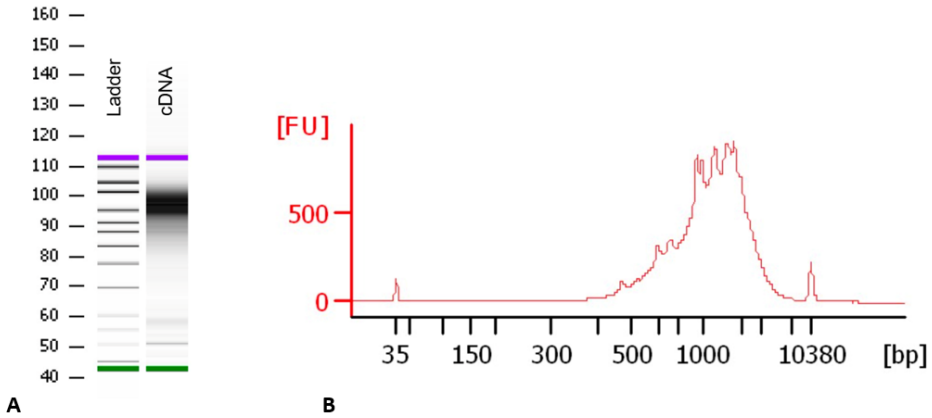

Figure 9: (A) Representative cDNA smear obtained by BioAnalyzer analysis. (B) Representative graph showing the size distribution obtained by the cDNA smear showed in (A).

Table V: cDNA concentrations of our samples and the number of PCR cycles subsequently selected for cDNA amplification and library construction are shown.

Patient	cDNAconcentration (ng/µl)	Number of PCR cycles
1	9.4	13
2	12.6	13
3	4.2	14

5.3.3 LIBRARY CONSTRUCTION

Only the 25% of the total cDNA mass generated during cDNA amplification was used for library preparation as library complexity saturates at a very low fraction of amplified cDNA and carrying more cDNA into library construction will not increase library complexity.

We proceeded to library preparation through the following steps (figure 11):

a. Enzymatic fragmentation and size selection (Double-Sided Size Selection-SPRIselect), that are used to optimize the cDNA amplicon size. As regard the Double-Sided Size Selection-SPRIselect, SPRI beads selectively bind DNA according to the ratio of SPRIselect reagent (beads). After the first SPRI, supernatant is transferred for a second SPRI while larger fragments are discarded (green area in figure 10). After the second SPRI, fragments on beads are eluted and kept while smaller fragments are discarded (blue area). Final sample has a tight fragment size distribution with reduced overall amount (black area).

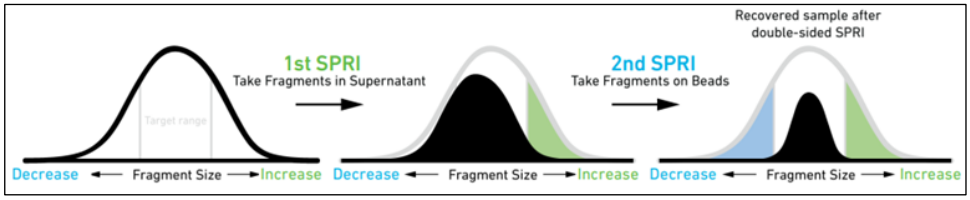

Figure 10: Schematic of Double-Sided Size Selection.
Modified from: Chromium Next GEM Single Cell 3' Reagent Kits – User guide, 10x Genomics.

b. End repair, A-tailing, and ligation of adaptors consisting of a partial Read 2 sequence.
c. Additional cleanup and priming.
d. Sample Index PCR reaction, which adds a unique Sample Index to each library and completes the addition of P5 and P7 adaptors needed for the Illumina Flow Cell sequencing.

The Sample Index PCR total cycles should be optimized based on 25% cDNA input calculated during Post-cDNA Amplification QC and Quantification.

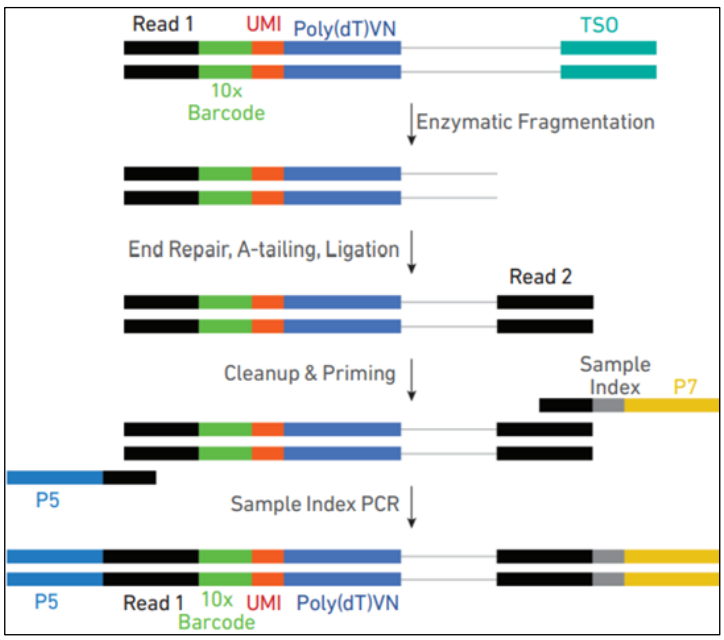

Figure 11: Schematic of 10x Genomics Chromium library construction
Modified from: Chromium Next GEM Single Cell 3' Reagent Kits – User guide, 10x Genomics.

The final construct contains:

- The insert of DNA, that is the fragment of double stranded DNA from the original sample.
- The two unique P5 and P7 adapters. They are on both ends of the DNA insert and contain the binding regions for clustering. P5 and the P7 are complementary to the flow cell oligo lawn, and these facilitate library hybridization to the flow cell.
- The binding sites for sequencing, that include the Read 1 sequencing binding site for the forward read and the Read 2 sequencing binding site for the reverse read.
- Sample Indexes, that allow for multiplexing, which is the ability to combine multiple samples in a single sequencing run. Indexes are short sequences of DNA, normally about 6 -10 bp that barcode or uniquely identify samples within a run. The presence of only one index signifies a single index library.

A second check was performed after library preparation to assess the quality and the concentration of the library prior to sequencing. As regards the quality control, the final library showed a size assay distribution between 300-600 bp, with an average fragment size around 450 bp. This control was performed using an Agilent BioAnalyzer. In figure 12 is shown a representative output of the library quality control and quantification of one of our samples in the form of an electrophoretic run (A) and of a graph (B).

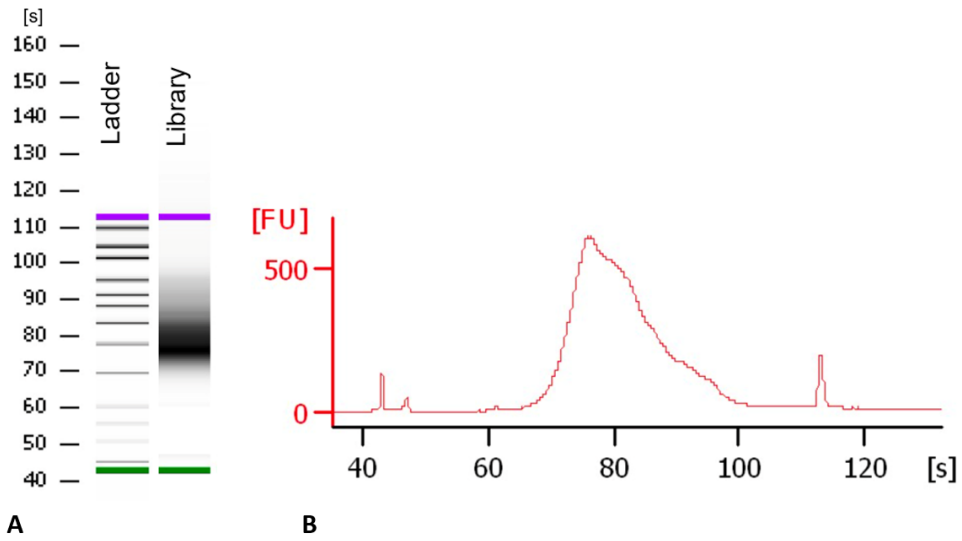

Figure 12: (A) Representative electrophoretic image of the library obtained by BioAnalyzer analysis. (B) Representative graph showing the size distribution and quantification of the fragments of the library, obtained from the smear showed in (A).

5.4 SINGLE-CELL RNA SEQUENCING

Single Cell 3' Gene Expression Library are compatible with most of the Illumina sequencers. For Single Cell 3' Gene Expression libraries we performed a paired-end, single indexed sequencing configuration, with a minimum of 20.000 raw read pairs per cell.

Each Chromium Single Cell 3' Gene Expression library comprises standard Illumina paired-end constructs which begin with P5 and end with P7. P5 and P7 sequences are critical for hybridization to the Illumina Flow Cell and sequencing (Figure 13).

16 bp 10x Barcodes are encoded at the start of TrueSeq Read 1, while sample index sequence is incorporated as the i7 index read. During indexed sequencing, the index is sequenced in a separate read called the Index Read, where a new sequencing primer is annealed.

TruSeq Read 1 and TruSeq Read 2 are standard Illumina sequencing primer sites used in paired-end sequencing. TrueSeq Read 1 is used to sequence 16 bp 10x Barcodes and 12 bp UMI, while Read 2 is used to sequence the cDNA fragment. Because these libraries are single indexed there is no i5 Index sequence.

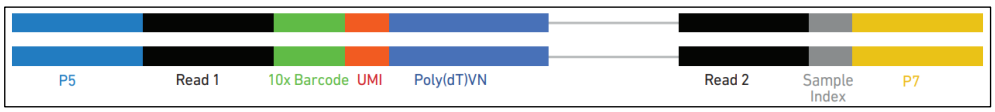

Figure 13: Schematic of the final construct that will be sequenced through an Illumina sequencer. Modified from: Chromium Next GEM Single Cell 3' Reagent Kits – User guide, 10x Genomics.

The 3' Gene Expression libraries were pooled for sequencing, taking into account the differences in cell number and per-cell read depth requirements between each library.

Once bound to the flow cell oligos (complementary to the P5/P7 regions), the target strands undergo clustering. Clustering is designed to increase the number of target strands in a single location such that the fluorescence signal resulting from fluorescent labeled nucleotide incorporation will be sufficiently above background noise confidently determine which base is incorporated [25].

Single Index Read protocol is based on three reads (Figure 14): Read 1, Index 1 and Read 2. The Read 1 primer is first hybridized to the template strand and the Read 1 is sequenced. The Read 1 primer and sequence are then washed away. The Index 1 primer is hybridized while the strand is still in the forward direction and Index 1 is sequenced. The Index 1 primer and sequence are then washed away. After that, paired-end

turnaround occurs. After a paired-end turnaround, the forward strand is washed away and the Read 2 primer is hybridized and extended to sequence Read 2.

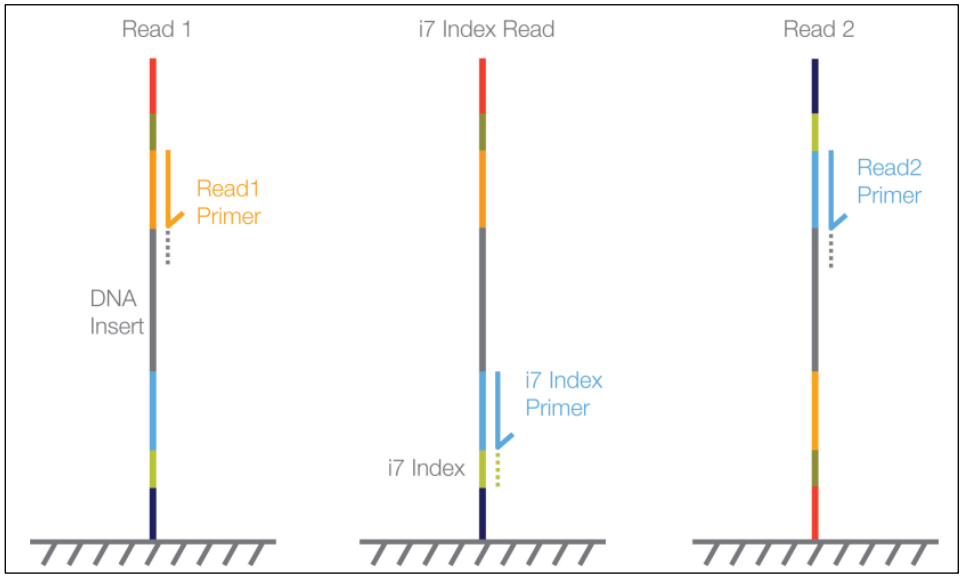

Figure 14: Schematic of single-indexed sequencing.
Modified from: Indexed Sequencing Overview Guide - Illumina®.

The sequencing and the data analyses of our samples is currently ongoing.

5.5 CELL SORTING

Three of all samples (also one that was processed through scRNA-sequencing) were sorted using a FACSAria flow cytometer to isolate breast luminal cells, whose expression profiles will then investigated with a bulk-RNA-seq.

FACS is an abbreviation for fluorescence-activated cell sorting, which is a flow cytometry technique that further adds different functions. By utilizing highly specific antibodies labeled with fluorescent conjugates, FACS analysis allows to simultaneously collect data on, and sort a biological sample by a nearly unlimited number of different parameters.
Just like in conventional flow cytometry, forward-scatter, side-scatter, and fluorescent signal data are collected. FACS is the process by which individual cells are separated from the bulk population through multiparameter fluorescent and light scatter analysis of the cell population. After defining the parameters on how cells should be sorted, the machine imposes an electrical charge on each cell so that cells will be sorted by charge (using electromagnets) into separate vessels upon exiting the flow chamber.

FACS can be used to remove dead cells and debris, remove cell clumps, and enrich or purify target populations defined by the expression of one or more cell surface receptor markers labelled with fluorescent labelled antibodies.

Fluorescently conjugated antibodies have commonly been used to label specific structures on the cell for flow cytometric analysis.

Fluorophores used for the detection of target proteins emit light after excitation by a laser of compatible wavelength. These fluorescently stained cells or particles can be detected individually. Each type of fluorescent dye or label has its own characteristic excitation and emission spectrum which is important for designing flow cytometry experiments.

After single cell dissociation of samples, performed as described above (see Methods section 4.2), the cell pellet was incubated with the following antibodies:

- CD31-APC/Cy7
- CD45-ECD
- EpCAM FITC
- CD49f-PE-Cy5

After incubation with antibodies, the cells were resuspended with sorting solution and subjected to FACS.

During FACS, $CD31^+$ and $CD45^+$ cells were excluded. As mentioned above, CD31 and CD45 are commonly used markers to exclude endothelial and hematopoietic cells from sorted cell populations.

Flow cytometry based on CD49f and EpCAM staining separated breast tissue cells into stromal cells ($CD49^-$/$EpCAM^-$ cells, the blue cloud in figure 15), and epithelial cells (those in the bigger red cloud in figure 15). Stromal cells were excluded.

Epithelial population includes:

- Basal cells ($CD49f^+EpCAM^{lo/-}$).
- Luminal progenitor or LP ($CD49f^+EpCAM^+$).
- Mature luminal or LM ($CD49f^-EpCAM^+$).

To isolate all epithelial mammary cells, the gating strategy was set to comprise all three subpopulations (Figure 15) [20].

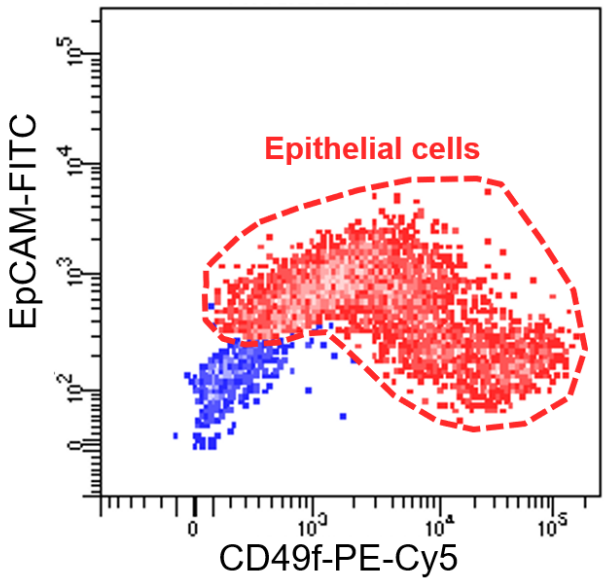

Figure 15: Representative FACS plot of the sorting strategy showing the epithelial cells (red), that were sorted together, and the double-negative stromal cells (blue), that were excluded.

The transcriptome of the selected cellular populations was then sequenced with a bulk approach.

RNA sequencing of a bulk population offers the possibility to obtain a different view from that of the single-cell RNA sequencing.

First of all, in single-cell transcriptome analysis we collect not only epithelial cells, but also stromal cells and cells that go to make up the microenvironment of the breast tissue.

While for the bulk approach, we select only epithelial cells, reducing the sample complexity, but allowing a higher resolution in terms of transcriptome analyses.

Indeed, single-cell transcriptome analyses permits to characterize the expression profiles at an individual-cell level with the drawback of a high frequency of dropout events. The numbers of expressed genes detected from single cells are typically lower compared with population-level ensemble measurements. In fact, only 10-20% of mRNA molecules in the cell lysate are detected. These dropout events are common due to low mRNA concentrations of less expressed genes that are not captured in the reverse transcription process, even if they are not less important.

So, single-cell analyses allow to know more and more complex information about cellular heterogeneity, intra-tumor differences (and not only inter-tumor differences), biological variability, and transcriptional states, that are missed by RNA-seq on pooled cells, but the image obtained can in same case be blurred.

While in bulk RNA-sequencing cell-to-cell differences are missed, cellular heterogeneity can be completely masked, and it often describes an inferred state in which none, or very

few of the cells really exist, but it offers a high output resolution thanks to a greater sequencing depth, in this case limited to sorted epithelial cells. Since we will obtain a characterization of the heterogeneity of the breast tissue ecosystem by the aforementioned sc-RNAseq analyses, we seek here to couple these analyses to bulk RNA sequencing to obtain a greater number of detected and sequenced transcripts: not only the more expressed transcripts, but also genes that are less expressed but equally important for the phenotypic characterization of the cellular populations of the breast (such as transcription factors).

5.6 RESULTS OF SC-RNA ANALYSES FROM BREAST CANCER BRAIN METASTASIS AND PRIMARY TUMOR

Here we report the data emerged from the single cell transcriptome analysis of two samples, a breast tumor brain metastasis, and a primary tumor of the breast, from two different patients.

5.6.1 SC-RNA ANALYSIS OF A BREAST CANCER BRAIN METASTASIS

This brain metastasis was surgically removed from a patient who had previously been diagnosed a breast cancer. The surgical specimen was then processed following the same protocol that we used for the processing of normal samples (Chromium Next GEM), and the single-cell RNA sequencing data were visualized using the so-called Uniform Manifold Approximation and Projection (UMAPs), that is often used for visualization by reducing data to 2-dimensions. After genome expression analysis and comparison, the software recognizes different clusters of cells, organizes them, and gives a number to each of them (Figure 16). The algorithm used to obtain this plot is the Seurat algorithm, that calculates highly variable genes and focuses on these for downstream analysis, collecting cells with similar genome expression within the same cloud of cells.

Through annotation, each cluster can be associated with the respective name of the cellular population that make it up (Figure 17). The algorithm used to do this compares the transcriptome of the analyzed cells with reference transcriptomes. Each cellular cluster of the brain metastasis is composed by a variable number of cells; the numerosity of each cluster is indicated in table VI.

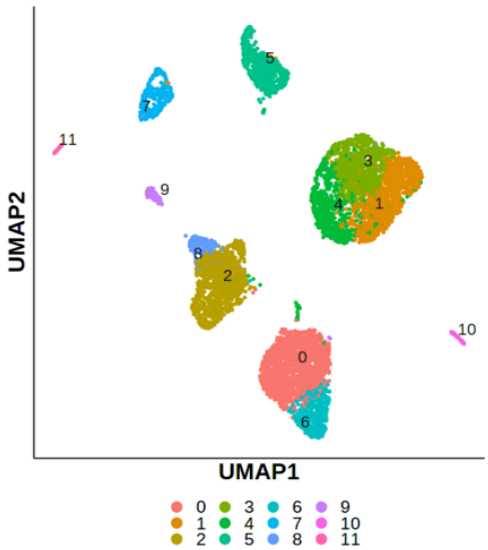

Figure 16: UMAP plot of the brain metastatic sample. Different cellular clusters are identified and plotted based on transcriptional similarities and differences.

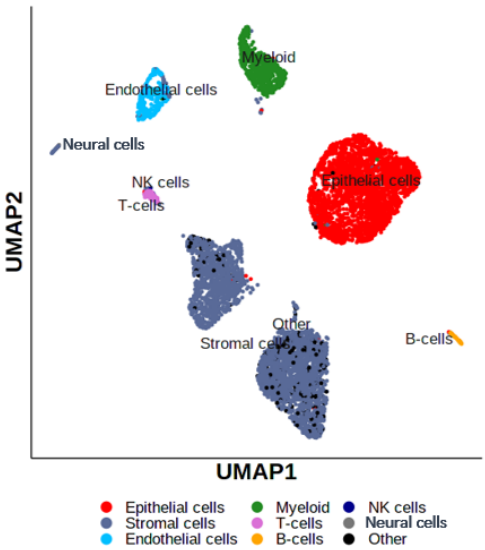

Figure 17: UMAP plot of the brain metastatic sample after annotation. Each cluster is associated with a specific cellular subtype.

Table VI: Number of cells that compose each different cellular cluster of the brain metastatic tissue.

Cellular group	Number of cells
Epithelial cells	2702
Stromal cells	3165
Endothelial cells	297
Myeloid	586
T - cells	117
B - cells	62
NK cells	2
Neural cells	12
Other	247

To characterize the different subpopulations that compose the brain metastasis we questioned the system about the single markers expressed specifically by different type of cells.

Through the analysis of EpCAM expression the epithelial cluster was identified. It is made up by 2.702 cells and it appears as the cluster of cells with the highest expression of EpCAM, as shown in figure 17.

COL1A2 and PDGFRB expression is typical of the stromal cells. COL1A2 gene encodes the pro-alpha2 chain of type I collagen, while PDGFRB gene encodes a protein called platelet-derived growth factor receptor beta (PDGFRβ), a receptor tyrosine kinase. In figures18 and 19 it is possible to find two different clusters that show higher expression of stromal markers. These cells are those most probably correspondent of those that were defined as $CD49^-/EpCAM^-$ cells during flow cytometry (Figure 6).

PECAM1 (Platelet and Endothelial Cell Adhesion Molecule 1, also known as CD31) and FLT1 analysis allows to detect the endothelial subpopulation. FLT1 gene encodes a protein called VGFR or vascular endothelial growth factor receptor 1, an important signaling protein involved in both vasculogenesis and angiogenesis. As regards PECAM1/CD31 (the same marker used during flow cytometry and cell-sorting), it is expressed in two different clusters (Figure 20). The highest expression is revealed by the endothelial cluster (in the upper-left zone of the plot in figure 20). A lower signal is shown by the upper-central cluster of the plot (figure 20), that represent the myeloid cluster. On the other hand, FLT1 is highly expressed only by the endothelial cluster (Figure 21).

Nucleated hematopoietic lineage cells can be selected using the PTPRC gene (Protein Tyrosine Phosphatase Receptor Type C), that is known also as the CD45 antigen. Highest expression of PTPRC was found in both the myeloid cluster and the T/NK cell cluster (Figure 22).

Through CD68 expression, it is possible to differ these two clusters. Those characterized by the expression of CD68 are the myeloid cells. While the little cluster, that was highly positive in the previous graph (Figure 22), and who doesn't show any expression of CD68 in figure 23, represents the T/NK cell cluster

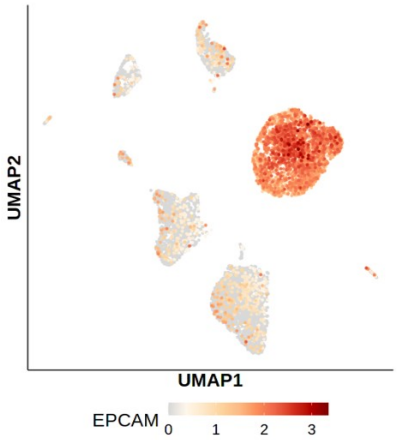

Figure 17: UMAP plot of the brain metastasis showing the expression levels of EpCAM as a heatmap. It is highly expressed within the epithelial cluster.

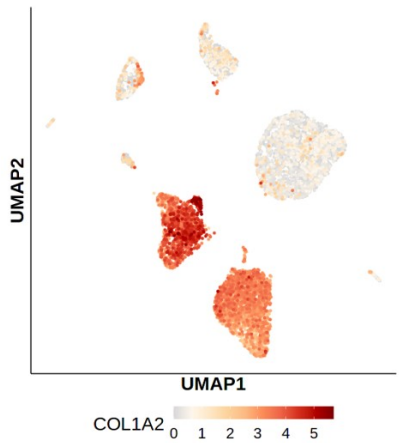

Figure 18: UMAP plot of the brain metastasis showing the expression levels of COL1A2 as a heatmap. It is highly expressed within stromal cells.

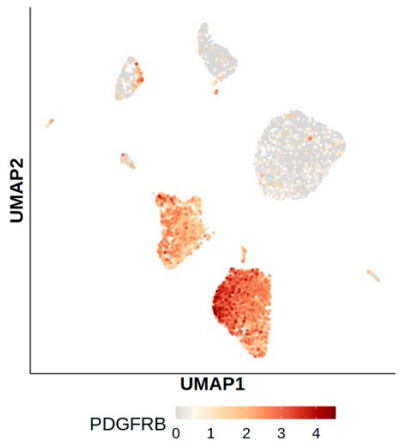

Figure 19: UMAP plot of the brain metastasis showing the expression levels of PDGFRB as a heatmap. It is highly expressed within stromal cells.

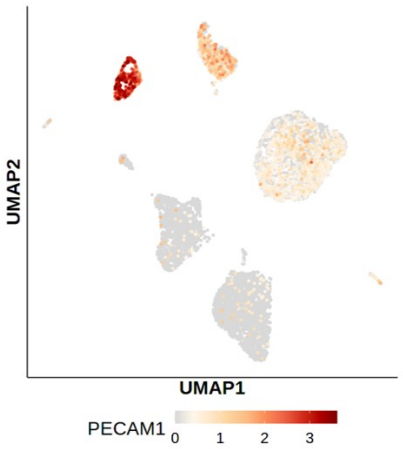

Figure 20: UMAP plot of the brain metastasis showing the expression levels of PECAM1 as a heatmap. It is highly expressed within endothelial and myeloid cells.

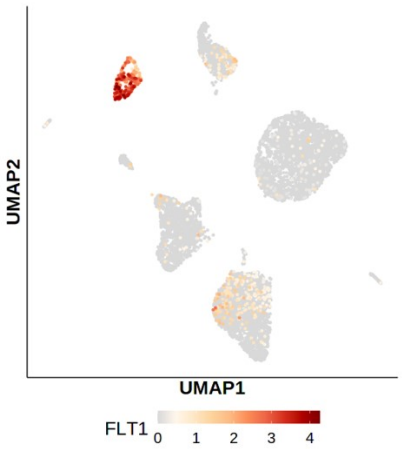

Figure 21: UMAP plot of the brain metastasis showing the expression levels of FLT1 as a heatmap. It is highly expressed within endothelial cells.

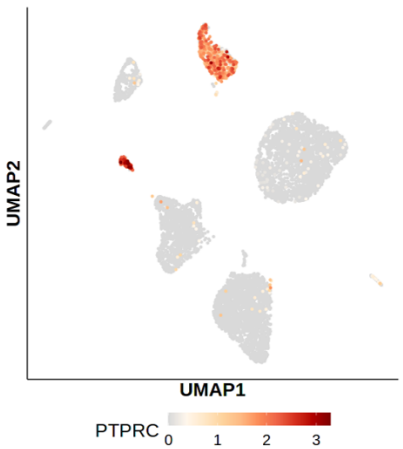

Figure 22: UMAP plot of the brain metastasis showing the expression levels of PTPRC as a heatmap. It is highly expressed within myeloid and T/NK cells.

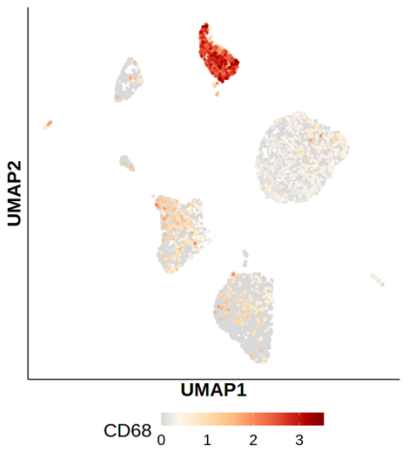

Figure 23: UMAP plot of the brain metastasis showing the expression levels of CD68 as a heatmap. It is highly expressed within myeloid cells.

In sum, we were able to confidently validate the cell population labeling of our analyses.
After identifying the different populations that compose the brain metastasis, and naming each of them, the next step was to detect, at a molecular level, which of these were the tumor cells.
For this, we employed InferCNV. This is used to explore single cell RNA-seq data to identify evidence for large-scale chromosomal copy number variations, such as gains or deletions of entire chromosomes or large segments of chromosomes, typically found in tumor cells. This is done by exploring expression intensity of genes across positions of the genome in comparison to a set of reference normal cells. This reference set is usually based on the cells that are present in the tumor sample, but that have a high probability of not being cancer cells. In our case macrophages and fibroblasts made up the “reference normal cells”.
A heatmap is generated illustrating the relative RNA expression intensities across each chromosome, and it often becomes readily apparent as to which regions of the genome are over-abundant or less-abundant as compared to normal cells.
What is analyzed during InferCNV is the mean RNA-synthesis from different chromosome regions, allowing to infer the presence of copy number variation in those regions were the RNA-levels are so high, or so low, that this cannot be explained by an increased or a decreased gene expression alone. There must be amplifications or deletions of those regions, respectively.

Figure 24 shows the InferCNV of the brain metastasis analyzed. The reference normal cells are represented in the top heatmap, in which no area of high or low RNA-expression is detectable, but only an average expression of RNA. In the bottom heatmap are plotted the tested cells.
The rows of both the heatmaps correspond to the cells, while the columns are genes, ordered by chromosome position.
The normal cells in the top heatmap define baseline expression for genes in normal cells. Removing this baseline of normal expression signal from the tumor cells should reveal those chromosomal regions that have significantly more or less expression than the normal cells, highlighting likely amplified or deleted whole chromosomes or large chromosomal regions.
Some cell clusters of the bottom box don’t show any over or down expression of RNA. These are the ones that refer to the T-cell, the endothelial cells, and the macrophages.

The row of macrophages shows a redder area in correspondence of the column of chromosome 6. This represents the transcripts codified by the MHC locus (major histocompatibility complex), that is physiologically highly expressed from these cells.

Epithelial cells (clusters 1, 3, and 4 in figure 24) are characterized by many amplified and deleted regions, as shown by the more colored areas. This data allows to identify the epithelial cluster as the tumor cells, as it has a typical property of cancer cells: the presence of aneuploidies. So, on the basis of the aneuploidies, that were found through InferCNV, tumor cells concentrate within the epithelial clusters (Figure 25).

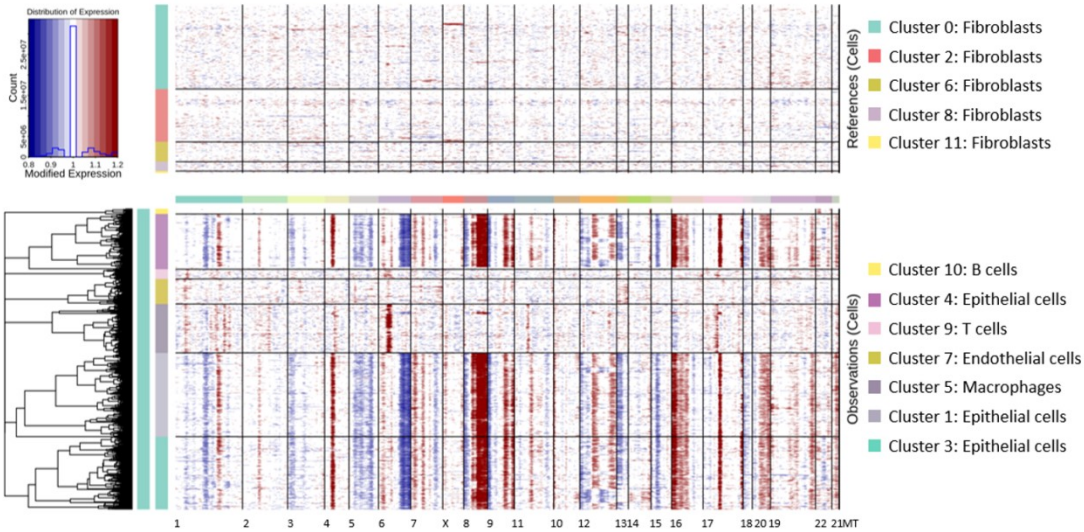

Figure 24: Representative figure showing the InferCNV resulting output of the brain metastasis as a heatmap.

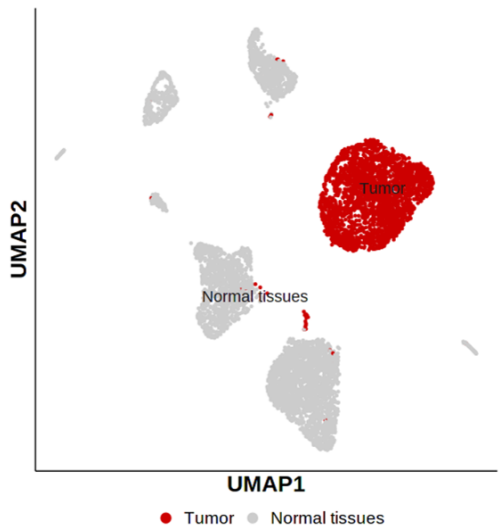

Figure 25: UMAP plot of the brain metastasis showing the cells with high CNV levels as reconstructed by InferCNV. Inferred tumor cells are shown in red.

The next step was to look for the expression of specifical breast cancer markers inside the cancer cells of the brain metastasis, such as the expression of HER2, ER, and PR. As shown in the graphs below (Figure 26) tumor epithelial cells are characterized by a high expression of ERBB2 gene, that encodes the HER2 protein (human epidermal growth factor receptor2). While ESR1, that encodes the estrogen receptor protein, and PGR, that encodes the progesterone receptor, are not highly expressed in this cluster. It is possible to find a low PGR expression within the stromal cluster.

This transcriptomic information allows to underline that this brain metastasis maintains the molecular features of the primary tumor, that was also annotated as HER2+ and HR- at the anatomopathological evaluation. Of note, consistently with our transcriptomic analyses, the anatomopathological characteristics of brain metastatic tissue also confirmed that it was a HER2+/HR- brain metastasis.

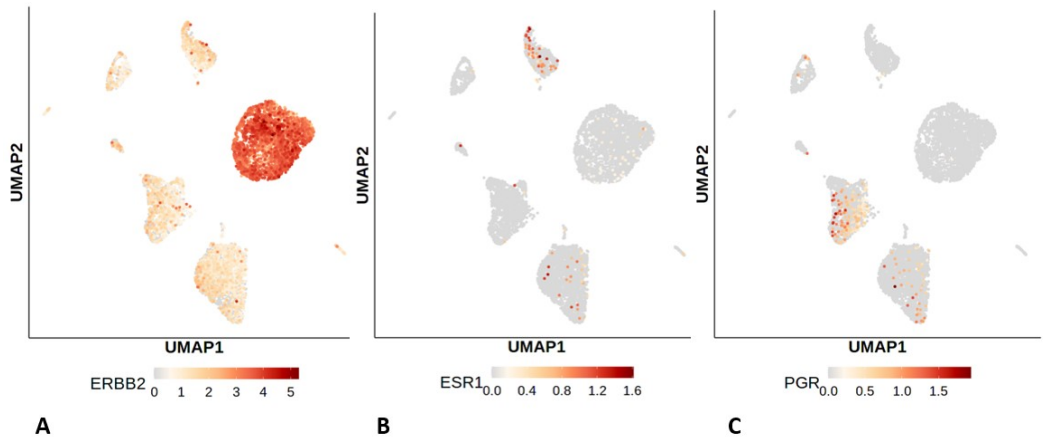

Figure 26: UMAP plots of the brain metastasis showing the expression level of different cellular markers. The epithelial cells of the brain metastasis show a high expression of ERBB2 (A), while ESR1 (B), and PGR (C) are not highly expressed.

Two different cytokeratins (KRT5, KRT18) and mucin 1 (MUC1) were tested to distinguish if epithelial cells of this brain metastasis were of basal or luminal lineage. KRT5 gene encodes the Keratin 5 protein, and its expression is consistent with a basal-like phenotype [26]. This gene is not expressed in our sample (Figure27, UMAP plot A).

KRT18 gene encodes the Keratin 18 protein, that is a luminal-specific cytokeratin [21] [27] [28]. As shown in figure 27, UMAP plot B, this brain metastasis express KRT18, in particular within the epithelial cluster.

As regards MUC1 gene, it codifies for one of the transmembrane mucins, expressed abundantly on the apical surfaces of glandular epithelial cells and some hematopoietic

cells, and it is another marker of breast cancer luminal subtype. This is also expressed by brain metastasis epithelial cells (Figure 27, UMAP plot C).

Of note, HER2 positive breast cancers are typically composed of luminal-like cells [28] and as such, thanks to the detection of HER2, KRT18 and MUC1 expression, it is possible to define this metastasis as a subclone of the primary tumor localized in a distal locus, that maintains the characteristics of the breast primary tumor.

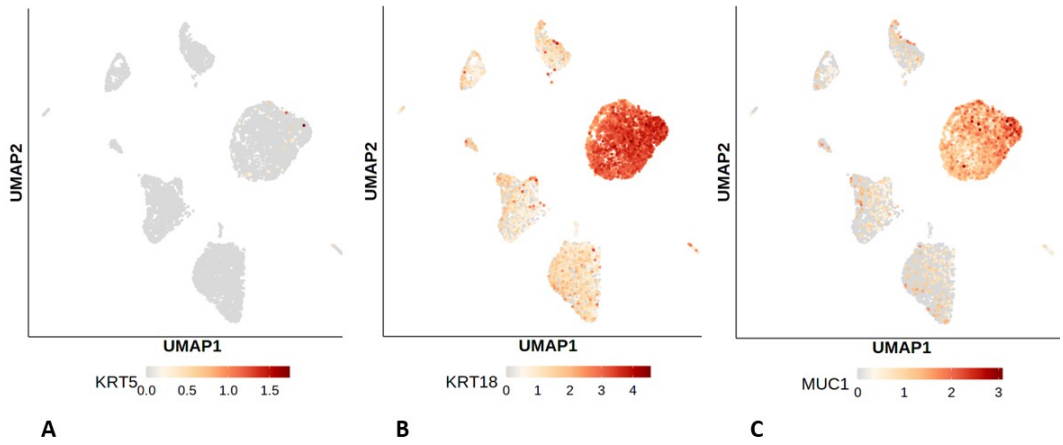

Figure 27: UMAP plots of the brain metastasis showing the expression level of two different cytokeratins and mucin 1. Within the epithelial tumor cells cluster KRT5 is not highly expressed (A), while KRT18 (B) and MUCIN1 (C) are highly expressed.

5.6.2 SC-RNA ANALYSIS OF A PRIMARY BREAST CANCER

We performed single-cell RNA sequencing and data analysis also for a breast primary tumor obtained from a different patient than the previous one.

Chromium Next GEM protocol, Seurat algorithm and UMAP plots were used also in this case.

Figure 28 shows the various numbered clusters of cells that constitutes the tumoral sample. In figure 29, thanks to the annotation, it is possible to associate each cluster with the correspondent cellular subtype.

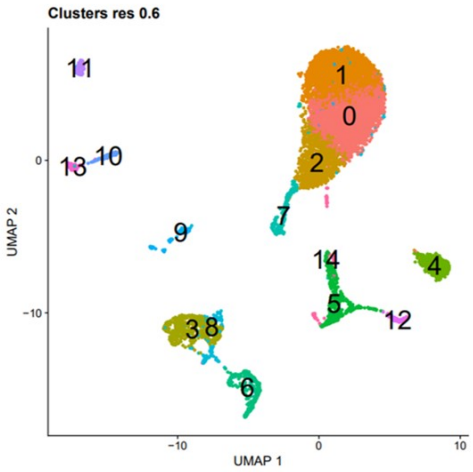

Figure 28: UMAP plot of the breast primary tumor. Different cellular clusters are identified and plotted based on transcriptional similarities and differences.

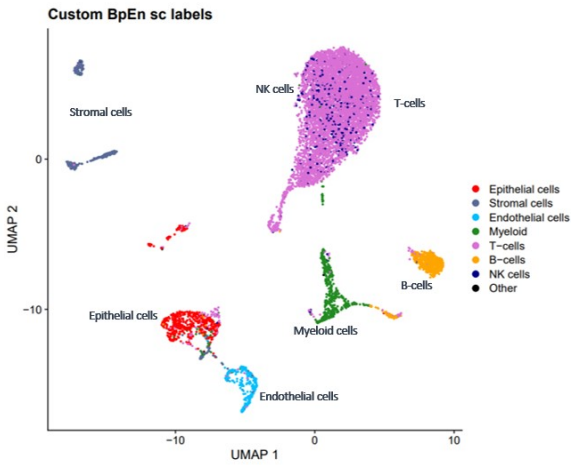

Figure 29: UMAP plot of the breast primary tumor after annotation. Each cluster is associated with a specific cellular subtype.

As regards the markers expressed by the different subpopulations, the expression of same previous genes was tested also in this case.

EPCAM expression is higher inside two different epithelial clusters (Figure 30), one bigger than the other one: cluster 3 and 9, respectively (Figure 28).

The cloud composed by stromal cells expresses the COL1A2 gene (Figure 31). PECAM1 positivity defines the endothelial cluster (Figure 32). The myeloid cluster is identified by CD68 expression (Figure 33). While T-cells and B-cells are detected by the positivity of CD3D (Delta Subunit Of T-Cell Receptor Complex) and IGHM (C region of the mu heavy chain of the immunoglobulins), respectively.

Comparing these graphs (Figure 29 and 34) with that of the brain metastasis (Figure 17), it is possible to realize that there is a clear difference between the numerosity of the T-cell cloud within the primary tumor and that of the brain metastasis. The brain metastasis T-cell population is smaller than that one of the primary breast tumor. This might be explained by the glucocorticoid treatment that the metastatic patient has received before neurosurgery. Indeed, this treatment damages the T-cell population, that is very sensitive to this type of therapy

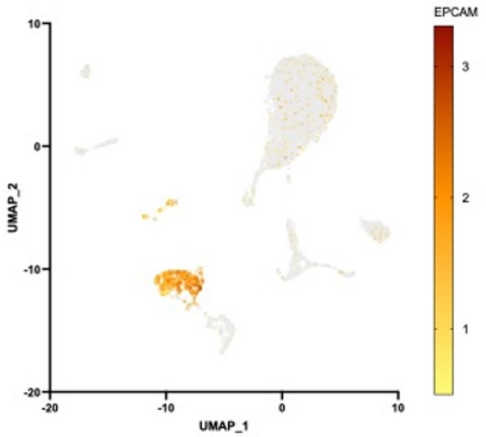

Figure 30: UMAP plot of breast primary tumor showing the expression levels of EpCAM as a heatmap. It is highly expressed within the epithelial cluster.

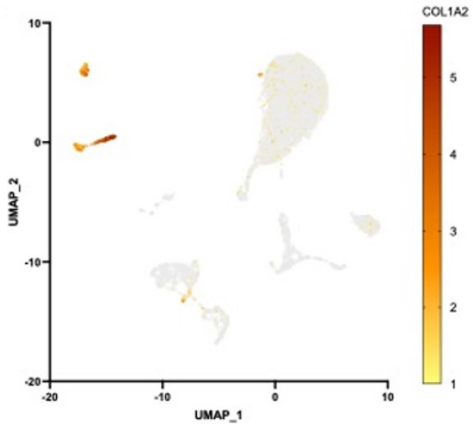

Figure 31: UMAP plot of breast primary tumor showing the expression levels of COL1A2 as a heatmap. It is highly expressed within stromal cells.

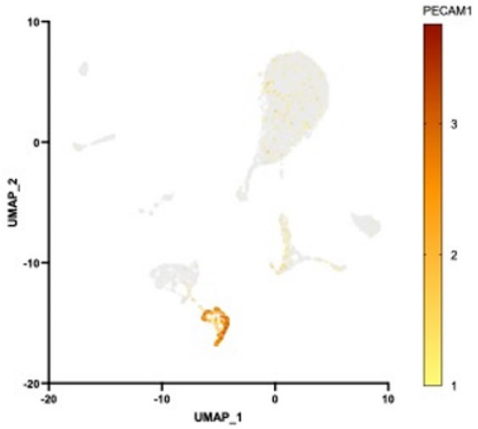

Figure 32: UMAP plot of breast primary tumor showing the expression levels of PECAM1 as a heatmap. It is highly expressed within endothelial cells.

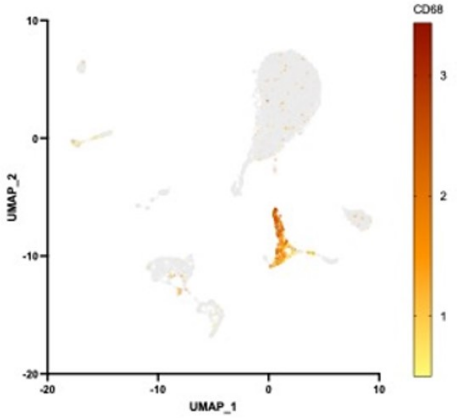

Figure 33: UMAP plot of breast primary tumor showing the expression levels of CD68 as a heatmap. It is highly expressed within myeloid cells.

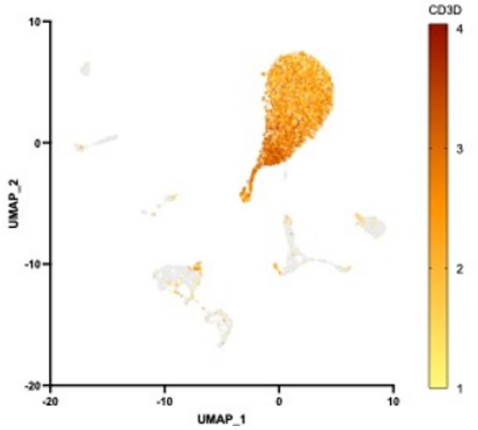

Figure 34: UMAP plot of breast primary tumor showing the expression levels of CD3D as a heatmap. It is highly expressed within T-cells.

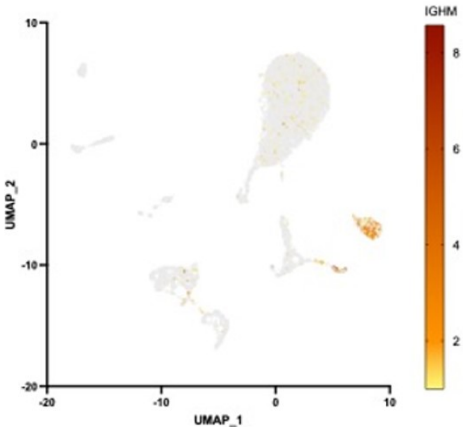

Figure 35: UMAP plot of breast primary tumor showing the expression levels of IGHM as a heatmap. It is highly expressed within B- cells.

InferCNV was performed also for this sample to define the identity of the cancer cells inside the primary tumor. As shown in figure 36, clusters 6 (endothelial cells), 10, 11, and 13 (stromal cells) are used as the reference normal cells. While clusters 3 and 9 (the wo epithelial clusters) are the tested clusters. Cluster 3 shows a high number of chromosomal aberrations (amplifications in red, and deletions in blue), inferred by the RNA-expression levels. While cluster 9 shows only a few alterations.

Consequently, it is possible to say that a big number of tumor cells is present within cluster 3, with a high confidence, because of the high number of aneuploidies. On the other hand, it is not possible to make the same statement for cluster 9, because of low chromosomal alterations.

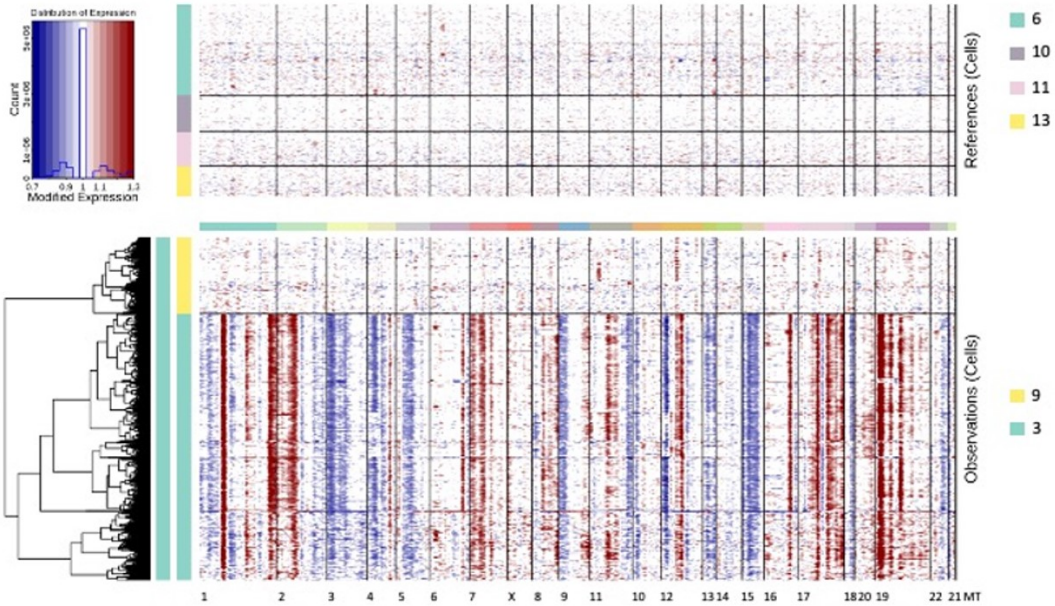

Figure 36: Representative figure showing the InferCNV resulting output of the breast primary tumor as a heatmap.

Analyzing ERBB2 (HER2), cluster 3 shows a high expression of this marker, while cluster 9 presents only a very slight enrichment. (Figure 37).

Considering the InferCNV, and a score limited to ERBB2 expression, we can say that this gene is surely amplified in cluster 3. Differently, in cluster 9 the score is lower, consistently with the lower observed expression (Figure 38).

These results are compatible with two different scenarios: cluster 9 might be composed of early tumor cells, with a low rate of chromosomal aberrations, and a low expression of ERBB2 gene, or cluster 9 might be composed of normal cells that express physiological levels of ERBB2. To discriminate between these two, it will be essential to complete the comparison with our normal tissue sc-RNAseq data. Only through comparing data coming

from a really normal control from healthy patients, and not from hypothetical “reference normal cells”, it will be possible to uncover differences and similarities between normal and tumoral tissue, and to understand what aberrations lead to breast cancer development.

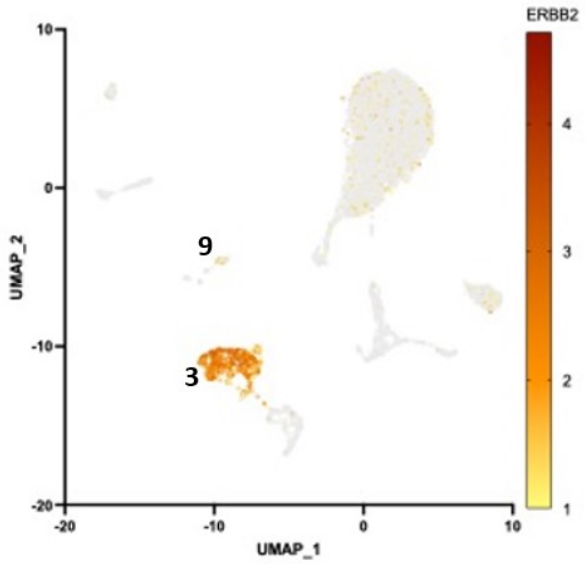

Figure 37: UMAP plot of the breast primary tumor showing the expression levels of ERBB2 as a heatmap. It is highly expressed in the epithelial cluster 3, while it is slightly expressed in the epithelial cluster 9.

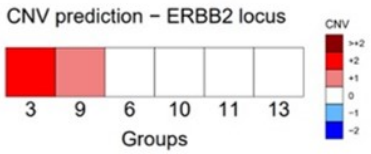

Figure 38: Prediction score for the ERBB2 locus in the clusters indicated by numbers above.

Primary tumor cells were also tested for KRT14 and KRT18 expression. KRT14 encodes a protein called keratin 14, and it is one of the markers of the basal-like subtype of breast cells [4]. In this sample it is not expressed by the cluster 3, while it is expressed within cluster 9 (Figure 39). KRT18, a breast luminal marker, is highly expressed in both clusters (Figure 40). So, within cluster 3 epithelial cells are characterized only by luminal markers, while inside cluster 9 both KRT14 and KRT18 are expressed (both basal-like and luminal markers). These results are compatible with the view that cluster 3, that shows higher amplification of HER2 (Figure 37) is composed of cancer cells that show exclusively luminal phenotype, while cluster 9 is composed of more normal epithelial cells of both basal and luminal lineage

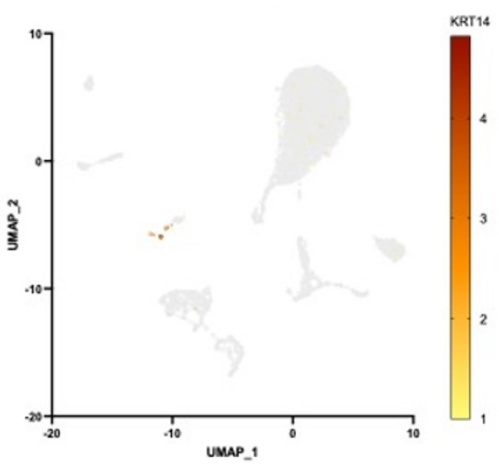

Figure 39: UMAP plot of breast primary tumor showing the expression levels of KRT14 as a heatmap. It is expressed within epithelial cluster 9 only.

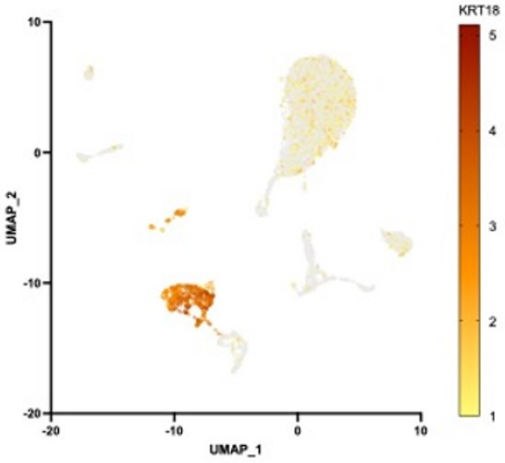

Figure 40: UMAP plot of breast primary tumor showing the expression levels of KRT18 as a heatmap. It is highly expressed within both epithelial cluster 3 and 9.

6.0 CONCLUSIONS AND DISCUSSION

Here we collected normal mammary gland samples from five healthy patients. These tissue samples were processed through mechanical and enzymatic dissociation to reach single cell suspensions. The normal cellular composition of the breast was confirmed thanks to flow cytometer assessment, and the different epithelial cellular clusters were identified: basal cells, luminal progenitor, and mature luminal [19] [20] [21]. Each sample was then subjected to bulk RNA sequencing after cell sorting, or to single cell capture for scRNA. During the scRNA sequencing workflow we carried out the cDNAs and library quality checks and quantification, provided for in the Chromium Next GEM protocol.
RNA sequencing and data analyses of our healthy samples is still in progress.

In this thesis we also introduced the data coming from two pathological samples, that were processed as the healthy samples mentioned above: a breast cancer brain metastasis and a breast primary tumor obtained from two different patients.
Through scRNA sequencing and data analysis we were able to identify the various cellular clusters that made up the pathological tissues and to investigate the expression levels of different markers, that are known as specifically expressed by distinct cellular clusters. Among these, HER2, ER (estrogen receptor), PR (progesterone receptor), and cytokeratins (keratin 5, 14, 18) expression was also investigated, to identify the molecular characterization of our sample.
Tumor cells within the pathologic samples were identified exploring single cell RNA-Seq data to analyze chromosomal copy number variations, thanks to InferCNV. In this way, cellular clusters with high levels of chromosomal aberrations, that are typical of cancer cells, were described as the tumor cells of the sample.
The presence of high RNA synthesis levels or highly expressed specific markers inside a cluster of cells, allows us to infer that this particular cluster exhibits tumor features.
On the other hand, the presence of not highly expressed markers, or an intermediate RNA expression increase does not allow us to draw any conclusion about the healthy or pathological “cell state”.
We could uncover it comparing these data with those coming from the healthy mammary gland samples.

The purpose of our thesis was to collect samples of human mammary glands from healthy patients. Having data coming from healthy controls allows to develop further studies, which will try to give an answer to many aspects that are still unclear.

1) Healthy controls and their data output offer the ideal “normal reference” to identify tumor cells within breast cancer primary tumors and brain metastases.
2) It will be possible to investigate and compare the breast epithelial transcriptional programs between normal tissue, primary tumor and metastasis, uncovering if tumor and metastatic cells keep memory about their origin.
3) As regards metastatic breast cancer development, it will be possible to study if spreading cells, that reach a different tissue than the one from which they are originated, maintain an expression profile quite similar to that one of the tissue of origin, or if they completely change their transcriptional patterns. Or, another point, can be the analyses of putative specific programs that could be activated to allow metastatic seeding in distant sites; maybe, thanks to the acquired capacity to respond to various growth factors, because of the expression of specific receptors that were previous absent.
4) Metastatic microenvironment will be analyzed to discover if the different cells (endothelial, lymphohematopoietic, stromal cells) that compose it, show some differences from those of primary tumors or normal tissues.
5) Another question is if, during metastatic growth, metastatic cells guide the microenvironment of the surrounding tissue where they are seeded to become more similar to that of the primary tumor.
6) Single cell transcriptomic analyses of normal breast tissue allow to infer the development trajectory of the cells, through quantification of spliced and not yet spliced transcripts. It will be probably possible to know if there are specific subclusters more inclined to become tumor cells and to give rise to a breast cancer or metastasis.
7) Crosstalk between breast cells could be investigated, to understand if metastatic cells continue to need integration of signal with the other cells or if they become independent of these.

In conclusion, further studies will be certainly continued to make useful what has been collected up until now, with the purpose of trying to understand the molecular mechanisms and the various pathways that guide breast primary tumors development and spreading within the human body.

7.0 BIBLIOGRAPHY

1 – Stranding S: Gray's Anatomy. The Anatomical Basis of Clinical Practice (forty-first edition) Elsevier Publ, 2016, pp 945-949.

2 – Arendt LM, Rudnick JA, Keller PJ, Charlotte Kuperwasser C: Stroma in Breast Development and Disease; Semin Cell Dev Biol, 2010; 21 (1): 11–18.

3 – Biswas SK, Banerjee S, Baker GW, Kuo CY, Chowdhury I: The Mammary Gland: Basic Structure and Molecular Signaling during Development. Int J Mol Sci, 2022; 23 (7): 3883.

4 – Laakso M, Loman N, Borg A, Isola J: Cytokeratin 5/14-positive breast cancer: true basal phenotype confined to BRCA1 tumors. Mod Pathol, 2005; 18 (10): 1321-1328.

5 – Hall JE Guyton and Hall Textbook of Medical Physiology (twelfth edition). Saunders eds, Elsevier Publ, 2011, p 1014.

6 - Lauby-Secretan B, Scoccianti C, Loomis D, Benbrahim-Tallaa L, Bouvard V, Bianchini F, Straif K: Breast-Cancer Screening — Viewpoint of the IARC Working Group. N Engl J Med, 2015, 372;24: 2353-2358.

7 - Gray GK, Li CMC, Rosenbluth JM, Selfors LM, Girnius N, Lin JR, Schackmann RCJ, Goh WL, Moore K, ShapiroHK, Mei S, D'Andrea K, Nathanson KL, Sorger PK, Santagata S, Regev A, Garber JE, Dillon DA, Brugge JS: A human breast atlas integrating single-cell proteomics and transcriptomics. Dev Cell, 2022; 57 (11): 1400-1420.

8 – Biganzioli L, calabrese M, Conte B, Cortesi L, Criscitiello C, Del Mastro L, Fiorentino A, Levaggi A, Montemurro F, Marchiò C, Tinterri C, Zambelli A: AIOM Breast Neoplasms Guidelines. 2018, p 22.

9 – AIOM, Registri Tumori Italiani, Siapec-Iap, Fondazione AIOM, Passi d'argento, Osservatorio Nazionale Screening: I numeri del cancro in Italia 2021. Available from: https://www.salute.gov.it/imgs/C_17_notizie_5681_0_file.pdf

10 – AIOM, Associazione ITALIANA Radioterapia e Oncologia clinica, A.N.I.S.C., Siapec-Iap, SICO, SIMG, Società Italiana di Radiologia Medica e Interventistica: Linee guida Neoplasie della mammella (edizione 2021). Sistema Nazionale Linee Guida, 2021.

11 – Cserni G: Histological type and typing of breast carcinomas and the WHO classification changes over time. Pathologica, 2020; 112: 25-41.

12– Alessandrelli L, Chiara D: Breast cancer staging - Modalita' di riparto e requisiti di utilizzo del fondo per i test genomici ormonoresponsivo per il carcinoma mammario in stadio precoce. GU Serie Generale n.161, 2021, Allegato 2, 23.

13 - Giuliano AE, Hurvitz SA: Carcinoma of the Female Breast. In: Papadakis MA, McPhee SJ, Rabow MW, McQuaid KR. eds. Current Medical Diagnosis & Treatment 2022. McGraw Hill; 2022.

14 – Hammond MEH, Hayes DF, Wolff AC, Mangu PB, Temin S: American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer. Journal of oncology practice, 2012; 6 (4): 196.

15 - Giuliano AE, Hurvitz SA: Carcinoma of the Female Breast. In: Papadakis MA, McPhee SJ, Rabow MW, McQuaid KR. eds. Current Medical Diagnosis & Treatment 2022. McGraw Hill; 2022.

16 – Suvà ML, Tirosh I: Single-Cell RNA Sequencing in Cancer: Lessons Learned an Emerging Challenges. Molecular Cell 75, 2019; Elsevier.

17 - Chromium Next GEM Single Cell 3' Reagent Kits v3.1 User Guide, Rev C. 10X Genomics, 2019. Available from: CG000204_ChromiumNextGEMSingleCell3_v3.1_Rev_D.pdf (ctfassets.net)

18 – 10x Genomics® Single Cell Protocols - Cell Preparation Guide. 10x Genomics, 2017. Available from: CG000053_CellPrepGuide_RevC.pdf (ctfassets.net)

19 – Eirew P, Stingl J, Raouf A, Turashvili G, Aparicio S, Emerman JT, Eaves CJ: A method for quantifying normal human mammary epithelial stem cells with in vivo regenerative ability. Nature medicine, 2008; 14 (12): 1384-1389.

20 – Pal B, Chen Y, Vaillant F, Capaldo BD, Joyce R, Song X, Bryant VL, Penington JS, Di Stefano L, Ribera NT, Wilcox S, Mann GB, Fab KC, Papenfuss AT, Lindeman GJ, Smyth GK, Visvader JE: A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. The EMBO Journal, 2021; 1-23.

21 – Linnemann JR, Miura H, Meixner LK, Irmler M, Kloos UJ, Hirschi B, Bartsch HS, Sass S, Beckers J, Theis FJ, Gabka C, Sotlar K, Scheel CH: Quantification of regenerative potential in primary human mammary epithelial cells. The company of biologists, 2015; 142: 3239-3251.

22 – Donovan JA, Koretzky GA: CD45 and the immune response. J Am Soc Nephrol, 1993; 4 (4): 976-985.

23 – Vanchinathan V, Mizramani N, Kantipudi R, Schwartz EJ, Sundram UN: The Vascular Marker CD31 Also Highlights Histiocytes and Histiocyte-Like Cells Within Cutaneous Tumors. Am J Clin Pathol, 2015; 143 (2): 177–185.

24 - Krebsbach PH, Villa-Dia LG: The Role of Integrin α6 (CD49f) in Stem Cells: More than a Conserved Biomarker. Stem Cells Dev, 2017; 26(15): 1090–1099.

25 – Salomon R, Kaczorowski D, Valdes-Mora F, Nordon RE, Neild A, Farbehi N, Bartonicek N, Gallego-Ortega D: Droplet-based single cell RNAseq tools: a practical guide. Lab Chip, 2019; 19 (10): 1706-1727.

26 - Zombori T, Cserni G:Immunohistochemical Analysis of the Expression of Breast Markers in Basal-like Breast Carcinomas Defined as Triple Negative Cancers Expressing Keratin 5. Pathol Oncol Res, 2018; 24(2):259-267.

27 - Walker LC, Harris GC, Holloway AJ, McKenzie GW, Wells JE, Robinson BA, Morris CM: Cytokeratin KRT8/18 expression differentiates distinct subtypes of grade 3 invasive ductal carcinoma of the breast. Cancer Genet Cytogenet, 2007; 178(2): 94-103.

28 - Dai X, Cheng H, Bai Z, Jia Li J: Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. J Cancer, 2017; 8(16): 3131–3141.