Narodowy Instytut Onkologii
im. Marii Skłodowskiej-Curie
Państwowy Instytut Badawczy

Rozprawa doktorska

Mgr Małgorzata Szostakowska-Rodzoś

Molecular analysis and clinical value of circulating tumor cells (CTCs) identified via CytoTrack system in metastatic breast cancer patients.

Molekularna analiza oraz znaczenie kliniczne krążących komórek nowotworowych (CTCs) w zaawansowanym raku piersi zidentyfikowanych z zastosowaniem systemu CytoTrack.

Praca wykonana pod kierunkiem:
Promotor: dr hab. Ewa Grzybowska
Promotor pomocniczy: dr Anna Fabisiewicz

2023

Podziękowania

Przede wszystkim chciałabym podziękować moim promotorom, dr hab. Ewie Grzybowskiej oraz dr Annie Fabisiewicz za wieloletnią opiekę merytoryczną oraz mentorską. Umiejętności, wiedza oraz zapał nabyte w trakcie współpracy z moimi przełożonymi przyczyniły się nie tylko do powstania niniejszej pracy doktorskiej, ale również były kluczowe w samorozwoju oraz nabyciu samodzielności naukowej.

Chciałabym również podziękować wszystkim osobom zaangażowanym w część eksperymentalną oraz analityczną tego projektu: dr Urszuli Śmietance, dr Maciejowi Wakule, lek. Sylwii Tabor oraz dr Łukaszowi Szafronowi. Bez wsparcia tych osób powstanie tej pracy nie byłoby możliwe.

Dziękuję również przyjaciołom oraz członkom rodziny, zwłaszcza mężowi, którego anielska cierpliwość oraz niezłomne wsparcie codziennie mnie motywowały i dawały siłę by iść naprzód.

Bez wsparcia wyżej wymienionych osób oraz wielu innych powstanie tej pracy nie byłoby możliwe.

Streszczenie

Wstęp: Luminalny rak piersi jest najczęściej diagnozowanym podtypem biologicznym i charakteryzuje się najlepszym rokowaniem. Jednak u około 40% pacjentek luminalnych pojawiają się odległe przerzuty, radykalnie pogarszające rokowania pacjentek. Monitorowanie zaawansowanego raka piersi (MBC) w celu wykrycia progresji jest ważnym etapem leczenia choroby. Wykrywanie krążących komórek nowotworowych (CTCs) i ich charakterystyka molekularna zyskują na znaczeniu jako narzędzie diagnostyczne, ale nie stanowią standardu klinicznego. Ich znaczenie prognostyczne nie zostało jeszcze ustalone. Celem tej pracy jest ustalenie wartości klinicznej CTCs w zaawansowanym, luminalnym raku piersi oraz ich charakterystyka molekularna.

Metodyka: CTCs wykryto i wyizolowano przy użyciu nowatorskiego, niezależnego od ekspresji EpCAM systemu CytoTrack. W badaniu wykorzystano materiał biologiczny pobrany od pacjentek leczonych w Narodowym Instytucie Onkologii im. Marii Skłodowskiej-Curie – Państwowym Instytucie Badawczym w okresie od czerwca 2018 r. do października 2020 r. Łącznie włączono 237 pacjentek ze zdiagnozowanym zaawansowanym, luminalnym rakiem piersi. Jądrzaste komórki krwi wyizolowane z pełnej krwi barwiono i poddano dalszej analizie za pomocą systemu CytoTrack. Zidentyfikowane CTCs wizualizowano, a uzyskane obrazy użyto do analizy ekspresji EpCAM. Następnie wyizolowano pojedyncze CTC za pomocą mikromanipulacji i poddano amplifikacji całego genomu, a następnie analizie genetycznej z sekwencjonowaniem nowej generacji (NGS).

Wyniki i wnioski: Wykazano, że obecność ≥5 CTCs jest ważnym czynnikiem prognostycznym, pogarszającym rokowania pacjentek. Co więcej, rosnącą liczbę CTC podczas leczenia również wskazano jako niekorzystny czynnik ryzyka. Ponadto rozpoznano, że utrzymująca się niska liczba CTCs (<5 CTC) podczas leczenia jest pozytywnym czynnikiem prognostycznym dla pacjentek z zaawansowanym rakiem piersi z przerzutami. Odkrycia te mają ogromne znaczenie dla poprawy rokowania w przypadku raka piersi z przerzutami oraz mogą pomóc klinicystom w monitorowaniu pacjentów podczas leczenia systemowego. Ponadto opisano różnicę w ekspresji EpCAM między klastrami CTCs

a pojedynczymi CTC. Wysoka heterogenność CTCs w statusie EpCAM podkreśla fenotypową plastyczność pojedynczych komórek. Dodatkowo potwierdzono ogromną heterogenność genomową krążących komórek nowotworowych. Wyniki tego badania podkreślają użyteczność kliniczną wykrywania i zliczania CTCs podczas leczenia chorych na raka piersi z przerzutami.

Abstract

Background: Luminal breast cancer is the most common subtype of breast cancer and has the best prognosis. However, approximately 40% of patients with this subtype can develop distant metastases, dramatically worsening the patient's survival. Monitoring metastatic Breast Cancer (MBC) for signs of progression is an important part of disease management. Circulating tumor cells (CTCs) detection and molecular characteristics gain importance as a diagnostic tool, but do not represent a clinical standard and its value as a predictor of progression is not yet established. This work aims to establish clinical value of the CTCs and their molecular characteristics.

Methods: CTCs were detected and isolated using novel image-based and EpCAM-independent system CytoTrack. In total, 237 patients with diagnosed luminal MBC were enrolled in the study between June 2018 and October 2020. The PBMCs isolated from whole blood samples were stained and further analyzed via CytoTrack system. Identified CTCs were visualized and gained images were used for EpCAM expression analysis. Next, the single CTCs were isolated via micromanipulation and subjected to whole genome amplification, followed by genetic analysis with next generation sequencing (NGS).

Results and Conclusions: The prognostic value of high CTCs count (≥5 CTCs) was maintained during the observation period. Moreover, the rising counts of CTCs during treatment were also identified as the unfavorable risk factor. Furthermore, the constant low CTCs count (<5 CTCs) during treatment was identified as strong favorable factor for prognosis in metastatic breast cancer patients. These findings are highly relevant for improving prognostication in metastatic breast cancer and in helping clinicians monitor patients during systemic therapy. Moreover, in our study we described the difference in EpCAM expression between CTCs clusters and single CTCs. The high heterogeneity of CTCs in EpCAM status, highlights the phenotypic plasticity of single cells. Additionally, this work also confirmed the great genomic heterogeneity of circulating tumor cells. The result of this study highlights the clinical utility of the CTCs detection and enumeration during the treatment in metastatic breast cancer patients.

Abbreviations

ABD	Adaptor-Binding Domain
AF1/AF2	Activation Factor 1 domain/Activation Factor 2 domain
AI	Aromatase Inhibitor
APBI	Accelerated partial breast irradiation
BC	Breast Cancer
BCS	Breast-conserving surgery
BSA	Bovine Serum Albumin
CDK4/6	Cyclin-dependent kinases 4 and 6
CHTH	Chemotherapy
CI	Confidence Intervals
CNV	Copy Number Variation
CT	Computer Tomography
CTC	Circulating Tumor Cell
CTCs	Circulating Tumor Cells
DAPI	4,6-diamidino-2-phenylindole
DMEM	Dulbecco's Modified Eagle Medium
DOP-PCR	Degenerate Oligonucleotide–Primed Polymerase Chain Reaction
DTCs	Disseminated Tumor Cells
EBRT	External beam radiation therapy
ECICB	European Commission Initiative on Breast Cancer
EDTA	Ethylenediaminetetraacetic acid
EMP	Epithelial-Mesenchymal Plasticity
EMT	Epithelial to Mesenchymal Transition
EpCAM	epithelial Cell Adhesion Molecule
ER	Estrogen Receptor

ESMO	European Society for Medical Oncology
FBS	Fetal Bovine Serum
FDA	Food and Drug Administration
FFPE	Formalin-fixed, Parrafin-embedded
HER2	Human epidermal growth factor receptor 2
HR	Hazard Ratio
HTH	Hormone (anti-estrogen) therapy
Hybrid E/M	hybrid Epithelial-Mesenchymal phenotype
IGF1R	Insulin-like Growth Factor-1 Receptor
IHC	Immunohistochemistry
LBD	Ligand Binding Domain
LH	luteinizing hormone
LHRHas	luteinizing hormone-releasing hormone agonists
MALBAC	Multiple Annealing and Looping–Based Amplification Cycles
MBC	metastatic Breast Cancer
MDA	Multiple Displacement Amplification
MMR	Mammography
MRI	Magnetic Resonance Imaging
NGS	Next Generation Sequencing
NST	No Special Type
OS	Overal Survival
PAM50	Prediction Analysis of Microarray 50
pan-CK	pan-cytokeratin
PBS	Phosphate Buffered Saline
PBS	Phosphate-buffered saline
PCR	Polymerase Chain Reaction

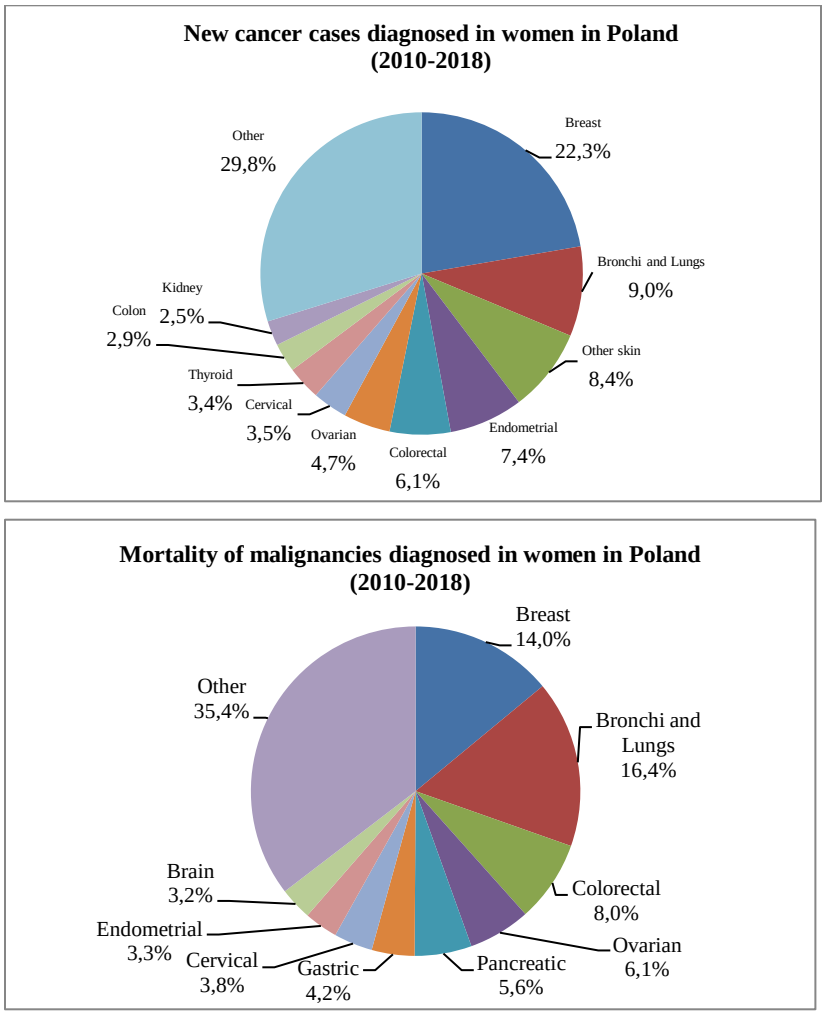
PDK1	phosphatidylinositol 3-dependent kinase-1
PFS	Progression-Free Survival
PI3K	phosphatidylinositol 3-kinase
PIP2	phosphatidylinositol-4,5-biphosphate
PIP3	phosphatidylinositol-3,4,5-triphosphate
PMT	photomultiplier tube
PR	Progesterone Receptor
PT	Primary Tumor
RBD	Ras-binding domain
ROI	Region of Interest
RTH	Radiotherapy
SERD	Selective Estrogen Receptor Downregulator
SERM	Selective Estrogen Receptor Modulator
SNV	Single Nucleotide Polymorphism
TNM	Tumor, Nodules, Metastases
USG	Ultrasonography
VAF	Variant Allele Frequency
VEGF	Vascular Endothelial Growth Factor
WBRT	Whole-breast irradiation
WGA	Whole Genome Amplification

1. Introduction

1.1.Breast cancer

1.1.1. Epidemiology

Breast cancer (BC) is one of the most common female tumors diagnosed worldwide. In Poland in the years 2010-2018 breast cancer represented 22,3% of all cancer cases in women. Metastatic breast malignancy was also the second most fatal cancer in women (Graph 1) [1].



Graph 1. Graph showing morbidity and mortality of malignancies diagnosed in women in Poland in years 2010-2018 [1].

1.1.2. Risk factors and screening

The breast cancer carcinogenesis is still unknown in majority of cases. The most important risk factors are: older age, mutations in particular genes (mainly *BRCA1* and *BRCA2*), family history of breast cancer, long-term hormone replacement therapy, overweight and obesity [2]. The basic imagining diagnostic and screening test for breast cancer is mammography (MMR). MMR is recommended by European Commission Initiative on Breast Cancer (ECIBC) for screening: every 2 or 3 years for women after 45 years of life, every 2 years for women after 50 and every 3 years for women after 70 years of life [3]. Ultrasonography (USG) is recommended by European Society for Medical Oncology (ESMO) as a supplementary screening method for high-density breasts, typical of young women where the value of the MMR is limited. Magnetic resonance imaging (MRI) is not standard screening method; however, it might be valuable supplementary screening method for patients with familiar breast cancer. Therefore, ESMO also recommends annual MRI as a screening method for women with a strong familial history of breast cancer, with or without proven *BRCA1/BRCA2* mutations [4].

1.1.3. Diagnosis

The diagnosis of breast cancer is based on clinical examination in combination with imaging and confirmed by pathological assessment. The clinical examination consists bimanual palpation of the breasts and regional lymph nodes and assessment of risk for distant metastases. Imagining in diagnosis is based on MMR and USG. MRI is recommended only in some cases however, should be considered for patients with: *BRCA1/BRCA2* mutations, strong family history, lobular cancers, large discrepancies between conventional imaging and clinical examination or when the findings of conventional imaging are inconclusive [4]. For right diagnosis microscopic pathomorphological examination is crucial. Material for pathomorphological examination should be collected before start of any treatment. The material should be collected via core needle biopsy with USG guidance. If core needle biopsy is inapplicable, it is advisable to perform an open biopsy and as a last resort fine needle aspiration biopsy. The pathomorphological diagnosis should assess the histological type and grade (Table 1), status of: estrogen and progesterone receptors, Ki67 and HER2, and biological subtype (Table 2). Final microscopic diagnosis should include the pTNM classification

determined by examining all material removed. In cases where the surgery was preceded by a systemic surgery treatment, postoperative stage assessment feature y (ypTNM) must be specified [2]. Diagnosis of metastatic breast cancer (MBC), both newly-diagnosed and recurrent should be done with biopsy form metastatic lesion, to confirm tumors' histological and biological subtype [5]. However, biopsies of bone metastases should be avoided, due to technical limitations of biomarker detection in decalcified tissue. If there are important differences in hormonal receptors status between the primary tumor and the metastasis lesion, it is still not known which biological features should be crucial for the treatment decision making. The final decision for therapy should take into account not only the primary tumor biological characteristics, but also estimated degree of hormonal receptor heterogeneity between metastases and primary tumor. Moreover, during decision making it is also crucial to take into account if the type of received treatment could potentially induce a selection of clones resistant to a specific targeted therapy [5].

Table 1. Histological classification of breast cancer, benign and precursor changes [6].

Type of pathomorphological changes	Category of pathomorphological changes	Histological type	ICD-O
Epithelial benign and precursor changes	Benign epithelial proliferations	Usual Ductal Hyperplasia	
		Atypical ductal hyperplasia	
		Columnar cell lesions including flat epithelial atypia	
	Adenosis and benign sclerosing lesions	Sclerosing adenosis	
		Apocrine adenoma	8401/0
		Micro glandular adenosis	
		Radial scar / complex sclerosing lesion	
	Papillary neoplasms	Intraductal papilloma	8503/0
	Adenomas	Tubular adenoma not otherwise specified	8211/0
		Lactating adenoma	8204/0
		Duct adenoma not otherwise specified	8503/0

	Noninvasive lobular neoplasia	Atypical lobular hyperplasia	
Breast carcinoma *in situ*	Lobular carcinoma *in situ*	Lobular carcinoma *in situ*, not otherwise specified	8520/2
		Lobular carcinoma *in situ*, pleomorphic	8519/2
	Ductal carcinoma *in situ*	Ductal carcinoma *in situ*, non-infiltrating, not otherwise specified	8500/2
	Papillary carcinoma *in situ*	Ductal carcinoma *in situ*, papillary	8503/2
		Solid papillary carcinoma *in situ*	8509/2
Epithelial Breast cancers	Invasive Breast Carcinoma	Infiltrating duct carcinoma, not otherwise specified	8500/3
		Oncocytic carcinoma	8290/3
		Lipid rich carcinoma	8314/3
		Glycogen rich carcinoma	8315/3
		Sebaceous carcinoma	8410/3
		Lobular carcinoma, not otherwise specified	8520/3
		Tubular carcinoma	8211/3
		Cribriform carcinoma, not otherwise specified	8201/3
		Mucinous adenocarcinoma	8480/3
		Mucinous cystadenocarcinoma, not otherwise specified	8480/3
		Invasive micropapillary carcinoma of breast	8507/3
		Metaplastic carcinoma, not otherwise specified	8575/3
	Papillary cancers	Encapsulated papillary carcinoma	8504/2
		Encapsulated papillary carcinoma with invasion	8504/3

		Solid papillary carcinoma with invasion	8509/3
		Intraductal papillary adenocarcinoma with invasion	8503/3
Mesenchymal Breast Cancers	Vascular cancers	Angiosarcoma and Postradiation angiosarcoma	9120/3
		Angioma	9120/0
	Fibroblastic and myofibroblastic tumors	Nodular fasciitis	8828/0
		Myofibroblastoma	8825/0
		Desmoid type fibromatosis	8821/1
		Inflammatory myofibroblastic tumor	8825/1
	Fibroepithelial cancers	Fibroadenoma not otherwise specified	9010/0
		Phyllodes tumor not otherwise specified	9020/1
		Phyllodes tumor, malignant,	9020/3

1.1.4. Biological subtypes

Biological subtypes are classified according to the expression of steroid receptors (estrogen (ER) and progesterone (PR)) and human epidermal growth factor receptor 2 (HER2). Cancers that are positive for estrogen receptor (ER+) and/or progesterone receptor (PR+) are classified as luminal. Currently, more accurate molecular expression profiling is available (PAM50 assay) that enables to distinguish more accurate risk groups (high recurrence risk subtypes) [7, 8].

Table 2. Biological subtypes of breast cancer distinguished by IHC.

Biological subtype	**ER**	**PR**	**HER-2**	**Frequency**	**References**
Luminal A	+ (-)	+/- (+)	- 	40-50% (<3%)	[9-11]
Luminal B	+ (-)	+/- (+)	+/-	20-30% (<3%)	
HER-2 enriched	-	-	+	20-30%	[9, 12]
Basal-like	-	-	-	~15%	[9, 13]

Luminal cancers are distinguished into two subgroups (A and B) according to the expression of ER, PR, HER-2 and Ki-67 (proliferation marker) established by immunohistochemistry (IHC) [14]. Luminal A breast cancers are characterized by high expression of luminal epithelial genes, low expression of Ki-67 and distinct methylation profile of more than 40 genes [15]. The luminal B breast cancers are characterized by higher expression of Ki-67 and lower ER expression, genomic instability, and a higher frequency of *TP53* gene mutations than luminal A cancers. Therefore, they are associated with worse prognosis and higher risk of relapse than luminal A breast cancers. Progress in genetic profiling allowed an additional luminal subtype to be distinguished: type C, which is unrecognizable via IHC. Luminal C subtype is characterized by the overexpression of genes that are characteristic for non-luminal breast cancers, like transferrin receptor (CD71), *MYB*, nuclear protein P40, *SQLE* and *GGH* [7, 9].

The HER2-enriched breast cancers show a gene signature that is closer to the luminal subtypes than basal-like cancers [16]. This subtype is characterized by overexpression of HER-2, high expression of proliferation-related genes like *GRB7* and low expression of basal-related genes like *FOXC1* gene. HER2-enriched breast cancers have also high

genomic instability, with highest ratio of *TP53* and *PIK3CA* mutations from all breast cancer subtypes [17]. Moreover, some studies highlighted that the surface expression of HER2 might play an important role in regulating the luminal cancer stem cell population [18-20].

Basal-like cancers are believed to be most malignant among all subtypes. The metastatic disease in basal-like cancers is more common than in other types of breast cancer. Basal-like cancers also include the triple-negative breast cancers and special histopathological subtypes such as medullary and adenoid cystic tumors [21]. The basal-like subtype shows the second highest number of mutations after the HER-enriched tumors, with many presenting *TP53* and *PIK3CA* mutations. This subtype is characterized by the high expression of proliferation-related genes like *MKI67* and keratins 5, 14, and 17 usually expressed by the basal layer of the skin, and very low expression of luminal-related genes [15]. Moreover, this breast cancer subtype is believed to be most undifferentiated. The ratio of cancer cells with expression of stem-like markers like CD44 and ALDH1 is highest in basal-like cancers [22].

1.1.5. Treatment

The choice of treatment strategy should be based on the tumor burden, location, number of lesions, extent of lymph node involvement and biologic subtype, including biomarkers and genetic signature. Also, patients' age, menopausal and general health status, as well as preferences of the patient should be a foundation for choosing treatment strategy [4].

1.1.5.1. Early Breast Cancer Treatment Recommendations

Treatment of early breast cancer is complex and involves combination of local treatment, systemic anticancer treatments, and supportive measures, delivered in diverse sequences [4].

1.1.5.1.1. Local Treatment

Breast-conserving surgery (BCS) is the primary choice for patients with breast cancer. However, increasing numbers of breast cancer patients are opting for contralateral prophylactic mastectomy (CPM) rather than the breast conservation and mammographic surveillance of the irradiated breast. Most CPMs are performed in patients in low risk of developing contralateral cancer [23]. Most probably, it is associated with patients fears and

overestimating their own level of risk. Many patients cite that the choosing CPM helped them achieving “piece of mind”. Other reasons patients cite for choosing CPM include avoidance of ongoing surveillance, diagnostic procedures or desire for breast symmetry following reconstruction [24, 25]. Despite, the patients beliefs regarding mastectomy and BCS, nowadays studies highlights that the BCS combined with radiotherapy (RTH) for early breast cancer patients is associated with better overall survival than mastectomy, especially for patients with diagnosed cancer in staging T1N0 [26-29]. After the surgery it is crucial to determine the negativity of margins. Careful histological assessment of resection margins is essential. No tumor at the inked margin is required. At least 2mm negative margin for *in situ* disease is preferred [4].

Mastectomy is the procedure of the whole breast/s removal. During the simple mastectomy (total mastectomy) the entire breast is removed. Simple mastectomy is usually conducted on the patients with larger and/or more advanced primary breast cancer. It is also performed in cases where BCS would be difficult to perform for example: large tumor in small breast. This type of mastectomy was also found to be preferred by patients with early breast cancer [23-25].

Types of mastectomy:

- Standard (simple/total) mastectomy – removal of all breast tissue and most of the skin covering it, including nipple.
- Skin-sparing mastectomy – removal of all breast tissue, including nipple, but most of the skin covering breast is left.
- Nipple-sparing mastectomy – it is a type of skin-sparing mastectomy when nipple is not removed.
- Modified radical mastectomy – removal of all breast tissue and skin covering it, including nipple, with additional removal of the lymph nodes in the armpit.
- Radical mastectomy – removal of all breast tissue and skin covering it, including nipple, with additional removal of the large muscle behind breast and the removal of the lymph nodes in the armpit.

1.1.5.1.2. Radiotherapy

Radiotherapy (RTH) is recommended after local treatment. Whole-breast irradiation (WBRT) is most common practice after BCS and mastectomy. The standard whole-breast irradiation dose is 50Gy in 25 fractions over 5 weeks. The NCT02295033 phase III clinical trial was set to test the benefits of boosting the whole-breast RTH with 16Gy extra radiation dose. Results from this clinical trial indicated that there was no significant difference between OS and PFS between patients treated with boosted RTH and no boosted RTH. However, it has shown that younger patients, age <50, treated with boosted RTH had lower risk of local or ipsilateral recurrence [30]. Accelerated partial breast irradiation (APBI) is an alternative RTH approach applied after BCS. APBI can be delivered using: brachytherapy, intraoperative RTH or conformal external beam radiation therapy (EBRT) (reviewed: [31]). Clinical trials NCT02104895 and NCT00402519 have shown that brachytherapy and EBRT after BCS are not significantly different from WBRT in terms of PFS, OS and local recurrence for low-risk early breast cancer patients. But both techniques were beneficial for patients in terms of health-related quality of life [32, 33]. Therefore, APBI might be considered for patients with low-risk early breast cancer.

1.1.5.1.3. Neoadjuvant therapy

Neoadjuvant therapy is the administration of therapeutic agents before a main treatment. It should be based on the predicted sensitivity to particular treatment types, the benefit from their use and an individual's risk of relapse. This should take into account biological subtype of breast cancer, staging and other clinical features of the primary tumor (Table 3). Selection of most beneficent therapy should also incorporate short- and long-term toxicities of proposed treatment, the patient's biological age, general health status, comorbidities and preferences. Data has shown that neoadjuvant treatments efficacy decreases when it is administered >12 weeks after surgery [34]. Therefore, it is important that the therapy would be started without any undue delays.

Table 3. Recommendations on neoadjuvant treatment in early breast cancer.

Biological subtype	Recommendations	Additional recommendations	References
Luminal A	Endocrine therapy	Chemotherapy if: >grade 3 tumor or higher >involvement of 4 or more lymph nodes > high 21-gene RS, if available > 70-gene high risk status, if available	[4, 35]
Luminal B (HER2-)	Chemotherapy followed by Endocrine therapy		[4]
Luminal B (HER2+)	Chemotherapy with anti-HER2 treatment followed by Endocrine therapy	If contraindications for the use of Chemotherapy, consider Endocrine therapy with anti-HER2 treatment	[4]
HER2-enriched	Chemotherapy with anti-HER2 treatment		[4]
Basal-like	Chemotherapy		[4]
Special histological types			
Endocrine responsive: cribriform, tubular and mucinous	Endocrine therapy		[35]
Endocrine non-responsive: apocrine, medullary, adenoid cystic and metaplastic	Chemotherapy		[35]

1.1.5.2. Luminal Metastatic Breast Cancer Treatment Recommendations

Luminal breast cancer is diagnosed in the majority of newly diagnosed breast cancer patients. Patients with diagnosed early luminal breast cancer are treated with endocrine (anti-estrogen) therapy. However, as this subtype is diagnosed in >70% of patients, luminal breast cancer metastatic relapse sheer numbers larger than other breast cancer subtypes. It has been shown that frequency of distant metastases in luminal breast cancer patients is even 2-times higher than in other breast cancer subtypes [36]. Mainly, it is associated with disease progression and resistance to endocrine treatment. For metastatic breast cancer patients, there are two types of resistance to endocrine treatment: primary (*de novo*) and secondary (acquired). Primary endocrine resistance is considered for patients who relapse during the first 2 years of neoadjuvant anti-estrogen treatment or who will have progression within the first 6 months of first-line anti-estrogen treatment. Secondary endocrine resistance is defined as relapse during neoadjuvant endocrine therapy, that occurs after the first 2 years of treatment or relapse within 12 months of completing treatment or progression at least 6 months after initiating endocrine treatment for MBC [37].

Standard of care first-line treatment for metastatic luminal breast cancer are CDK4/6 inhibitors combined with endocrine treatment. The CDK4/6 inhibitors combined with endocrine treatment are characterized by similar or even better efficiency than chemotherapy for luminal MBC. Moreover, this treatment is also associated with less toxicity, making it the preferred treatment for most patients [38, 39]. CDK4/6 inhibitors were found to be effective in *de novo* or recurrent MBC patients, in cases of primary or secondary endocrine resistance, in postmenopausal and premenopausal women [5]. According to ESMO guideline endocrine therapy alone in the first-line setting should be reserved for the small group of patients with comorbidities or a performance status that prevents the use of CDK4/6 inhibitors. However, there are no clinical or biomarker data that can help to identify patients suitable for only endocrine treatment. In patients who required first-line chemotherapy due to imminent organ failure or who did not have access to a CDK4/6 inhibitors in the first-line setting, it is clinically acceptable to use anti-estrogen treatment plus a CDK4/6 inhibitor as a subsequent therapy in cases of progressive disease [5].

The optimal sequence of endocrine-based therapy is uncertain after progression on CDK4/6 inhibitors treatment. It should be dependent on what agents were used previously, duration of response to previous therapy, disease burden, patient preference and treatment availability. In patients who relapse after first-line endocrine + CDK4/6 inhibitors treatment determination of *PIK3CA* and *ESR1*, as well as *BRCA1* and *BRCA2* mutations status should be assessed [5]. According to ESMO recommendations, for patients with detected mutations in *PIK3CA* (exons 7, 9 or 20) and previous AI treatment, the recommended therapy option should be alpelisibe-fulvestrant treatment. For patients with *BRCA1/BRCA2* mutations PARP inhibitors monotherapy should be considered [5].

1.2.Endocrine therapy

1.2.1. Estrogen receptor

In luminal breast cancer, the crucial pathways associated with tumor growth are regulated by estrogen receptor α (ERα). The estrogen receptor α acts as a ligand-dependent (estrogen-dependent) transcription factor for genes associated with cell survival, proliferation and tumor growth, like: genes for insulin-like growth factor-1 receptor (IGF1R), cyclin D1, anti-apoptotic BCL-2 protein, vascular endothelial growth factor (VEGF) [40]. ERα is built of several functional (N-terminal domain, DNA-binding domain, ligand-binding domain) and activation (AF1, AF2) domains. Activation domains are involved in the recruitment of co-activators and co-repressors for certain target genes. Activation domain AF2 located at the functional ligand-binding domain (LBD) is the main “activation switch” for tumor growth pathways. AF2 is built of four helices (H9-H12) that are placed the end of ligand-binding domain. Those helices are flexible, and their conformation is dependent on the ligand presence [41, 42]. In the presence of the ligand helix 12 (H12) bends in the conformation that activates binding of the co-activators and co-repressors of pathways regulated by ERα, including pathways associated with tumor growth (Figure 1).

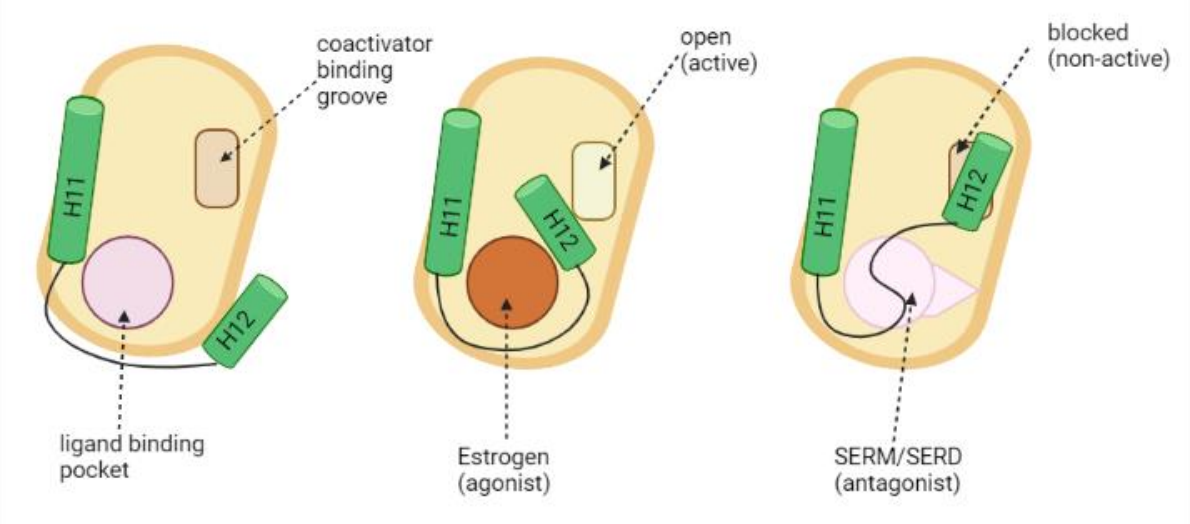


Figure 1. Estrogen receptor α in three conformations: without ligand, when ligand binding pocket is empty and coactivator binding groove is inactivated (no H12 influence); with bounded agonist (estrogen) in ligand binding pocket, the H12 orientation activates and opens coactivator binding groove; with agonist (SERM/SERD) bounded in ligand binding domain, when H12 orientation blocks coactivator binding groove.

The main aim of the endocrine therapy is to stop the ERα-regulated pathways by blocking ERα function on receptor level or on estrogen secretion level. Endocrine therapy is standard first-line treatment for early luminal breast cancer patients, as well as for luminal metastatic breast cancer patients.

1.2.2. Selective estrogen receptor modulators (SERMs)

While the DNA-binding domain of ERs is highly conservative, the ligand-binding domain shows different specificities depending on the tissue [43]. Selective estrogen receptor modulators (SERMs) bind to the ligand-binding domain of ERs. They can act as agonists, partial agonists or antagonists of ERs [44] depending on the tissue and specificity of LBD. In the breast, SERMs act as ERα antagonists [45].

Tamoxifen is considered to be the first targeted therapy and the first SERM [46]. Nowadays, it is a "gold standard" method of treatment for ER+ breast cancer. Tamoxifen reduces the risk of breast cancer recurrence by 50% and the risk of mortality by 25%. It is the only SERM used in the treatment of premenopausal patients [47].

Toremifene differs from tamoxifen by only 1 chlorine atom. It is metabolized differently and is less potent when binding to ER than tamoxifen. It is recommended in the treatment of postmenopausal patients with ER+ or ER unknown metastatic breast cancer [48].

1.2.3. Selective estrogen receptor down-regulators (SERDs)

Fulvestrant is the only FDA approved SERD. While SERMs could act as agonists or partial agonists of ER depending on the tissue (e.g., bones, endometrium, liver), fulvestrant is a pure ER antagonist. Fulvestrant induces conformational changes, leading to ER degradation [49]. It is efficient in the treatment of breast tumors resistant to SERM or aromatase inhibitors. It is recommended for the treatment of postmenopausal patients with advanced breast malignancy [50].

1.2.4. Aromatase inhibitors (AIs)

Aromatase inhibitors (AIs) block the aromatase enzyme, which catalyzes the conversion of androgens to estrogens. AIs are used as adjuvants or first-line treatments in postmenopausal patients with advanced and/or metastatic ER+ breast cancers [51]. As AIs reduce estrogen secretion, female patients could report side effects associated with a lack of estrogen (e.g., weight gain, infertility) or even premature menopause. For this reason, AI treatment is usually combined with luteinizing hormone-releasing hormone agonists (LHRHas). LHRHas enhance the concentration of luteinizing hormone (LH), which has a positive impact on ovarian functions when estrogen synthesis is inhibited. LHRHa increases disease-free survival (DFS) and reduces the risk of infertility and premature ovarian failure in breast cancer patients treated with AIs [51].

Anastrozole and **letrozole** are non-steroidal aromatase inhibitors. They are type II inhibitors, and as such are generally reversible; estrogen blockade is dependent on the continuous presence of the drug [52]. Anastrozole can significantly reduce the risk of metastatic breast cancer in 53–65% of high risk cases [53].

Exemestane is a steroidal aromatase inhibitor. This inhibitor interacts with the substrate-binding site of the enzyme and is identified as a type I inhibitor. The inactivation

of aromatase is irreversible and remains until a new enzyme is secreted [52]. The efficiency of treatment with exemestane is comparable to anastrozole [53].

1.3.Resistance to endocrine therapy

Despite endocrine therapy being one of the most effective treatment strategies for luminal breast malignancy, up to 40–50% of luminal breast cancer patients will eventually relapse [54]. The resistance to endocrine therapy is associated with the activation of an ERα-independent proliferation mechanisms.

1.3.1. ESR1 modifications

Aberrations in the *ESR1* gene can lead to changes in the functionality and activity of ERα. Point mutations in exons encoding ligand-binding domains can affect the receptor's ability to bind the ligand and/or activity of AF2. The crucial structural element for estrogen-dependent AF2 activity is an α-helix, named 'helix 12' (H12). When LBD is bound to the estrogen, H12 unveils the binding place for ERα co-activators, leading to tumor growth. However, when bound with SERM or SERD, H12 changes its orientation and covers the binding place for ERα co-activators [55, 56], which prevents them from binding. The length of helix 12 is different for agonist- and antagonist-bound structures; agonist-bound structures start at D538 and antagonist-bound structures at L536 [55]. Substitutions of amino acids in positions 536, 537, 538 can lead to changes in H12 orientation, leading to constant activation of ERα.

Point missense mutations in the ligand-binding domain result in conformational changes, placing H12 in the agonist (activated) position. This leads to constitutive activity of ERα in the absence of the ligand. Moreover, the position of H12 is not only activating the ERα, but also is decreasing its ability to bind the ligand. A decrease in ERα ability to bind the ligand is associated with the stability of the mutated conformation [55-57]. If the conformation of H12 is stable, the ligand presence would not disrupt the LBD structure. However, less stable conformation may break when ligand would try to bound. The most common mutations in *ESR1* associated with H12 misplacement occur between 536-538AA of the ERα protein (Figure 2). The point mutation c.1609T>A in exon 8 results

in the substitution of Tyr→Asn at position 537 of the full protein (Y537N) [58] while c.1610A>C in exon 8 leads to the Y537S (Tyr537Ser) substitution [57]. These changes result in different curves of helix 12. High stability of mutated conformations causes a decrease in the ERα ability to bind the ligand and SERMs. The efficiency of ERα activity in Y537S mutants is equal to 100% of maximum activity of non-mutated ER bound to E2 [57]. Another missense mutation in LBD, c.1613A>G (Asp538Gly, D538G), was recently detected in both patients and cancer cells *in vitro* [56, 59]. While aspartic acid contains a fairly large negatively charged sidechain, glycine contains only a hydrogen substituent as its sidechain. Thus, the D538G change affects the tertiary structure of the ERα and decreases the ability of ERα to bind estrogen and SERMs or SERD [56]. The mutation c.1608TC>AG leads to the substitution L536Q (Leu536Gln). This aberration induces overall disorder in the secondary structure of H12 and a decrease in its helical content. While the mechanism and conformational changes are different from the D538G aberration, the consequences are similar [60].

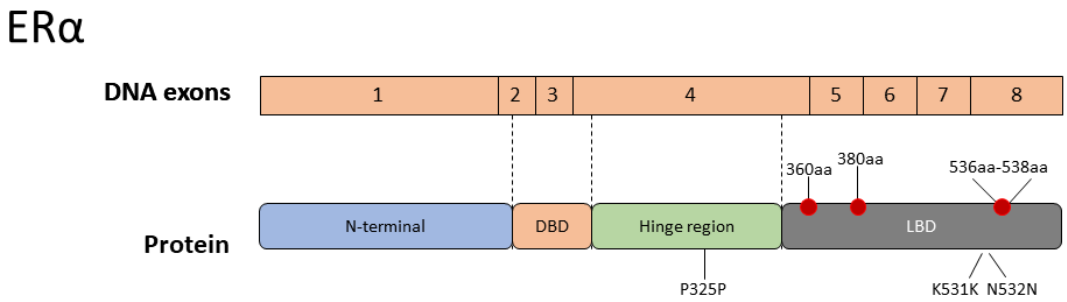


Figure 2. Scheme of the exons in *ESR1* gene and full protein (ERα) domains. The main polymorphisms and main mutation hot spots are highlighted.

Table 4. Frequency of most common *ESR1* mutations in breast cancer patients [61].

***ESR1* mutation**	**AA substitution**	**Frequency in breast cancer patients (%)**	**References**
1609 T>A	Y537N (Tyr537Asn)	5-33	[58, 60]
1610 A>C	Y537S (Tyr537Ser)	13-22	[57, 60]
1613 A>G	D538G (Asp538Gly)	14-36	[59, 60, 62]
1608 TC>AG	L536Q (Leu536Gln)	<5	[60]
1138G>C	E380Q (Glu380Gln)	<7	[62]

1.3.2. PI3K-AKT-mTOR pathway

The phosphatidylinositol 3-kinase/protein kinase B/mammalian target of rapamycin (PI3K-AKT-mTOR) pathway plays a crucial role in cell growth, proliferation, motility, survival and angiogenesis. Therefore, activation of this pathway is usually associated carcinogenesis and resistance to the many types of treatment [63].
The phosphatidylinositol 3-kinases (PI3K) are family of intracellular lipid kinases that are activated by cell stress and/or growth factors. The class I PI3Ks are most studied, because of their association with carcinogenesis and resistance. Class I PI3Ks are heterodimers consisting of a p85 regulatory subunit and a p110 catalytic subunit (p110α, p110β, p110γ, or p110δ). The role of the regulatory subunit p85 is to inhibit catalytic subunit p110. Regulatory subunit p85 contains SH2 domain, divided to 3 parts: nSH2, iSH2 and cSH2. All catalytic subunits are composed of N-terminal adaptor-binding domain (ABD), the Ras-binding domain (RBD), the putative membrane-binding domain (C2), and the helical and kinase domains. The role of iSH2 of p85 unit is to bind ABD of p110 unit. This tight binding alone is not enough to inhibit the p110 subunit. Therefore, the stabilization from nSH2 and cSH2, that are binding to phosphorylated Tyr-X-X-Met motifs plays crucial role in activity regulation of PI3K. It both inhibits the basal activity and facilitates activation by binding phosphotyrosine peptides [64]. After activation catalytic subunit of PI3K phosphorylate phosphatidylinositol-4,5-biphosphate (PIP2) to produce phosphatidylinositol-3,4,5-triphosphate (PIP3). Then, PIP3 binds to a number of downstream mediators, like AKT and phosphatidylinositol 3-dependent kinase-1 (PDK1) [65]. Upon activation, AKT

can phosphorylate substrates that are part of the cell cycle regulation system, leading to increased proliferation and cell motility. Moreover, AKT also activates the mTORC1 and mTORC2, which are the regulators of the protein synthesis in cell [66]. In luminal breast cancer activation of PI3K-AKT-mTOR pathway is associated with cellular growth independent from ERα – by alternative pathway (Figure 3).

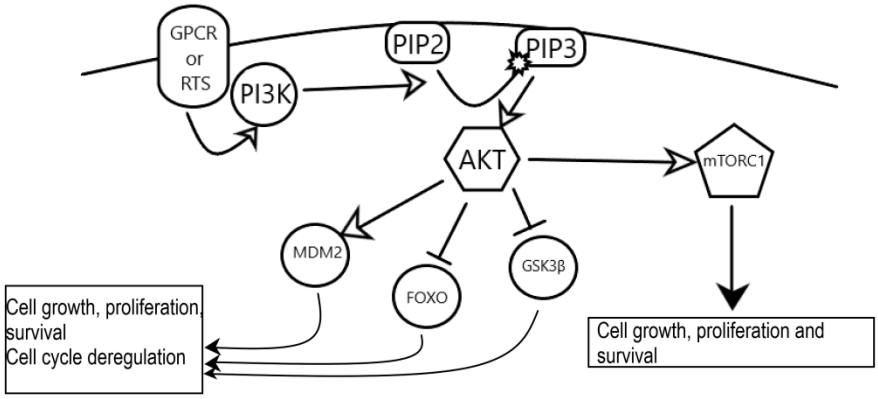


Figure 3. PI3K-AKT-mTOR pathway. PI3K phosphorylates PIP2 to PIP3, which induces AKT activation. AKT phosphorylates various proteins in this pathway, including MDM2, FOXO, GSK3β, and mTORC1. Phosphorylation by AKT inhibits FOXO and GSK3β, while it activates MDM2 and mTORC1, leading to cell cycle deregulation, cell growth and proliferation.

PI3K subunit p110α, encoded by *PIK3CA* gene is one of the most frequently mutated genes in breast cancer. Mutations in this gene occurs in 13-61% of patients, dependent on advancement level and detection method [67, 68]. Around 80% of mutations in *PIK3CA* are located in three hotspots: 542aa, 545 aa (in helical domain) and 1047aa (in kinase domain) [69] (Figure 4). Mutations in these two domains were found to act synergistically but independently [70, 71]. Substitutions in the helical domain: E542K and E545K have similar mechanism. The negatively charged glutamic acid is changed by positively charged lysine leading to change in polarity of the amino acid. The amino acids 542-545 are located near the place when p110α subunit is bound to SH2 of p85 subunit. Change of the charge in this area leads to different initial conformation of the protein [72, 73] where p85 regulatory unit is shifted away from p110α. The substitution in the kinase domain: H1047R is known to overactivate the PI3 kinase by dynamic changes in kinase domain. These changes increase basal ATPase activity as well as expose the membrane binding regions [74, 75]. It has also been shown that the C-terminal region, where kinase domain resides, is essential

for catalysis. The C-terminal region of the protein enhances membrane binding, while inhibits the activity of the enzyme in the absence of the membrane [76]. Mutation in C-terminal region H1047R is associated with higher membrane binding as the mutant protein accumulates positive charge in regions that contact the cell membrane. Moreover, slight change in the polarity and conformation associated with this substitution is causing lost in crucial intracellular interactions between catalytic and regulatory part of protein [77].

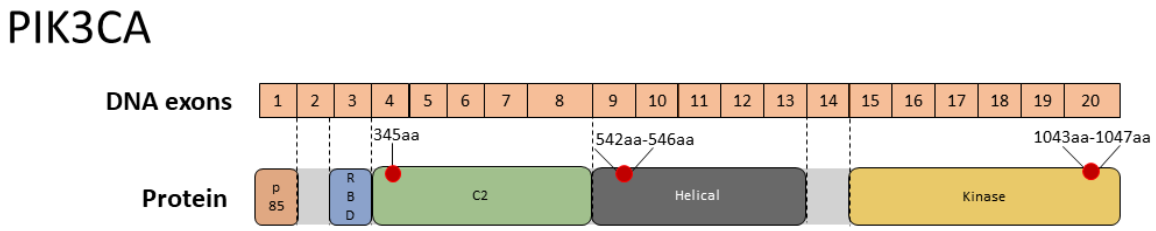


Figure 4. Scheme of the exons in *PIK3CA* gene and full protein domains. The main mutation hot spots are highlighted.

Other frequently mutated gene in PI3K-AKT-mTOR pathway is *AKT1* gene, encoding AKT1 kinase. AKT1 is one of the 3 closely related serine/threonine-protein kinases (AKT1, AKT2 and AKT3) called the AKT kinase. These kinases are involved in regulation process of cell metabolism, proliferation, survival, growth and angiogenesis [78-80]. The AKT1 kinase contains 3 domains: PH domain, protein kinase domain and C-terminal domain. The frequency of the most common *AKT1* mutation: E17K in breast cancers ranges from 4-8%. The E17K mutation results in the modified charge on the 17aa in the PH domain of AKT1 protein, resulting in change of lipid specificity by greatly increasing the affinity for PIP2 [81].

1.4.Progression

There are two main models describing progression in breast cancer: parallel progression model and linear progression model (Figure 5). Parallel progression model presumes that invasion is associated with passive shading. This may take place at early stages of development, like the angiogenic switch. In contrast linear progression model states the evolving primary tumor gives rise to metastases due to increasingly aggressive and invasive phenotypes of tumor cells inside primary tumor. This model is in agreement with the crucial role of the epithelial-to-mesenchymal transition (EMT) in cancer cells. Linear progression assumes that invasion is caused by cells that gained ability of active stroma degradation and mesenchymal subtype, that enables cells to migrate in a bloodstream [82-85]. Although linear progression model seems to be well established, there are growing evidence indicating that the onset of cancer metastasis occurs much earlier in tumor development than is generally indicated by clinical staging of primary tumors [86]. It is most likely that both mechanisms: passive shading and active EMT are associated with metastases formation, depending on the advancement of tumorigenesis.

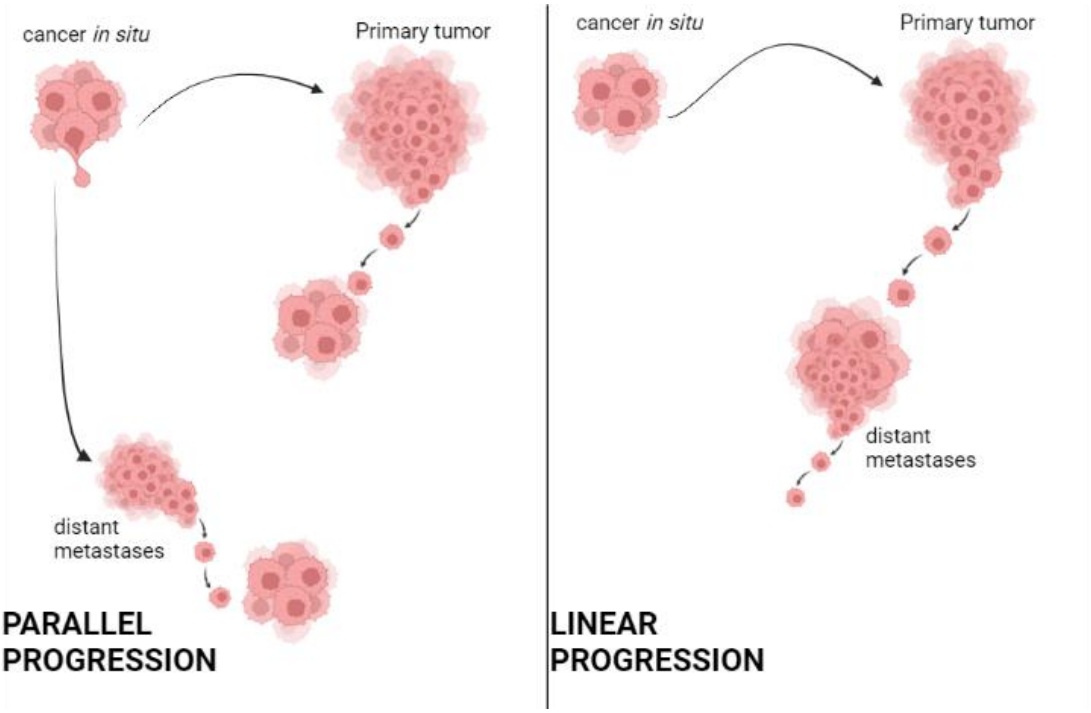


Figure 5. Parallel and linear progression model scheme.

1.4.1. Luminal breast cancer dormancy

Luminal breast cancer (ERα+) is known to have more favorable prognosis in ten-year survival that other subtypes. However, ten-year survival does not fully reflect the long-term survival and mortality of these cancers. In particular, in patients with luminal breast cancers late relapses are common. Late relapse in woman with ERα+ breast cancers is associated with the activation of dormant tumor cells at metastatic sides [87]. Dormant tumor cells display the growth arrest, which precludes proliferation, which lowers their sensitivity to the cytotoxic treatment. The presence of dormant, disseminated tumor cells (DTCs) has been shown to be associated with poorer prognosis, with 40-60% of patients with DTCs detected in bone marrow suffering from metastatic disease [88, 89]. Thus, the detection of DTCs may be important for prognosis. However, the main challenge is ability to correctly assess probability of DTCs re-activation, leading to metastatic progression. Detailed mechanism of dormant cells reactivation is still unknown, but proposed models identify many potential triggers for this process, like: hormonal signaling, angiogenesis switch, immunosurveillance, neutrophil extracellular trap (NET) formation, interaction with extracellular matrix (ECM) and a crosstalk with stromal cells [90-93].

1.4.2. Epithelial to Mesenchymal Transition (EMT)

The epithelial-to-mesenchymal transition (EMT) is a developmental program that enables stationary epithelial cells to migrate and invade distant lesions. It is a multistage process associated with dynamic changes in morphology, cytoskeleton and adhesion. Epithelial cells are characterized by strong cell-cell interactions and extracellular matrix-cell interactions that are related to adhesive molecules, like: E-cadherin or EpCAM, called epithelial markers. During EMT the level of expression of epithelial markers decreases. At the same time, the expression level of mesenchymal markers, like N-cadherin and Vimentin, is increasing. Molecular changes during EMT are controlled by several signaling pathways (TGF-β1, WNT, NNOTCH, HEDGEHOG) and transcription factors (TWIST1, SNAIL1, SLUG, ZEB1 and/or FOXC1/2) [94, 95].

Cells after EMT are more motile, invasive and more resistant to anoikis. Thus, it is believed that EMT process is crucial for CTCs survival in bloodstream and migration to distant lesions. However, recent research demonstrated that EMT may not be essential for metastasis in breast cancer [96]. Recently, there is growing evidence that EMT process might be more complex and that cells that undergo EMT might in fact represent different spectra of epithelial and mesenchymal properties. Therefore, epithelial-mesenchymal plasticity (EMP) is considered to better describe the complexity of the molecular and phenotypical changes that cancer cells undergo [97-99]. EMP enables cells to gain mesenchymal traits without necessary losing the epithelial phenotype. This phenotype when cells co-express epithelial and mesenchymal markers was described in literature as hybrid epithelial-mesenchymal phenotype (hybrid E/M). Hybrid E/M phenotype was found to have more metastatic potential than full mesenchymal or epithelial phenotype [100, 101]. Distribution of epithelial and mesenchymal markers is linked to the breast cancer phenotype. Circulating tumor cells (CTCs) identified in luminal breast cancer patients are predominantly epithelial, while for the other subtypes mesenchymal [100].

Interestingly, while the association of epithelial-mesenchymal changes with cancer progression seems to be obvious, it is unclear which phenotypes may be prognostic for treatment response, progression free survival (PFS) and overall survival (OS). Some studies highlighted that the presence of CTCs with mesenchymal phenotype is associated with poor outcome and shorter PFS [102, 103]. On the contrary, other studies highlighted that presence of epithelial CTCs might be prognostic for breast cancer patients [104, 105].

1.4.3. Circulating Tumor Cells (CTCs)

Formation of metastases is associated with tumor cells being able to migrate and settle in distant parts of the body. Cancer cells migrating through the bloodstream are called circulating tumor cells (CTCs). These cells may origin not only from primary tumor but also from metastatic lesions. Therefore, CTCs may reflect changes in disease dynamics, including changes in sensitivity/resistance to treatment. Also, the sheer number of CTCs circulating in the blood of a patient may be statistically significant factor in assessing the outcome. In breast cancer, CTCs are detected in about 20–30% of early breast cancer patients and around 35-60% of advanced breast cancer patients [106, 107]. Most probably, the higher percentage

of CTCs detected in advanced patients is associated with more metastases lesions that might be the source of CTCs presence in blood. Interestingly, breast cancer biological subtype was found to have no impact on absolute CTC number as well as CTC positivity rate. However, very high CTC counts are observed more frequently in luminal A breast cancer subtype [107].

CTCs might migrate through bloodstream in two forms: single CTCs and CTC clusters. Single CTCs are very rare and occur in the amounts of even 1 CTC in 10^5–10^7 mononuclear blood cells [108]. Most of CTCs occur as circulating single cells however, they might also appear as multicellular (>3 cells) groups, called clusters, composed exclusively of tumor cells (homotypic) or tumor cells and immune/stromal cells (heterotypic). CTC clusters has been shown to be even ten times rarer events than single CTCs, but most probably they have higher metastatic potential [109]. Detailed mechanism of CTC clusters origin is still unknown. Research with the use of intravital microscopic imaging showed aggregation of individual tumor cells resulting in cluster formation. These aggregates were characterized by expression of the breast cancer stem cell marker CD44 [110]. In contrary, Aceto et al. [109] proposed that CTC clusters do not origin from single cell clustering, but rather from passive shedding of tumor fragments, firmly connected by cell-cell junctions.

Because CTCs are derived from both primary and metastasis lesions, analysis of genotype and phenotype changes and differences between CTCs may reflect the dynamic changes and intra-tumor heterogeneity. Currently, genotype analyses of single CTCs are focused on the most common tumor driver mutations and hotspot mutations associated with resistance to targeted therapies, like hormonal therapy for luminal breast cancer patients. According to COSMIC database most frequently mutated genes in luminal breast cancer patients include: *PIK3CA*, *TP53*, *GATA3*, *ESR1*, *CDH1* and *AKT1* [111]. It has been shown that mutations in single CTCs compared to primary tumor reflect intra-tumor heterogeneity, with highest number of somatic mutations detected in *TP53* gene [112, 113]. Moreover, phenotypical changes in single CTCs have been detected. It has been demonstrated that for 77% of metastatic luminal breast cancer patients, with ERα+ primary tumor, detected CTCs were ERα- [114]. For luminal breast cancer patients switch in ERα expression might be associated with acquired resistance to hormonal treatment and activation of estrogen-independent mechanism of proliferation and survival.

1.4.3.1. CTCs detection

Despite the growing knowledge about CTCs, there are still not many techniques that allow them to be counted and analyzed effectively. Even though the new methods are being developed there is still lack of standardization, compatibility, and reproducibility in CTCs research. Currently, only CellSearch [115] and Parsortix [116] systems of CTCs detection are approved by Food and Drug Administration (FDA) for clinical use in metastatic breast cancer patients. The CellSearch method is based on immunoaffinity to EpCAM, which might be lower for cells after EMT or in cells with hybrid E/M phenotype. Therefore, there is a need for development and validation of CTCs detection methods independent of EpCAM expression. Nowadays many methods aiming this approach are verified. One of the newly developed methods is CytoTrack system. This system combines flow cytometry and fluorescent microscopy and enables to create custom protocols. In *in vitro* settings CytoTrack and CellSearch were found to have similar sensitivity of CTCs detection [117].

Immunoaffinity based methods are using positive or negative selection for CTCs identification and isolation. These methods harvest only enriched cellular fraction, so they cannot be used for single cell analysis. In contrary, imagining-based approaches do not include any specific enrichment apart from erythrocyte removal. They are based on immunofluorescence staining with antibodies of choice and are as specific as the antibodies used to detect CTCs. There are also methods based on the physical properties of CTCs, like Parsortix [116, 118]., usually based on the size differentiation, as in most solid tumors tumor cells generally are bigger than blood cells. However, these methods do not filter the bigger blood cells, like macrophages. Also, smaller CTCs subpopulations might be lost using size-filter methods [119-121].

1.5. Single cell analysis

Solid tumors, especially breast cancers, are characterized by high intra-tumor heterogeneity. As primary tumor is composed of various cancer cell subpopulations it is a great challenge to select subpopulation with most metastatic potential. Currently, mostly bulk downstream analyses of cancer primary lesions are routine. However, this approach doesn't enable to identify different cancer cell subpopulations. Therefore, single cell research seems to be the key to a better understanding the intra-tumor heterogeneity. Development of new technologies like next generation sequencing (NGS) and whole genome amplification (WGA) enabled the downstream high throughput analyses on single-cell level.

1.5.1. Whole Genome Amplification (WGA)

Typically, single cancer cell contains ~6-12pg of DNA and ~10-15pg of RNA, depending on the cell type and ploidy [122]. Therefore, the first step in single cell sequencing analyses is single cell whole genome amplification (WGA). However, WGA techniques may introduce bias and artifacts compared to an unamplified material. Many different methods for this procedure have been published to date [123-126] , only few are used widely: Degenerate Oligonucleotide–Primed Polymerase Chain Reaction (DOP-PCR), Multiple Displacement Amplification (MDA) and Multiple Annealing and Looping–Based Amplification Cycles (MALBAC).

The principle of DOP-PCR is to use degenerate primers that contain random six-bases at the 3′ end and constant 5' end sequence. During initial amplification, the primers bind to the DNA template at a low annealing temperature. After initial amplification annealing temperature is elevated. At this stage products which form initializing phase are amplified with the primers targeting the constant 5′ sequence at a higher annealing temperature. The concentrations of the primers and polymerase directly affect the result of DOP-PCR [127]. DOP-PCR WGA often yields low genome coverage, because of the risk of PCR bias, like overamplification and/or underamplification of some regions in the genome [128].

The principle of MDA method is to use random hexamers as starters and ϕ29 DNA polymerase in the amplification reaction. The ϕ29 DNA polymerase is a highly processive

DNA polymerase with strong strand displacement activity, 3′→5′ exonuclease activity and proofreading activity [129]. During MDA reaction random hexamers are elongated producing branched structures, which are further extended by other primers and eventually form multi-branched structures. MDA has better genome coverage than DOP-PCR. However, like DOP-PCR, is also an exponential amplification process, which results in sequence-dependent bias, causing overamplification in certain genomic regions and underamplification in other regions [130].

MALBAC is a unique method of quasi-linear amplification, which reduces the sequence-dependent bias exacerbated by exponential amplification. The clue of MALBAC method is to amplify only the original genomic DNA template by protecting the amplification products during reaction. In the first step of the reaction semi-amplicons are formed by extension of the primers. After that semi-amplicon are melted from a template. Next, the hairpin structure is made thank to complimentary 5’ ends of primers. This process is repeated 8-12 times in pre-amplification reaction. The quasi-linear amplification at these first few cycles is critical for avoiding the sequence-dependent bias exacerbated by exponential amplification. The final amplification is done by standard PCR with the use of primers complementary to 5’ ends of pre-amplification primers. MALBAC method is unique because of the quasi-linear pre-amplification step, which results in two major advantages comparing to DOP-PCR and MDA: accuracy for CNV detection and a low false negative rate for SNV detection [123, 130].

1.6. Clinical approaches for CTCs characterization

Observing and analyzing the number and phenotypes of CTCs during treatment may reveal the nature of tumor lesions, disease evolution and treatment response. In breast cancer, the potential prognostic value of CTCs number was demonstrated by many studies [107, 131-133]. Meta-analyses of CTCs detection studies concluded that CTCs numbers represent a significant risk factor for both PFS and OS in metastatic breast cancer patients [132, 134-136]. For most studies the cut-off value for clinical utility is ≥5CTCs detected. However, some research highlighted that for HER2+ breast cancer patients, seer presence of CTCs might have prognostic value [137]. Moreover, the clinical value of detected clusters is still under validation. Recent study showed that longitudinal evaluation of CTC and CTC

clusters improves prognostication and monitoring in patients with metastatic breast cancer. The prognostic value of CTC and CTC clusters presence was also found to increase over time, suggesting that changes in CTCs numbers during treatment might be clinically relevant [138-140].

As mentioned before phenotype of CTCs might also reflect intratumor heterogeneity and response to treatment. Therefore, some studies are focused on finding the association between CTCs phenotypes and clinical outcome. Some studies highlighted that the presence of CTCs with mesenchymal phenotype is associated with poor outcome and shorter PFS [102, 103]. However, other studies highlighted that presence of epithelial CTCs might be prognostic for breast cancer patients [104, 105]. More advanced downstream techniques like single cell sequencing enables to study not only phenotype but also genotype and expression patterns in single CTCs. This approach might be very useful for identification of circulating cancer stem cells, metastasis-initiating cells in a bloodstream or genetic changes associated with resistance to therapy, like *ESR1* or *PIK3CA*. Novel research showed that mutations in *ESR1* found in CTCs are associated with shorter PFS in luminal breast cancer patients [141]. As the mutations is *ESR1* gene are known to be associated with resistance to hormonal treatment identification of patients with these mutations might guide treatment decisions for another line of treatment.

Tumor heterogeneity is a great challenge for therapeutic decisions, as it is unknown which cells from all primary tumor subpopulations will migrate to invade distant lesions. Moreover, according to the clonal evolution hypothesis tumor cells constantly evolve, possibly in response to the treatment [142, 143]. This evolution might also involve only small population of migrating cells. Thus, the analysis on the single cell level might be helpful in recognition of small cancer cells subpopulations with activated resistance systems. Therefore, treatment decisions should be updated according to changing phenotypes of the cancer cells. Nowadays the treatment decisions are mostly based on analysis of the primary tumor or eventually on biopsy material. However, there are metastases sites that are inaccessible or with limited access due to the invasive procedures. In these cases, the analysis of biopsy material might be impossible. Therefore, analysis of circulating tumor markers, like CTCs, may provide solutions and improve treatment decisions.

2. Aims of the project

The aims of this project are to:

1. Optimize and asses the recovery ratio of the EpCAM independent CTCs detection system CytoTrack.
2. Analyze of the phenotype and genotype of single CTCs and CTC clusters detected via CytoTrack system.
3. Estimation of prognostic value of CTCs and changes in the CTCs numbers during the treatment.
4. Analysis of intra- and interpatient heterogeneity of CTCs.

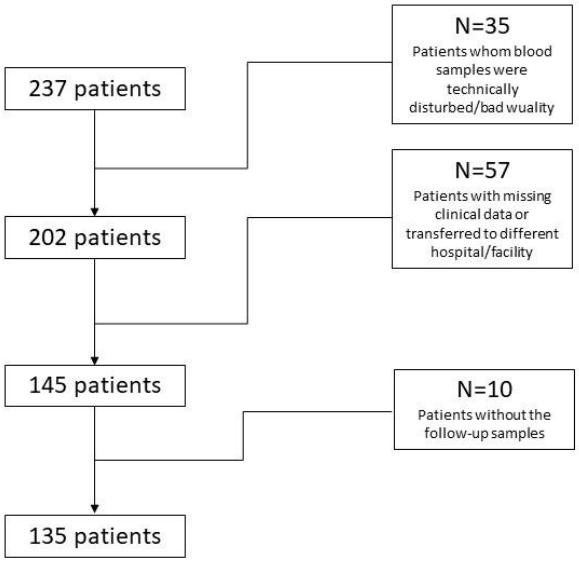
3. Materials

3.1.Biological material

3.1.1. Human samples

Patients diagnosed with luminal metastatic breast cancer (MBC) were selected for the study. Patients' selection was performed in close cooperation with the experienced clinicians from the Department of Breast Cancer and Reconstructive Surgery of National Research Institute of Oncology. The study protocol was approved by the National Research Institute of Oncology Ethics Committee (34/2016). The inclusion criteria for patients were age >18, ongoing hormonal treatment and identification of distant metastases. All participants signed informed consent. The blood collections were done three times during treatment in the 3-month intervals. Overall, 539 samples were collected from 237 patients. For final analysis we included only patients with good quality samples and complete medical record (Graph 2).

For spike-in experiments blood was collected from heathy volunteers. Formalin-fixed, paraffin-embedded (FFPE) primary cancer samples were collected form the archive of the Pathology Department in National Research Institute of Oncology.



Graph 2. Criteria of selecting patients for final analysis.

3.1.2. Cell culture

- MCF-7 (HTB-22) cell line (ATCC) human epithelial cell line isolated from the breast tissue of a 69-year-old, white female patient with metastatic adenocarcinoma
- Dulbecco's Modified Eagle Medium (DMEM), low glucose (Thermo-Fisher, cat. no. 11885084)
- Fetal Bovine Serum (FBS), heat inactivated (Thermo-Fisher, cat. no. A3840102)
- Trypsin-EDTA (0.25%) (Thermo-Fisher, cat. no. 25200072)

3.2. CytoTrack analysis

- BD FACS™ Lysing Solution (BD Biosciences, cat. no. 349202)
- Pan Cytokeratin Monoclonal Antibody (AE1/AE3) conjugated with Alexa Fluor™ 488, (eBioscience™, Thermo-Fisher, cat. no. #53-9003-82)
- EpCAM (CD326) Monoclonal Antibody (323/A3) conjugated with PE (Thermo-Fisher, cat. no. #MA5-3871)
- CD45 Monoclonal Antibody (HI30) conjugated with APC (eBioscience™, Thermo-Fisher, cat. no. #17-0459-42)
- 4',6-Diamidino-2-Phenylindole Dilactate (DAPI) (Thermo-Fisher, cat. no. D3571)
- Bovine Serum Albumin (BSA), heat shock fraction (Merck, cat. no. A7906)
- Saponin (Merck, cat. no. A7906)
- Glycerol, for fluorescent microscopy (Merck, cat. no. 56-81-5)
- N-Propyl Gallate, for microscopy (Merck, cat. no. 02370)
- Fixogum, rubber cement (Marabu, Germany)
- Phosphate Buffered Saline (PBS), pH=7,4 (137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4)
- Tris-HCl, pH=8,5
- Blocking Buffer: 1xPBS, 1% BSA
- Permeabilization buffer: 1xPBS, 0,25% Saponin, 0,5% BSA
- Antibody Dilution buffer: 1xPBS, 0,5% BSA, 0,2% Saponin
- Moun Medium: 1:4 Tris-HCl: Glycerol, 0,5% N-Propyl Gallate

3.3.Single Cell analysis

- MALBAC Single Cell WGA kit (Yikon Genomics, cat. no. KT110700150)
- Agilent DNA 1000 Kit for Bioanalyzer (Agilent, cat. no. 5067-1504)
- Agilent High Sensitivity DNA Kit for Bioanalyzer (Agilent, cat. no. 5067-4626)
- AmpliSeq™ Library PLUS (Illumina, cat. no. 20019102)
- Custom Library (Illumina, cat. no. 20020495)
- AmpliSeq™ CD Indexes Set A (Illumina, cat. no. 20019105)
- MiniSeq™ Mid Output Kit (300-cycles) (Illumina, cat. no. FC-420-1004)
- Phix Control v3 (Illumina, cat. no. FC-110-3001)
- MagSi-NGS Prep PLUS (5 ml) (Magtivio, cat. no. MDKT00010005)
- QuantiFluor® ONE dsDNA System (Promega, cat. no. E4871)

3.4.FFPE analysis

- QIAamp DNA FFPE Tissue Kit (Qiagen, cat. no. 56404)
- BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo-Fisher, cat. no. 4337455)
- PCR primers (Table 5)

Table 5. Primer sequence for sequencing PCR

Gene	Forward primer	Reverse primer	Product length
ESR1 8 exon	5'- TCTGTGTCTTCCCACCTACAGT-3'	5'- ATGCGATGAAGTAGAGCCCG-3'	200bp
ESR1 5 exon	5'- GCTTTGTGGATTTGACCCTCCA-3'	5'- AGAGCAAGTTAGGAGCAAACAG-3'	135bp
ESR1 4 exon	5'-TGAAACACAAGCGCCAGAGA-3'	5'- CCAGGTTGGTCAGTAAGCCC-3'	255bp
PIK3CA 9 exon	5'- AGCTAGAGACAATGAATTAAGGGA -3'	5'- TCCATTTTAGCACTTACCTGTGAC -3'	130bp
PIK3CA 20 exon	5'- AACTGAGCAAGAGGCTTTGGA -3'	5'- CAATCGGTCTTTGCCTGCTG -3'	200bp

4. Methods

4.1. Cell culture

For optimization of the staining and analysis condition as well as reproducibility and sensitivity of the method the cells from the cancer cell line MCF-7 were used. The cells were cultured in DMEM Low-Glucose Medium with 10% FBS, in 37°C and 5% CO_2. The medium was changed every 2 days and after gaining ~80% density the cells were passaged. The cells used for the spike-in experiments were between 17-25 passage. The cells were counted in Cell Counter (Bio-Rad). The counted cells were further used in spike-in optimization experiments.

4.2. CytoTrack analysis

CytoTrack method was chosen, because at the start of the study it was the only method without enrichment associated bias. In this method all nuclear blood cells without previous enrichment are stained, mounted on the glass disk and scanned with an automated fluorescence scanning microscope. During this procedure, the stained cells are excited with a laser at 488nm. The signals are detected by a photomultiplier tube (PMT) and the positions on the disk with possible CTCs are recorded as hotspots [144] (Figure 6).

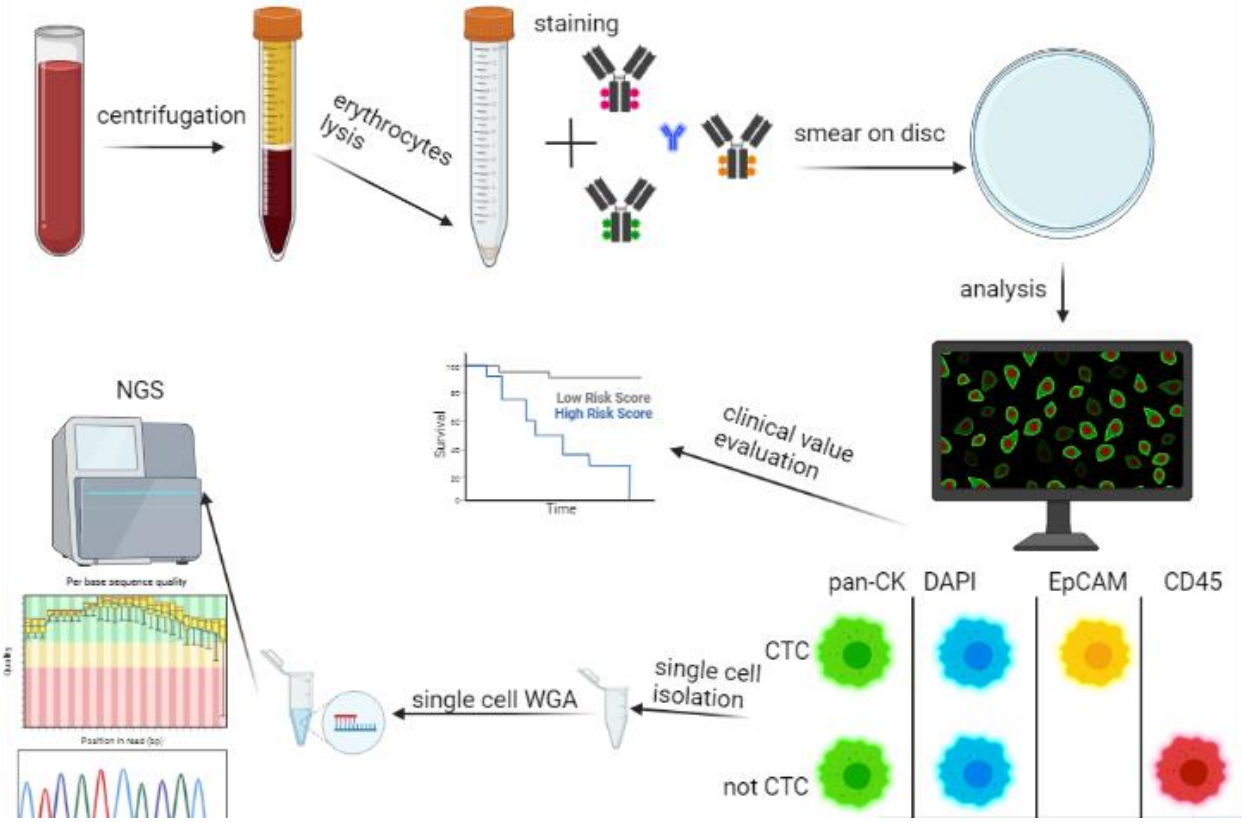


Figure 6. Methodology scheme for blood samples analysis.

4.2.1. Sample processing

Blood samples were collected to the 9 ml EDTA tubes and processed in 2h after collection. Samples were centrifuged at 2500×g for 15 min. The buffy coat containing nuclear cells, including tumor cells, were transferred to a new 15ml tube. The residual erythrocytes were lysed with the use of FACS Lysing solution (BD Biosciences) and with incubation time 15 minutes in the room temperature. Then the samples were centrifuged at 3000×g for 5 minutes. The lysis step was repeated, with the incubation time 10 minutes in room temperature and centrifugation at 3000×g for 5 minutes. Thereafter, cells were permeabilized with the permeabilization buffer, with incubation time 15 minutes in the room temperature and centrifugation at 3000×g for 5 minutes. Permeabilized cells were stained with the use of: Alexa Fluor 488-conjugated pan-cytokeratin (pan-CK) antibody (1:25), APC conjugated CD45 antibody (1:10), PE conjugated EpCAM antibody (1:10) and 4,6-diamidino-2-phenylindole (DAPI) (1:1000). Incubation time for staining was 60 minutes in the dark, in the 4°C. Next,

stained cells were washed two times with blocking buffer and centrifugated at 3000×g for 5 minutes. Washed and stained cells were resuspended in 1ml of ultra-pure H_2O and smeared on the glass disc in sterile conditions. The discs were left to dry overnight. Dried discs were mounted and fixed with the use of fixogum. Prepared samples were stored in -80°C until CytoTrack analysis.

4.2.2. Scanning and CTCs detection

All patient's samples were processed as described above and analyzed on CytoTrack system. For analysis the glass disc containing stained cells was mounted in the mounting arm with a spring-lock mechanism. Focus plan was obtained based on the DAPI channel, at eight places on the disc. Scanning was performed with 488nm Argon-Neon laser, a spiral pattern with a bandwidth of 10μm, a process taking ∼5min. All signals from the Alexa Fluor 488 emission channel were recorded and listed in the hotspot table. Recorded events were visually inspected by the operator. The criteria for CTC identification were set as: nearly round size with ≥6μm diameter, visible nucleus, pan-CK signal, CD45 negative. Detected cells with the above criteria were identified as CTCs. Clusters were defined as: group of ≥3 cells, with at least 3 visible nuclei in DAPI channel, with at least 3 cells identified as CTCs. Homotypic clusters were defined as clusters composed only from cancer cells. Heterotypic clusters were defined as clusters composed with both, PBMCs and cancer cells. All samples were scanned and cells meeting the criteria for CTCs were counted. Classified CTCs were photographed using the CytoTrack system. Sets of images for every fluorescent channel were taken for further EpCAM expression analysis. All coordinates of identified CTCs were saved for micromanipulation isolation of single cells.

4.3.EpCAM expression analysis

EpCAM heterogeneity analysis was performed using images from the CytoTrack system and ImageJ dedicated macro. Image analysis started with the identification and separation of regions of interest (ROI) by subjecting the image to the Otsu's threshold clustering algorithm on the pan-CK channel. This process enabled to automatically designate the potential CTC. Next, the mean fluorescence signal from each ROI was corrected

by subtracting from it the mean fluorescence of its direct surrounding area (background signal for each individual ROI). This process was repeated independently for two separate fluorescent channels corresponding to pan-CK and EpCAM staining, respectively. Finally, the EpCAM to pan-CK staining ratio was calculated for each ROI using the background corrected signals of the respective fluorescent channels. ROIs with large areas were manually checked for cluster identification. Low quality images or high background noise images were excluded from the analysis as Otsu's threshold clustering was unable to calculate proper ROI area.

4.4.Molecular analysis

4.4.1. Cell picking and single cell whole genome amplification (WGA)

Single CTCs were isolated form discs via CytoPicker (micromanipulator). After scanning discs with positive CTCs counts were further proceeded for cell picking. Cover glasses of discs were detached by overnight incubation in 1xPBS buffer. After detaching, discs were left for air-drying. Glass discs without cover glasses were set into the machine and saved coordinates of the identified CTCs were used for CTCs retracing. Cells were picked individually in maximum 5µl volume of 1xPBS buffer. Picked single cells were transferred into PCR tubes and whole genome amplification (WGA) was performed. Single cell WGA was performed with MALBAC WGA kit according to the manufacturer's instructions. Single cell WGA products were visualized on 1% agarose gel and analyzed on Bioanalyzer with the use of High Sensitivity DNA kit. Only single cell WGA products with visible products with distribution lengths between 200-3000bp were chosen to the next generation sequencing (NGS) (Figure 7).

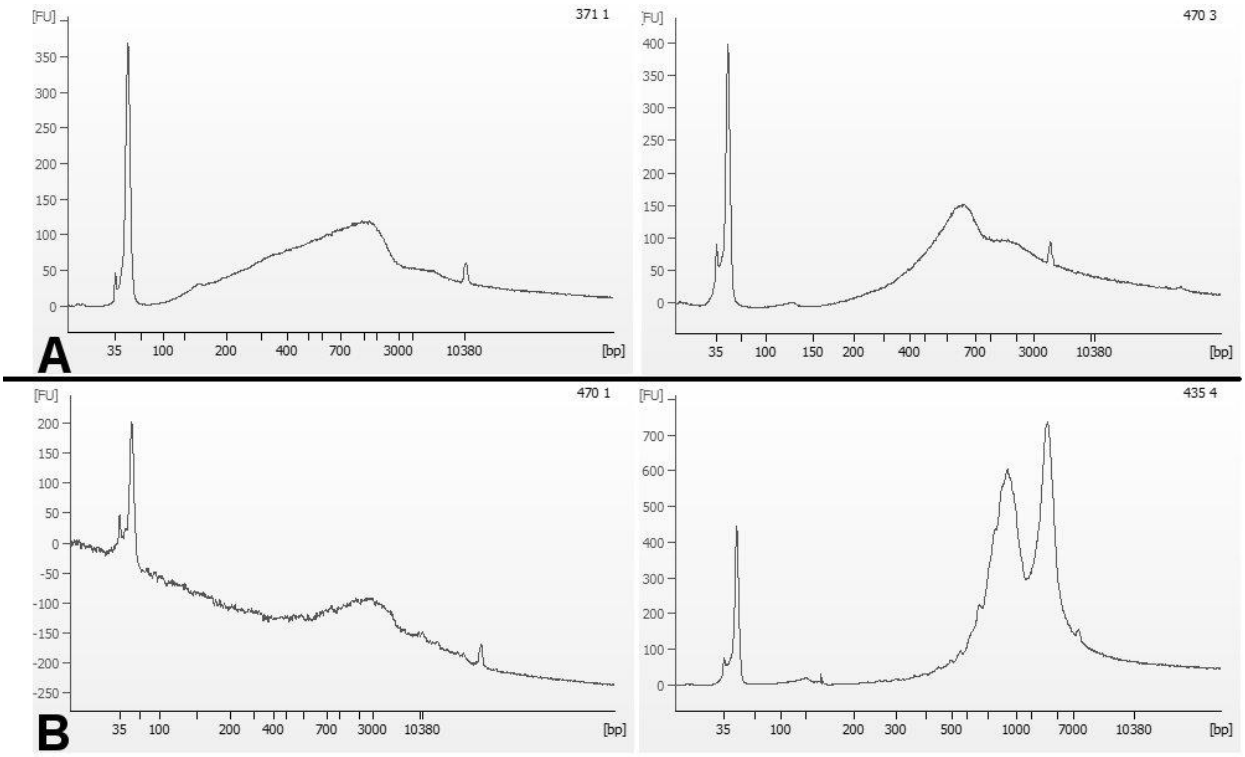


Figure 7. Results from Bioanalyzer for WGA products; A: Distribution of products between 200-3000bp, acceptable for NGS analysis, B: Distribution of products below or above 200-3000bp, not acceptable for NGS analysis.

4.4.2. Next Generation Sequencing (NGS)

The NGS was performed using custom library designed in Illumina Design Studio. The amplification sequencing method was chosen. Designed library covered main hotspots in *ESR1* (exons: 4, 5, 8), *PIK3CA* (exons: 9, 20) and *AKT1* (exon 4) genes and all coding regions for genes: *TP53*, *GATA3*, *ESR2*, *AKT2*. The coverage of the amplicons building the library was 99,84%. Total size of the library was 8368bp and average amplicon length was 136bp. Designed library contained two pools of primers, with 70 primer pairs per pool. Libraries were prepared using Illumina Library PLUS according to manufacturer's instruction. The libraries were then quantified via Quantus (Promega) using QuantiFluor dsDNA ONE system, purity and size of the libraries were established using High Sensitivity DNA kit for Bioanalyzer. As libraries were prepared with the use of WGA product as a template, in some libraries the WGA artifacts were notices. Libraries with WGA artifacts were additionally purified using BluePipin system (Sage Science). Libraries after BluePipin were

measured on Quantus and Bioanalyzer again, to confirm extraction of WGA artifacts (Figure 8). Prepared libraries were sequenced with the use of MiniSeq Mid Output Kit. The depth of the sequencing was set as x300 with bases higher than Q30 ~95% and 1,4GB generated data.

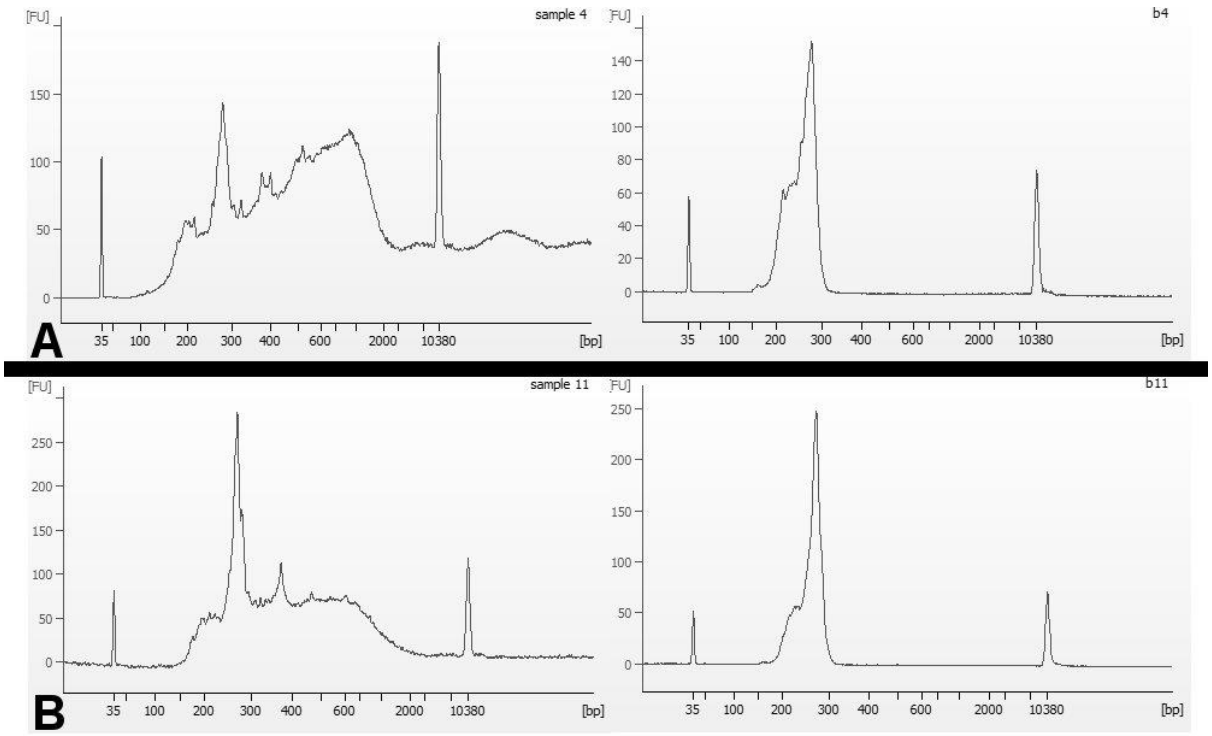


Figure 8. Bioanalyzer results for libraries prepared from WGA material; A-B: Library contaminated with artificial WGA amplicons (long fragments) on the left and library after BluePippin cleaning on the right.

4.4.3. Single Cell variant analysis

For each sample, the reads were extracted from the original BAM file using the cell-specific barcodes and were aggregated to generate a sub-BAM file. Mutect2 (v.4.1.0.0) was then applied to the sub-BAM files to identify somatic point variants. Then, outputs were run through pipeline for filtering and annotation. As the single cell NGS coverage was small we used calculator for diagnostic NGS settings for detection of subclonal mutations for minimal depth and coverage calculations [145]. For our sequencing parameters and library specification, the minimal depth for hotspots was set as 20x. Variants with coverage lower than 20x and variant allele frequency (VAF)<0,2 was discarded, unless they were present in the primary tumor sample.

4.4.4. FFPE DNA isolation

The FFPE samples obtained from Pathology Department were cut with thinness 10μm. Up to 8 scratches were used to the DNA isolation. DNA was isolated with the use of QIAamp DNA FFPE Tissue Kit according to manufacturers' instruction. Isolated genetic material was measured and checked for purity using NanoDrop spectrophotometer. In this system nucleic acid concentrations are determined by measuring the absorbance of ultraviolet light. Derived from the Beer-Lambert law, the amount of light absorbed at 260nm is proportional to the concentration of nucleic acid in solution. Moreover, as proteins absorbs the light at 280nm and EDTA, ethanol and polysaccharides absorb light at 230nm, ratios A260/A280 and A260/A230 are used as purity check for isolated nucleic acid. For DNA the ratio at A260/A280 should be around 1,8 and A260/A230 ratio should be around 2. Only pure DNA with concentration at least 50ng/μl were used for sequencing.

4.4.5. Polymerase Chain Reaction (PCR)

Polymerase chain reaction was conducted with standard PCR GoldTaq Polymerase with the primers concentration 250nM and 200nM for *ESR1* and *PIK3CA*, respectively. The reaction was done with the use of GeneAmp PCR System 9700 Thermal Cycler (Applied Biosystmes) (Table 6). Conditions of the amplification reactions are presented in the Table 8. PCR products were further sequenced using BigDye™ Terminator v3.1 Cycle Sequencing Kit (Thermo-Fisher) according to manufacturers' guide.

Table 6. Reaction conditions for PCR amplification of FFPE samples.

PCR Stage	Temperature	Time	No. of cycles
Initial Denaturation	95°C	10min	-
Denaturation	95°C	15s	×40
Primer Hybridization	55°C	30s	
Elongation	72°C	30s	
Final Elongation	72°C	5min	-
Storage	4°C	∞	-

4.5.Statistical analysis

Categorized quantitative data at different time points were compared using the Mann-Whitney U test or, if there were more than two categories, the Kruskal-Wallis test.

The primary end point was overall survival (OS), the secondary end points were progression-free survival (PFS)and progression versus non-progression in relation to CTCs numbers/presence and the CTCs dynamics. The progression during observation time was defined as progression assets with the use of MRI and/or CT. If patient was progressing for few imagining assets in the row it was treated as one long progression. The new progression was defined as progression after achieving stable disease in imagining assets. Time from the date of the blood collection to progression or death from any cause was calculated. If an outcome was not reached during the observation time variables were censored. Kaplan-Meier plots and the log-rank tests were used to illustrate and compare survival between subgroups. Survival analysis of variables measured at follow-up collections was performed by landmark analysis. Univariable and multivariable hazard ratios (HRs) for selected potential predictors of PFS and OS were determined by Cox proportional hazards regression. Model fit was measured using Harrell's C-index, and the fit of nested prognostic models was compared using log-likelihood ratio (G squared) test. The association between CTCs presence and dynamics with primary and secondary end points were also tested by logistic regression. All data were analyzed using the GraphPad Prism 9 software.

5. Results

5.1. CytoTrack method optimization and recovery ratio

The spike-in experiments were done with the use of prepared, counted cell pellets and blood collected from healthy individuals. For establishing the most effective dilution of selecting antibody, anti-pan-CK, the series of optimization experiments were conducted. The whole blood samples with spiked in ~100 MCF-7 cells were proceed according to the protocol with different pan-CK dilutions, from 1:10 to 1:100. The experiments were done in 2 repetitions for every antibody dilution and negative control (healthy donor whole blood sample without MCF-7 cells spiked in). Recovery ratio was set as ratio of the number of MCF-7 cells detected to the number of MCF-7 cells spiked into blood. The recovery ratio of method was ~65% for the lowest antibody dilution. The highest recovery ratio (>83%) was observed for pan-CK dilution of 1:25 (Table 7).

After establishing the most optimal antibodies dilution, the mean recovery ratio and specificity of the method were established. Spike-in experiments with 100, 50, 20 and 10 MCF-7 cells were done the same way as optimization experiments. All experiments were done in 2 repetitions. The recovery ratio of the method was set up between 75-85% (Table 8), with the mean from all conditions ~79%. To establish the specificity of the method and potential false positive ratio the additional experiment with healthy donors' blood was conducted. The blood samples collected from 3 healthy donors were stained with established protocol. No cells meeting the CTCs criteria were detected in healthy donor samples. This highlights the high specificity of the CytoTrack method.

Table 7. Recovery ratio for different antibody dilution

Antibody dilution	Mean recovery ratio
1:100	65%
1:50	70%
1:25	85%
1:10	83%
Negative control	0 cells

Table 8. Recovery ratio for different spike-in experiments

Spike-in	Mean recovery ratio
100 cells	84%
50 cells	80%
20 cells	78%
10 cells	75%
Negative control	0 cells

5.2. Patients' characteristics

Overall, the 135 patients were enrolled to the final analysis. Patient and tumor characteristics are summarized in Table 9. The median follow-up time from 1st collection was 20,93 months (range 1–25) for patients alive at the last medical visit before the cutoff date of 31 July 2022. The median age of patients at the 1st blood collection was ~65 years (range 37-90). All patients were identified with distant metastases, as it was one of the inclusion criteria. The 35 patients (23,7%) had one meta site at the beginning of observation (bones metastases only), 43,9% of patients were identified with 2 distant metasites and 33,3% of patients with ≥3 distant metasites. The patients included in the study were very advanced patients with majority treated with hormonal treatment combined with chemotherapy (HTH+CHTH). Only 12,5% of patients were treated with CDK4/6 inhibitors combined with hormonal therapy (HTH + CDK4/6 inhibitors) and 25,18% of patients were treated with only hormonal therapy (HTH). The majority of patients were diagnosed with NST (ductal) subtype of breast cancer, 76,29%. 17 patients were diagnosed with lobular subtype and 9 with mixed ductal-lobular subtype. 9 patients were diagnosed with other, rare histopathological subtypes, including: tubulare, cribriform, mucinous (Table 9).

Table 9. The clinicopathological characterization of group of patients enrolled in the study.

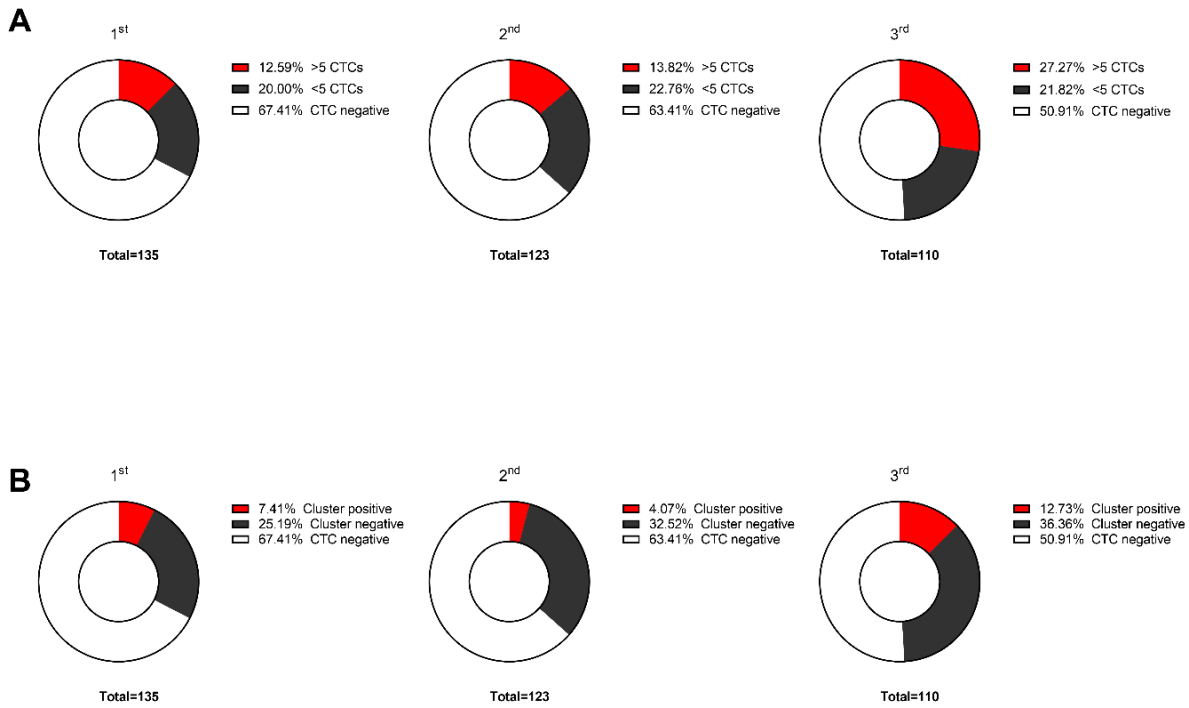
Clinical feature	**No. of patients**
Age at the 1st blood collection	
<65	68
≥65	67
HER2 status	
HER2-	132
HER2+	3
TNM	
TxN0M0	35
TxNxM0	51
TxNxMx	49
No. of distant metastases sites	
1	32
2	58
≥3	45
Metastases sites	
Bones	114
Liver	43
Lungs	40
Other	57
Treatment during study	
HTH+CHTCH	84
HTH	34
HTH+CDK4/6inh	17
Histological type	
NST (ductal)	103
Lobular	17
Ductal-Lobular	6
Other	9
BRCA1/2	
WT	129
Mutant	6

5.3. CTCs detection and clinical value of CTC count

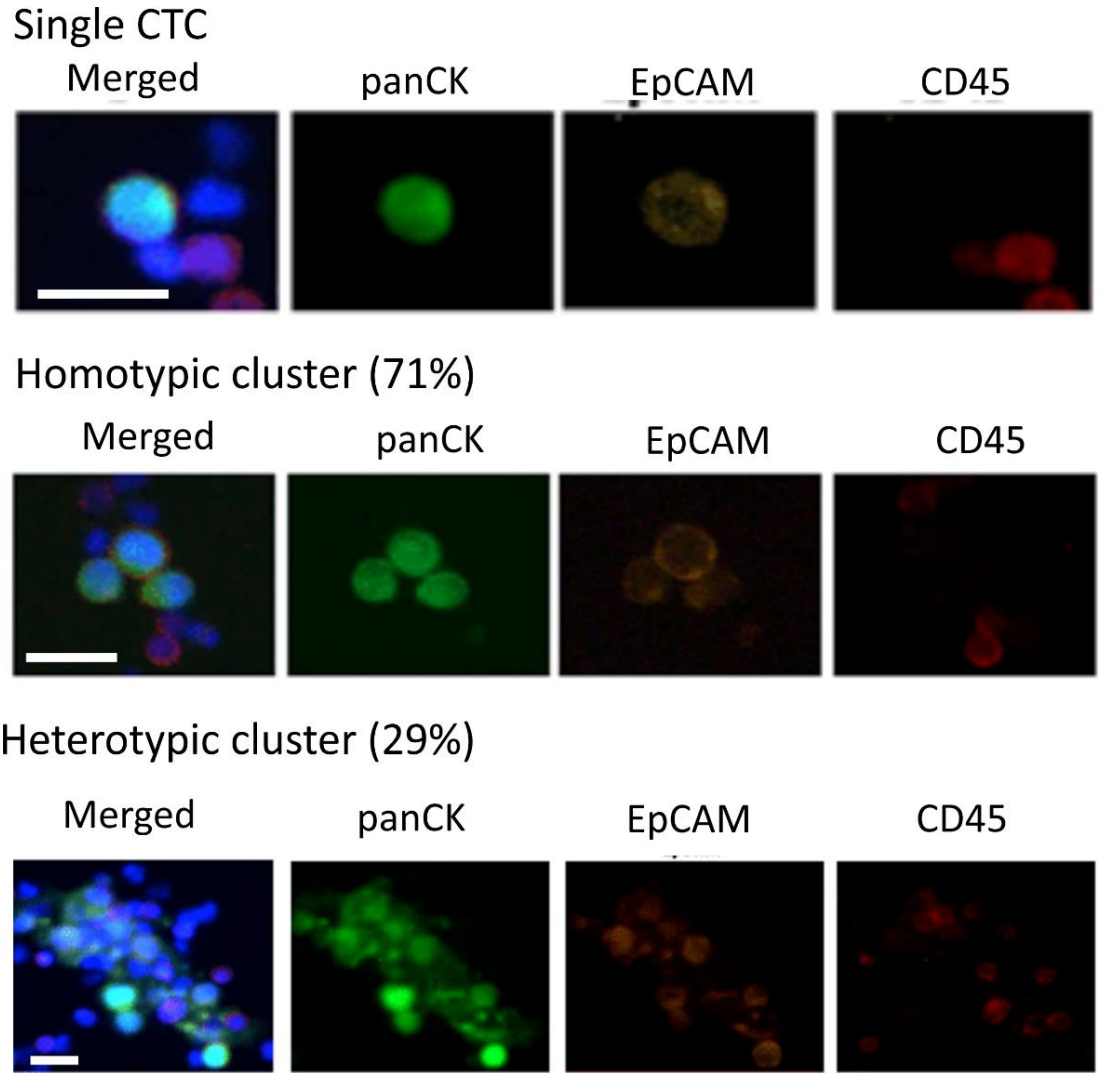
5.3.1. CTC counts

The CTCs were detected overall in 143 samples collected from 90 patients during the observation time. The CTCs were detected in the 44 (32.59%) patients in 1st blood collection, 45 (36.58%) patients in 2nd blood collection and 54 (49.09%) patients in 3rd blood collection. Moreover, ≥5 CTCs were detected in 17 patients in 1st and 2nd blood collection (12.59% and 13.82%, respectively) and in 30 (27.27%) patients in 3rd blood collection (Graph 3). The clusters were detected in the material collected from 23 patients. Overall, we identified over 160 clusters in all samples with heterotypic clusters fraction ~29%. However, as the clusters were mainly observed in the last blood collection the presence of clusters was not analyzed further in the clinical analysis (Graph 4).

The observed trend with increasing number of CTC positive patients during the observation time might be associated with increasing advancement of the disease. Taking this into account we further explored the potential prognostic value of CTCs numbers and dynamics during treatment.

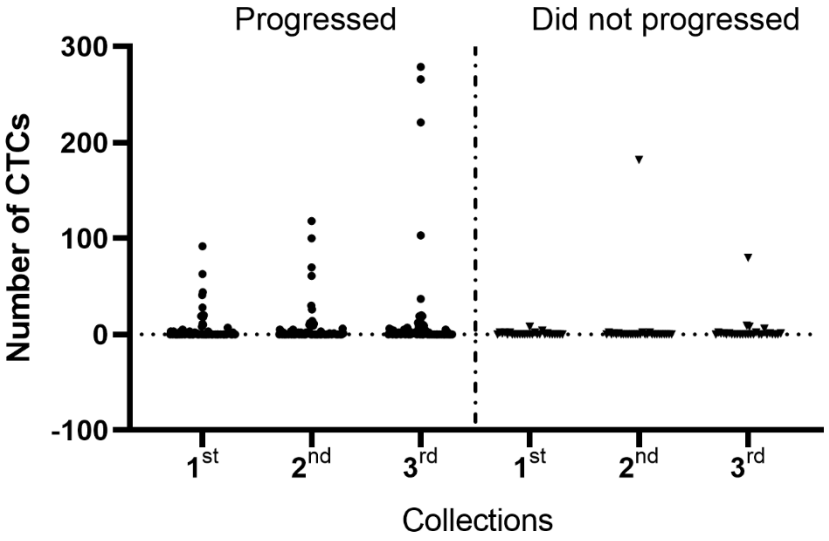


Graph 3. The % of detection of CTCs and clusters in separate collections. A) Detection of ≥5CTCs, <5CTCs and no CTCs; B) Detection of clusters in CTCs positive patients.



Graph 4. The images of identified: single CTC, homotypic cluster and heterotypic cluster in CytoTrack analysis.

The patients were divided according to the progression status during treatment. 105 patients progressed during the observation time and 29 patients did not progress during the observation time. The number of CTCs detected in every collection for patients were visualized in the Graph 5. The number of CTCs detected in the group that progressed was higher, however, the difference was not significant. In the groups of patients with progression during treatment there is visible trend associated with rising number of CTCs. To further validate the value of the CTCs dynamics we did the survival analyses.



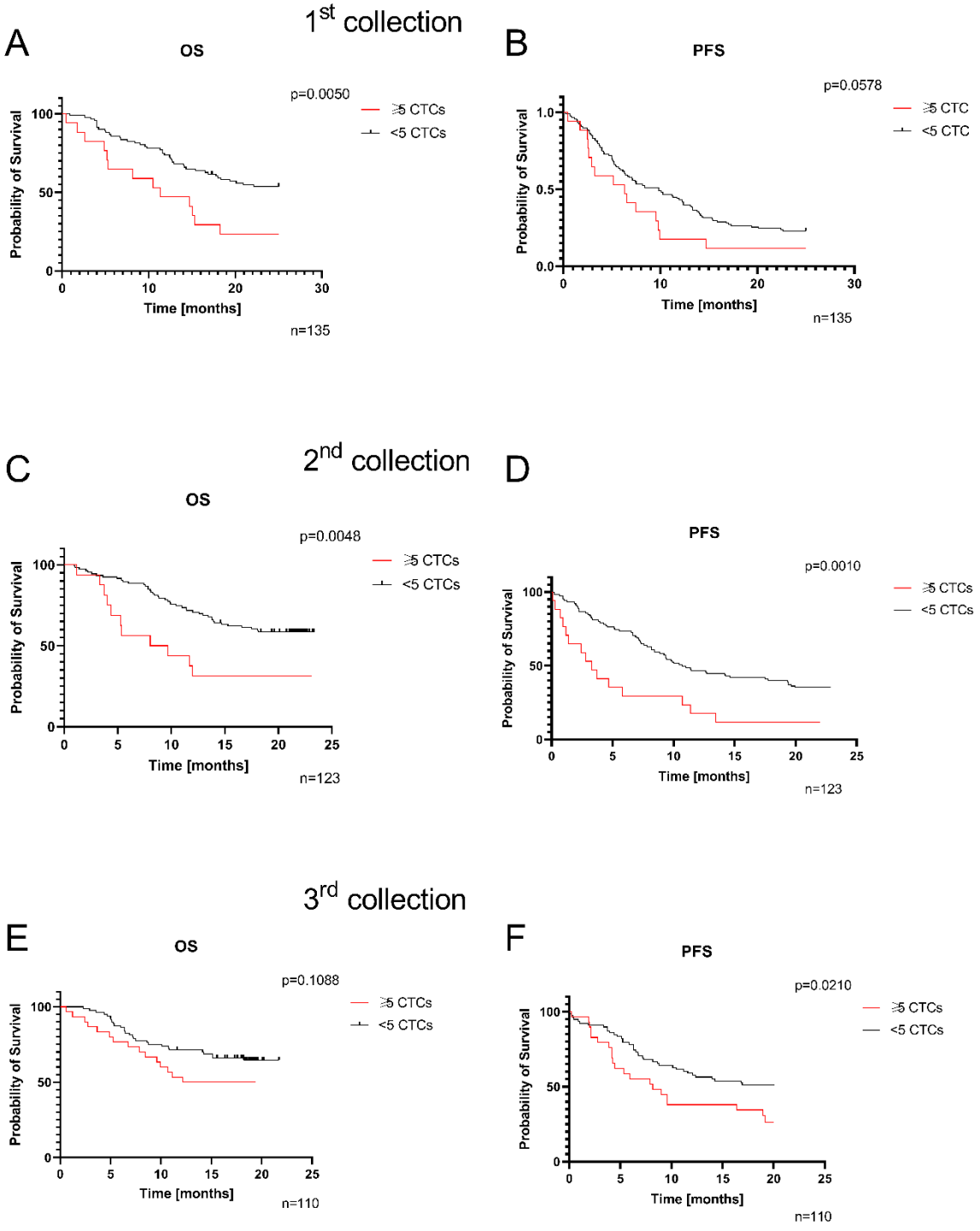
Graph 5. The number of CTCs detected in all collection in patients with and without progression during observation time. The difference in CTCs number is not significant (Kruskal-Wallis; p>0.05).

5.3.2. Clinical value of detected CTCs

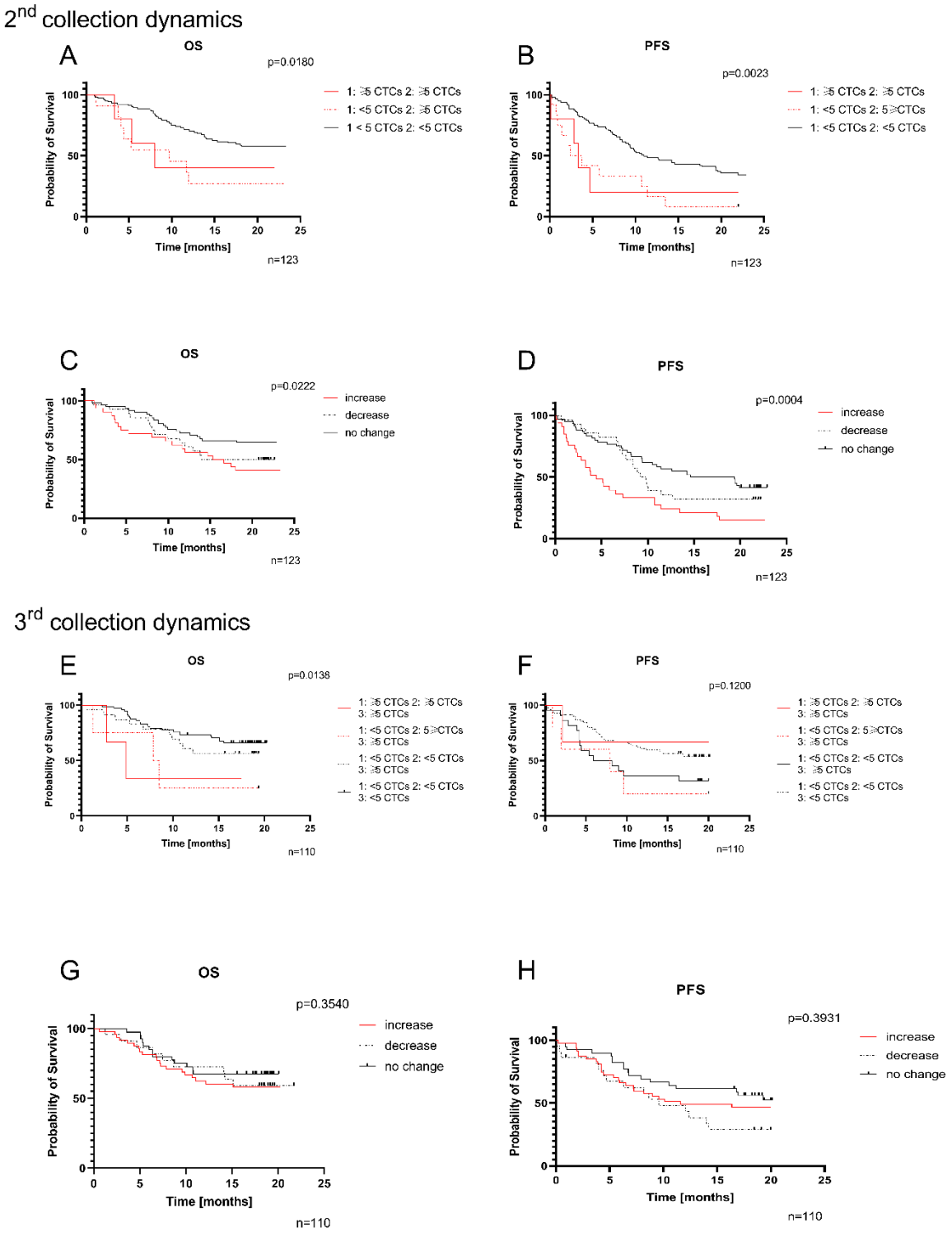
To establish the clinical value of CTCs detected via CytoTrack system the Kaplan-Maier survival analysis was done with landmark analysis for follow-up collections (Graph 6). CTCs presence in 1st collection was not found to be significant for PFS and OS. However, the presence of ≥5 CTCs was identified to be significant for OS (log-rank test p<0.05) and strong trend for PFS was also observed (log-rank test p=0.0578) (Graph 6). We observed that the presence of ≥5CTCs is the strong survival predictor in all collections. In 2nd collection the identification of ≥5CTCs was found to be strong predictor of PFS and OS (log-rank test p<0.05) (Graph 6). In 3rd collection the presence of ≥5CTCs was significant predictor only for PFS (log-rank test p<0.05), but with the visible trend for OS (Graph 6). As the 3rd blood collection was performed in the most advanced group of patients, we decided to also confirm value of ≥5CTCs for PFS and OS with Gehan-Breslow-Wilcoxon test, which gives more weight to events/deaths at early time points. For PFS we still observed significant curve separation (p-value=0.0314), but for OS statistical significance was not observed (p-value p=0.0873).

To explore previously described trend of higher CTC numbers during treatment we also performed more detailed analysis of changes in CTCs numbers during treatment, further called/referred to as CTCs dynamics (Graph 7). For 2nd blood collection the CTCs dynamics in comparison to 1st collection was significant predictor for PFS and OS (log-rank p-value<0.05). The patients with increase in CTCs count in 2nd blood collection were characterized with shorter median survival for PFS (4.47 months) and OS (15.97 months) than patients with decrease or no changes in CTCs numbers (PFS-9.5 months, OS-18.3 months and PFS-16.985 months, OS-undefined, respectively). For the patients with rising number of CTCs in 3rd collection this dynamic was found to be significantly inferior predictor of PFS and OS (Graph 7). It might be associated with the advancement of the disease.

To further investigate the potential value of the CTCs dynamics we investigated the certain patterns of CTCs counts in blood collections (Graph 7). For 2nd blood collection the constant high CTCs counts (≥5 CTCs) and high CTC (≥5 CTCs) count in 2nd blood collection were identified as significant unfavorable predictors for PFS and OS (Graph 7a). The similar trend was observed for data from 3rd collection. Patients with constant high CTCs counts and persistent high CTCs counts from 2nd collection were identified with shorter median survival for OS (4,9 months and 8,17 months, respectively), while for patients with high CTCs counts only in 3rd blood collection and patients with constant low CTCs numbers (<5 CTCs) the median survival for OS was undefined. The data for PFS from 3rd blood collection were not significant (log-rank p>0.05) however, we observed the strong trend that patients with constant low CTCs counts were observed with longer PFS (Graph 7).



Graph 6. The Kaplan-Mayer survival curves with p-value from log-rank tests. The patients were distinguished in groups with ≥5 CTCs detected and <5 CTCs detected: A-B) Data from 1st collection; C-D) data from 2nd collection landmark; E-F) data from 3rd collection landmark.



Graph 7. The Kaplan-Mayer analysis of changes in CTCs numbers from 2nd and 3rd collections landmarks with p-values from log-rank tests. The patients were divided into subgroups according to the: A-B; E-F) CTCs counts in every blood collection and C-D; G-H) sore changes in numbers.

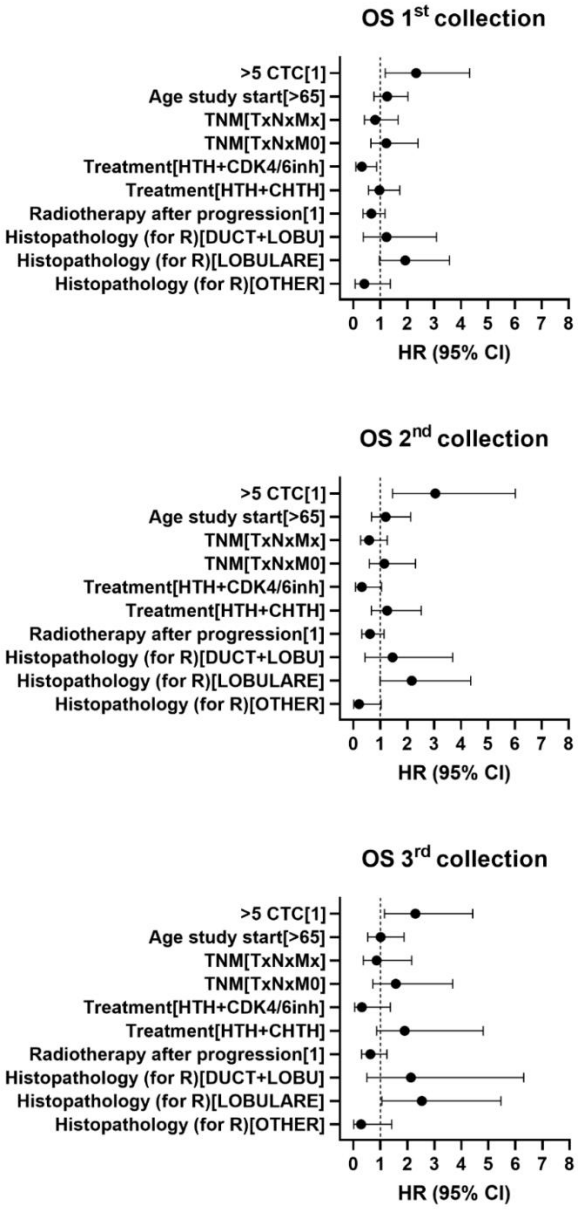
To furhter investigate obtained data we also conducted univariable and multivariable Cox proportional hazard regression models. For multivariable clnicopathological model the following covariates were used: histopatological type, RTH during the observation, TNM status in the time of diagnosis, type of treatment and age. The data used for the Cox hazard regression were categorical variables with reference level defined in the model. All clinicopathological data with their reference levels were listed in the Table 10.

Table 10. The clinical and pathological data with reference levels used for multivariable model.

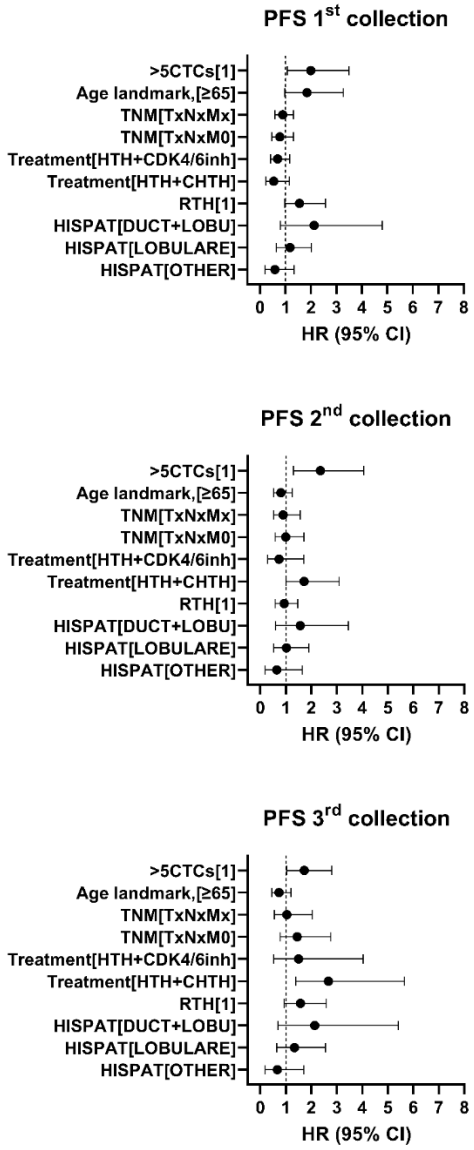
Variable	Refrence level	Other levels
Histopatological subtype	NST	LOBULARE OTHER DUCTALE-LOBULARE
RTH during observation	No RTH during observation	RTH during observation
TNM status	T_XN0M0	T_XN_XM0 $T_XN_XM_X$
Type of treatment	HTH	HTH + CDK4/6 inhibitors HTH + CHTH
Age	<65	≥65

The presence of ≥5CTCs in 1st blood collection was found to be strong predictor for OS (HR_{OS}=2.3; 95% CI: 1.201 - 4.083, p-value<0.05) in univariable analysis, but not for PFS (HR_{PFS}=1.672; 95% CI: 0.9277 - 2.806; p-value=0.0668). However, in the multivariable model including the clinicopathological features of patients the high count of CTCs in 1st collection was identified as strong predictor for both OS (HR_{OS}=2.323; 95% CI: 1.175 - 4.320; p-value<0.05) and PFS (HR_{PFS}=1.987; 95% CI: 1.067 - 3.493; p-value<0.05). The presence of ≥5CTCs remained strong predictor for OS and PFS in univariable (HR_{OS}=2.724; 95% CI: 1.370 - 5.010; p-value<0.05 and HR_{PFS}=2.539; 95% CI: 1.421 - 4.263; p-value<0.05) and multivariable (HR_{OS}=3.004; 95% CI: 1.453 - 6.016; p-value<0.05 and HR_{PFS}=2.359; 95% CI: 1.296 - 4.051; p-value<0.05) analyses for 2nd blood collection landmark. Interestingly, the high CTCs count in 3rd collection was not significant predictor in univariable analysis. However, after adjusting to chosen clinicopathological model the high counts of CTCs (≥5CTCs)

in 3rd collection were found to be predictive for OS (HR_{OS}=2.29; 95% CI: 1.145 - 4.417; p-value<0.05) and PFS (HR_{PFS}=1.72; 95% CI: 1.027 - 2.804; p-value<0.05) (Table 11). Overall, we found that in presented clinicopathological model the ≥5CTCs count is strong OS and PFS predictor regardless of blood collection time-point (Graphs 8-9).



Graph 8. The Hazard Ratio (HR) of predictors used in multivariable Cox proportional hazard regression for OS.



Graph 9. The Hazard Ratio (HR) of predictors used in multivariable Cox proportional hazard regression for PFS.

To investigate the potential value of the changes in CTCs numbers during the treatment the survival data from 2nd blood collection and 3rd blood collection landmarks were analyzed in regards of specific CTCs dynamics. We confirmed that the rising counts of CTCs in 2nd blood collection are unfavorable predictors for OS and PFS. The patients with rising CTCs counts to ≥5 CTCs were characterized with 2.933 (95% CI: 1.340 - 5.723;

p-value<0.05) times higher probability of death than patients with other dynamics in univariable analysis. Moreover, patients with this dynamic were also characterized with 2.516 (95% CI: 1.298 - 4.457; p-value<0.05) times shorter PFS than patients with other dynamics in univariable analysis. The strong predictive value of this dynamic remains significant in the multivariable analysis (HR_{OS}=3.067; 95% CI: 1.343 - 6.348; p-value<0.05 and HR_{PFS}=2.261; 95% CI: 1.149 - 4.100; p-value<0.05) (Table 11). Further, we confirmed, previously observed in Kaplan-Meier analysis, predictive value of persistent high counts of CTCs from 2nd blood collection. The patients with continual ≥5CTCs in 2nd and 3rd blood collection were characterized with 3.728 (95% CI: 1.112 - 9.346; p-value<0.05) higher probability of death than patients with other dynamics, in univariable analysis from 3rd collection landmark. Moreover, when adjusted to chosen clinicopathological model, the hazard ratio for OS, for this dynamic was increased to 7.001 (95% CI: 1.744 - 23.33; p-value<0.05) (Table 11). This dynamic was not found to be significant predictor for PFS. Overall, we observed the strong trend for worse outcomes for patients with increase to ≥5 CTCs during the observation. Furthermore, the patients in which the CTCs count stayed high (≥5 CTCs) in next blood collection were also identified with worsen OS (HR_{OS}=3.728; 95% CI: 1.112 - 9.346; p-value<0.05 for univariable analysis and HR_{OS}=7.001; 95% CI: 1.744 - 23.33; p-value<0.05) (Table 11).

Interestingly, in all Cox univariable and multivariable analyses we did not confirm the significance of constant high CTCs numbers as strong predictors for OS and PFS. However, we did find that constant low number of CTCs (<5 CTCs in all collections) is strong favorable predictor for both OS and PFS in all analyses (Table 11).

Table 11. The results of Cox proportional hazard regression models.

	OS		p-value	PFS		p-value
	HR	95% CI		HR	95% CI	
1st collection						
Univariable						
≥5 CTCs	2.3	1.201 - 4.083	0.0071	1.672	0.9277 - 2.806	0.0668
Multivariable						
≥5 CTCs	2.323	1.175 - 4.320	0.0105	1.987	1.067 - 3.493	0.0223
2nd collection landmark analysis						
Univariable						
≥5 CTCs	2.724	1.370 - 5.010	0.0022	2.539	1.421 - 4.263	0.0008
Constant ≥5 CTCs	1.834	0.4461 - 4.986	0.3076	2.147	0.6531 - 5.188	0.1387
Constant <5 CTCs	0.3686	0.2126 - 0.6633	0.005	0.4813	0.3046 - 0.7893	0.0025
Increase from <5 CTCs to ≥5 CTC	2.933	1.340 - 5.723	0.0033	2.516	1.298 - 4.457	0.0031
Decrease from ≥5 CTCs to <5 CTCs	1.943	0.7439 - 4.197	0.1257	1.265	0.5306 - 2.545	0.5508
General increase in CTCs number	1.806	1.026 - 3.087	0.0343	2.174	1.383 - 3.342	0.0005
General decrease in CTCs number	1.173	0.6181 - 2.094	0.6047	1.057	0.6391 - 1.680	0.8214
Multivariable						
≥5 CTCs	3.004	1.453 - 6.016	0.002	2.359	1.296 - 4.051	0.003
Constant ≥5 CTCs	2.086	0.4738 - 6.426	0.5243	2.377	0.6461 - 6.904	0.1437
Constant <5 CTCs	0.3338	0.1798 - 0.6334	0.0006	0.4423	0.2717 - 0.7429	0.0014

Increase from <5 CTCs to ≥5 CTC	3.067	1.343 - 6.348	0.0042	2.261	1.149 - 4.100	0.0112
Decrease from ≥5 CTCs to <5 CTCs	1.869	0.6958 - 4.237	0.1671	1.684	0.6696 - 3.667	0.2231
General increase in CTCs number	1.571	0.8631 - 2.800	0.1301	1.99	1.247 - 3.116	0.0031
General decrease in CTCs number	1.183	0.6160 - 2.144	0.5952	1.163	0.6917 - 1.882	0.553
3rd collection landmark						
Univariable						
≥5 CTCs	1.661	0.8648 - 3.067	0.1129	1.559	0.9543 - 2.476	0.0667
Constant ≥5 CTCs	3.296	0.5355 - 10.79	0.1008	0.8709	0.2123 - 2.354	0.8156
Constant <5 CTCs	0.541	0.2969 - 0.9941	0.0447	0.6245	0.4011 - 0.9838	0.0389
Persistent ≥5CTCs from 2nd collection	3.728	1.112 - 9.346	0.0128	1.772	0.6215 - 3.965	0.2164
Increase to ≥5CTCs only in 3rd collction	0.8898	0.3627 - 1.880	0.7774	1.522	0.8334 - 2.603	0.1455
General increase in CTCs number in 2nd and 3rd collection (constant increase)	1.959	0.7428 - 4.309	0.1268	1.275	0.6135 - 2.377	0.4774
General decrease in CTCs numbers in 2nd and 3rd collection	1.63	0.3942 - 4.485	0.4148	1.92	0.5798 - 4.707	0.2097

(constant decrease)						
Multivariable						
≥5 CTCs	2.29	1.145 - 4.417	0.0154	1.72	1.027 - 2.804	0.0336
Constant ≥5 CTCs	3.279	0.4968 - 12.54	0.1298	0.9485	0.2225 - 2.771	0.9322
Constant <5 CTCs	0.4434	0.2246 - 0.8725	0.0181	0.5867	0.3671 - 0.9468	0.0267
Persistent ≥5CTCs from 2nd collection	7.001	1.744 - 23.33	0.0027	2.251	0.7468 - 5.552	0.1054
Increase to ≥5CTCs only in 3rd collction	1.157	0.4568 - 2.552	0.7354	1.55	0.8211 - 2.751	0.1529
General increase in CTCs number in 2nd and 3rd collection (constant increase)	1.929	0.7056 - 4.513	0.1574	1.45	0.6878 - 2.753	0.2887
General decrease in CTCs numbers in 2nd and 3rd collection (constant decrease)	1.233	0.2844 - 3.703	0.7409	1.605	0.4743 - 4.089	0.3767

To evaluate CTCs as the rapid progression predictor, CTCs count and CTCs dynamics were analyzed in relation to the clinical outcome (dead or progression/dead) occuring within 3 months from the collection. In logistic regression model the ≥5 CTCs detected in 1st blood collection was not statistically significant predictor of rapid progression. However, for 2nd and 3rd collection detection of ≥5 CTCs was associated with significantly higher odds

for progression or dead within 3 months (OR=6.143; 95% CI: 2.111-18.88; p-value<0.05 and OR=3.273; 95% CI: 1.088 – 9.916; p-value<0.05, respectively).

In logistic regression models we observed that patients with contant low numbers of CTCs had lower rapid progression odds than patients with other dynamics. Interestingly, this trend was observed for data from 2nd collection (OR=0.2437; 95% CI: 0.09441-0.6230; p-value<0.05) and 3rd collection (OR =0.3646; 95% CI: 0.1197-1.069; p-value=0.0657). This results highlights the importance of low CTCs counts as favorable predictor (Table 12).

The group of patients with persistent high CTCs numbers did not have significantly different death and/or progresion odds within 3 months from the collection that patients with other dynamics. Interestingly, the patients with CTCs numbers contantly rising in every blood collection were characterized with higher odds of progression/death accuring within 3 months from 3rd blood collection than patients with other dynamics (OR=8.091; 95% CI: 1.976 - 33.75; p-value<0.05). This results emphasizes the importance of constantly rising number of CTCs as PFS predictor.

Table 12. The results of logistic regression modeling.

	Progression or dead accuring within 3 months from the collection		
	OR	95% CI	p-value
1st collection			
≥5 CTCs	1.721	0.5070 – 5.161	0.3634
2nd collection			
≥5 CTCs	6.143	2.111 - 18.88	0.0009
Increase to ≥5 CTCs	5.359	1.571 - 19.61	0.0078
Constant <5 CTCs	0.2437	0.09441 - 0.6230	0.0034
Constant ≥5 CTCs	5.056	0.7995 - 39.89	0.0832
3rd collection			
≥5 CTCs	3.273	1.088 - 9.916	0.0352
Persistent ≥5CTCS from 2nd collection	4.333	0.5358 - 28.45	0.152
Increase to ≥5 CTCs inly in 3rd collection	2.597	0.7293 - 8.396	0.1341
Constant <5 CTCs	0.3646	0.1197 - 1.069	0.0657
Constant ≥5 CTCs	3.067	0.1376 - 33.99	0.4063
General increase in CTCs number in 2nd and 3rd collection (constant increase)	8.091	1.976 - 33.75	0.0045

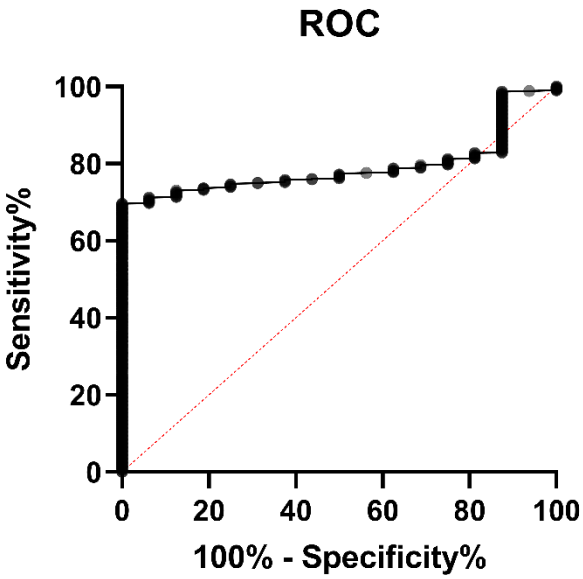
5.4. CTCs heterogeneity

5.4.1. EpCAM expression heterogeneity

For EpCAM expression analysis we analyzed only samples with more than 5 good quality images. Those samples were further grouped according to patients' ID. We observed that patients-derived cells were characterized by dynamic changes of both EpCAM and pan-CK expression, which was not observed in MCF-7 reference cell line. Therefore, in further analysis we did not divide the CTCs according to the raw EpCAM expression itself, but according to the EpCAM/pan-CK ratio. Overall, we analyzed 1 303 cells originated from biological material collected from 31 patients. We observed high heterogeneity in EpCAM/pan-CK ratio of detected CTCs between patients. For some patients there were visible two subpopulations of CTCs with high and low EpCAM/pan-CK ratio (patients: #1, #4, #14, #17, #20, #29) (Graph 11A).

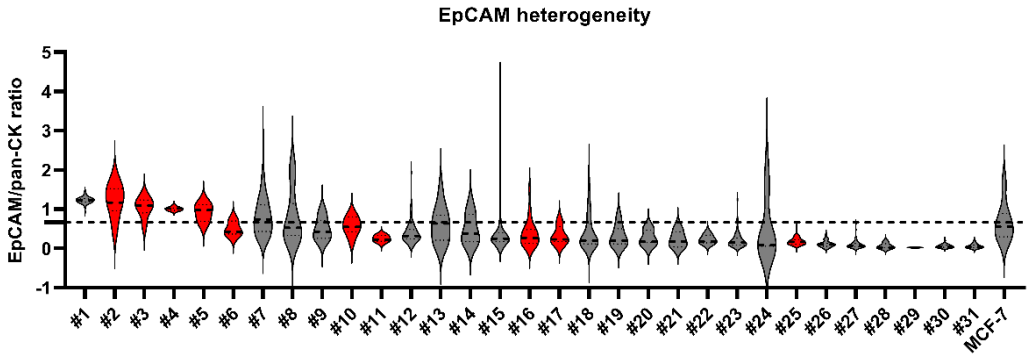
The CTCs EpCAM status was classified according to EpCAM/pan-CK ratio. For cut-off value estimation the ROC curve with the data from patients' samples compared with MCF-7 cells as control samples was used (Graph 10). The cut-off value was determined using the calculated Likelihood ratio based on the highest sensitivity and specificity. The cut-off value for $EpCAM^{low}$ cells was calculated from the ROC curve using the single cells data. The cut-off value was set as 0.2213. From all analyzed CTCs 44% were identified as $EpCAM^{low}$ and 0,5% were identified with no EpCAM expression.

Most patients were identified with at least two subpopulations of CTCs according to their EpCAM status. In patients #21 (n=8), #22 (n=7), #23 (n=26), #26 (n=16) we observed only $EpCAM^{low}$ cells and in patient #16 (n=16) we identified only $EpCAM^{high}$ CTCs. Interestingly, patients with highest percentage of $EpCAM^{high}$ CTCs: #10, #11, #15, #16 were patients identified with more than 50% clusters identified in all samples and patients #11 and #15 were identified with clusters only (Graph 11B). This trend might suggest that CTCs clusters are characterized with higher EpCAM than single CTCs.

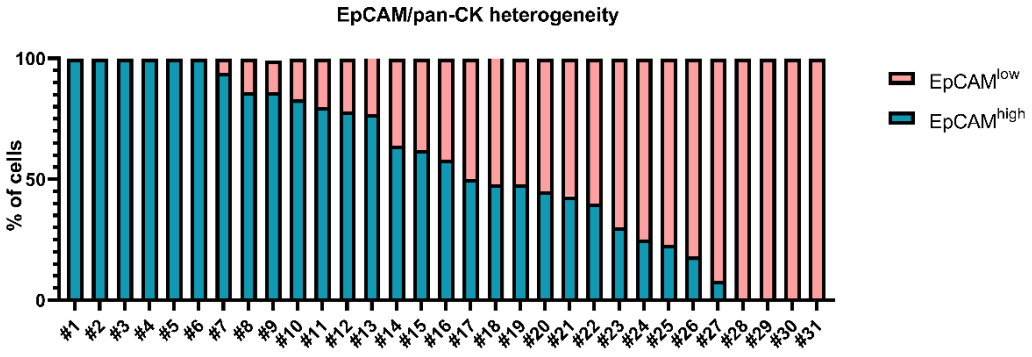


Graph 10. ROC curve for establishing the EpCAM/pan-CK ratio cut-off values. ROC curve from patients' data with the MCF-7 cells (controls). Area under the curve: 0,7910; SD=0,02448; p-value <0,0001.

A



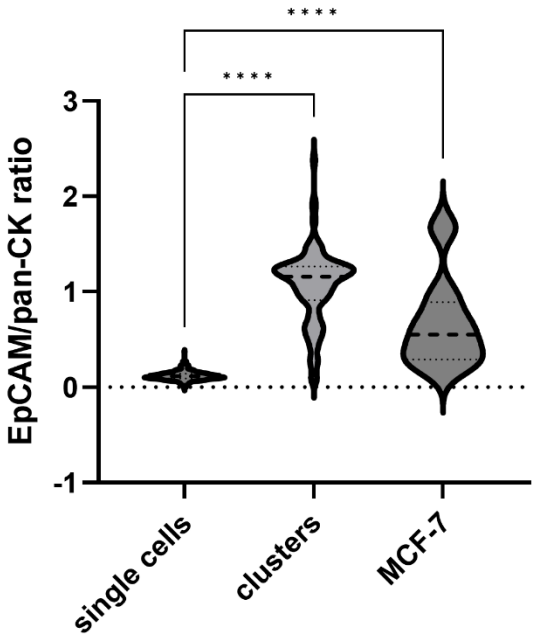
B



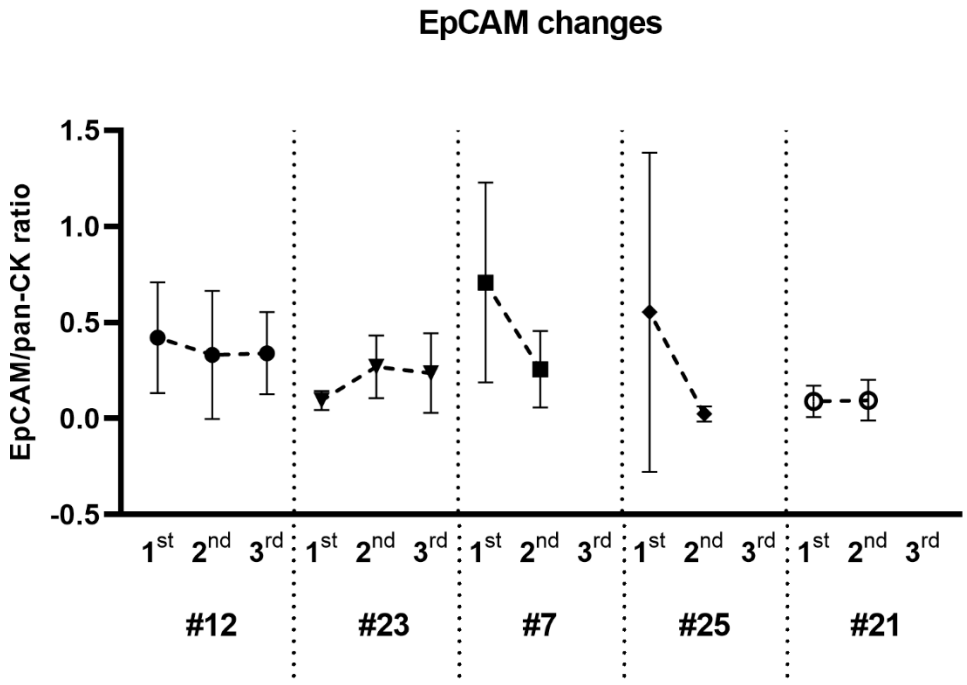
Graph 11. EpCAM/pan-CK heterogeneity among patients. A) The presentation of EpCAM/pan-CK ratio calculated for CTCs in patients' samples. Each number represents samples from one patient. MCF-7 reference cell line was used for the comparison. The line in the graph represents mean EpCAM/pan-CK ratio calculated for single cells from MCF-7 cell line. Patients with clusters in at least one sample were marked with red color. B) The percentage of $EpCAM^{low}$ and $EpCAM^{high}$ CTCs detected in patients' samples.

To further analyze this phenomenon, we compared the EpCAM expression, defined as EpCAM/pan-CK ratio, of single cells and clusters from all the samples. Single CTCs were characterized with significantly lower EpCAM expression than CTCs clusters and MCF-7 cells (Graph 12).

For patients with more than 5 CTCs analyzed for EpCAM/pan-CK ratio in at least 2 different samples we compared of EpCAM/pan-CK ratio changes during the treatment. We did not observe any trend in EpCAM changes during treatment. Two patients were identified with the decreasing level of EpCAM/pan-CK during the treatment. Other patients were identified with the similar EpCAM expression during the treatment (Graph 13).



Graph 12. The comparison of EpCAM expression, defined as EpCAM/pan-CK ratio in single CTCs and CTC clusters identified in patients. A) Comparison for all detected CTCs and all detected CTC clusters, merged data from all samples; **** represents $p<0.0005$ Kruskal-Walli's test.

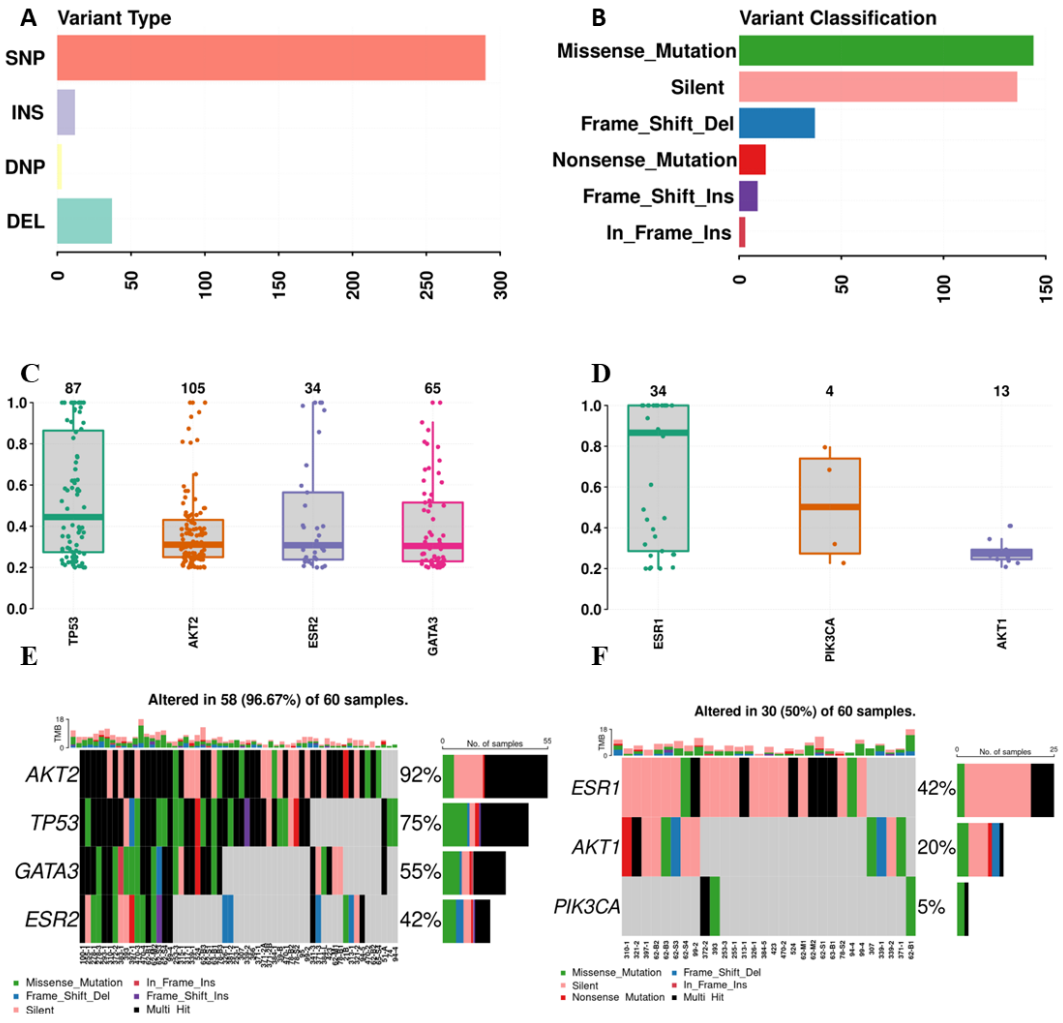


Graph 13. Changes in EpCAM/pan-CK ratio in patients during the treatment.

5.4.2. Molecular heterogeneity

5.4.2.1. Interpatient molecular heterogeneity

The molecular heterogeneity of single CTCs was established via Next Generation Sequencing (NGS). The WGA material from 66 single cells originated from 35 different patients was analyzed. We excluded the mutations occurring in 3'UTR and introns as their association with final protein amino acid sequence and conformation is uncertain. We identified 342 variants in exons of 60 of 66 sequenced single cells (Graph 14). In further mutational analysis we focused on the single nucleotide variants (SNVs) changes. Therefore, further part will be focused on SNVs occurring in exons.

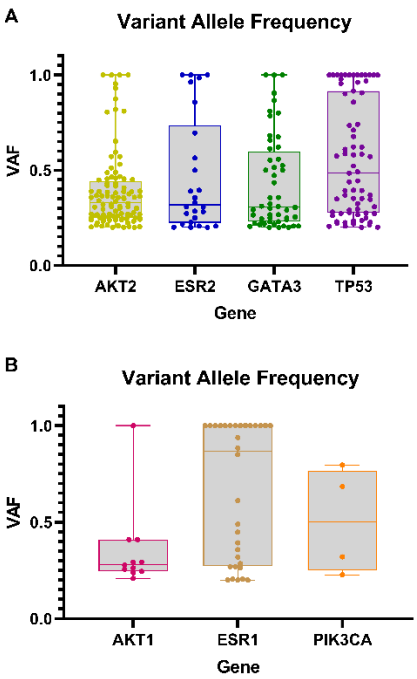


Graph 14. Summary of variants detected in single CTCs via Next Generation Sequencing. A) Variant types: single nucleotide polymorphism (SNP), insertion (INS), double nucleotide polymorphism (DNP), deletion (DEL); B) Variants classification according to functional changes: missense, silent, frame shift, nonsense and in frame mutations; C) The variant allele frequency (VAF) in genes from whole gene sequencing in NGS; D) The variant allele frequency (VAF) from hotspot gene sequencing in NGS; E) The summary of variants detected in exons in whole gene sequencing in NGS, presented as oncoplot; F) The summary of variants detected in exons in hotspot sequencing, presented as oncoplot. Graph generated using maftools [146].

We detected 296 SNVs in exons of 60 sequenced single cells. Further we explored the difference between number of identified SNVs in sequenced genes: *AKT1*, *AKT2*, *ESR1*, *ESR2*, *GATA3*, *PIK3CA*, *TP53*. We identified: 11 SNVs in *AKT1*, 100 SNVs in *AKT2*, 32 SNVs in *ESR1*, 26 SNVs in *ESR2*, 52 SNVs in *GATA3*, 4 SNVs in *PIK3CA*, 71 SNVs in *TP53*. The Variant Allele Frequency (VAF) of detected SNVs was shown

on Graph 15. The median VAF value for most of the genes was between 0,2-0,4 suggesting that detected variants were most probably heterozygotic or occurred in amplified gene. For *ESR1* the median VAF value was the highest >0,8, which is associated with occurrence of 13 variants with VAF=1 (Graph 15). Interestingly, all these variants were associated with the common populational polymorphism P325P. All detected SNVs in all genes were visualized on the Supplementary Graph. The detected variants were marked on the full protein scheme and the numbers of these variants identified in all samples were shown.

To establish the genetic heterogeneity of single CTCs we divided the detected variants into two groups: unique variants occurring in only one single cell and common variants occurring in at least two single cells. Common variants were listed in Table 13. For *AKT2* gene p.R23R silent polymorphism was not described before. However, for *ESR1* gene is associated with common populational polymorphism. The unique variants were the majority of all exon SNVs, with frequency ~67%. The high number of unique variants among all SNVs highlights the heterogeneity of single CTCs.



Graph 15. The variant allele frequency (VAF) of SNVs detected in the exons of sequenced genes. The lines in the box plot corresponds with median values of VAF. A) The VAF in whole gene sequencing; B) The VAF in hotspot gene sequencing. Graph generated using maftools [146]

Table 13. Mutations detected in more than one single cell.

Gene	Mutation	No. Of cells identified
AKT1	p.H354H	4
AKT2	p.K111K	2
AKT2	p.S398S	2
AKT2	p.P453L	2
AKT2	p.R467Q/F	2
AKT2	p.R23R	41
ESR1	p.K531K	2
ESR1	p.N532N	2
ESR1	p.P325P	15
ESR2	p.V328V	2
GATA3	p.P135S/A/S	2
GATA3	p.T418T	3
TP53	p.C182Y	2
TP53	p.P4S	3
TP53	p.P72R	23

5.4.2.2. Intrapatient molecular heterogeneity

To further explore the intrapatient heterogeneity of single CTCs we investigated the SNVs in single cells originated from the same patient. We analyzed 39 single cells originated from material collected from 13 different patients (Table 14). We observed that for 9 patients some single CTCs are sharing the same variants. Similarly, to previously observed trend for common variants, most of the variants occurring in more than one single cell were associated with silent polymorphisms (Table 15). For 4 patients all variants identified in single CTCs were exceptional and did not occur in other cells collected from those patients.

Table 14. Number of analyzed single CTCs originated from one patient.

Patients no.	**No. of CTCs**
1	4
2	8
4	2
8	2
10	2
11	3
17	3
20	2
21	3
22	2
23	2
24	4
35	2

Table 15. Variants detected in at least two cells originated from one patient.

Gene	**Patient no.**	**Mutation**	**Variant Type**	**No. of cells with variant/No. of cells analyzed**
AKT2	1	p.R23R	Silent	3/4
AKT2	2	p.R23R	Silent	8/8
ESR1	2	p.P325P	Silent	3/8
TP53	2	p.P72R	Missense	4/8
AKT2	8	p.R23R	Silent	2/2
TP53	10	p.P72R	Missense	2/2
AKT2	11	p.R23R	Silent	2/3
TP53	20	p.P72R	Missense	2/2
GATA3	23	p.T418T	Silent	2/2
AKT2	24	p.R23R	Silent	3/4
AKT2	35	p.R23R	Silent	2/2

5.4.3. Detection of mutations associated with the resistance to hormonal therapy.

The NGS libraries were designed to identify the mutations in hotspot regions of *ESR1* and *PIK3CA* genes. Overall, we detected 32 SNVs variants in *ESR1* gene and 5 variants in *PIK3CA* gene. In *ESR1* most of the variants were silent mutations. We identified 8 missense mutations in 8 different single CTCs. For *PIK3CA* gene 4 detected mutations were missense variants (Table 17). One of these variants was identified as p.H1047R mutation, which is widely described as the mutation associated with the HTH resistance. Interestingly, the missense mutations in *ESR1* gene were mainly detected in different cells originating from the same sample. This highlights the intrapatient heterogeneity of *ESR1* missense mutations. For *PIK3CA* all missense mutations originated from different samples, collected from different patients, highlighting the interpatient heterogeneity of *PIK3CA* missense mutations.

As the mutations in *ESR1* and *PIK3CA* genes were detected in few patients we did not perform any survival analysis associated with the occurrence of presented variants.

Table 17. Missense mutations detected in *ESR1* and *PIK3CA* genes.

ESR1		
Mutation	Sample no.	Cell ID.
p.L536I	62	M2
p.M522I	62	S1
p.D374N	62	S4
p.S527N	94	4
p.Q375E	94	5
p.N304Y	313	1
p.W292S	313	1
p.W393R	524	1
PIK3CA		
Mutation	Sample no.	Cell ID.
p.Q1042E	62	B1
p.H1047R	372	2
p.H1060L	393	1

5.4.4. Comparison with the primary tumor (FFPE)

For comparison with the single cell NGS data, we sequenced FFPE samples from 19 patients with starters specific for *ESR1* and *PIK3CA* regions (Sanger sequencing). We had chosen these genes as they association with the resistance to the anti-estrogen therapy is well described in the literature.

We identified 13 different mutations in *ESR1* gene and 4 different mutations in *PIK3CA* gene (Table 18) in the FFPE samples. From mutations detected in *ESR1* gene we identified 3 silent mutations, 1 nonsense mutation and 8 missense mutations. The missense mutations were identified in material from 11 different patients. One FFPE

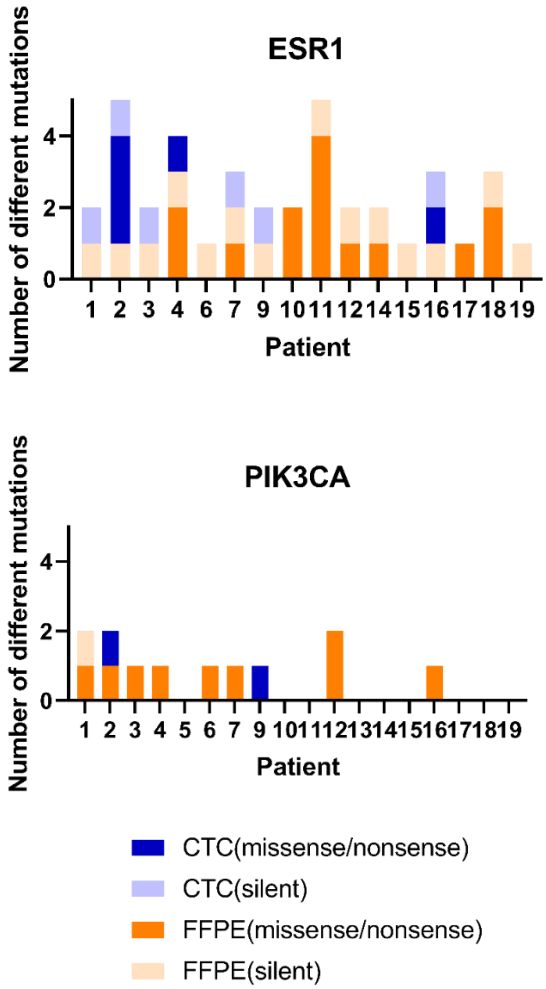
sample #11 was recognized with 5 different missense mutations. All mutations detected in *PIK3CA* gene were missense mutations.

Next, we compared the presence of mutations identified in FFPE with variants identified in the single CTCs. The only mutations that were present in both FFPE and CTCs material were the silent mutations in *ESR1* gene (Supplementary Table 1). For further analysis of changes in mutational status between FFPE and CTCs material we focused on the missense mutations as these variants are associated with the potential resistance to anti-estrogen therapy.

Table 18. Summary of mutations identified in primary tumor (FFPE) material.

Gene	**Mutation**	**Mutation type**	**No. of FFPE identified**	**Frequency**
ESR1	p.P325P	Silent	15	78,95%
ESR1	p.K401*	Nonsense	3	15,79%
ESR1	p.S396T	Missense	2	10,53%
ESR1	p.D369Y	Missense	1	5,26%
ESR1	p.D569Y	Missense	1	5,26%
ESR1	p.Q565P	Missense	2	10,53%
ESR1	p.R548R	Silent	1	5,26%
ESR1	p.H547Y	Missense	1	5,26%
ESR1	p.H550Y	Missense	1	5,26%
ESR1	p.P552Q	Missense	1	5,26%
ESR1	p.S559S	Silent	1	5,26%
ESR1	p.D545E	Missense	1	5,26%
PIK3CA	p.Q546H	Missense	1	5,26%
PIK3CA	p.D549H	Missense	3	15,79%
PIK3CA	p.E517D	Missense	5	26,32%
PIK3CA	p.H1047R	Missense	1	5,26%

To analyze the changes in primary tumor (FFPE) and single CTCs data we compared the number of missense SNVs detected in the corresponding materials from 19 patients. 3 patients were identified with missense mutations in CTCs derived material, while there were no missense mutations in the primary tumor samples. Interestingly, one patient was identified with mutations occurring in single CTCs in both *PIK3CA* and *ESR1* genes. 11 patients were identified with missense mutations present in the primary tumor samples in *ESR1* gene and 9 patients were identified with the missense mutations in the *PIK3CA* in the primary tumor samples. Interestingly, none of the missense mutations detected in FFPE material were found in the CTCs samples (Graph 16).



Graph 16. Distribution of different mutations detected in the primary tumor and CTC of the same patient (n=19)

Because of the small number of patients with the identified missense variants in *ESR1* and *PIK3CA* genes we did not perform any survival analysis associated with the mutational status.

5.4.5. Detection of new mutations in patients' samples

The molecular analysis of the FFPE and single CTCs samples enabled to detect new, unreported mutations - absent in NCBI and COSMIC databases. These mutations are presented in Table 19.

Table 19. New mutations detected during the molecular analysis.

Gene	**Mutation**	**Position**	**Substitution**	**Detected in**
ESR1	p.S396T	152011742	T>A	FFPE
AKT2	p.R23R	40257034	T>G	Single CTCs

6. Discussion

The main aim of this work was to evaluate the EpCAM-independent CTCs detection method CytoTrack as the new approach for the CTCs detection and characterization. The optimization of this method enabled to detect the CTCs with mean recovery ratio ~79%. High CTCs count ≥5CTCs was identified as significant prognostic factor for OS and PFS in multivariate analysis in all collections. Moreover, changes in CTCs counts during treatment correlated with the survival of patients. Importantly, the constant low CTCs counts (<5 CTCs) were identified as the significant favorable predictor for both OS and PFS.

The high counts of CTCs in metastatic breast cancer patients were previously reported as prognostic for PFS and OS [107, 132, 133, 136]. In our study, the high CTCs count (≥5CTCs) was identified as prognostic risk factor regardless of the follow-up time. This data confirms previous reports for CTCs count as the independent prognostic factor during treatment regardless of blood collection time [132].

Most studies in CTCs area are planned as only one blood collection at the beginning of a new treatment or with one follow-up during the treatment. Despite growing evidence of CTCs clinical utility as early recurrence markers, there is still small number of studies utilizing CTCs detection for consistent monitoring during treatment. Therefore, the knowledge about clinical utility of CTCs dynamics during treatment is still insufficient. In our research we analyzed the changes in CTCs number during nine months treatment and their prognostic significance. The patients with increased CTCs count from <5CTCs to ≥5CTCs in 2nd blood collection were characterized with worse PFS and OS. In further collection, the patients with persistent high CTCs counts were characterized with shorter OS. This supports previous reports considering prognostic value of CTCs changes during treatment [138-140].

In our analysis patients with consistent low CTCs numbers (<5CTCs) were characterized with longer PFS and OS than other patients. Additionally, we identified constant low CTCs counts as significant favorable prognostic factor for PFS and OS in multivariable Cox model. This data supports previous studies [139, 140] highlighting the importance of consistent CTCs monitoring. We also observed previously reported trend that patients with constant ≥5CTCs

counts were characterized with shorter OS and PFS [139]. However, this was not reflected in the proportional Cox hazard regression model. Most probably, it is associated with low number of patients with constant ≥5CTCs in the analyzed group of patients.

We also approached validation of the CTCs and CTCs dynamics as rapid progression markers. The ≥5CTCs count was identified as significant rapid progression marker in 2^{nd} and 3^{rd} blood collections, but not in 1^{st}. This suggests that its utility as rapid progression marker might be associated with the advancement of the disease. Interestingly, other data from this analysis support this hypothesis, as we identified the rising numbers of CTCs during treatment as significant prognostic factors for rapid progression. The increase in the CTCs counts in 2^{nd} blood collection was significantly rising the odds for progression or dead during next 3 months. Moreover, constantly rising numbers of CTCs in all collections were associated with very high odds of progression or death during 3 months from last collection. These data highlight the importance and utility of CTCs monitoring during therapy. This data supports previous report of CTCs utility as an early predictor for progression [138].

The only methods for CTCs identification approved by FDA are EpCAM-dependent CellSearch system and size-separation system Parsortix. These systems might lack the specificity for identification of CTCs with low EpCAM and small size cancer cells. Therefore, alternate methods are still being developed. In our analysis we used EpCAM independent CytoTrack system that enabled us to characterize the EpCAM status of CTCs. In our EpCAM analysis, we identified ~44% of $EpCAM^{low}$ cells and ~0.5% of EpCAM negative cells, which might be in EpCAM-dependent systems. Furthermore, our data highlights the EpCAM expression heterogeneity of CTCs, not only in interpatient but also in intrapatient level. Moreover, our data emphasizes the difference in epithelial status between CTCs clusters and single CTCs. CTCs clusters were identified with higher EpCAM expression than single CTCs, suggesting that firm cell-cell junctions are crucial for cluster survival in a bloodstream. Other studies highlighted that $EpCAM^{high}$ CTCs are more predictive for breast cancer patients than $EpCAM^{low}$ CTCs [104, 105]. Similar conclusions had been drawn for CTCs cluster presence [109, 110]. In our studies we did not perform any prognostic analysis. However, considering other works in this area the

observed trend of high EpCAM expression in clusters compared with single CTCs seem to be important improvement to current knowledge.

The tumor heterogeneity is one of the main challenges in cancer studies. In this study, we analyzed single cell material from circulating tumor cells isolated from metastatic breast cancer patients' blood. In most cells we detected the unique variants, which occurred only in one cell, highlighting the great heterogeneity of breast cancer. Variants that were shared between the cells from different patients and/or the same patient were associated with common polymorphisms previously described in the literature [147-149]. The R23R silent polymorphism in *AKT2* gene is the only undescribed polymorphism we detected. Interestingly, it was detected in ~67% of the analyzed single cells, being the most frequent variant occurring in patients. Therefore, we suspect that this silent variant might be undescribed common populational polymorphism or silent gene polymorphism associated with higher breast cancer risk. The larger genomic study should be done to clarify the role of R23R polymorphism in *AKT2* as potential breast cancer risk factor.

The occurrence of the mutations in hotspot regions of *ESR1* and *PIK3CA* genes is known to be associated with resistance to the endocrine therapy [150-152]. In our study we detected missense mutations in *ESR1* and *PIK3CA* genes in both FFPE and CTCs material of different patients. Interestingly, none of the mutations occurring in primary tumor samples were found in CTCs material. Similarly, none of the mutations detected in CTCs were found in primary tumor samples. We also did not observe previously reported trend for higher occurrence of *ESR1* and/or *PIK3CA* mutations in CTCs material compared to FFPE samples [143]. Interestingly, we detected different missense mutations in distinct CTCs originated from the same samples. It might indicate that the analyzed CTCs originated from different metastases or from a different subclone of the same lesion. This data also supports previously reported heterogeneity of CTCs in *ESR1* gene [141, 143].

7. Conclusions

First of all, we optimized the novel EpCAM-independent method of CTCs detection. The mean recovery ratio for this method was established as ~79% with very high specificity, as we did not identify any false positives in negative controls and healthy donors' samples. This new method enabled us to assess the EpCAM status heterogeneity in CTCs.

The prognostic value of high CTCs count (≥5 CTCs) was maintained during the observation period. Moreover, the rising counts of CTCs during treatment were also identified as the unfavorable risk factor. Furthermore, the constant low CTCs count (<5 CTCs) during treatment was identified as strong favorable factor for metastatic breast cancer patients. These findings are highly relevant for improving prognostication in metastatic breast cancer and in helping clinicians monitor patients during systemic therapy. Moreover, in this study describes the difference in EpCAM expression between CTCs clusters and single CTCs. The high heterogeneity of CTCs in EpCAM status, highlights the phenotypic plasticity of single cells. Additionally, we also confirmed the great genomic heterogeneity of circulating tumor cells. The results of this study highlight the clinical utility of the CTCs detection and enumeration during the treatment in metastatic breast cancer patients.

8. Ethics statement

All patients enrolled in this study signed the informed consent. The study protocol was approved by the Ethics Committee of the National Cancer Research Institute (34/2016).

9. Funding

This work was done as the part of the scientific grant from National Science Center (NCN: 2016/21/B/NZ2/03473)

10. NGS data availability

All data from next generation sequencing analysis were deposited in the NCBI SRA repository under the accession number PRJNA1021042.

11. References

1. Wojciechowska Urszula, D.J. *Zachorowania i zgony na nowotwory złośliwe w Polsce. Krajowy Rejestr Nowotworów*. [cited 2021 20-10]; Available from: http://onkologia.org.pl/raporty/.
2. Jassem, J. and M. Krzakowski, *Breast cancer.* Oncology in Clinical Practice, 2018. **14**(4): p. 171-215.
3. Cancer, E.C.I.o.B. *Screening ages and frequencies*. [cited 2021 21-10]; Available from: https://healthcare-quality.jrc.ec.europa.eu/european-breast-cancer-guidelines/screening-ages-and-frequencies.
4. Cardoso, F., et al., *Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-updagger.* Ann Oncol, 2019. **30**(8): p. 1194-1220.
5. Gennari, A., et al., *ESMO Clinical Practice Guideline for the diagnosis, staging and treatment of patients with metastatic breast cancer ☆.* Annals of Oncology, 2021. **32**(12): p. 1475-1495.
6. Agarwal I, B.L. *WHO classification*. [cited 2021 25-10]; Available from: https://www.pathologyoutlines.com/topic/breastmalignantwhoclassification.html.
7. Nielsen, T.O., et al., *Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma.* Clin Cancer Res, 2004. **10**(16): p. 5367-74.
8. Wallden, B., et al., *Development and verification of the PAM50-based Prosigna breast cancer gene signature assay.* BMC Med Genomics, 2015. **8**: p. 54.
9. Kittaneh, M., A.J. Montero, and S. Gluck, *Molecular profiling for breast cancer: a comprehensive review.* Biomark Cancer, 2013. **5**: p. 61-70.
10. Itoh, M., et al., *Estrogen receptor (ER) mRNA expression and molecular subtype distribution in ER-negative/progesterone receptor-positive breast cancers.* Breast Cancer Res Treat, 2014. **143**(2): p. 403-9.
11. Ignatiadis, M. and C. Sotiriou, *Luminal breast cancer: from biology to treatment.* Nat Rev Clin Oncol, 2013. **10**(9): p. 494-506.
12. Jelovac, D. and A.C. Wolff, *The adjuvant treatment of HER2-positive breast cancer.* Curr Treat Options Oncol, 2012. **13**(2): p. 230-9.
13. Sharma, P., *Biology and Management of Patients With Triple-Negative Breast Cancer.* Oncologist, 2016. **21**(9): p. 1050-62.
14. Matsumoto, A., et al., *Prognostic implications of receptor discordance between primary and recurrent breast cancer.* Int J Clin Oncol, 2015. **20**(4): p. 701-8.
15. Dumitrescu, R.G., *Interplay Between Genetic and Epigenetic Changes in Breast Cancer Subtypes.* Methods Mol Biol, 2018. **1856**: p. 19-34.
16. Prat, A., et al., *PAM50 assay and the three-gene model for identifying the major and clinically relevant molecular subtypes of breast cancer.* Breast Cancer Res Treat, 2012. **135**(1): p. 301-6.
17. Prat, A., et al., *Clinical implications of the intrinsic molecular subtypes of breast cancer.* Breast, 2015. **24 Suppl 2**: p. S26-35.
18. Rodriguez, C.E., et al., *Breast cancer stem cells are involved in Trastuzumab resistance through the HER2 modulation in 3D culture.* J Cell Biochem, 2018. **119**(2): p. 1381-1391.
19. Korkaya, H. and M.S. Wicha, *HER2 and breast cancer stem cells: more than meets the eye.* Cancer Res, 2013. **73**(12): p. 3489-93.
20. Ithimakin, S., et al., *HER2 drives luminal breast cancer stem cells in the absence of HER2 amplification: implications for efficacy of adjuvant trastuzumab.* Cancer Res, 2013. **73**(5): p. 1635-46.

21. Prat, A., et al., *Molecular characterization of basal-like and non-basal-like triple-negative breast cancer.* Oncologist, 2013. **18**(2): p. 123-33.
22. Park, S.Y., et al., *Distinct patterns of promoter CpG island methylation of breast cancer subtypes are associated with stem cell phenotypes.* Mod Pathol, 2012. **25**(2): p. 185-96.
23. King, T.A., et al., *Clinical management factors contribute to the decision for contralateral prophylactic mastectomy.* J Clin Oncol, 2011. **29**(16): p. 2158-64.
24. Han, E., et al., *Increasing incidence of bilateral mastectomies: the patient perspective.* Am J Surg, 2011. **201**(5): p. 615-8.
25. Katz, S.J. and M. Morrow, *Contralateral prophylactic mastectomy for breast cancer: addressing peace of mind.* JAMA, 2013. **310**(8): p. 793-4.
26. van Maaren, M.C., et al., *10 year survival after breast-conserving surgery plus radiotherapy compared with mastectomy in early breast cancer in the Netherlands: a population-based study.* The Lancet Oncology, 2016. **17**(8): p. 1158-1170.
27. Hwang, E.S., et al., *Survival after lumpectomy and mastectomy for early stage invasive breast cancer: the effect of age and hormone receptor status.* Cancer, 2013. **119**(7): p. 1402-11.
28. Lagendijk, M., et al., *Breast conserving therapy and mastectomy revisited: Breast cancer-specific survival and the influence of prognostic factors in 129,692 patients.* Int J Cancer, 2018. **142**(1): p. 165-175.
29. Gentilini, O.D., M.J. Cardoso, and P. Poortmans, *Less is more. Breast conservation might be even better than mastectomy in early breast cancer patients.* Breast, 2017. **35**: p. 32-33.
30. Bartelink, H., et al., *Whole-breast irradiation with or without a boost for patients treated with breast-conserving surgery for early breast cancer: 20-year follow-up of a randomised phase 3 trial.* Lancet Oncol, 2015. **16**(1): p. 47-56.
31. Goldberg, M. and T.J. Whelan, *Accelerated Partial Breast Irradiation (APBI): Where Are We Now?* Curr Breast Cancer Rep, 2020. **12**(4): p. 275-284.
32. Meattini, I., et al., *Accelerated partial breast irradiation using intensity modulated radiotherapy versus whole breast irradiation: Health-related quality of life final analysis from the Florence phase 3 trial.* Eur J Cancer, 2017. **76**: p. 17-26.
33. Strnad, V., et al., *5-year results of accelerated partial breast irradiation using sole interstitial multicatheter brachytherapy versus whole-breast irradiation with boost after breast-conserving surgery for low-risk invasive and in-situ carcinoma of the female breast: a randomised, phase 3, non-inferiority trial.* Lancet, 2016. **387**(10015): p. 229-38.
34. Lohrisch, C., et al., *Impact on survival of time from definitive surgery to initiation of adjuvant chemotherapy for early-stage breast cancer.* J Clin Oncol, 2006. **24**(30): p. 4888-94.
35. Goldhirsch, A., et al., *Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013.* Ann Oncol, 2013. **24**(9): p. 2206-23.
36. van Maaren, M.C., et al., *Ten-year recurrence rates for breast cancer subtypes in the Netherlands: A large population-based study.* Int J Cancer, 2019. **144**(2): p. 263-272.
37. Cardoso, F., et al., *ESO-ESMO 2nd international consensus guidelines for advanced breast cancer (ABC2).* The Breast, 2014. **23**(5): p. 489-502.
38. Martin, M., et al., *Palbociclib in combination with endocrine therapy versus capecitabine in hormonal receptor-positive, human epidermal growth factor 2-negative, aromatase inhibitor-resistant metastatic breast cancer: a phase III randomised controlled trial—PEARL ☆.* Annals of Oncology, 2021. **32**(4): p. 488-499.
39. Park, Y.H., et al., *Palbociclib plus exemestane with gonadotropin-releasing hormone agonist versus capecitabine in premenopausal women with hormone receptor-positive, HER2-*

negative metastatic breast cancer (KCSG-BR15-10): a multicentre, open-label, randomised, phase 2 trial. The Lancet Oncology, 2019. **20**(12): p. 1750-1759.
40. Arpino, G., et al., *Crosstalk between the estrogen receptor and the HER tyrosine kinase receptor family: molecular mechanism and clinical implications for endocrine therapy resistance.* Endocr Rev, 2008. **29**(2): p. 217-33.
41. Kumar, R. and E.B. Thompson, *The structure of the nuclear hormone receptors.* Steroids, 1999. **64**(5): p. 310-9.
42. Pike, A.C., *Lessons learnt from structural studies of the oestrogen receptor.* Best Pract Res Clin Endocrinol Metab, 2006. **20**(1): p. 1-14.
43. Kuiper, G.G., et al., *Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta.* Endocrinology, 1997. **138**(3): p. 863-70.
44. Jordan, V.C. and M. Morrow, *Tamoxifen, raloxifene, and the prevention of breast cancer.* Endocr Rev, 1999. **20**(3): p. 253-78.
45. Barkhem, T., et al., *Differential response of estrogen receptor alpha and estrogen receptor beta to partial estrogen agonists/antagonists.* Mol Pharmacol, 1998. **54**(1): p. 105-12.
46. Jordan, V.C., *Tamoxifen: catalyst for the change to targeted therapy.* Eur J Cancer, 2008. **44**(1): p. 30-8.
47. Cronin-Fenton, D.P., P. Damkier, and T.L. Lash, *Metabolism and transport of tamoxifen in relation to its effectiveness: new perspectives on an ongoing controversy.* Future Oncol, 2014. **10**(1): p. 107-22.
48. Vogel, C.L., et al., *Toremifene for breast cancer: a review of 20 years of data.* Clin Breast Cancer, 2014. **14**(1): p. 1-9.
49. Robertson, J.F., et al., *A good drug made better: the fulvestrant dose-response story.* Clin Breast Cancer, 2014. **14**(6): p. 381-9.
50. Ciruelos, E., et al., *The therapeutic role of fulvestrant in the management of patients with hormone receptor-positive breast cancer.* Breast, 2014. **23**(3): p. 201-8.
51. Lambertini, M., et al., *Ovarian suppression using luteinizing hormone-releasing hormone agonists during chemotherapy to preserve ovarian function and fertility of breast cancer patients: a meta-analysis of randomized studies.* Ann Oncol, 2015. **26**(12): p. 2408-19.
52. Ahmad, I. and Shagufta, *Recent developments in steroidal and nonsteroidal aromatase inhibitors for the chemoprevention of estrogen-dependent breast cancer.* Eur J Med Chem, 2015. **102**: p. 375-86.
53. Olin, J.L. and M. St Pierre, *Aromatase inhibitors in breast cancer prevention.* Ann Pharmacother, 2014. **48**(12): p. 1605-10.
54. Ma, C.X., C.G. Sanchez, and M.J. Ellis, *Predicting endocrine therapy responsiveness in breast cancer.* Oncology (Williston Park), 2009. **23**(2): p. 133-42.
55. Skafar, D.F., *Formation of a powerful capping motif corresponding to start of "helix 12" in agonist-bound estrogen receptor-alpha contributes to increased constitutive activity of the protein.* Cell Biochem Biophys, 2000. **33**(1): p. 53-62.
56. Merenbakh-Lamin, K., et al., *D538G mutation in estrogen receptor-alpha: A novel mechanism for acquired endocrine resistance in breast cancer.* Cancer Res, 2013. **73**(23): p. 6856-64.
57. Weis, K.E., et al., *Constitutively active human estrogen receptors containing amino acid substitutions for tyrosine 537 in the receptor protein.* Mol Endocrinol, 1996. **10**(11): p. 1388-98.
58. Zhang, Q.X., et al., *An estrogen receptor mutant with strong hormone-independent activity from a metastatic breast cancer.* Cancer Res, 1997. **57**(7): p. 1244-9.

59. Toy, W., et al., *ESR1 ligand-binding domain mutations in hormone-resistant breast cancer.* Nat Genet, 2013. **45**(12): p. 1439-45.
60. Pavlin, M., et al., *A Computational Assay of Estrogen Receptor alpha Antagonists Reveals the Key Common Structural Traits of Drugs Effectively Fighting Refractory Breast Cancers.* Sci Rep, 2018. **8**(1): p. 649.
61. Szostakowska, M., et al., *Resistance to endocrine therapy in breast cancer: molecular mechanisms and future goals.* Breast Cancer Res Treat, 2019. **173**(3): p. 489-497.
62. Fribbens, C., et al., *Plasma ESR1 Mutations and the Treatment of Estrogen Receptor-Positive Advanced Breast Cancer.* J Clin Oncol, 2016. **34**(25): p. 2961-8.
63. McKenna, M., S. McGarrigle, and G.P. Pidgeon, *The next generation of PI3K-Akt-mTOR pathway inhibitors in breast cancer cohorts.* Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 2018. **1870**(2): p. 185-197.
64. Miled, N., et al., *Mechanism of two classes of cancer mutations in the phosphoinositide 3-kinase catalytic subunit.* Science, 2007. **317**(5835): p. 239-42.
65. Ghayad, S.E. and P.A. Cohen, *Inhibitors of the PI3K/Akt/mTOR pathway: new hope for breast cancer patients.* Recent Pat Anticancer Drug Discov, 2010. **5**(1): p. 29-57.
66. Linke, M., et al., *mTORC1 and mTORC2 as regulators of cell metabolism in immunity.* FEBS Lett, 2017. **591**(19): p. 3089-3103.
67. Banerji, S., et al., *Sequence analysis of mutations and translocations across breast cancer subtypes.* Nature, 2012. **486**(7403): p. 405-9.
68. Anderson, E.J., et al., *A Systematic Review of the Prevalence and Diagnostic Workup of PIK3CA Mutations in HR+/HER2- Metastatic Breast Cancer.* Int J Breast Cancer, 2020. **2020**: p. 3759179.
69. Samuels, Y., et al., *High frequency of mutations of the PIK3CA gene in human cancers.* Science, 2004. **304**(5670): p. 554.
70. Carson, J.D., et al., *Effects of oncogenic p110alpha subunit mutations on the lipid kinase activity of phosphoinositide 3-kinase.* Biochem J, 2008. **409**(2): p. 519-24.
71. Zhao, L. and P.K. Vogt, *Helical domain and kinase domain mutations in p110alpha of phosphatidylinositol 3-kinase induce gain of function by different mechanisms.* Proc Natl Acad Sci U S A, 2008. **105**(7): p. 2652-7.
72. Leontiadou, H., et al., *Insights into the mechanism of the PIK3CA E545K activating mutation using MD simulations.* Sci Rep, 2018. **8**(1): p. 15544.
73. Kalsi, N., et al., *Biophysical aspect of phosphatidylinositol 3-kinase and role of oncogenic mutants (E542K & E545K).* J Biomol Struct Dyn, 2016. **34**(12): p. 2711-2721.
74. Mandelker, D., et al., *A frequent kinase domain mutation that changes the interaction between PI3Kalpha and the membrane.* Proc Natl Acad Sci U S A, 2009. **106**(40): p. 16996-7001.
75. Burke, J.E., et al., *Oncogenic mutations mimic and enhance dynamic events in the natural activation of phosphoinositide 3-kinase p110alpha (PIK3CA).* Proc Natl Acad Sci U S A, 2012. **109**(38): p. 15259-64.
76. Vadas, O., et al., *Structural basis for activation and inhibition of class I phosphoinositide 3-kinases.* Sci Signal, 2011. **4**(195): p. re2.
77. Gkeka, P., et al., *Investigating the structure and dynamics of the PIK3CA wild-type and H1047R oncogenic mutant.* PLoS Comput Biol, 2014. **10**(10): p. e1003895.
78. Ronnstrand, L., *Signal transduction via the stem cell factor receptor/c-Kit.* Cell Mol Life Sci, 2004. **61**(19-20): p. 2535-48.
79. Heron-Milhavet, L., et al., *Akt1 and Akt2: differentiating the aktion.* Histol Histopathol, 2011. **26**(5): p. 651-62.

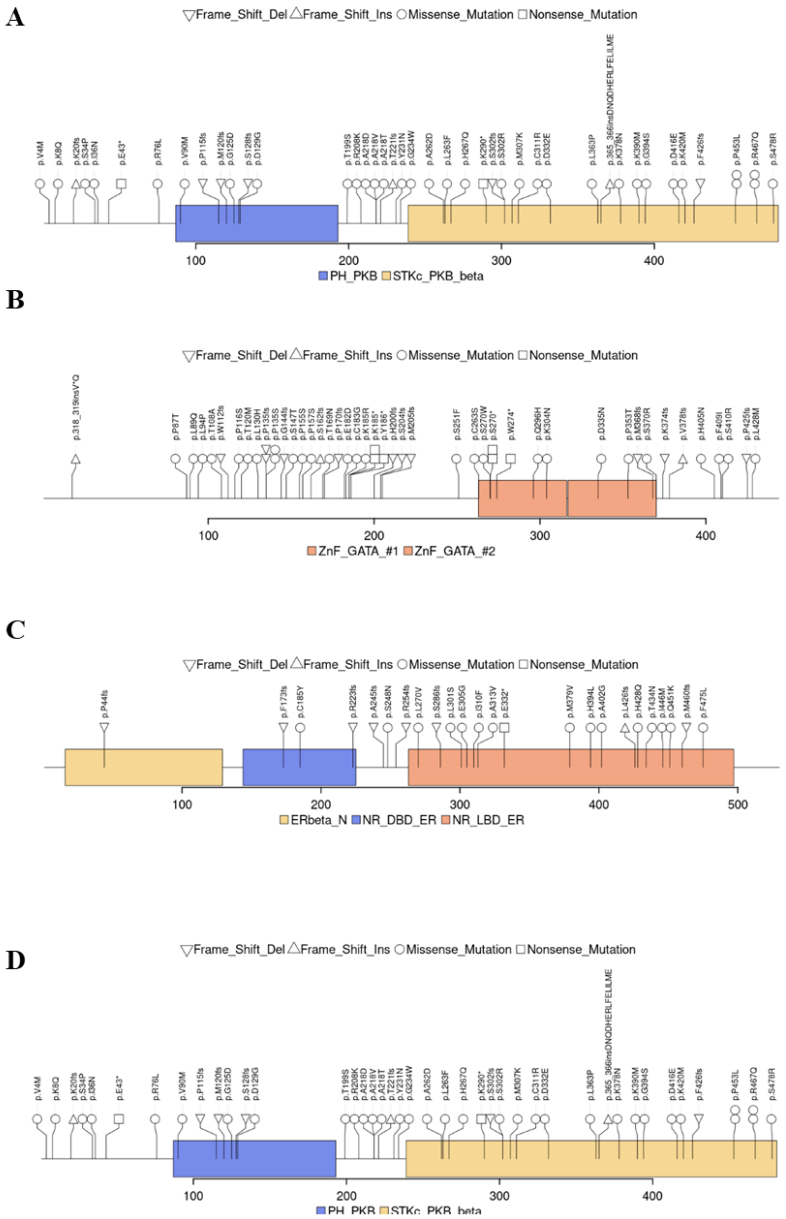
80. Hers, I., E.E. Vincent, and J.M. Tavare, *Akt signalling in health and disease.* Cell Signal, 2011. **23**(10): p. 1515-27.
81. Landgraf, K.E., C. Pilling, and J.J. Falke, *Molecular mechanism of an oncogenic mutation that alters membrane targeting: Glu17Lys modifies the PIP lipid specificity of the AKT1 PH domain.* Biochemistry, 2008. **47**(47): p. 12260-9.
82. Bockhorn, M., R.K. Jain, and L.L. Munn, *Active versus passive mechanisms in metastasis: do cancer cells crawl into vessels, or are they pushed?* Lancet Oncol, 2007. **8**(5): p. 444-8.
83. Folkman, J., *Role of angiogenesis in tumor growth and metastasis.* Semin Oncol, 2002. **29**(6 Suppl 16): p. 15-8.
84. Klein, C.A., *Parallel progression of primary tumours and metastases.* Nat Rev Cancer, 2009. **9**(4): p. 302-12.
85. Klein, C.A. and D. Holzel, *Systemic cancer progression and tumor dormancy: mathematical models meet single cell genomics.* Cell Cycle, 2006. **5**(16): p. 1788-98.
86. Turajlic, S. and C. Swanton, *Metastasis as an evolutionary process.* Science, 2016. **352**(6282): p. 169-75.
87. Copson, E., et al., *Prospective observational study of breast cancer treatment outcomes for UK women aged 18-40 years at diagnosis: the POSH study.* J Natl Cancer Inst, 2013. **105**(13): p. 978-88.
88. Banys-Paluchowski, M., et al., *Circulating and Disseminated Tumor Cells in Breast Carcinoma: Report from the Consensus Conference on Tumor Cell Dissemination during the 39th Annual Meeting of the German Society of Senology, Berlin, 27 June 2019.* Geburtshilfe Frauenheilkd, 2019. **79**(12): p. 1320-1327.
89. Banys, M., N. Krawczyk, and T. Fehm, *The role and clinical relevance of disseminated tumor cells in breast cancer.* Cancers (Basel), 2014. **6**(1): p. 143-52.
90. Ogba, N., et al., *Luminal breast cancer metastases and tumor arousal from dormancy are promoted by direct actions of estradiol and progesterone on the malignant cells.* Breast Cancer Res, 2014. **16**(6): p. 489.
91. Aguirre-Ghiso, J.A., *Models, mechanisms and clinical evidence for cancer dormancy.* Nat Rev Cancer, 2007. **7**(11): p. 834-46.
92. Cools-Lartigue, J., et al., *Neutrophil extracellular traps sequester circulating tumor cells and promote metastasis.* J Clin Invest, 2013. **123**(8): p. 3446-58.
93. Zhang, X.H., et al., *Metastasis dormancy in estrogen receptor-positive breast cancer.* Clin Cancer Res, 2013. **19**(23): p. 6389-97.
94. Skovierova, H., et al., *Molecular regulation of epithelial-to-mesenchymal transition in tumorigenesis (Review).* Int J Mol Med, 2018. **41**(3): p. 1187-1200.
95. Lamouille, S., J. Xu, and R. Derynck, *Molecular mechanisms of epithelial-mesenchymal transition.* Nat Rev Mol Cell Biol, 2014. **15**(3): p. 178-96.
96. Fischer, K.R., et al., *Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance.* Nature, 2015. **527**(7579): p. 472-6.
97. Pastushenko, I., et al., *Identification of the tumour transition states occurring during EMT.* Nature, 2018. **556**(7702): p. 463-468.
98. Jolly, M.K., et al., *Stability of the hybrid epithelial/mesenchymal phenotype.* Oncotarget, 2016. **7**(19): p. 27067-84.
99. Williams, E.D., et al., *Controversies around epithelial-mesenchymal plasticity in cancer metastasis.* Nat Rev Cancer, 2019. **19**(12): p. 716-732.
100. Yu, M., et al., *Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition.* Science, 2013. **339**(6119): p. 580-4.

101. Saitoh, M., *Involvement of partial EMT in cancer progression.* J Biochem, 2018. **164**(4): p. 257-264.
102. Guan, X., et al., *The prognostic and therapeutic implications of circulating tumor cell phenotype detection based on epithelial-mesenchymal transition markers in the first-line chemotherapy of HER2-negative metastatic breast cancer.* Cancer Commun (Lond), 2019. **39**(1): p. 1.
103. Markiewicz, A., et al., *Spectrum of Epithelial-Mesenchymal Transition Phenotypes in Circulating Tumour Cells from Early Breast Cancer Patients.* Cancers (Basel), 2019. **11**(1).
104. de Wit, S., et al., *EpCAM(high) and EpCAM(low) circulating tumor cells in metastatic prostate and breast cancer patients.* Oncotarget, 2018. **9**(86): p. 35705-35716.
105. Liu, X., et al., *Epithelial-type systemic breast carcinoma cells with a restricted mesenchymal transition are a major source of metastasis.* Sci Adv, 2019. **5**(6): p. eaav4275.
106. Stefanovic, S., et al., *The Lack of Evidence for an Association between Cancer Biomarker Conversion Patterns and CTC-Status in Patients with Metastatic Breast Cancer.* Int J Mol Sci, 2020. **21**(6).
107. Peeters, D.J., et al., *Detection and prognostic significance of circulating tumour cells in patients with metastatic breast cancer according to immunohistochemical subtypes.* Br J Cancer, 2014. **110**(2): p. 375-83.
108. Bostick, P.J., et al., *Limitations of specific reverse-transcriptase polymerase chain reaction markers in the detection of metastases in the lymph nodes and blood of breast cancer patients.* J Clin Oncol, 1998. **16**(8): p. 2632-40.
109. Aceto, N., et al., *Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis.* Cell, 2014. **158**(5): p. 1110-1122.
110. Liu, X., et al., *Homophilic CD44 Interactions Mediate Tumor Cell Aggregation and Polyclonal Metastasis in Patient-Derived Breast Cancer Models.* Cancer Discov, 2019. **9**(1): p. 96-113.
111. Tate, J.G., et al., *COSMIC: the Catalogue Of Somatic Mutations In Cancer.* Nucleic Acids Res, 2019. **47**(D1): p. D941-D947.
112. Pestrin, M., et al., *Heterogeneity of PIK3CA mutational status at the single cell level in circulating tumor cells from metastatic breast cancer patients.* Mol Oncol, 2015. **9**(4): p. 749-57.
113. De Luca, F., et al., *Mutational analysis of single circulating tumor cells by next generation sequencing in metastatic breast cancer.* Oncotarget, 2016. **7**(18): p. 26107-19.
114. Aktas, B., et al., *Comparison of estrogen and progesterone receptor status of circulating tumor cells and the primary tumor in metastatic breast cancer patients.* Gynecol Oncol, 2011. **122**(2): p. 356-60.
115. Neumann, M.H., et al., *Isolation and characterization of circulating tumor cells using a novel workflow combining the CellSearch((R)) system and the CellCelector().* Biotechnol Prog, 2017. **33**(1): p. 125-132.
116. Templeman, A., et al., *Analytical performance of the FDA-cleared Parsortix((R)) PC1 system.* J Circ Biomark, 2023. **12**: p. 26-33.
117. Hillig, T., et al., *In vitro detection of circulating tumor cells compared by the CytoTrack and CellSearch methods.* Tumour Biol, 2015. **36**(6): p. 4597-601.
118. Miller, M.C., et al., *The Parsortix Cell Separation System-A versatile liquid biopsy platform.* Cytometry A, 2018. **93**(12): p. 1234-1239.
119. Andreopoulou, E., et al., *Comparison of assay methods for detection of circulating tumor cells in metastatic breast cancer: AdnaGen AdnaTest BreastCancer Select/Detect versus Veridex CellSearch system.* Int J Cancer, 2012. **130**(7): p. 1590-7.

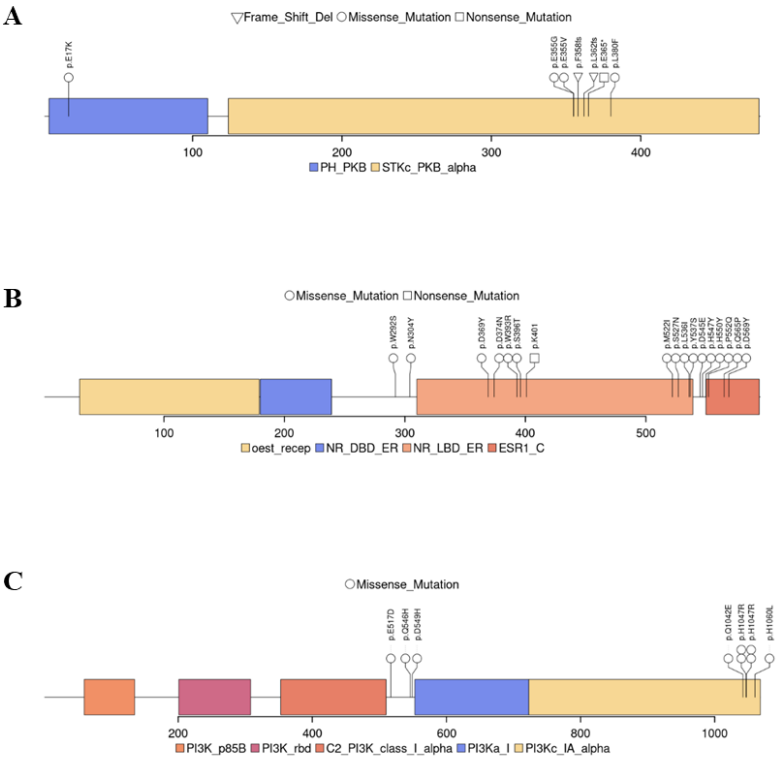
120. Drucker, A., et al., *Comparative performance of different methods for circulating tumor cell enrichment in metastatic breast cancer patients.* PLoS One, 2020. **15**(8): p. e0237308.
121. Ma, Y.C., L. Wang, and F.L. Yu, *Recent advances and prospects in the isolation by size of epithelial tumor cells (ISET) methodology.* Technol Cancer Res Treat, 2013. **12**(4): p. 295-309.
122. Livesey, F.J., *Strategies for microarray analysis of limiting amounts of RNA.* Brief Funct Genomic Proteomic, 2003. **2**(1): p. 31-6.
123. Zong, C., et al., *Genome-wide detection of single-nucleotide and copy-number variations of a single human cell.* Science, 2012. **338**(6114): p. 1622-6.
124. Klein, C.A., et al., *Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells.* Proc Natl Acad Sci U S A, 1999. **96**(8): p. 4494-9.
125. Dean, F.B., et al., *Comprehensive human genome amplification using multiple displacement amplification.* Proc Natl Acad Sci U S A, 2002. **99**(8): p. 5261-6.
126. Telenius, H., et al., *Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer.* Genomics, 1992. **13**(3): p. 718-25.
127. Cheung, V.G. and S.F. Nelson, *Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA.* Proc Natl Acad Sci U S A, 1996. **93**(25): p. 14676-9.
128. Navin, N.E., *Cancer genomics: one cell at a time.* Genome Biol, 2014. **15**(8): p. 452.
129. Blanco, L., et al., *Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication.* J Biol Chem, 1989. **264**(15): p. 8935-40.
130. Huang, L., et al., *Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications.* Annu Rev Genomics Hum Genet, 2015. **16**: p. 79-102.
131. Bidard, F.C., C. Proudhon, and J.Y. Pierga, *Circulating tumor cells in breast cancer.* Mol Oncol, 2016. **10**(3): p. 418-30.
132. Zhang, L., et al., *Meta-analysis of the prognostic value of circulating tumor cells in breast cancer.* Clin Cancer Res, 2012. **18**(20): p. 5701-10.
133. Cristofanilli, M., et al., *Circulating tumor cells, disease progression, and survival in metastatic breast cancer.* N Engl J Med, 2004. **351**(8): p. 781-91.
134. Lv, Q., et al., *Prognostic value of circulating tumor cells in metastatic breast cancer: a systemic review and meta-analysis.* Clin Transl Oncol, 2016. **18**(3): p. 322-30.
135. Yan, W.T., et al., *Circulating tumor cell status monitors the treatment responses in breast cancer patients: a meta-analysis.* Sci Rep, 2017. **7**: p. 43464.
136. Martin, M., et al., *Circulating tumor cells following first chemotherapy cycle: an early and strong predictor of outcome in patients with metastatic breast cancer.* Oncologist, 2013. **18**(8): p. 917-23.
137. Giordano, A., et al., *Circulating tumor cells in immunohistochemical subtypes of metastatic breast cancer: lack of prediction in HER2-positive disease treated with targeted therapy.* Ann Oncol, 2012. **23**(5): p. 1144-1150.
138. Larsson, A.M., et al., *Longitudinal enumeration and cluster evaluation of circulating tumor cells improve prognostication for patients with newly diagnosed metastatic breast cancer in a prospective observational trial.* Breast Cancer Res, 2018. **20**(1): p. 48.
139. Hayes, D.F., et al., *Circulating tumor cells at each follow-up time point during therapy of metastatic breast cancer patients predict progression-free and overall survival.* Clin Cancer Res, 2006. **12**(14 Pt 1): p. 4218-24.
140. Bidard, F.C., et al., *Clinical validity of circulating tumour cells in patients with metastatic breast cancer: a pooled analysis of individual patient data.* Lancet Oncol, 2014. **15**(4): p. 406-14.

141. Sundaresan, T.K., et al., *Evaluation of endocrine resistance using ESR1 genotyping of circulating tumor cells and plasma DNA.* Breast Cancer Res Treat, 2021. **188**(1): p. 43-52.
142. Greaves, M. and C.C. Maley, *Clonal evolution in cancer.* Nature, 2012. **481**(7381): p. 306-13.
143. Franken, A., et al., *Detection of ESR1 Mutations in Single Circulating Tumor Cells on Estrogen Deprivation Therapy but Not in Primary Tumors from Metastatic Luminal Breast Cancer Patients.* J Mol Diagn, 2020. **22**(1): p. 111-121.
144. Hillig, T., et al., *In vitro validation of an ultra-sensitive scanning fluorescence microscope for analysis of circulating tumor cells.* APMIS, 2014. **122**(6): p. 545-51.
145. Petrackova, A., et al., *Standardization of Sequencing Coverage Depth in NGS: Recommendation for Detection of Clonal and Subclonal Mutations in Cancer Diagnostics.* Front Oncol, 2019. **9**: p. 851.
146. Mayakonda, A., et al., *Maftools: efficient and comprehensive analysis of somatic variants in cancer.* Genome Res, 2018. **28**(11): p. 1747-1756.
147. De Souza, C., et al., *Effect of the p53 P72R Polymorphism on Mutant TP53 Allele Selection in Human Cancer.* J Natl Cancer Inst, 2021. **113**(9): p. 1246-1257.
148. Zhang, Y., et al., *Association Between ESR1 PvuII, XbaI, and P325P Polymorphisms and Breast Cancer Susceptibility: A Meta-Analysis.* Med Sci Monit, 2015. **21**: p. 2986-96.
149. Zavratnik, A., et al., *Exonic, but not intronic polymorphisms of ESR1 gene might influence the hypolipemic effect of raloxifene.* J Steroid Biochem Mol Biol, 2007. **104**(1-2): p. 22-6.
150. Herzog, S.K. and S.A.W. Fuqua, *ESR1 mutations and therapeutic resistance in metastatic breast cancer: progress and remaining challenges.* Br J Cancer, 2022. **126**(2): p. 174-186.
151. Tolaney, S.M., et al., *Clinical Significance of PIK3CA and ESR1 Mutations in Circulating Tumor DNA: Analysis from the MONARCH 2 Study of Abemaciclib plus Fulvestrant.* Clin Cancer Res, 2022. **28**(8): p. 1500-1506.
152. Mosele, F., et al., *Outcome and molecular landscape of patients with PIK3CA-mutated metastatic breast cancer.* Ann Oncol, 2020. **31**(3): p. 377-386.

12. Supplementary Materials



Supplementary graph 1. The SNVs variants associated with mutations detected in A) *TP53;* B) *GATA3;* C) *ESR2;* D) *AKT2* genes. The bar on the left (Y axis) shows the number of identified mutations. The X axis is the visualization of the full protein. Detected variants are marked on the protein scheme and their number corresponds to the heigh of the dot. Graphs generated using maftools [146].



Supplementary graph 2. The SNVs variants associated with mutations detected in A) *AKT1;* B) *ESR1;* C) *PIK3CA* genes. The bar on the left (Y axis) shows the number of identified mutations. The X axis is the visualization of the full protein. Detected variants are marked on the protein scheme and their number corresponds to the heigh of the dot. Graphs generated using maftools [146].

Supplementary table 1a. The mutations identified in the FFPE material and their presence in the corresponding CTCs.

Gene	FFPE ID	Position	Substitution	Protein change	Variant Type	Present in corresponding CTCs	Corresponding Disc ID
ESR1	1	151944387	G>C	p.P325P	Silent	yes	78
ESR1	1	152098788	A>C	p.Y537S	Missense	no	78
PIK3CA	1	179234297	A>G	p.H1047R	Missense	no	78
ESR1	2	151944387	G>C	p.P325P	Silent	yes	62
PIK3CA	2	179218221	A>T	p.E517D	Missense	no	62
ESR1	3	151944387	G>C	p.P325P	Silent	yes	63
ESR1	3	152098788	A>C	p.Y537S	Missense	no	63
PIK3CA	3	179218221	A>T	p.E517D	Missense	no	63
ESR1	4	151944387	G>C	p.P325P	Silent	no	94
ESR1	4	152011752	A>T	p.K401*	Nonsense	no	94
ESR1	4	152011663	G>T	p.D369Y	Missense	no	94
PIK3CA	4	179218221	A>T	p.E517D	Missense	no	94
ESR1	5	151944387	G>C	p.P325P	Silent	no	95
ESR1	6	151944387	G>C	p.P325P	Silent	yes	310
ESR1	6	152098788	A>C	p.Y537S	Missense	no	310
PIK3CA	6	179218315	G>C	p.D549H	Missense	no	310
ESR1	7	151944387	G>C	p.P325P	Silent	no	397
ESR1	7	152011663	G>T	p.D569Y	Missense	no	397
ESR1	7	152098788	A>C	p.Y537S	Missense	no	397
PIK3CA	7	179218308	G>A	p.Q546H	Missense	no	397
ESR1	9	151944387	G>C	p.P325P	Silent	yes	372
ESR1	9	152098788	A>C	p.Y537S	Missense	no	372
ESR1	10	152011752	A>T	p.K401*	Nonsense	no	253
PIK3CA	10	179218221	A>T	p.E517D	Missense	no	253
ESR1	11	151944387	G>C	p.P325P	Silent	yes	371
ESR1	11	152098823	C>T	p.R548R	Missense	no	371
ESR1	11	152098817	C>T	p.H547Y	Missense	no	371
ESR1	11	152098826	C>A	p.H550Y	Missense	no	371
ESR1	11	152098833/4	CC>AA	p.P552Q	Missense	no	371
ESR1	11	152098854	C>A	p.S559S	Silent	no	371
ESR1	11	152098873	A>T	p.Q565P	Missense	no	371
ESR1	12	151944387	G>C	p.P325P	Silent	no	366
ESR1	12	152098812	C>G	p.D545E	Missense	no	366
PIK3CA	12	179218221	A>T	p.E517D	Missense	no	366
PIK3CA	12	179218315	G>C	p.D549H	Missense	no	366
ESR1	13	151944387	G>C	p.P325P	Silent	no	293
ESR1	14	152011742	T>A	p.S396T	Missense	no	381
ESR1	15	151944387	G>C	p.P325P	Silent	no	317
ESR1	16	151944387	G>C	p.P325P	Silent	yes	313

PIK3CA	16	179218315	G>C	p.D549H	Missense	no	313
ESR1	17	152011742	T>A	p.S396T	Missense	no	384
ESR1	18	151944387	G>C	p.P325P	Silent	no	322
ESR1	18	152011752	A>T	p.K401*	Nonsense	no	322
ESR1	18	152098872	A>C	p.Q565P	Missense	no	322
ESR1	19	151944387	G>C	p.P325P	Silent	no	21

Suplementary table 1b. The mutations idetified in CTCs and their presence in corresponding FFPE.

Gene	Disc ID	Position	Substitution	Mutation	Variant Type	Present in corresponding FFPE	Corresponding FFPE ID
ESR1	62	152098784	C>A	p.L536I	Missense	no	2
ESR1	62	152011679	G>A	p.D374N	Missense	no	2
ESR1	62	151944287	G>T	p.M522I	Missense	no	2
ESR1	62	152098744	G>C	p.P325P	Silent	yes	2
ESR1	62	152098758	G>C	p.P325P	Silent	yes	2
ESR1	62	151944387	G>C	p.P325P	Silent	yes	2
ESR1	62	151944387	G>C	p.P325P	Silent	yes	2
PIK3CA	62	151944387	C>G	p.Q1042E	Missense	no	2
ESR1	63	151944387	G>C	p.P325P	Silent	yes	3
ESR1	78	151944387	G>C	p.P325P	Silent	yes	1
ESR1	94	151944387	G>A	p.S527N	Missense	no	4
ESR1	310	151944387	G>C	p.P325P	Silent	yes	6
ESR1	313	151944387	G>C	p.W292S	Missense	no	16
ESR1	313	151944387	G>C	p.P325P	Silent	yes	16
ESR1	372	151944387	G>C	p.P325P	Silent	yes	9
PIK3CA	372	179234297	A>G	p.H1047R	Missense	no	9
ESR1	397	179234281	G>C	p.P325P	Silent	yes	7