

**AVALIAÇÃO MOLECULAR DE LESÕES PRÉ-INVASIVAS DE
CARCINOMA DUCTAL DE MAMA PARA DETECTAR
CANDIDATOS A BIOMARCADORES DE PREDIÇÃO DE
PROGRESSÃO DE CARCINOMA DUCTAL *IN SITU***

PAULO HENRIQUE BALDAN PINEDA

**Tese apresentada à Fundação Antônio Prudente para obtenção
do título de Doutor em Ciências**

Área de concentração: Oncologia

Orientadora: Dra. Dirce Maria Carraro

São Paulo

2022

FICHA CATALOGRÁFICA

Baldan-Pineda, Paulo Henrique.

Avaliação molecular de lesões pré-invasivas de carcinoma ductal de mama para detectar candidatos a biomarcadores de predição de progressão de carcinoma ductal *in situ* / Paulo Henrique Baldan Pineda – São Paulo, 2022.
98f.

Tese (Doutorado)-Fundação Antônio Prudente. Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.
Orientador: Dirce Maria Carraro.

1. Carcinoma ductal *in situ* de mama. 2. Avaliação molecular. 3. Biomarcadores de progressão de CDIS.

CDU 616

FOLHA DE APROVAÇÃO

Paulo Henrique Baldan Pineda

Avaliação molecular de lesões pré-invasivas de carcinoma ductal de mama para detectar candidatos a biomarcadores de predição de progressão de carcinoma ductal *in situ*

Aprovado em: 20/05/2022

Banca Examinadora

Orientador: Dra. Dirce Maria Carraro
Instituição: Fundação Antônio Prudente

Membro da banca: Dra. Anamaria Aranha Camargo
Instituição: Hospital Sírio-Libanês

Membro da banca: Dra. Marcia Maria Chiquitelli Marques Silveira
Instituição: Hospital de Amor

Membro da banca: Dra. Maria Aparecida Nagai
Instituição: Faculdade de Medicina da Universidade de São Paulo

Membro da banca: Dra. Marina De Brot Andrade
Instituição: Fundação Antônio Prudente

*Dedico este esforço a minha mãe Maria Elza e a minha esposa
Fernanda por contribuírem na formação do meu caráter, e
pelo incentivo e suporte durante toda esta jornada!*

AGRADECIMENTOS

A minha orientadora, **Dra. Dirce Carraro**, pela orientação e pela oportunidade de desenvolver um estudo que apresentou vários desafios durante a sua execução, os quais me proporcionaram uma melhor capacitação técnica e consequentemente um amadurecimento pessoal.

A **Dra. Eloisa Olivieri** pelo suporte na seleção de amostras da casuística do estudo.

Aos colaboradores do estudo, **Dra. Cynthia Osorio** e **Dr. Almir Bitencourt** pelo suporte com a indicação de pacientes para comporem a casuística do estudo.

Ao técnico de pesquisa **Severino Ferreira** pelo suporte com os cortes histológicos das amostras parafinadas da casuística do estudo.

A técnica de pesquisa **Ana Paula Suenaga** pelo scrape das áreas CDIS e pelo suporte com a extração de DNA tumoral das amostras congeladas e parafinadas da casuística do estudo.

A técnica de pesquisa **Louise Mota** pela extração de DNA de leucócitos das amostras da casuística do estudo.

A **Dra. Cynthia Osorio** pela revisão anatomopatológica dos pacientes que compuseram a casuística do estudo.

A **Dra. Vanina Elias** pelo suporte na etapa de análise de expressão gênica, com a customização dos ensaios TLDA e análise dos dados.

A **Dra. Giovana Torrezan** pelo suporte na etapa de análise de expressão gênica com o manuseio do equipamento de PCR em tempo real, e na etapa de avaliação mutacional com os critérios de chamada de variantes e análise dos dados de sequenciamento.

Ao **Dr. Rafael Brianese** pelo suporte em várias etapas deste estudo, principalmente na etapa da avaliação mutacional, com o sequenciamento de DNA e com a análise dos dados de sequenciamento.

A **Dra. Fabiana Makdissi** e **Dra. Maria Amorim** pela avaliação dos relatórios do doutorado e acompanhamento deste estudo, contribuindo para a melhoria dele.

Aos colegas do **grupo de genômica e biologia molecular do A. C. Camargo Cancer Center** pelo suporte em várias etapas do estudo.

A **pós-graduação da Fundação Antônio Prudente**, por terem me acolhido como seu aluno durante o meu mestrado e doutorado, e pelo suporte ao ensino.

A instituição **A. C. Camargo Cancer Center**, em especial ao **Centro Internacional de Pesquisa e Ensino (CIPE)**, pela infraestrutura que proporciona um ambiente de excelência no ensino e pesquisa.

As agências **FAPESP** e **CNPq** pelo suporte financeiro ao estudo.

A **minha família e amigos** pelos pensamentos positivos e contínuo estímulo aos meus estudos.

A **minha esposa Fernanda Grande** pelo incentivo, ombro amigo e suporte na formatação desta tese.

A **minha mãe Maria Elza Baldan** pela sua dedicação profissional que deram suporte aos meus estudos.

RESUMO

Baldan-Pineda, Paulo Henrique. Avaliação molecular de lesões pré-invasivas de carcinoma ductal de mama para detectar candidatos a biomarcadores de predição de progressão de carcinoma ductal *in situ*. [Tese]. São Paulo; Fundação Antônio Prudente; 2022.

INTRODUÇÃO: O câncer de mama é o tipo de câncer mais frequente na população feminina mundial. Cerca de 75% dos pacientes são do tipo carcinoma ductal, podendo se apresentar como carcinoma ductal *in situ* (CDIS puro) ou carcinoma mamário invasivo tipo não especial (CMI-TNE), sendo que este último pode conter o componente *in situ* associado (componente *in situ* de CMI-TNE). O número de pacientes de CDIS diagnosticados tem aumentado nas últimas décadas devido aos programas de rastreamento, estimulando a busca por biomarcadores capazes de estratificar adequadamente as lesões com maior ou menor risco de progredir ou recidivar, e consequentemente poupar as pacientes com menor risco do tratamento convencional, atualmente considerado excessivo para essas pacientes. **OBJETIVO:** Assim, o objetivo do presente estudo foi caracterizar molecularmente lesões pré-invasivas de carcinoma ductal de mama a fim de identificar candidatos a biomarcadores de predição de progressão ou recidiva de CDIS. **METODOLOGIA:** Amostras de RNA amplificado de células epiteliais neoplásicas e mioepiteliais capturadas por microdissecção a laser dos dois grupos de lesões pré-invasivas (CDIS puro e componente *in situ* de CMI-TNE) foram submetidas a uma análise de expressão gênica com ensaios TLDA customizados e o perfil de expressão dos genes diferencialmente expressos (GDE) ($fold\ change \geq |2|$) foi submetido a uma análise de clusterização para avaliar a capacidade em distinguir os dois tipos de lesões pré-invasivas. Na análise mutacional, dois painéis multigenes (409 genes inteiros e regiões *hotspots* de 50 genes) foram utilizados no sequenciamento de DNA (plataforma *Ion Torrent*) de uma casuística composta por pacientes com laudo anatomopatológico de CDIS puro na biópsia e que apresentaram diferentes desfechos: grupo 1 - diagnóstico de CDIS puro na biópsia e peça cirúrgica e que não manifestaram progressão ou recidiva da doença num período mínimo de 3 anos de seguimento; grupo 2 - CDIS puro na biópsia e que foi subdividido em dois subgrupos: grupo 2.1 - diagnóstico de CMI-TNE com componente *in situ* na peça cirúrgica; grupo 2.2 - diagnóstico de CDIS puro na peça cirúrgica também e que apresentaram progressão ou recidiva da doença no período de seguimento. **RESULTADOS:** Na análise de expressão gênica foram analisados 11 pacientes representados por 15 amostras de células epiteliais neoplásicas ou mioepiteliais (5 CDIS puro e 10 componente *in situ* de CMI-TNE). Confirmamos a diferença de expressão de 49% (26/53) dos genes analisados. O perfil de expressão de 12 genes (*ANAPC13*, *BAALC*, *DNAJB6*, *ERCC8*, *FLCN*, *NCF4*, *RARRES3*, *RBM47*, *REC8*, *RRP1*, *SRA1* e *TMEM45A*) foi capaz de discriminar 80% (4/5) das amostras de CDIS puro de 100% (10/10) das amostras de componente *in situ* de CMI-TNE. Na análise mutacional foram avaliados 34 pacientes (43 amostras) divididos em dois grupos. O grupo 1 (CDIS sem progressão) com 20 pacientes e o grupo 2 (CDIS com progressão) com 14 pacientes, sendo que no grupo 2.1 e 2.2 foram 7 pacientes cada. Dos 34 pacientes, 23 pacientes tiveram 29 amostras com sequenciamento qualificado para a análise mutacional (15 e 8 pacientes dos grupos 1 e 2, respectivamente). Dos 23 pacientes, 16 apresentaram variantes potencialmente somáticas, 53% (8/15) e 100% (8/8) dos grupos 1 e 2,

respectivamente. Das 106 variantes potencialmente somáticas detectadas, 45 foram em 40 oncogenes ou genes supressores tumorais, os quais apresentaram variantes oncogênicas ou provavelmente oncogênicas em 33% (5/15) e 75% (6/8) dos grupos 1 e 2, respectivamente. Variantes potencialmente somáticas no gene *TP53* foram identificadas em 6,6% (1/15) e 25% (2/8) dos grupos 1 e 2, respectivamente; e em *PIK3CA* em 0% e 25% (2/8) dos grupos 1 e 2, respectivamente. **CONCLUSÃO:** A assinatura gênica baseada nos 12 genes tem potencial de prever a progressão de CDIS para a doença invasiva. Além disso, variantes oncogênicas em oncogenes e genes supressores de tumor foram identificadas em maior frequência nos pacientes com diagnóstico de CDIS puro na biópsia e que sofreram progressão/recidiva ou discordância com o diagnóstico da peça cirúrgica.

Descritores: carcinoma ductal *in situ* de mama; avaliação molecular; biomarcadores de progressão de CDIS.

ABSTRACT

Baldan-Pineda, Paulo Henrique. Molecular assessment of preinvasive lesions of breast ductal carcinoma for detecting potential biomarkers of ductal carcinoma *in situ* progression. [Tese]. São Paulo; Fundação Antônio Prudente; 2022.

INTRODUCTION: Breast cancer is the most common cancer in women worldwide. Ductal carcinoma accounts for approximately 80% of all breast cancers and is represented by ductal carcinoma *in situ* (DCIS) and invasive breast carcinoma of no special type (IBC-NST), the last can be detected concurrent with an *in situ* component (*in situ* component of IBC). The diagnosis of DCIS has been increased in the last decades due to an increase in screening mammography programs, encouraging the seek for molecular markers able to stratify patients according to risk of progression or relapse, and consequently to identify patients at a lower risk who could avoid an overtreatment. **OBJECTIVE:** The aim of this study was to perform a molecular characterization of preinvasive lesions of breast ductal carcinoma, to identify potential biomarkers of DCIS progression or relapse. **METHODOLOGY:** Amplified RNA from epithelial and myoepithelial cells captured by laser microdissection of both groups of preinvasive lesions (pure DCIS and *in situ* component of IBC) were submitted to gene expression analysis by TLDA assays and gene expression profile. It was considered genes differentially expressed those with fold change $\geq |2|$. It was performed an unsupervised hierarchical clustering analysis with Euclidian distance and average linkage to evaluate the ability of the expression profile of GDE to correctly discriminate between preinvasive lesions (DCIS and *in situ* component of IBC). Somatic mutational profiling was performed by 2 multigene panels (409-gene panel and 50-gene hotspot regions panel) for NGS sequencing on Ion Torrent platform. The cohort was composed essentially by cases diagnosed as pure DCIS at anatomopathological report of biopsy and presenting different outcomes afterwards. Group 1: pure DCIS – pure DCIS at anatomopathological report of both biopsy and surgical specimen and with no manifestation of progression or relapse after at least 3-year follow-up. Group 2 – pure DCIS at anatomopathological report of biopsy and it was divided into 2 groups. Group 2.1 - pure DCIS at anatomopathological report of biopsy and invasive ductal carcinoma of no special type with concurrent *in situ* component at anatomopathological report of surgical specimen (DCIS/invasive). Group 2.2 – pure DCIS at anatomopathological report of both biopsy and surgical specimen and with manifestation of progression or relapse during the follow-up. **RESULTS:** Gene expression analysis was performed for 11 patients which were represented by 15 samples of epithelial or myoepithelial cells (5 DCIS and 10 *in situ* component of IBC). We confirmed a differential expression for 49% (26/53) of genes. The expression profile of 12 selected genes (*ANAPC13*, *BAALC*, *DNAJB6*, *ERCC8*, *FLCN*, *NCF4*, *RARRES3*, *RBM47*, *REC8*, *RRP1*, *SRA1* e *TMEM45A*) was able to discriminate 80% (4/5) of pure DCIS samples from 100% (10/10) of *in situ* component of IBC samples. Somatic mutational profiling was performed for 34 patients (43 samples) divided into 2 groups. Group 1 (DCIS with no progression) was composed by 20 patients and group 2 (DCIS with progression) by 14 patients, of which 7 patients were from each of group 2.1 and group 2.2. From 34 patients, 23 patients (29 samples) presented sequencing data qualified for single nucleotide variant analysis (15 and 8 patients from groups 1 and 2, respectively). From

23 patients, a total of 16 patients was found carrying at least one potentially somatic variant, 53% (8/15) and 100% (8/8) of the patients from groups 1 and 2, respectively. A total of 106 potentially somatic variants were found, of which 45 variants were found in 40 oncogenes or tumor suppressor genes. Oncogenic and probably oncogenic variants were found in 33% (5/15) and 75% (6/8) of patients from groups 1 and 2, respectively. Potentially somatic variants in gene *TP53* were found in 6,6% (1/15) and 25% (2/8) of patients from groups 1 and 2, respectively; and in *PIK3CA* in 0% and 25% (2/8) of patients from groups 1 and 2, respectively. **CONCLUSION:** The gene signature based on the 12 genes shows a potential to predict DCIS progression to invasive disease. Additionally, oncogenic variants in oncogenes and tumor suppressor genes were found in a higher proportion in patients who were diagnosed with pure DCIS at anatomopathological report of biopsy and manifested progression/relapse or invasion at anatomopathological report of surgical specimen.

Keywords: ductal carcinoma *in situ* of the breast; molecular assessment; biomarkers of DCIS progression.

LISTA DE FIGURAS

Figura 1 - Anatomia mamária. Ilustração do lúmen, local onde ocorre o CDIS.	4
Figura 2 - Padrões arquiteturais de CDIS.....	5
Figura 3 - Fluxograma ilustrando o processo de seleção de amostras de tecido tumoral fixado em formalina e emblocado em parafina.	23
Figura 4 - Fluxograma ilustrando o processo de seleção de amostras de tecido tumoral congelado.	24
Figura 5 - Método de coleta das áreas CDIS.	27
Figura 6 - Delineamento experimental do estudo.....	40
Figura 7 - Genes diferencialmente expressos entre amostras de CDIS puro e de componente <i>in situ</i> de CMI-TNE confirmados por RT-qPCR usando Low Taqman Density Arrays (TLDA).	45
Figura 8 - Clusterização hierárquica baseada no perfil de expressão de 12 genes diferencialmente expressos entre amostras CDIS puro e componente <i>in situ</i> de CMI-TNE.	46
Figura 9 - Rede de interação gênica.....	50
Figura 10 - Frequência de pacientes com variantes potencialmente somáticas.	60
Figura 11 - Repertório de variantes em CDIS.	64

LISTA DE QUADROS

Quadro 1 - Grau nuclear de CDIS..... 4

Quadro 2 - Lista de genes dos painéis multigenes utilizados no estudo. Lista com 409 genes cobertos pelo painel CCP..... 30

LISTA DE TABELAS

Tabela 1 - Principais características clínico-patológicas da casuística submetida a avaliação transcricional.....	42
Tabela 2 - Processos biológicos dos 12 genes cujo perfil de expressão foi capaz de discriminar amostras de CDIS puro de amostras de componente in situ de CMI-TNE...	48
Tabela 3 - Vias canônicas que estão enriquecidas pelos 12 genes cujo perfil de expressão foi capaz de discriminar amostras de CDIS puro de amostras de componente in situ de CMI-TNE.	49
Tabela 4 - Pacientes selecionados para sequenciamento de DNA por NGS.....	51
Tabela 5 - Características histológicas e moleculares das 34 pacientes de CDIS selecionados (43 amostras) e quantidade e qualidade do DNA obtido para sequenciamento por NGS.	53
Tabela 6 - Dados do sequenciamento de DNA das 29 amostras tumorais e dos 4 pools de leucócitos.	56
Tabela 7 - Variantes detectadas em oncogenes e genes supressores de tumor.....	58
Tabela 8 - Vias de reparo de DNA enriquecidas.	65

LISTA DE SÍMBOLOS E ABREVIATURAS

°C	graus Celsius
ABCA1	ATP binding cassette subfamily A member 1
ABL1	ABL proto-oncogene 1, non-receptor tyrosine kinase
ABL2	ABL proto-oncogene 2, non-receptor tyrosine kinase
ACCCC	Antônio Cândido Camargo Cancer Center
ACTB	actin beta
ACVR2A	activin A receptor type 2A
ADAMTS20	ADAM metalloproteinase with thrombospondin type 1 motif 20
AFF1	AF4/FMR2 family member 1
AFF3	AF4/FMR2 family member 3
AKAP9	A-kinase anchoring protein 9
AKT1	AKT serine/threonine kinase 1
AKT2	AKT serine/threonine kinase 2
AKT3	AKT serine/threonine kinase 3
ALK	ALK receptor tyrosine kinase
ALMS1	ALMS1 centrosome and basal body associated protein
ANAPC13	anaphase promoting complex subunit 13
APC	APC regulator of WNT signaling pathway
AR	androgen receptor
ARHGAP9	Rho GTPase activating protein 9
ARID1A	AT-rich interaction domain 1A
ARID2	AT-rich interaction domain 2
aRNA	amplified RNA
ARNT	aryl hydrocarbon receptor nuclear translocator
ASXL1	ASXL transcriptional regulator 1
ATF1	activating transcription factor 1
ATM	ATM serine/threonine kinase
ATP2A3	ATPase sarcoplasmic/endoplasmic reticulum Ca ²⁺ transporting 3
ATR	ATR serine/threonine kinase
ATRX	ATRX chromatin remodeler
AURKA	aurora kinase A
AURKB	aurora kinase B
AURKC	aurora kinase C
AXL	AXL receptor tyrosine kinase
AZGP1	alpha-2-glycoprotein 1, zinc-binding
BAALC	BAALC binder of MAP3K1 and KLF4
BAI3	adhesion G protein-coupled receptor B3
BAP1	BRCA1 associated protein 1
BCL10	BCL10 immune signaling adaptor
BCL11A	BAF chromatin remodeling complex subunit BCL11A
BCL11B	BAF chromatin remodeling complex subunit BCL11B
BCL2	BCL2 apoptosis regulator
BCL2L1	BCL2 like 1
BCL2L2	BCL2 like 2
BCL3	BCL3 transcription coactivator

BCL6	BCL6 transcription repressor
BCL9	BCL9 transcription coactivator
BCR	BCR activator of RhoGEF and GTPase
BIRC2	baculoviral IAP repeat containing 2
BIRC3	baculoviral IAP repeat containing 3
BIRC5	baculoviral IAP repeat containing 5
BLM	BLM RecQ like helicase
BLNK	B cell linker
BMP7	bone morphogenetic protein 7
BMPR1A	bone morphogenetic protein receptor type 1A
BRAF	B-Raf proto-oncogene, serine/threonine kinase
BRD3	bromodomain containing 3
BRIP1	BRCA1 interacting helicase 1
BTB	Bruton tyrosine kinase
BUB1B	BUB1 mitotic checkpoint serine/threonine kinase B
C15orf48	chromosome 15 open reading frame 48
CALR	calreticulin
CAMP	cathelicidin antimicrobial peptide
CARD11	caspase recruitment domain family member 11
CASC5	kinetochore scaffold 1
CBL	Cbl proto-oncogene
CC2D1A	coiled-coil and C2 domain containing 1A
CCND1	cyclin D1
CCND2	cyclin D2
CCNE1	cyclin E1
CCP	comprehensive cancer panel
CD79A	CD79a molecule
CD79B	CD79b molecule
CDC42BPG	CDC42 binding protein kinase gamma
CDC73	cell division cycle 73
CDH1	cadherin 1
CDH11	cadherin 11
CDH2	cadherin 2
CDH20	cadherin 20
CDH5	cadherin 5
CDIS	carcinoma ductal in situ
CDK12	cyclin dependent kinase 12
CDK4	cyclin dependent kinase 4
CDK6	cyclin dependent kinase 6
CDK8	cyclin dependent kinase 8
CDKN2A	cyclin dependent kinase inhibitor 2A
CDKN2B	cyclin dependent kinase inhibitor 2B
CDKN2C	cyclin dependent kinase inhibitor 2C
cDNA	DNA complementar
CEACAM5	CEA cell adhesion molecule 5
CEBPA	CCAAT enhancer binding protein alpha
CHEK1	checkpoint kinase 1

CHEK2	checkpoint kinase 2
CHP2	cancer hotspot panel v.2
CHRNA1	cholinergic receptor nicotinic beta 1 subunit
CIC	capicua transcriptional repressor
CID	código internacional de doenças
CKS1B	CDC28 protein kinase regulatory subunit 1B
CMI-TNE	carcinoma mamário invasivo tipo não especial
CMPK1	cytidine/uridine monophosphate kinase 1
COL1A1	collagen type I alpha 1 chain
COLEC12	collectin subfamily member 12
CPNE3	copine 3
Cq	cycle quantification
CRBN	cereblon
CREB1	cAMP responsive element binding protein 1
CREBBP	CREB binding protein
CRKL	CRK like proto-oncogene, adaptor protein
CRTC1	CREB regulated transcription coactivator 1
CSF1R	colony stimulating factor 1 receptor
CSMD3	CUB and Sushi multiple domains 3
CTNNA1	catenin alpha 1
CTNNB1	catenin beta 1
CTTNBP2NL	CTTNBP2 N-terminal like
CXCL10	C-X-C motif chemokine ligand 10
CYLD	CYLD lysine 63 deubiquitinase
CYP2C19	cytochrome P450 family 2 subfamily C member 19
CYP2D6	cytochrome P450 family 2 subfamily D member 6
DAXX	death domain associated protein
dbSNP	database of single nucleotide polymorphisms
DCC	DCC netrin 1 receptor
DCIS	ductal carcinoma in situ
DDB2	damage specific DNA binding protein 2
DDIT3	DNA damage inducible transcript 3
DDR2	discoidin domain receptor tyrosine kinase 2
DEGS1	delta 4-desaturase, sphingolipid 1
DEK	DEK proto-oncogene
DICER1	dicer 1, ribonuclease III
DIN	DNA integrity number
DNA	desoxyribonucleic acid
DNAJA3	DnaJ heat shock protein family (Hsp40) member A3
DNAJB6	DnaJ heat shock protein family (Hsp40) member B6
DNMT3A	DNA methyltransferase 3 alpha
DPYD	dihydropyrimidine dehydrogenase
DST	dystonin
EDN1	endothelin 1
EEF1D	eukaryotic translation elongation factor 1 delta
EGFR	epidermal growth factor receptor
EML4	EMAP like 4

emPCR	PCR em emulsão
ENSA	endosulfine alpha
EP300	E1A binding protein p300
EP400	E1A binding protein p400
EPB41L3	erythrocyte membrane protein band 4.1 like 3
EPHA3	EPH receptor A3
EPHA7	EPH receptor A7
EPHB1	EPH receptor B1
EPHB4	EPH receptor B4
EPHB6	EPH receptor B6
EPOR	erythropoietin receptor
ER	estrogen receptor
ERBB2	erb-b2 receptor tyrosine kinase 2
ERBB3	erb-b2 receptor tyrosine kinase 3
ERBB4	erb-b2 receptor tyrosine kinase 4
ERCC1	ERCC excision repair 1, endonuclease non-catalytic subunit
ERCC2	ERCC excision repair 2, TFIIH core complex helicase subunit
ERCC3	ERCC excision repair 3, TFIIH core complex helicase subunit
ERCC4	ERCC excision repair 4, endonuclease catalytic subunit
ERCC5	ERCC excision repair 5, endonuclease
ERCC8	ERCC excision repair 8, CSA ubiquitin ligase complex subunit
ERG	ETS transcription factor ERG
ESR1	estrogen receptor 1
ETS1	ETS proto-oncogene 1, transcription factor
ETV1	ETS variant transcription factor 1
ETV4	ETS variant transcription factor 4
EXT1	exostosin glycosyltransferase 1
EXT2	exostosin glycosyltransferase 2
EZH2	enhancer of zeste 2 polycomb repressive complex 2 subunit
FAM123B	APC membrane recruitment protein 1
FANCA	FA complementation group A
FANCC	FA complementation group C
FANCD2	FA complementation group D2
FANCF	FA complementation group F
FANCG	FA complementation group G
FAS	Fas cell surface death receptor
FBXW7	F-box and WD repeat domain containing 7
FFPE	formalin fixed paraffin embedded
FGFR1	fibroblast growth factor receptor 1
FGFR2	fibroblast growth factor receptor 2
FGFR3	fibroblast growth factor receptor 3
FGFR4	fibroblast growth factor receptor 4
FH	fumarate hydratase
FLCN	folliculin
FLI1	Fli-1 proto-oncogene, ETS transcription factor
FLT1	fms related receptor tyrosine kinase 1
FLT3	fms related receptor tyrosine kinase 3

FLT4	fms related receptor tyrosine kinase 4
FMNL1	formin like 1
FN1	fibronectin 1
FOXL2	forkhead box L2
FOXO1	forkhead box O1
FOXO3	forkhead box O3
FOXP1	forkhead box P1
FOXP4	forkhead box P4
FZR1	fizzy and cell division cycle 20 related 1
G6PD	glucose-6-phosphate dehydrogenase
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GATA1	GATA binding protein 1
GATA2	GATA binding protein 2
GATA3	GATA binding protein 3
GNDF	glial cell derived neurotrophic factor
GNA11	G protein subunit alpha 11
GNAQ	G protein subunit alpha q
GNAS	GNAS complex locus
GPR124	adhesion G protein-coupled receptor A2
GRB10	growth factor receptor bound protein 10
GRCh37	Genome Reference Consortium
GRM8	glutamate metabotropic receptor 8
GST	gene supressor tumoral
GUCY1A2	guanylate cyclase 1 soluble subunit alpha 2
GUSB	glucuronidase beta
HAT1	histone acetyltransferase 1
HCAR1	hydroxycarboxylic acid receptor 1
HCP5	HLA complex P5
HDC	histidine decarboxylase
HE	hematoxilina e eosina
HER2	human epidermal growth factor receptor 2
hg19	human genome 19
HIF1A	hypoxia inducible factor 1 subunit alpha
HIPK2	homeodomain interacting protein kinase 2
HLF	HLF transcription factor, PAR bZIP family member
HLTF	helicase like transcription factor
HNF1A	HNF1 homeobox A
HNRNPDL	heterogeneous nuclear ribonucleoprotein D like
HOOK3	hook microtubule tethering protein 3
HPRT1	hypoxanthine phosphoribosyltransferase 1
HR	hormone receptor
HRAS	HRas proto-oncogene, GTPase
HSP90AA1	heat shock protein 90 alpha family class A member 1
HSP90AB1	heat shock protein 90 alpha family class B member 1
IBC	invasive breast cancer
IC_DCIS-IBC	componente in situ de CMI-TNE
ICK	ciliogenesis associated kinase 1

IDH1	isocitrate dehydrogenase (NADP(+)) 1
IDH2	isocitrate dehydrogenase (NADP(+)) 2
IGF1R	insulin like growth factor 1 receptor
IGF2	insulin like growth factor 2
IGF2R	insulin like growth factor 2 receptor
IKB	ingenuity knowledge base
IKBKB	inhibitor of nuclear factor kappa B kinase subunit beta
IKBKE	inhibitor of nuclear factor kappa B kinase subunit epsilon
IKZF1	IKAROS family zinc finger 1
IL2	interleukin 2
IL21R	interleukin 21 receptor
IL6ST	interleukin 6 cytokine family signal transducer
IL7R	interleukin 7 receptor
ING4	inhibitor of growth family member 4
INPP1	inositol polyphosphate-1-phosphatase
IPA	ingenuity pathway analysis
IRF2BPL	interferon regulatory factor 2 binding protein like
IRF4	interferon regulatory factor 4
IRS2	insulin receptor substrate 2
ITGA10	integrin subunit alpha 10
ITGA9	integrin subunit alpha 9
ITGB2	integrin subunit beta 2
ITGB3	integrin subunit beta 3
JAK1	Janus kinase 1
JAK2	Janus kinase 2
JAK3	Janus kinase 3
JUN	Jun proto-oncogene, AP-1 transcription factor subunit
KAT6A	lysine acetyltransferase 6A
KAT6B	lysine acetyltransferase 6B
KDM5C	lysine demethylase 5C
KDM6A	lysine demethylase 6A
KDR	kinase insert domain receptor
KEAP1	kelch like ECH associated protein 1
KIAA0930	KIAA0930
KIT	KIT proto-oncogene, receptor tyrosine kinase
KLF6	Kruppel like factor 6
KMT2A	lysine methyltransferase 2A
KMT2C	lysine methyltransferase 2C
KMT2D	lysine methyltransferase 2D
KRAS	KRAS proto-oncogene, GTPase
LAMP1	lysosomal associated membrane protein 1
LCK	LCK proto-oncogene, Src family tyrosine kinase
LIFR	LIF receptor subunit alpha
LILRB1	leukocyte immunoglobulin like receptor B1
LIMS1	LIM zinc finger domain containing 1
LPHN3	adhesion G protein-coupled receptor L3
LPP	LIM domain containing preferred translocation partner in lipoma

LRP1B	LDL receptor related protein 1B
LSM4	LSM4 homolog, U6 small nuclear RNA and mRNA degradation associated
LTF	lactotransferrin
LTK	leukocyte receptor tyrosine kinase
LYRM2	LYR motif containing 2
MAF	MAF bZIP transcription factor
MAFB	MAF bZIP transcription factor B
MAGEA1	MAGE family member A1
MAGI1	membrane associated guanylate kinase, WW and PDZ domain containing 1
MALT1	MALT1 paracaspase
MAML2	mastermind like transcriptional coactivator 2
MAP2K1	mitogen-activated protein kinase kinase 1
MAP2K2	mitogen-activated protein kinase kinase 2
MAP2K4	mitogen-activated protein kinase kinase 4
MAP3K7	mitogen-activated protein kinase kinase kinase 7
MAPK1	mitogen-activated protein kinase 1
MAPK8	mitogen-activated protein kinase 8
MARK1	microtubule affinity regulating kinase 1
MARK4	microtubule affinity regulating kinase 4
MBD1	methyl-CpG binding domain protein 1
MCL1	MCL1 apoptosis regulator, BCL2 family member
MDM2	MDM2 proto-oncogene
MDM4	MDM4 regulator of p53
MEN1	menin 1
MET	MET proto-oncogene, receptor tyrosine kinase
mg	miligram
MGAT1	alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase
MITF	melanocyte inducing transcription factor
mL	mililitro
MLH1	mutL homolog 1
MLL	lysine methyltransferase 2A
MLL2	lysine methyltransferase 2D
MLL3	lysine methyltransferase 2C
MLLT10	MLLT10 histone lysine methyltransferase DOT1L cofactor
mm	milímetro
MMP2	matrix metalloproteinase 2
MN1	MN1 proto-oncogene, transcriptional regulator
MPL	MPL proto-oncogene, thrombopoietin receptor
MRE11	MRE11 homolog, double strand break repair nuclease
MRE11A	MRE11 homolog, double strand break repair nuclease
MSH2	mutS homolog 2
MSH6	mutS homolog 6
MTOR	mechanistic target of rapamycin kinase
MTR	5-methyltetrahydrofolate-homocysteine methyltransferase

MTRR	5-methyltetrahydrofolate-homocysteine methyltransferase reductase
MUC1	mucin 1, cell surface associated
MUTYH	mutY DNA glycosylase
MYB	MYB proto-oncogene, transcription factor
MYC	MYC proto-oncogene, bHLH transcription factor
MYCL1	MYCL proto-oncogene, bHLH transcription factor
MYCN	MYCN proto-oncogene, bHLH transcription factor
MYD88	MYD88 innate immune signal transduction adaptor
MYH11	myosin heavy chain 11
MYH9	myosin heavy chain 9
MYO1C	myosin IC
NBN	nibrin
NCF4	neutrophil cytosolic factor 4
NCOA1	nuclear receptor coactivator 1
NCOA2	nuclear receptor coactivator 2
NCOA4	nuclear receptor coactivator 4
NDFIP1	Nedd4 family interacting protein 1
NF1	neurofibromin 1
NF2	NF2, moesin-ezrin-radixin like (MERLIN) tumor suppressor
NFE2L2	NFE2 like bZIP transcription factor 2
NFKB1	nuclear factor kappa B subunit 1
NFKB2	nuclear factor kappa B subunit 2
ng	nanograma
NG	nuclear grade
NGS	next generation sequencing
NIN	ninein
NKX2-1	NK2 homeobox 1
NLRP1	NLR family pyrin domain containing 1
NOTCH1	notch receptor 1
NOTCH2	notch receptor 2
NOTCH4	notch receptor 4
NPM1	nucleophosmin 1
NRAS	NRAS proto-oncogene, GTPase
NSD1	nuclear receptor binding SET domain protein 1
NSD2	nuclear receptor binding SET domain protein 2
NTRK1	neurotrophic receptor tyrosine kinase 1
NTRK3	neurotrophic receptor tyrosine kinase 3
NUMA1	nuclear mitotic apparatus protein 1
NUP214	nucleoporin 214
NUP98	nucleoporin 98 and 96 precursor
P4HB	prolyl 4-hydroxylase subunit beta
PAK3	p21 (RAC1) activated kinase 3
PALB2	partner and localizer of BRCA2
PARP1	poly(ADP-ribose) polymerase 1
PAX3	paired box 3
PAX5	paired box 5
PAX7	paired box 7

PAX8	paired box 8
pb	par de bases
PBRM1	polybromo 1
PBX1	PBX homeobox 1
PCR	polymerase chain reaction
PDE4DIP	phosphodiesterase 4D interacting protein
PDGFB	platelet derived growth factor subunit B
PDGFRA	platelet derived growth factor receptor alpha
PDGFRB	platelet derived growth factor receptor beta
PER1	period circadian regulator 1
PEX11A	peroxisomal biogenesis factor 11 alpha
PGAP3	post-GPI attachment to proteins phospholipase 3
PHOX2B	paired like homeobox 2B
PIK3C2B	phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2 beta
PIK3CA	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha
PIK3CB	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit beta
PIK3CD	phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit delta
PIK3CG	gamma
PIK3R1	phosphoinositide-3-kinase regulatory subunit 1
PIK3R2	phosphoinositide-3-kinase regulatory subunit 2
PIM1	Pim-1 proto-oncogene, serine/threonine kinase
PKHD1	PKHD1 ciliary IPT domain containing fibrocystin/polyductin
PLAG1	PLAG1 zinc finger
PLAUR	plasminogen activator, urokinase receptor
PLCG1	phospholipase C gamma 1
PLEKHG5	pleckstrin homology and RhoGEF domain containing G5
PML	PML nuclear body scaffold
PMS1	PMS1 homolog 1, mismatch repair system component
PMS2	PMS1 homolog 2, mismatch repair system component
POLR3A	RNA polymerase III subunit A
POT1	protection of telomeres 1
POU5F1	POU class 5 homeobox 1
PPARG	peroxisome proliferator activated receptor gamma
PPP2R1A	protein phosphatase 2 scaffold subunit Aalpha
PR	progesterone receptor
PRDM1	PR/SET domain 1
PRKAR1A	protein kinase cAMP-dependent type I regulatory subunit alpha
PRKDC	protein kinase, DNA-activated, catalytic subunit
PSIP1	PC4 and SFRS1 interacting protein 1
PTAFR	platelet activating factor receptor
PTCH1	patched 1
PTEN	phosphatase and tensin homolog
PTGS2	prostaglandin-endoperoxide synthase 2
PTPN11	protein tyrosine phosphatase non-receptor type 11
PTPRD	protein tyrosine phosphatase receptor type D
PTPRT	protein tyrosine phosphatase receptor type T

RABEPK	Rab9 effector protein with kelch motifs
RAD50	RAD50 double strand break repair protein
RAF1	Raf-1 proto-oncogene, serine/threonine kinase
RALGDS	ral guanine nucleotide dissociation stimulator
RARA	retinoic acid receptor alpha
RARRES3	phospholipase A and acyltransferase 4
RB1	RB transcriptional corepressor 1
RBM47	RNA binding motif protein 47
REC8	REC8 meiotic recombination protein
RECQL4	RecQ like helicase 4
REEP1	receptor accessory protein 1
REL	REL proto-oncogene, NF-kB subunit
RELT	RELT TNF receptor
RET	ret proto-oncogene
RHOH	ras homolog family member H
RNA	ribonucleic acid
RNASEL	ribonuclease L
RNF103	ring finger protein 103
RNF2	ring finger protein 2
RNF213	ring finger protein 213
RNPC3	RNA binding region (RNP1, RRM) containing 3
ROBO1	roundabout guidance receptor 1
ROS1	ROS proto-oncogene 1, receptor tyrosine kinase
RPLP0	ribosomal protein lateral stalk subunit P0
rpm	rotações por minuto
RPS6KA2	ribosomal protein S6 kinase A2
RPS6KA4	ribosomal protein S6 kinase A4
RQ	relative quantification
RRM1	ribonucleotide reductase catalytic subunit M1
RRP1	ribosomal RNA processing 1
RT-qPCR	reverse transcription – quantitative polymerase chain reaction
RUNX1	RUNX family transcription factor 1
RUNX1T1	RUNX1 partner transcriptional co-repressor 1
S100A9	S100 calcium binding protein A9
SACM1L	SAC1 like phosphatidylinositide phosphatase
SAMD9	sterile alpha motif domain containing 9
SBDS	SBDS ribosome maturation factor
SCPEP1	serine carboxypeptidase 1
SDHA	succinate dehydrogenase complex flavoprotein subunit A
SDHB	succinate dehydrogenase complex iron sulfur subunit B
SDHC	succinate dehydrogenase complex subunit C
SDHD	succinate dehydrogenase complex subunit D
SDS	Sequence Detection System
SEPT9	septin 9
SETD2	SET domain containing 2, histone lysine methyltransferase
SF3B1	splicing factor 3b subunit 1
SGK1	serum/glucocorticoid regulated kinase 1

SH2D1A	SH2 domain containing 1A
SLC16A3	solute carrier family 16 member 3
SLC16A6	solute carrier family 16 member 6
SLC35G2	solute carrier family 35 member G2
SLC44A5	solute carrier family 44 member 5
SMAD2	SMAD family member 2
SMAD4	SMAD family member 4
SMARCA4	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4
SMARCB1	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1
SMO	smoothened, frizzled class receptor
SMUG1	single-strand-selective monofunctional uracil-DNA glycosylase 1
SOCS1	suppressor of cytokine signaling 1
SOX11	SRY-box transcription factor 11
SOX2	SRY-box transcription factor 2
SPG11	SPG11 vesicle trafficking associated, spatacsin
SPTBN2	spectrin beta, non-erythrocytic 2
SRA1	steroid receptor RNA activator 1
SRC	SRC proto-oncogene, non-receptor tyrosine kinase
SRRM1	serine and arginine repetitive matrix 1
SSX1	SSX family member 1
ST6GALNAC4	ST6 N-acetylgalactosaminide alpha-2,6-sialyltransferase 4
STK11	serine/threonine kinase 11
STK36	serine/threonine kinase 36
SUFU	SUFU negative regulator of hedgehog signaling
SYK	spleen associated tyrosine kinase
SYNE1	spectrin repeat containing nuclear envelope protein 1
TAE	Tris-acetato-EDTA
TAF1	TATA-box binding protein associated factor 1
TAF1L	TATA-box binding protein associated factor 1 like
TAL1	TAL bHLH transcription factor 1, erythroid differentiation factor
TBX22	T-box transcription factor 22
TCF12	transcription factor 12
TCF3	transcription factor 3
TCF7L1	transcription factor 7 like 1
TCF7L2	transcription factor 7 like 2
TCL1A	TCL1 family AKT coactivator A
TE	Tris-EDTA
TET1	tet methylcytosine dioxygenase 1
TET2	tet methylcytosine dioxygenase 2
TFE3	transcription factor binding to IGHM enhancer 3
TGFBR2	transforming growth factor beta receptor 2
TGM7	transglutaminase 7
THBS1	thrombospondin 1
TIMP3	TIMP metalloproteinase inhibitor 3
TLDA	Taqman Low Density Array

TLR4	toll like receptor 4
TLX1	T cell leukemia homeobox 1
TMEM45A	transmembrane protein 45A
TNFAIP3	TNF alpha induced protein 3
TNFRSF14	TNF receptor superfamily member 14
TNFRSF18	TNF receptor superfamily member 18
TNK2	tyrosine kinase non receptor 2
TOP1	DNA topoisomerase I
TP53	tumor protein p53
TPR	translocated promoter region, nuclear basket protein
TRIM24	tripartite motif containing 24
TRIM33	tripartite motif containing 33
TRIP11	thyroid hormone receptor interactor 11
TRPM2	transient receptor potential cation channel subfamily M member 2
TRRAP	transformation/transcription domain associated protein
TSC1	TSC complex subunit 1
TSC2	TSC complex subunit 2
TSHR	thyroid stimulating hormone receptor
TVC	torrent variant caller
U	unidade
UBR5	ubiquitin protein ligase E3 component n-recognin 5
UGT1A1	UDP glucuronosyltransferase family 1 member A1
USP9X	ubiquitin specific peptidase 9 X-linked
UTP20	UTP20 small subunit processome component
VAF	variant allele frequency
VCF	variant call format
VHL	von Hippel-Lindau tumor suppressor
WAS	WASP actin nucleation promoting factor
WHSC1	nuclear receptor binding SET domain protein 2
WRN	WRN RecQ like helicase
WT1	WT1 transcription factor
XAF1	XIAP associated factor 1
XPA	XPA, DNA damage recognition and repair factor
XPC	XPC complex subunit, DNA damage recognition and repair factor
XPO1	exportin 1
XRCC2	X-ray repair cross complementing 2
YKT6	YKT6 v-SNARE homolog
ZBP1	Z-DNA binding protein 1
ZNF384	zinc finger protein 384
ZNF521	zinc finger protein 521
μL	microlitro

SUMÁRIO

1. INTRODUÇÃO.....	1
1.1 O câncer de mama.....	1
1.2 Carcinoma ductal in situ	3
1.3 O processo de progressão de CDIS.....	6
1.4 Caracterização molecular do câncer de mama	8
1.5 Caracterização molecular de lesões pré-invasivas de mama	10
1.6 Biomarcadores de predição de risco de progressão ou recidiva de CDIS	14
1.7 Justificativa	15
2. OBJETIVOS	16
2.1 Objetivo geral	16
2.2 Objetivos específicos	16
3. METODOLOGIA	17
3.1 Aprovação do projeto no Comitê de Ética em Pesquisa	17
3.2 Confidencialidade dos dados.....	17
3.3 Financiamento do estudo	17
3.4 Avaliação transcricional de CDIS em células epiteliais neoplásicas capturadas por microdissecção a laser	17
3.4.1 Seleção da casuística	17
3.4.2 Avaliação da expressão gênica de CDIS por Taqman Low Density Array (TLDA).....	18
3.4.3 Análise de clusterização	20
3.4.4 Análise de enriquecimento funcional.....	21
3.5 Avaliação mutacional de CDIS	21
3.5.1 Caracterização da casuística	21
3.5.1.1 Critérios de inclusão	21
3.5.1.2 Critério de exclusão	22
3.5.2 Seleção da casuística	22
3.5.2.1 Amostras de tecido tumoral fixado em formalina e embocado em parafina.....	22
3.5.2.2 Amostras de tecido tumoral congelado	23
3.5.3 Extração de DNA das amostras do estudo	24
3.5.4 Critérios de controle de qualidade de DNA.....	28
3.5.5 Construção de bibliotecas para sequenciamento NGS com painéis multigenes	28
3.5.6 Sequenciamento de DNA.....	32
3.5.7 Sequenciamento de amplicon	33

3.5.8	Análise dos dados de sequenciamento de DNA para seleção das variantes potencialmente somáticas	34
3.5.9	Anotação das variantes potencialmente oncogênicas.....	35
3.5.10	Análise de associação de mutações oncogênicas e provavelmente oncogênicas potencialmente somáticas com os grupos de CDIS.....	37
3.5.11	Análise de enriquecimento funcional.....	38
4.	RESULTADOS.....	39
4.1	Avaliação molecular de CDIS	39
4.2	Avaliação transcricional de CDIS	41
4.2.1	Casuística selecionada	41
4.2.2	Análise das diferenças de expressão gênica entre CDIS puro e componente <i>in situ</i> de CMI-TNE	42
4.2.3	Análise de clusterização	46
4.2.4	Análise de enriquecimento funcional.....	47
4.3	Avaliação mutacional de CDIS	51
4.3.1	Casuística selecionada	51
4.3.2	Sequenciamento de DNA.....	54
4.3.3	Identificação de variantes potencialmente somáticas em CDIS	57
4.3.3.1	Variantes potencialmente somáticas classificadas como oncogênicas e provavelmente oncogênicas detectadas em oncogenes e GST em CDIS.....	62
4.3.3.2	Variantes potencialmente somáticas de significado clínico incerto (VUS) detectadas em oncogenes e GST em CDIS	65
4.3.4	Análise de enriquecimento funcional.....	65
5.	DISCUSSÃO	66
6.	CONCLUSÃO.....	83
7.	REFERÊNCIAS BIBLIOGRÁFICAS	84

ANEXO

Anexo 1 - Carta de aprovação do estudo pelo Comitê de Ética em Pesquisa

APÊNDICES

Apêndice 1 - Valores de expressão dos 12 genes nas etapas da progressão de CDIS

Apêndice 2 - Características anatomopatológicas e antecedentes familiares do grupo CDIS com progressão ou recidiva

Apêndice 3 - Relação dos 12 genes com mutações em CDIS

Apêndice 4 - Vias bioquímicas enriquecidas pelos 12 genes com mutações em CDIS

1. INTRODUÇÃO

1.1 O câncer de mama

A Organização Mundial da Saúde (OMS), através do projeto GLOBOCAN, estimou para 2020 a incidência global de novos pacientes de câncer em 19,3 milhões e a mortalidade em quase 10 milhões (Sung et al. 2021). De acordo com essas estimativas, o câncer de mama é atualmente a neoplasia maligna mais comum, apresentando uma incidência global de 2,3 milhões de novos pacientes, representando 11,7% de todas as neoplasias malignas (Sung et al. 2021). Na população feminina, o câncer de mama é o mais frequente tanto em países desenvolvidos quanto em países em desenvolvimento. A incidência na população norte-americana em 2022 foi estimada em 341.960 novos pacientes (Siegel et al. 2022) e a incidência anual na população brasileira no triênio 2020-2022 foi estimada em 66.280 novos pacientes (INCA 2020).

Os fatores de risco mais importantes para o desenvolvimento do câncer de mama compreendem fatores biológicos, tais como idade avançada e alta densidade mamária; fatores genéticos, tais como histórico familiar e predisposição genética devido a mutações nos genes BRCA1 e BRCA2; fatores hormonais e reprodutivos, tais como menarca precoce, menopausa tardia, idade avançada na primeira gestação, baixa paridade, pouco tempo de amamentação, terapia hormonal pós-menopausa e uso de contraceptivos orais por longo período; e fatores ambientais, tais como consumo de álcool, excesso de peso corpóreo e inatividade física (Cardoso et al. 2019; Sung et al. 2021).

Exames de imagem de rotina incluem a mamografia e o ultrassom, mas em pacientes específicos, como por exemplo em pacientes com um forte histórico familiar de câncer de mama, a ressonância magnética também pode ser utilizada. O diagnóstico de câncer de mama é realizado por biópsia seguida de uma avaliação patológica. As biópsias podem ser realizadas por punção aspirativa por agulha fina ou por agulha grossa, as quais podem ser guiadas por mamografia ou ultrassom, em função do tamanho e localização da lesão (Cardoso et al. 2019).

O câncer de mama é classificado em diversos tipos histológicos (Kim et al. 2020), podendo ser *in situ* ou invasivo. Em relação aos carcinomas *in situ*, o carcinoma ductal *in situ* (CDIS) é o tipo mais frequente das lesões pré-invasivas, representando cerca de 75% de todos os pacientes de carcinoma *in situ*, e 25% de todos os tipos de câncer de mama (Siegel et al.

2022). Em relação aos carcinomas invasivos, o carcinoma de mama invasivo tipo não especial (CMI-TNE) é o tipo mais frequente, representando cerca de 75% de todos os pacientes (Kim et al. 2020).

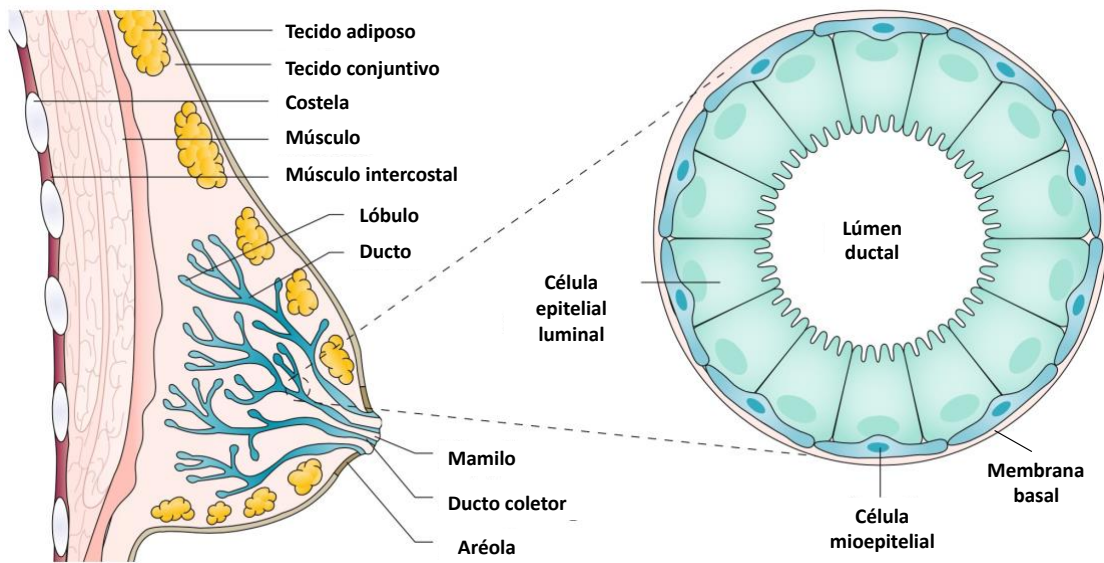
O estadiamento dos tumores mamários é realizado através do sistema TNM, publicado pela *American Joint Committee on Cancer* (AJCC) e pela *Union for International Cancer Control* (UICC). O sistema TNM agrupa diversos tipos tumorais de acordo com o tamanho do tumor (T), o envolvimento de nódulo linfático (N), e a ocorrência de metástase à distância do tumor primário (M). O estadiamento pode ser clínico (cTNM), realizado normalmente a partir de exame físico e diagnóstico por imagem (evidências anteriores ao processo cirúrgico), ou patológico (pTNM) quando é realizado por exame histopatológico da peça cirúrgica e dos linfonodos, para detectar a presença ou ausência de metástases em linfonodos regionais. De acordo com o tamanho do tumor, o número de linfonodos acometidos por metástase e a presença de metástase, o câncer de mama pode ser agrupado em um de 5 estádios: estadio 0: carcinoma *in situ*; estadio I: carcinoma invasivo e localizado; estadio II: carcinoma invasivo localmente limitado ou espalhado regionalmente; estadio III: carcinoma invasivo localmente extensivo ou espalhado regionalmente; estadio IV: carcinoma invasivo extensivo com presença de metástase à distância. Sendo que do estadio I ao IV é possível haver a presença do componente *in situ* associado ao componente invasivo (Lakahani et al. 2012).

O tratamento do carcinoma de mama invasivo pode ser local, com a realização de cirurgia e/ou radioterapia; ou pode ser sistêmico, com a realização de quimioterapia e/ou hormonioterapia (tratamento com modulador seletivo do receptor de estrógeno em pacientes com positividade de expressão imunoistoquímica para este receptor) e/ou terapia imunológica. A cirurgia pode ser do tipo conservadora (ressecção segmentar), onde é retirada a área comprometida pela neoplasia com margens cirúrgicas de segurança, ou não conservadora (mastectomia), podendo ter preservação de pele e/ou mamilo. O procedimento cirúrgico pode também incluir biópsia de linfonodo sentinela (LS), com ou sem esvaziamento dos linfonodos axilares (Osório et al. 2012). O tratamento do CDIS difere principalmente por não envolver um procedimento de quimioterapia. O tratamento convencional de CDIS inclui cirurgia, mais frequentemente a ressecção segmentar, a qual comumente é seguida de radioterapia, e terapia hormonal, para os pacientes com expressão proteica positiva dos receptores hormonais (receptores de estrógeno e progesterona).

1.2 Carcinoma ductal *in situ*

A propagação mundial de programas de rastreamento por mamografia tem contribuído significativamente para o diagnóstico de lesões não palpáveis, o que tem aumentado a taxa de detecção do câncer de mama em estágio inicial, como por exemplo, a lesão pré-invasiva carcinoma ductal *in situ* (CDIS) (Grimm et al. 2022). Contudo, existem poucos estudos epidemiológicos que estimaram a incidência de CDIS na população brasileira (Lemos et al. 2018), os quais encontraram valores discrepantes. Por exemplo, Gebrim e colaboradores (2014) avaliaram os dados de mais de 3 mil pacientes atendidos entre 2012 e 2014 no Hospital Perola Byington em São Paulo e encontraram uma incidência de CDIS de aproximadamente 8%. Enquanto Tsunoda e colaboradores (2013) avaliaram os dados de 2003-2010 do programa de rastreamento mamográfico do Hospital de Amor em Barretos e encontraram uma incidência de CDIS de aproximadamente 20%. Freitas e colaboradores (2016) sugerem que os valores discrepantes de incidência de CDIS podem ser devido à baixa e desigual cobertura nacional do rastreamento mamográfico. Em países desenvolvidos, como por exemplo nos Estados Unidos, a incidência estimada de CDIS era de 2-3% na era pré-mamografia (Rosner 1980), e atualmente corresponde a cerca de 15-20% dos novos pacientes anuais (DeSantis et al. 2017, 2019; Siegel et al. 2022).

O CDIS corresponde a uma proliferação de células epiteliais tumorais que permanecem restritas ao lúmen, circundadas pelas células mioepiteliais e pela membrana basal (Figura 1). O CDIS tem origem na unidade ducto lobular terminal, mas pode ser encontrado em qualquer nível da ramificação ductal mamária (Allred 2010).



Fonte: Adaptado de Harbeck et al. 2019.

Figura 1 - Anatomia mamária. Ilustração do lúmen, local onde ocorre o CDIS.

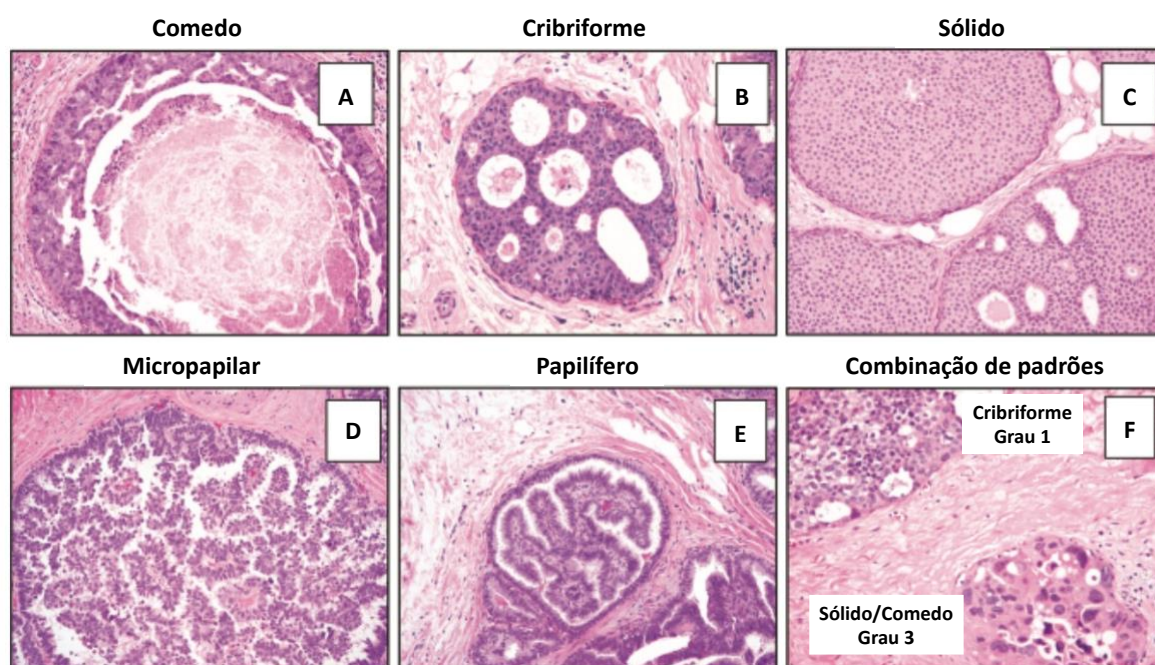
De acordo com as últimas recomendações da OMS, sua graduação histológica deve ser baseada nos aspectos morfológicos celulares e nucleares. Desta forma, a classificação histológica é determinada de acordo com o grau de pleomorfismo, tamanho do núcleo, padrão da cromatina, quantidade e padrão dos nucléolos, padrão mitótico e a orientação (polarização) celular. Baseando-se nesses aspectos, o CDIS é dividido em três graus histológicos: baixo grau (grau nuclear I), grau intermediário (grau nuclear II) e alto grau (grau nuclear III) (Quadro 1) (WHO 2019).

Quadro 1 - Grau nuclear de CDIS.

Característica	Baixo grau (grau nuclear I)	Grau intermediário (grau nuclear II)	Alto grau (grau nuclear III)
Pleomorfismo	Discreto	Intermediário	Acentuado
Tamanho nuclear	1,5–2x o tamanho de uma hemácia ou do núcleo de célula epitelial ductal	Intermediário	> 2,5x o tamanho de uma hemácia ou do núcleo de célula epitelial ductal
Cromatina	Dispersa	Intermediário	Grosseira ou vesicular
Nucléolos	Ausente ou imperceptível	Intermediário	Proeminente, frequentemente múltiplo
Mitoses	Raro	Intermediário	Frequente
Orientação	Polarizada em direção luminal	Intermediário	Geralmente não polarizada em direção luminal

Fonte: Adaptado de WHO 2019.

Embora o padrão arquitetural do CDIS não seja necessário para a sua graduação histológica, sua identificação e descrição no laudo anatomopatológico é importante e recomendada (WHO 2019). De acordo com o padrão arquitetural, o CDIS pode ser classificado nos tipos comedo (comedocarcinoma), sólido, cribriforme, micropapilar e papilífero (Figura 2). Embora, muitos pacientes de CDIS apresentem uma complexa combinação de padrões arquiteturais e características celulares (Figura 2F) (Allred 2010). O tipo comedo é caracterizado por um alto grau nuclear, e é associado a presença de necrose central, calcificação e à recorrências mais precoces (Badve et al. 1998; Badve and Gökmen-Polar 2019).



Fonte: Adaptado de Allred 2010.

Legenda: (A), comedocarcinoma; (B), Cribriforme; (C), Sólido; (D), micropapilar; (E), papilífero; (F), combinação de padrões.

Figura 2 - Padrões arquiteturais de CDIS.

O CDIS compreende um grupo de tumores que apresenta uma heterogeneidade biológica e clínica, podendo permanecer indolentes por muitos anos ou progredir para a forma invasiva da doença, ou ainda, apresentar recidiva (recorrência ipsilateral de CDIS). As características histológicas de CDIS que são associadas a maior risco de progressão/recidiva

são: grande tamanho tumoral, alto grau nuclear, padrão arquitetural do tipo comedo, presença de necrose central (comedonecrose) e margem cirúrgica positiva (visualização de tinta nanquim no espécime tumoral na avaliação microscópica) (Morrow et al. 2016; WHO 2019).

1.3 O processo de progressão de CDIS

O modelo clássico de progressão do carcinoma ductal de mama é um modelo simplista que propõe uma sucessão de estágios, iniciando com uma atipia epitelial plana, que progride para uma hiperplasia ductal atípica, a qual evolui para o CDIS, e este culmina no CMI-TNE, o último estágio da progressão e que corresponde a uma lesão invasiva e potencialmente letal, pois é representado por uma proliferação anormal de células que rompem a membrana basal dos ductos e invadem os tecidos adjacentes, podendo chegar à circulação sanguínea e linfática, e causar metástases (Wellings e Jensen 1973; Bombonati e Sgroi 2011). Neste contexto, o CDIS é considerado um precursor não-obrigatório do CMI-TNE, e o processo para isso ocorrer pode durar um longo período, possivelmente décadas (Allred 2010). Contudo, existem diversas evidências indiretas, porém convincentes, que sugerem que o CDIS seja mesmo um precursor do CMI-TNE. Por exemplo, aproximadamente 80% dos pacientes diagnosticados de CMI-TNE apresentam também um componente *in situ* associado (componente *in situ* de CMI-TNE) (WHO 2019; Cho et al. 2021); a maioria dos fatores de risco para CMI-TNE são os mesmos para CDIS; CMI-TNE e CDIS que estão presentes na mesma mama compartilham diversas alterações genéticas idênticas; modelos animais de câncer de mama geneticamente modificados demonstram progressão de CDIS para CMI-TNE (Allred 2010).

Alguns estudos têm identificado uma expressão diferencial de alguns genes específicos entre pacientes de CDIS de baixo e alto graus. Tais estudos observaram que os tumores *in situ* agrupam com os tumores invasivos de mesmo grau, corroborando que o CDIS seja mesmo um precursor do CMI-TNE (Ma et al. 2003; Bombonati e Sgroi 2011), e que devem existir duas vias de progressão de CDIS, a via de progressão de CDIS de baixo grau e a via de progressão de CDIS de alto grau.

A história natural de CDIS não é totalmente compreendida devido a uma escassez de dados, pois na maioria dos pacientes as pacientes recebem tratamento cirúrgico logo após o diagnóstico (Grimm et al. 2022), e conseqüentemente o percentual de pacientes de CDIS não

tratados que podem progredir para o CMI-TNE é impreciso (Carraro et al. 2014). Entretanto, estudos epidemiológicos e observacionais antigos que analisaram pacientes de pacientes com CDIS não submetidos a tratamento cirúrgico estimaram a taxa de ocorrência de lesão invasiva em aproximadamente 25% num período de 15 anos e em aproximadamente 30% em 30 anos (Page et al. 1982, 1995; Sanders et al. 2005, 2015). Outros estudos mais recentes, como os estudos clínicos randomizados que estão atualmente em andamento, como por exemplo, o *Comparison of Operative versus Medical Endocrine Therapy* (COMET) (Hwang et al. 2019), *Low Risk DCIS* (LORIS) (Francis et al. 2015) e *Low Risk DCIS* (LORD) (Elshof et al. 2015), buscam entender melhor a propensão do CDIS em progredir para a forma invasiva da doença. Esses estudos avaliam pacientes de CDIS tipicamente de baixo risco (não alto grau nuclear, ausência de comedonecrose, pacientes assintomáticos, sem histórico de câncer de mama) que não foram submetidos a tratamento cirúrgico, apenas a hormonioterapia (tamoxifeno ou inibidores de aromatase) e seguem em vigilância ativa por um período de até 10 anos (Grimm et al. 2022).

Uma das preocupações no recrutamento de pacientes para participarem de estudos de vigilância ativa é a possibilidade da existência de doença invasiva oculta. Um estudo do tipo meta-análise avaliou 52 estudos que incluíam 7350 pacientes, cujo material da biópsia foi diagnosticado como CDIS. Estes pacientes foram submetidos a procedimento cirúrgico, dos quais aproximadamente 26% tiveram o material cirúrgico diagnosticado contendo lesão invasiva (Brennan et al. 2011), demonstrando que nem sempre a biópsia representa o tumor como um todo em termos da presença do componente invasivo.

A taxa de recidiva em pacientes de CDIS submetidos a tratamento cirúrgico também tem sido investigada. Um amplo estudo do tipo revisão sistemática e meta-análise avaliou a taxa de recorrência de CDIS e CMI-TNE ipsilateral em mais de 9 mil pacientes de CDIS primário submetidos a diferentes formas de tratamento. Os autores observaram que a taxa de recidiva de CDIS e CMI-TNE num período médio de 10 anos em pacientes que realizaram mastectomia, cirurgia conservadora seguida de radioterapia, cirurgia conservadora sem radioterapia, e submetidas apenas a biópsia excisional foi de 2,6%, 13,6%, 25,5% e 27,8%, respectivamente (Stuart et al. 2015). Esse estudo mostrou uma associação entre uma maior intervenção local e um melhor desfecho num período médio de 10 anos (menor chance de recidiva de lesão invasiva); e mostrou também que uma taxa considerável de recidiva ainda existe mesmo após o CDIS ter sido submetido a tratamento cirúrgico. Também tem sido proposto que o tipo

histológico do CDIS está associado a taxa, tipo e tempo de recidiva. Wallis e colaboradores (2012) analisaram os dados de 15 anos de seguimento de 700 pacientes de CDIS primário. Aproximadamente 64% (449/700) dos pacientes foram submetidos a cirurgia conservadora, dos quais aproximadamente 63% (286/449) receberam terapia adjuvante (hormonioterapia e/ou radioterapia), e 35% (251/700) dos pacientes foram submetidos a mastectomia. Os autores observaram que o tempo médio de recidiva de CMI-TNE foi de 76 meses a partir de CDIS de alto grau e de 131 meses a partir de CDIS de baixo e intermediário graus. Além disso, os dados deste estudo sugerem que uma recidiva de lesão pré-invasiva ocorre mais precocemente do que uma recidiva de lesão invasiva, pois foi observado que o tempo médio de recidiva de CDIS foi de 15 meses e o de CMI-TNE foi de 60 meses.

1.4 Caracterização molecular do câncer de mama

A caracterização molecular do câncer de mama teve grandes avanços a partir do desenvolvimento de metodologias de avaliação transcricional em larga escala, como o cDNA *microarray* (Schena et al. 1995). Através desta metodologia, Perou e colaboradores (2000) avaliaram a expressão de 8102 genes em 65 tumores de mama invasivos para identificar os padrões de expressão gênica que ocorrem naturalmente em câncer de mama. Os autores identificaram um conjunto de 496 genes diferencialmente expressos, denominado “conjunto intrínseco de genes”, cujo padrão de expressão identificou quatro grupos tumorais diferentes, denominados "subtipos intrínsecos", uma vez que o padrão de expressão desses genes refletia as propriedades intrínsecas do câncer de mama. Um dos subtipos foi composto por tumores com expressão positiva do receptor de estrógeno, e foi denominado subtipo luminal, o qual pode ter um esquema terapêutico beneficiado com a inclusão de hormonioterapia com inibidor do receptor de estrógeno (tamoxifeno). Os demais subtipos apresentaram expressão negativa do receptor de estrógeno (subtipos não-luminais), e foram denominados subtipo semelhante à mama normal, subtipo basaloide e subtipo superexpressor de HER2, este último podendo ter um esquema terapêutico beneficiado com a inclusão de terapia-alvo inibidora de HER2 (trastuzumabe). Em um outro estudo do grupo, os autores utilizaram a mesma metodologia para avaliar a expressão de 456 genes em 78 tumores de mama (Sørli et al. 2001). O padrão de expressão desses genes identificou novamente os subtipos superexpressor de HER2, basaloide e semelhante à mama normal, e ainda dividiu o subtipo luminal em pelo menos 2 subtipos (luminal A e luminal B). Os subtipos moleculares apresentaram fortes

implicações prognósticas. Os pacientes com tumores dos subtipos basaloídes (enriquecidos de tumores-negativos (TN)), superexpressor de HER2 e luminal B mostraram-se com sobrevida significativamente menor quando comparados aos pacientes com tumores do subtipo luminal A. Esses achados mostraram ser bastante robustos, uma vez que foram reproduzidos por vários outros estudos (Sørli et al. 2003; Hu et al. 2006; Langerød et al. 2007; Naume et al. 2007; Parker et al. 2009). As abordagens de classificação dos tumores de mama em grupos moleculares distintos têm sido constantemente aprimoradas, com o intuito de se identificar assinaturas gênicas que possam subclassificar os tumores nos subtipos moleculares e também discriminá-los quanto aos seus diferentes prognósticos e riscos de recorrência, como o preditor PAM50, o qual é baseado na expressão de 50 genes e é bastante utilizado para classificar molecularmente os tumores de mama nos quatro principais subtipos (luminal A, luminal B, basal (TN) e HER2 enriquecido (Parker et al. 2009).

A caracterização molecular e genética do câncer de mama tem sido continuamente investigada. Em 2012, num trabalho do grupo *The Cancer Genome Atlas* (TCGA), Koboldt e colaboradores utilizaram 6 metodologias ômicas - sequenciamento completo de exoma, análise genômica do número de cópias de DNA, sequenciamento de miRNA, análise de metilação de DNA, análise de expressão de mRNA, e análise de expressão proteica - para avaliar um grupo diverso de tumores de mama oriundos de 825 pacientes. Os autores confirmaram a existência dos 4 grupos principais de câncer de mama, semelhantes aos subtipos moleculares intrínsecos previamente identificados nos estudos de expressão gênica, e identificaram que o subtipo luminal A foi o que apresentou o maior número de genes alterados. Os autores também observaram que os 3 genes com mais mutações adquiridas considerando os 4 subtipos moleculares foram *TP53* (37%), *PIK3CA* (36%) e *GATA3* (11%), sendo que os genes *PIK3CA* e *GATA3* são mais alterados nos subtipos luminais A e B e menos alterados nos subtipos superexpressor de HER2 e basaloíde, enquanto o gene *TP53* apresentou um padrão inverso de alterações. De forma detalhada, por subtipo molecular, os genes que tiveram mais mutações adquiridas foram: Luminal A - *PIK3CA* (45%), *GATA3* (14%), *MAP3K1* (13%), *TP53* (12%) e *CDH1* (9%); Luminal B - *PIK3CA* (29%), *TP53* (29%) e *GATA3* (15%); HER-enriquecido - *TP53* (72%) e *PIK3CA* (39%); e finalmente basaloíde - *TP53* (80%) (Koboldt et al. 2012).

1.5 Caracterização molecular de lesões pré-invasivas de mama

As lesões pré-invasivas de mama (CDIS) são comumente identificadas apresentando um tamanho reduzido e com a presença de células não tumorais que podem dificultar uma análise acurada e robusta dos dados de expressão gênica das células tumorais. Essas características incentivaram diversos grupos, incluindo o nosso (Ma et al. 2003; Schuetz et al. 2006; Castro et al. 2008; Elias et al. 2016), a avaliarem as lesões pré-invasivas de mama através de uma abordagem que integra a captura por microdissecção a laser (Emmert-Buck et al. 1996) das células epiteliais tumorais, a amplificação linear do mRNA por transcrição *in vitro* baseada na metodologia da T7 RNA polimerase (Castro et al. 2008), e a avaliação transcricional por *cDNA microarray* (Schena et al. 1995). Ao aplicar esta abordagem para avaliar e comparar o padrão de expressão gênica das lesões pré-invasivas e invasivas (CMI-TNE), foi identificada uma grande semelhança entre os perfis (Ma et al. 2003; Schuetz et al. 2006; Lee et al. 2012).

No trabalho do nosso grupo (Castro et al. 2008) foi avaliado o padrão de expressão gênica de células epiteliais neoplásicas capturadas de mama normal, do componente *in situ* dos dois tipos de lesões (CDIS puro e componente *in situ* de CMI-TNE) e de lesões invasivas (CMI-TNE). Entre os grupos de amostras tumorais, as maiores diferenças foram observadas nos padrões de expressão gênica dos dois tipos de lesões pré-invasivas, as quais foram capazes de distinguir as lesões CDIS puro e componente *in situ* de CMI-TNE, e também, foram observadas semelhanças nos padrões de expressão gênica entre o componente *in situ* de CMI-TNE e o componente invasivo (CMI-TNE) em amostras independentes. De forma geral os resultados do grupo apresentaram concordância com os demais estudos (Ma et al. 2003; Schuetz et al. 2006; Lee et al. 2012) que avaliaram o padrão de expressão das células epiteliais neoplásicas nos dois grupos – do componente *in situ* de CMI-TNE e das células de lesões puramente invasivas (CMI-TNE), demonstrando que esse padrão era bastante semelhante. No entanto, nosso grupo, o único que avaliou de forma separada as células dos dois tipos de lesões pré-invasivas (CDIS puro e componente *in situ* de CMI-TNE), foi capaz de demonstrar que a maior diferença no padrão de expressão das células epiteliais neoplásicas foi observada justamente nesse ponto da progressão, entre os dois grupos de lesões pré-invasivas CDIS puro com baixo potencial maligno e componente *in situ* de CMI-TNE com alto potencial maligno, uma vez que este já coexistia com células epiteliais neoplásicas do componente invasivo (Castro et al. 2008). Assim, os resultados sugeriram fortemente que a programação molecular já esteja estabelecida nas células do componente *in situ* de CMI-TNE antes que o evento de

invasão tumoral seja morfológicamente identificado (Castro et al. 2008), corroborando com a suposição de que programação molecular para a invasão já está presente nos estágios iniciais (Ma et al. 2003; Knudsen et al. 2012).

A partir desses achados, nosso grupo tem se empenhado em caracterizar as diferenças moleculares, especialmente as diferenças de expressão gênica existentes nas células epiteliais neoplásicas dos dois tipos de lesões pré-invasivas, vislumbrando identificar biomarcadores com potencial de prever a progressão de lesões CDIS. Em um dos estudos utilizamos diferentes abordagens metodológicas, tais como *cDNA microarray*, bibliotecas subtrativas de cDNA (método baseado em Jiang e colaboradores (2000) e em Baldan-Pineda (2007), que trabalhou com RNA amplificado) e RT-qPCR, para investigar os genes diferencialmente expressos (GDE) em células epiteliais neoplásicas capturadas a laser dos dois tipos de lesões pré-invasivas e avaliamos a capacidade do padrão de expressão baseado nos GDE das células epiteliais neoplásicas (GDEe) em discriminar corretamente os dois grupos. Identificamos um grupo de 20 GDE, cujo perfil de expressão foi capaz de discriminar relativamente bem os dois grupos de lesões pré-invasivas, sendo que 78% das amostras de células epiteliais neoplásicas de CDIS puro e 89% das amostras de células epiteliais neoplásicas de componente *in situ* de CMI-TNE foram corretamente discriminadas (Elias et al. 2016), demonstrando um potencial dessa assinatura molecular baseada no padrão de expressão de 20 genes ser capaz de discriminar células epiteliais neoplásicas com potenciais malignos distintos.

Diversos estudos têm demonstrado o papel das células mioepiteliais na manutenção e no desenvolvimento dos ductos mamários (Lakhani e O'Hare 2001), assim como seu possível papel como supressor de invasão tumoral (Polyak and Hu 2005). Tem sido observado que uma diminuição na expressão de marcadores de diferenciação do mioepitélio, como por exemplo, calponina-1, está correlacionado com a progressão de CDIS (Russell et al. 2015; Mitchell et al. 2020). Desta forma, com o intuito de avaliar se o padrão de expressão baseado nos GDE das células mioepiteliais (GDEm), reconhecidas como supressoras de invasão, poderia ser mais robusto em discriminar os dois grupos de lesões pré-invasivas (CDIS puro e componente *in situ* de CMI-TNE), nosso grupo avaliou por RNAseq células epiteliais neoplásicas e mioepiteliais capturadas a laser dessas lesões. Nesse estudo, o método de construção de biblioteca e sequenciamento de RNAseq foi adaptado de Carraro et al. (2011) e Ferreira et al. (2015) para que o mesmo pudesse ser utilizado a partir de quantidades pequenas de RNA obtido de amostras microdissecadas a laser. Assim o RNA poliadenilado obtido das células epiteliais

neoplásicas e mioepiteliais capturadas a laser das lesões pré-invasivas foi convertido em cDNA e submetido a dois ciclos de amplificação linear por transcrição *in vitro* para poder ser avaliado pela metodologia de RNAseq (RNAseq-3'). Foram identificados 464 GDEe e 338 GDEm, cujos padrões de expressão foram capazes de discriminar os dois tipos de lesões pré-invasivas. A partir destes resultados, investigou-se os GDE que eram comuns aos GDEe e GDEm. Foram identificados 132 GDE comuns, cujo padrão de expressão também foi capaz de discriminar os dois tipos de lesões pré-invasivas. Com isso, esse trabalho demonstrou que os padrões de expressão dos GDE em ambas as células epiteliais neoplásicas e mioepiteliais foram igualmente capazes em discriminar amostras de CDIS puro e de componente *in situ* de CMI-TNE (Elias et al., em preparação). Além disso, esses resultados indicaram que tanto as células epiteliais neoplásicas quanto as mioepiteliais das lesões pré-invasivas refletem a programação para a invasão tumoral.

Um dos grandes desafios da conduta clínica de CDIS é justamente estratificar as pacientes de acordo com o risco da lesão CDIS progredir para a doença invasiva ou apresentar recidiva de lesão *in situ*, no caso de pacientes com maior risco, submetê-las aos tratamentos padrões que normalmente são agressivos e considerados excessivos por grande parte da comunidade médica e científica; e poupar aquelas pacientes com menor risco para que recebam tratamentos mais brandos, garantindo os altos níveis de cura do CDIS.

A semelhança molecular entre as amostras de componente *in situ* de CMI-TNE e amostras de componente invasivo de CMI-TNE também já foi relatada em estudos que avaliaram seus perfis mutacionais. Num estudo do grupo de Reis-Filho (Hernandez et al. 2012), foram utilizadas as metodologias de Hibridização Genômica Comparativa baseada em Arranjos (*Array Comparative Genomic Hybridization* (a-CGH)) e sequenciamento de DNA por NGS com um painel de 19 genes para avaliar o padrão de mutações em 12 pacientes pareados de componente *in situ* e invasivo das lesões. Os autores identificaram uma grande semelhança no padrão de ganhos e perdas cromossômicas entre o componente *in situ* e invasivo, assim como, mutações no gene *PIK3CA* em 50% dos pacientes (Hernandez et al. 2012). No estudo de Pang e colaboradores (2017) os autores avaliaram um painel de 107 genes numa coorte de 20 pacientes de CDIS puro e também identificaram mutações em *PIK3CA* em 55% dos pacientes. Além disso, também encontraram os genes *TP53* e *GATA3* mutados em 30% e 45% dos pacientes, respectivamente. Por outro lado, em outro estudo do grupo de Reis-Filho (Pareja et al. 2020), os autores realizaram o sequenciamento completo de exoma (*Whole-*

Exome Sequencing (WES)) de 7 pacientes de CDIS puro, 25 amostras de componente *in situ* de CMI-TNE e 24 tumores invasivos (CMI-TNE). Os autores confirmaram uma grande concordância no repertório mutacional entre o componente *in situ* e invasivo, dos quais 52% e 54% apresentaram mutação no gene *TP53*, 41% e 42% em *PIK3CA* e 26% e 23% em *GATA3*, respectivamente. Ao compararem o repertório mutacional do CDIS puro (sem progressão por um período médio de 6 anos de acompanhamento) com o componente *in situ*, encontraram que o CDIS puro apresenta uma frequência menor de mutações nos genes *TP53* (14% vs 52%) e *PIK3CA* (0% vs 41%). Além disso, ao investigarem a relação entre o repertório mutacional de CDIS (puro e componente *in situ* de lesão invasiva juntos) e o grau nuclear, os autores observaram que o CDIS de grau nuclear 3 (n=19) apresentou mais variantes somáticas do que o CDIS de grau 2 (n=15). E identificaram com significância estatística que o CDIS de grau nuclear 3 (n=20) apresentou uma frequência de mutações no gene *TP53* maior do que o CDIS de grau nuclear 2 (n=13; 75% vs 0%) (Pareja et al. 2020).

Esforços também têm sido realizados na tentativa de se identificar diferenças mutacionais em pacientes de CDIS puro que apresentaram diferentes desfechos (Lin et al. 2019) através de NGS avaliaram um painel com 223 genes numa coorte formada por 125 pacientes de CDIS: 70 CDIS puro (5 pacientes de CDIS que apresentaram recidiva e 65 pacientes de CDIS que não apresentaram recidiva num período médio de 13 anos) e 55 pacientes de componente *in situ* de CMI-TNE. Os autores encontraram os genes *PIK3CA* e *TP53* como frequentemente mutados tanto nos pacientes de CDIS puro (37% e 13%, respectivamente) quanto nos pacientes de componente *in situ* de CMI-TNE (26% e 21%, respectivamente). Além disso, nenhuma diferença na frequência mutacional foi detectada nos pacientes de CDIS puro com e sem recidiva. Mais recentemente, Cosimo e colaboradores (2020) utilizaram um painel de 409 genes (*Ion AmpliSeq Comprehensive Cancer Panel*) e avaliaram os tumores primários de 27 pacientes de CDIS puro, dos quais 14 apresentaram pelo menos uma mutação. Os genes *PIK3CA* e *TP53* foram encontrados como alterados em 26% e 11% dos pacientes, respectivamente. E em 3 dos 27 pacientes as pacientes apresentaram recidiva, sendo que em 2 desses pacientes foram encontradas as mesmas mutações no tumor primário e na recidiva. Além disso, os autores detectaram no plasma de 71% (10/14) das pacientes, antes da cirurgia, as mesmas mutações identificadas nos tumores, mostrando que mesmo tumores CDIS puro liberam DNA livre no plasma.

1.6 Biomarcadores de predição de risco de progressão ou recidiva de CDIS

A estratificação de pacientes com CDIS de acordo com o risco de progressão para a doença invasiva e de recidiva de CDIS ainda é um dos grandes desafios clínicos. Alguns testes moleculares estão disponíveis comercialmente, como o *Oncotype DX Breast DCIS Score* (Genomic Health) (Solin et al. 2013) e o *DCISionRT* (Prelude Dx) (Bremer et al. 2018), e métricas validadas, como o Índice Prognóstico de Van Nuys (Silverstein 2003) e o nomograma de CDIS do *Memorial Sloan-Kettering Cancer Center* (Rudloff et al. 2010), os quais têm sido desenvolvidos para identificar esse risco, de acordo com diferentes fatores clínico-patológicos dos tumores e na expressão de determinados marcadores moleculares.

O teste *Oncotype DX Breast DCIS Score* (Solin et al. 2013), o qual é baseado na avaliação por RT-qPCR de 12 genes, sendo 7 genes relacionados ao câncer e 5 genes referência, foi desenvolvido a partir do teste de 21 genes *Oncotype DX Breast Recurrence Score* (Paik et al. 2004), onde foram mantidos os genes do grupo de proliferação e excluídos os genes dos grupos de invasão e do grupo relacionado ao HER2. Este é um teste que tem valor prognóstico, pois prediz o risco de recidiva tanto de CDIS como de lesão invasiva em pacientes de CDIS de baixo risco que foram submetidos somente a cirurgia conservadora. O teste gera um escore, uma variável contínua em escala de 0 a 100, que estratifica as pacientes em grupos de baixo (escore < 39), intermediário (escore 39-54) e alto risco (escore $\geq$ 55), independentemente dos fatores clínico-patológicos, tais como idade, tamanho do tumor, grau, necrose, multifocalidade e subtipo. O teste já foi validado por estudo clínico (Rakovitch et al. 2015) para determinar o risco de recidiva de CDIS e de lesão invasiva (CMI-TNE) em 10 anos.

O teste *DCISionRT* é baseado numa assinatura biológica que contempla fatores moleculares, tais como a avaliação por imunoistoquímica de 7 proteínas, e fatores clínico-patológicos, tais como idade, tamanho do tumor, margem cirúrgica e palpabilidade. Este teste tem valor prognóstico e preditivo, pois prediz o risco de recidiva tanto de CDIS como de lesão invasiva, e prediz o benefício de incluir a radioterapia no tratamento. O teste gera um escore, uma variável contínua em escala de 0 a 10, que estratifica as pacientes num grupo de baixo risco (escore $\leq$ 3) ou num grupo de elevado risco (escore > 3). Este teste já teve o seu valor prognóstico validado por estudo clínico (Weinmann et al. 2020).

1.7 Justificativa

O CDIS constitui um grupo heterogêneo de tumores mamários pré-invasivos que podem recidivar e progredir para o carcinoma de mama invasivo, que é a forma potencialmente letal da doença. O número de pacientes de CDIS diagnosticados aumentou bastante nas últimas décadas devido aos programas de rastreamento, o que tem estimulado a contínua busca por biomarcadores robustos com potencial de estratificar adequadamente as pacientes em maior ou menor risco. Mesmo com todos os esforços ainda não é possível distinguir de maneira acurada as pacientes diagnosticadas com CDIS com baixo potencial de progredir ou de recidivar, e consequentemente, que poderiam receber modalidades de tratamento mais leves e ainda assim manter as excelentes taxas de sobrevida do tratamento convencional, que hoje é considerado um tratamento excessivo para algumas pacientes. Neste estudo nossa proposta é investigar candidatos a biomarcadores de predição de progressão para o carcinoma de mama invasivo ou de recidiva de lesão *in situ*, através da avaliação do perfil transcricional e mutacional de lesões pré-invasivas de carcinoma ductal de mama com diferentes potenciais malignos (CDIS e componente *in situ* de lesão invasiva) e com diferentes desfechos (CDIS com e sem progressão ou recidiva).

2. OBJETIVOS

2.1 Objetivo geral

Caracterizar molecularmente lesões pré-invasivas de carcinoma ductal de mama a fim de identificar candidatos a biomarcadores de predição de progressão ou recidiva de carcinoma ductal *in situ*.

2.2 Objetivos específicos

- Selecionar 40 amostras de carcinoma ductal *in situ* de mama, que atendam aos critérios de inclusão do estudo;
- Avaliar pela metodologia de RT-qPCR com o uso da plataforma *Taqman Low Density Array* (TLDA) 78 genes selecionados em estudos prévios do grupo (Elias et al., 2016 e Elias et al., em preparação) como diferentemente expressos entre células epiteliais neoplásicas capturadas de lesões pré-invasivas com diferentes potenciais malignos;
- Identificar uma assinatura gênica cujo perfil de expressão seja capaz de discriminar os dois grupos de amostras pré-invasivas de carcinoma de mama;
- Caracterizar o repertório mutacional de CDIS com diferentes desfechos;
- Avaliar a associação entre as mutações identificadas no DNA do tumor e os grupos de CDIS com e sem progressão.

3. METODOLOGIA

3.1 Aprovação do projeto no Comitê de Ética em Pesquisa

O presente projeto de pesquisa correspondeu a um projeto afiliado do projeto temático intitulado “Aspectos moleculares envolvidos no risco, desenvolvimento e progressão do carcinoma ductal de mama: busca de novos genes de susceptibilidade e investigação da progressão do carcinoma *in situ* e do papel da mutação em *BRCA1* no tumor triplo negativo”, o qual foi coordenado pela Dra. Dirce Maria Carraro (processo FAPESP 2013/23277-8). O presente estudo teve caráter retrospectivo e foi aprovado pelo Comitê de Ética em Pesquisa do A. C. Camargo Cancer Center (número 1746/13F).

3.2 Confidencialidade dos dados

O risco de perda de confidencialidade dos dados coletados dos participantes da pesquisa é mínimo, pois houve a preservação da privacidade dos mesmos cujos dados foram coletados. As informações foram utilizadas única e exclusivamente para a execução do estudo e as informações que possam ser divulgadas, sempre serão realizadas de forma anônima, não sendo usadas iniciais ou quaisquer outras indicações que possam identificar os participantes da pesquisa.

3.3 Financiamento do estudo

Este estudo foi financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) com verba proveniente do projeto temático ao qual ele é afiliado (processo FAPESP do projeto temático: 2013/23277-8) e Projeto INCT (FAPESP 2014/50943-1 e CNPq 465682/2014-6).

3.4 Avaliação transcricional de CDIS em células epiteliais neoplásicas capturadas por microdissecação a laser

3.4.1 Seleção da casuística

A avaliação transcricional de CDIS foi realizada com amostras de tecido tumoral congelado que foram selecionadas de estudos prévios do nosso grupo (Castro et al. 2008; Pinilla-Fernández 2015).

Os critérios de inclusão foram: amostras submetidas a microdissecção a laser das células epiteliais neoplásicas e mioepiteliais de CDIS puro ou componente *in situ* de CMI-TNE; pacientes não submetidas a tratamento quimioterápico neoadjuvante.

Todas as amostras já haviam sido submetidas a microdissecção a laser das células epiteliais neoplásicas e mioepiteliais com o uso do equipamento PixCell II LCM system (Arcturus Engineering), e realizada a extração do RNA total e amplificação do RNA mensageiro (mRNA). De forma resumida, aproximadamente 1500-3000 células epiteliais neoplásicas e mioepiteliais foram capturadas em dispositivos plásticos CapSure HS LCM Caps (Arcturus Engineering), de acordo com as recomendações do fabricante. O RNA foi extraído com o uso do kit PicoPure™ RNA Isolation kit (Arcturus Engineering), de acordo com as recomendações do fabricante, e incluiu uma etapa de tratamento com DNase com o uso do kit RNase-Free DNase Set (Qiagen), de acordo com as recomendações do fabricante. O RNA foi amplificado por 2 ciclos de transcrição *in vitro* e o cDNA sintetizado com o uso do kit Arcturus RiboAmp PLUS Kit (Arcturus Engineering), de acordo com as recomendações do fabricante. O cDNA foi submetido a reação de PCR com sondas SYBR-green para amplificação dos genes *ACTB*, *HPRT1* e *GAPDH*, com a finalidade de avaliar sua integridade.

As amostras degradadas foram excluídas da seleção.

3.4.2 Avaliação da expressão gênica de CDIS por Taqman Low Density Array (TLDA)

A avaliação do nível de expressão gênica foi realizada através da metodologia quantitativa da reação em cadeia da polimerase por transcrição reversa (*reverse transcription – quantitative polymerase chain reaction* - RT-qPCR). Nesta abordagem foram customizados 94 ensaios TLDA (ThermoFisher Scientific) em placas de 384 poços, de acordo com as recomendações do fabricante. Desta forma foi possível avaliar simultaneamente o nível de expressão de 94 genes. Cada placa TLDA foi configurada contendo 6 genes referências (*ACTB*, *BCR*, *GAPDH*, *GUSB*, *HPRT1* e *RPLP0*) (candidatos a normalizadores), 20 genes que compuseram uma assinatura gênica identificada em um trabalho prévio do grupo (*ALMS1*, *ANAPC13*, *ARHGAP9*, *AZGP1*, *CAMP*, *CHRNA1*, *CPNE3*, *CTTNBP2NL*, *EDN1*, *EPOR*, *GRB10*, *HLTF*, *INPP1*, *LSM4*, *MAPK8*, *P4HB*, *RABEPK*, *RARRES3*, *REC8* e *UTP20*) (Elias et al. 2016) e 68 genes obtidos do estudo anterior (Elias et al., em preparação) que utilizou um método customizado de sequenciamento de RNA da porção 3' dos transcritos (RNAseq 3'), adaptado de (Carraro et al. 2011; Ferreira et al. 2015). Dos 68 genes, 58 correspondem aos genes interrogados que

foram selecionados entre os genes diferencialmente expressos entre dois grupos de amostras de células epiteliais neoplásicas e mioepiteliais microdissecadas a laser (CDIS puro e componente *in situ* de CMI-TNE) (*ABCA1, ATP2A3, BAALC, BMP7, C15orf48, CC2D1A, CDC42BPG, CEACAM5, COLEC12, CXCL10, DEGS1, DNAJA3, DNAJB6, EPB41L3, ERCC8, FLCN, FLT4, FMNL1, HAT1, HCP5, HDC, HIPK2, IRF2BPL, LILRB1, LIMS1, LOC285593, MYB, MYO1C, NCF4, NDFIP1, NOTCH4, PEX11A, PLAUR, POLR3A, PTAFR, RBM47, REEP1, RELT, RNPC3, ROBO1, RPS6KA4, RRP1, S100A9, SACM1L, SCPEP1, SLC16A3, SLC16A6, SLC35G2, SLC44A5, SPTBN2, SRA1, SRRM1, ST6GALNAC4, TMEM45A, TNFRSF18, TRPM2, XAF1 e ZBP1*) e 10 genes correspondem aos controles internos obtidos dos dados de RNAseq 3' (*CALR, EEF1D, ENSA, HNRNPDL, KIAA0930, LYRM2, MGAT1, RNF103, SPG11 e YKT6*).

Os ensaios TaqMan foram customizados de acordo com os seguintes critérios: (1) sondas direcionadas à região 3' de transcritos humanos; (2) sondas compreendendo dois éxons diferentes (quando possível) e (3) sondas complementares ao maior número possível de variantes transcricionais.

Foram utilizados 5 ng de cDNA para a avaliação de cada amostra nos ensaios TLDA, os quais foram sintetizados a partir de um RNA previamente submetido a 2 ciclos de amplificação por transcrição *in vitro*, seguindo as recomendações do fabricante do kit Arcturus RiboAmp HS PLUS (Thermo Fisher Scientific). Estes 5 ng de cDNA foram diluídos em 50 µL de água, com a adição de 50 µL de *master mix*. A reação (100 µL de volume final) foi distribuída na placa TLDA para todas as 94 sondas previamente desenhadas. Os ensaios foram realizados no equipamento Applied Biosystems 7900HT Fast Real-Time PCR System (Thermo Fisher Scientific). As condições foram: 50°C por 2 minutos, 95°C por 10 minutos, 40 ciclos a 97°C por 30 segundos seguido de 60°C por 1 minuto. Os valores de expressão gênica foram calculados pelo programa Sequence Detection System (SDS) (Thermo Fisher Scientific).

Os genes referência candidatos a normalizadores tiveram o seu nível de expressão avaliados pelo programa *qBase* (Hellemans et al. 2007) para a identificação dos genes mais estáveis para sua utilização na etapa de normalização dos dados de expressão ($M < 1.5$). Para avaliarmos de maneira conjunta todos os dados de expressão gênica, os arquivos SDS foram importados no aplicativo *Relative Quantification* (RQ) (www.apps.thermofisher.com) (Thermo Fisher Scientific), uma plataforma online e gratuita de quantificação relativa de expressão gênica. Assim conseguimos garantir que o valor de *threshold* de expressão de um gene fosse

o mesmo para todos os diferentes experimentos (diferentes placas). O valor de *threshold* de expressão que é gerado automaticamente pelo equipamento foi utilizado para a maioria dos ensaios. Entretanto, para alguns ensaios o ajustamos manualmente para que estivesse na região das curvas de amplificação correspondente a fase de produção exponencial dos amplicons. Além disso, ao visualizarmos o padrão dessas curvas de amplificação, também foi possível excluir ensaios que não apresentaram um padrão exponencial.

Com o intuito de analisarmos somente dados de expressão gênica com qualidade, estabelecemos os seguintes critérios de controle de qualidade: (1) inspeção visual dos gráficos de magnitude da intensidade de sinal de fluorescência (Δ -Rn) *versus* o ciclo de amplificação, para selecionarmos apenas as curvas que apresentassem um claro padrão exponencial; (2) seleção de genes-alvo que foram expressos em pelo menos 50% das amostras em cada grupo (CDIS puro ou componente *in situ* de CMI-TNE); (3) seleção de amostras nas quais pelo menos 50% dos genes-alvo foram expressos.

O nível de expressão gênica normalizada foi calculado através do método comparativo de Cq ($\Delta\Delta$ Cq), calculando-se a razão entre 2^{-Cq} do gene-alvo e o fator de normalização (média geométrica de 2^{-Cq} dos genes-referência mais estáveis identificados pelo plugin GeNorm do programa qBase (Hellemans et al. 2007). A diferença de expressão foi calculada através da razão entre a média de expressão no grupo de amostras CDIS e a média de expressão no grupo de amostras de componente *in situ* de CMI-TNE. O critério adotado para considerar que houve uma diferença de expressão foi *fold change* $\geq |2|$.

A análise de significância estatística das diferenças de expressão gênica e os gráficos *dotplot* para ilustrarem os valores de expressão gênica foram realizados pelo programa *GraphPad Prism* (<https://www.graphpad.com/scientific-software/prism/>). O teste estatístico utilizado foi o teste “t de *Student*” não pareado e a significância estatística foi considerada com $P < 0,05$.

3.4.3 Análise de clusterização

A análise de clusterização foi realizada através do programa *Multiple Experiment Viewer* (MeV), versão 4.4.1 (Saeed et al. 2003). Inicialmente, os valores de expressão normalizados foram transformados em escore-Z, em que para cada gene o seu valor de expressão numa determinada amostra foi subtraído da média de sua expressão em todas as

amostras e dividida pelo desvio-padrão. Estes valores de expressão foram submetidos a uma clusterização hierárquica não supervisionada com distância Euclidiana e média das distâncias. A confiabilidade da clusterização foi avaliada através da técnica de *bootstrapping*, considerando 1000 permutações.

3.4.4 Análise de enriquecimento funcional

Os genes diferencialmente expressos foram submetidos a uma análise de enriquecimento funcional através do programa *Ingenuity Pathway Analysis* (IPA) (Qiagen). Utilizamos parâmetros padrão: a população de genes referência utilizada foi o banco de dados *Ingenuity Knowledge Base* (IKB), um banco de dados manualmente curado por especialistas; as interações gênicas consideradas foram diretas e indiretas; um número máximo de 35 genes por rede de interação; número máximo de 25 redes de interação por análise; a significância estatística dos resultados foi considerada baseando-se em evidências da literatura; e P-valor < 0,05 foi considerado para identificar enriquecimento de redes de interação gênica e vias bioquímicas canônicas e seus estados preditos como ativadas ou inativadas.

3.5 Avaliação mutacional de CDIS

3.5.1 Caracterização da casuística

3.5.1.1 Critérios de inclusão

Foram selecionadas pacientes cujas biópsias tiveram um laudo anatomopatológico de CDIS puro, com pelo menos 20% de células tumorais (estimativa visual na lâmina HE do percentual de CDIS dentro da área marcada pelo patologista) e que atendiam aos critérios descritos abaixo. A casuística foi dividida em dois grupos, o grupo 1 foi denominado CDIS sem progressão e o grupo 2 foi denominado CDIS com progressão.

Para os pacientes do grupo CDIS sem progressão, os critérios de inclusão foram: laudo anatomopatológico de carcinoma ductal *in situ* (CDIS puro) de mama tanto no material da biópsia quanto no material da cirurgia e sem progressão da doença num período mínimo de seguimento de 3 anos (Grupo 1: CDIS sem progressão – CDIS puro).

Os pacientes do grupo CDIS com progressão foram organizados em 2 subgrupos, os critérios de inclusão para o grupo 2.1 (CDIS/invasivo) foram: laudo anatomopatológico de carcinoma ductal *in situ* no material da biópsia e laudo anatomopatológico de carcinoma de mama invasivo tipo não especial (CMI-TNE) no material da cirurgia; e para o grupo 2.2

(CDIS/progressão/recidiva) os critérios de inclusão foram: pacientes que tiveram laudo anatomopatológico de carcinoma ductal *in situ* tanto no material da biópsia quanto no material da cirurgia e que apresentaram progressão (nova lesão ipsilateral invasiva (CMI-TNE)) ou recidiva (nova lesão ipsilateral *in situ* (CDIS)) durante o seguimento.

3.5.1.2 Critério de exclusão

Os pacientes excluídos do estudo atenderam ao seguinte critério: pacientes cujas amostras de biópsia tiveram laudo anatomopatológico de CDIS, mas que durante a revisão anatomopatológica foi identificada a presença de microinvasão (presença de foco (s) de invasão tumoral com ≤ 1 mm em sua maior dimensão) nos novos cortes histológicos deste material.

3.5.2 Seleção da casuística

3.5.2.1 Amostras de tecido tumoral fixado em formalina e embocado em parafina

Inicialmente foi realizado um levantamento de pacientes na plataforma eletrônica *Recruit*, um sistema desenvolvido pelo Departamento de Informática Médica do A. C. Camargo Cancer Center (ACCCC), para identificar amostras tumorais de pacientes que foram tratadas na instituição. Os parâmetros de busca foram ajustados da seguinte maneira: Data de diagnóstico: 2008 a 2018; Sexo: feminino; Diagnóstico CID-10: mama; CID-O: carcinoma ductal *in situ*. O limite de 2018 foi considerado para garantir um acompanhamento da paciente diagnosticada com CDIS de no mínimo 3 anos. Pacientes anteriores a 2008 não foram considerados, pois materiais emblocados em parafina e armazenados por muito tempo apresentam RNA/DNA com pior qualidade, dificultando a obtenção de resultados robustos.

A etapa seguinte foi realizada a partir da lista gerada pelo *Recruit* contendo os códigos referentes ao registro geral hospitalar (RGH) dos pacientes. A partir dos RGHs foi possível analisar os prontuários eletrônicos disponíveis na plataforma *H2TC*, outra plataforma eletrônica desenvolvida pelo Departamento de Informática Médica do ACCCC, para avaliar o histórico e evolução dessas pacientes com o intuito de identificar quais delas atendiam aos nossos critérios de inclusão. Além disso, nesta etapa também já foi possível identificar a topografia correta do bloco de parafina daqueles pacientes em que esta informação estava disponível no prontuário eletrônico. Em seguida, foram identificados os pacientes em que os blocos de parafina estavam disponíveis, através da solicitação deles ao Arquivo da Patologia.

A etapa de identificação dos pacientes a serem incluídos no estudo também teve a colaboração da Dra. Cynthia Osorio do departamento de Patologia e do Dr. Almir Bitencourt do departamento de Imagem do ACCCC.

Após esses pacientes terem sido submetidos a uma reavaliação anatomopatológica, a qual foi realizada pela Dra. Cynthia Osorio, foi possível selecionar os pacientes para serem submetidos a extração de DNA da área tumoral de interesse. A Figura 3 apresenta um fluxograma com uma breve descrição das principais etapas realizadas durante o processo de seleção desses pacientes.

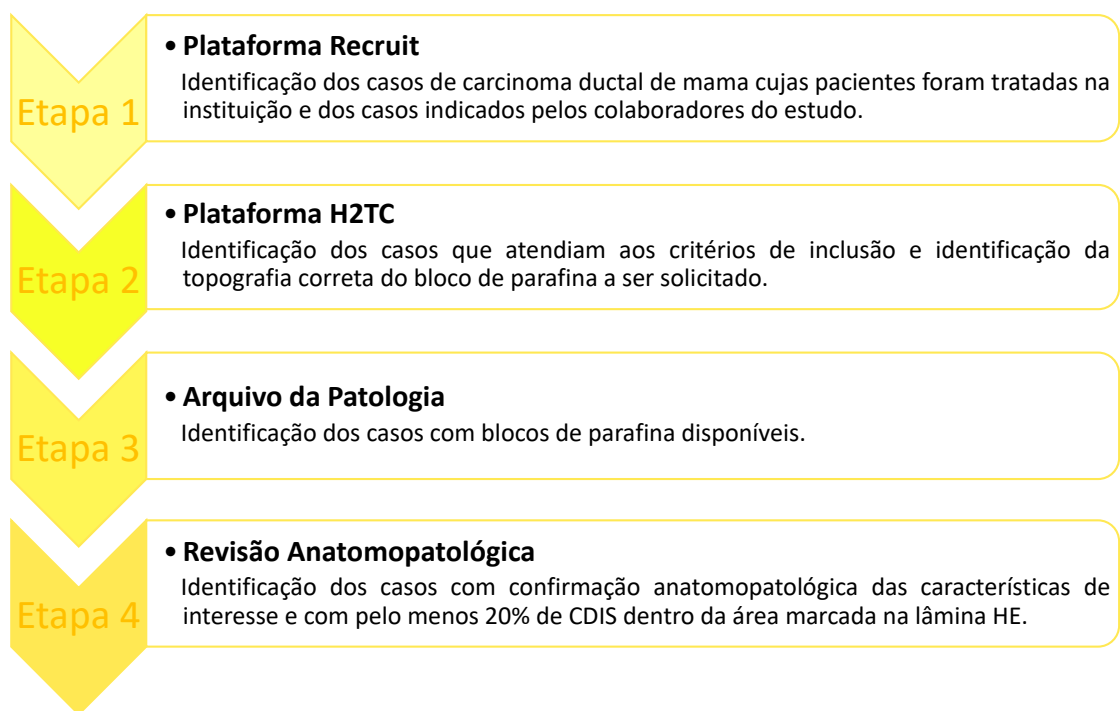


Figura 3 - Fluxograma ilustrando o processo de seleção de amostras de tecido tumoral fixado em formalina e embocado em parafina.

3.5.2.2 Amostras de tecido tumoral congelado

Inicialmente foi realizado um levantamento de pacientes na plataforma eletrônica *Biobank*, um sistema desenvolvido pelo Departamento de Informática Médica do ACCCC, para identificar amostras tumorais congeladas e armazenadas no biobanco da instituição.

A etapa seguinte foi realizada seguindo o mesmo fluxo que foi realizado com as amostras parafinadas. A Figura 4 apresenta um fluxograma com uma breve descrição das principais etapas realizadas durante o processo de seleção desses pacientes.

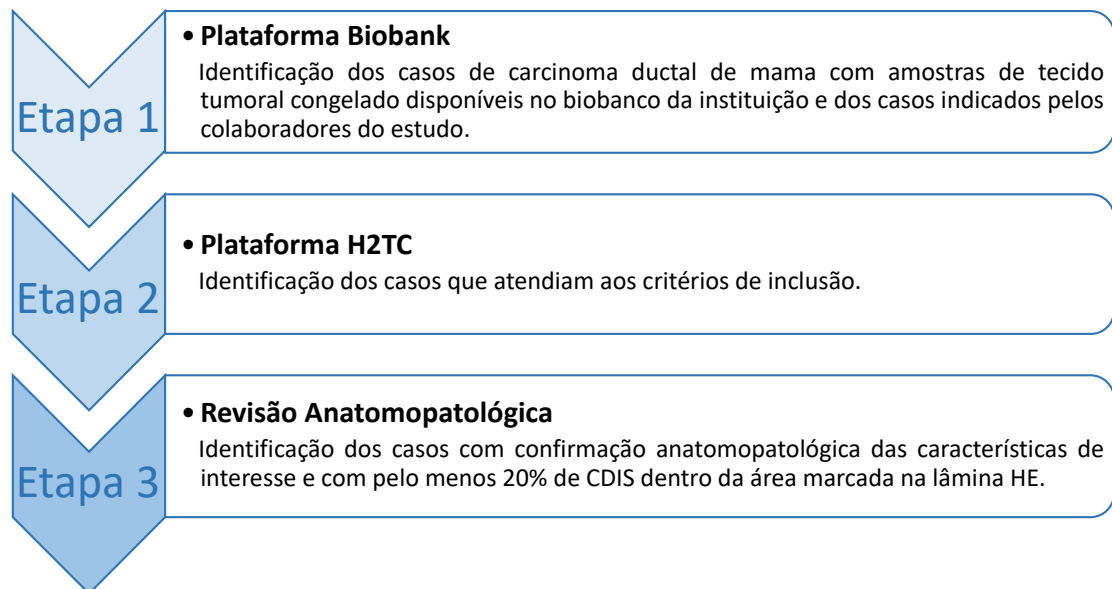


Figura 4 - Fluxograma ilustrando o processo de seleção de amostras de tecido tumoral congelado.

3.5.3 Extração de DNA das amostras do estudo

Todos os pacientes de biópsia de CDIS selecionados foram submetidos a uma revisão anatomopatológica pela patologista Dra. Cynthia Osório para garantir que não havia a presença de focos de microinvasão tumoral. As lâminas HE com as áreas CDIS marcadas foram utilizadas como referência para posterior dissecação manual das áreas CDIS nas lâminas não coradas, e o conteúdo de células tumorais nas áreas de CDIS foi estimado visualmente pela patologista, e só foram incluídas as amostras com no mínimo 20% de células tumorais. Nas amostras de peças cirúrgicas em que foram identificados foco (s) de invasão tumoral (Grupo 2.1 – CDIS/invasivo) somente as áreas de CDIS foram macrodissecadas e submetidas a extração de DNA.

Para as amostras de tecido tumoral em FFPE que não havia escassez de material, submetemos os blocos de parafina de 4 pacientes a 4 a 8 *punches* de 1 mm de diâmetro nas regiões dos blocos correspondentes às áreas CDIS das lâminas HE (Figura 5). Optamos pelo *punch* como método de macrodissecação, considerando que esta abordagem iria proporcionar

uma quantidade maior de material. Este material foi submetido a desparafinização através de 3 banhos de xilol. Em cada um deles o material foi incubado a 57°C e em agitação a 300 rpm por 5 minutos. Após centrifugação em velocidade máxima por 1 minuto, o sobrenadante foi descartado. Em seguida, fez-se 3 lavagens com etanol 100%, com homogeneização por inversão. O *pellet* foi mantido em estufa a 42°C por 2 minutos para completa evaporação do etanol. Em seguida, o pellet foi ressuspendido em 180 µL de tampão ATL e a digestão proteica ocorreu com a adição de 30 µL da enzima proteinase K. A reação foi incubada a 56°C e em agitação a 500 rpm por cerca de 20 horas. Para as amostras cuja digestão não havia ocorrido completamente, foram adicionados mais 20 µL de proteinase K e a reação foi incubada a 56°C e em agitação a 500 rpm por cerca de 30 minutos. Em seguida, a reação foi incubada a 90°C por 1 hora, depois de resfriar em temperatura ambiente foram adicionados 2 µL de RNase A (100 mg/mL) e a reação foi incubada a temperatura ambiente por 2 minutos. Em seguida, foram adicionados 200 µL de tampão AL, seguido de mais 200 µL de etanol 100%. O lisado foi transferido para uma coluna MinElute (Qiagen), a qual foi centrifugada a 8000 rpm por 1 minuto. A purificação teve início com a adição de 500 µL de tampão AW1 e centrifugação a 8000 rpm por 1 minuto, seguido da adição de 500 µL de tampão AW2 e centrifugação a 8000 rpm por 1 minuto, seguido da adição de 500 µL de tampão álcool 70% e centrifugação a 8000 rpm por 1 minuto. Em seguida, a coluna foi centrifugada em velocidade máxima por 3 minutos. A eluição seguiu com a adição de 20 ou 30 µL de TAE para material de biópsia ou peça cirúrgica, respectivamente. A coluna foi mantida em temperatura ambiente por 5 minutos e depois centrifugada em velocidade máxima por 1 minuto. O eluato foi quantificado através de espectrofotometria por Nanodrop (Thermo Fisher Scientific) e por fluorescência por Qubit (Thermo Fisher Scientific).

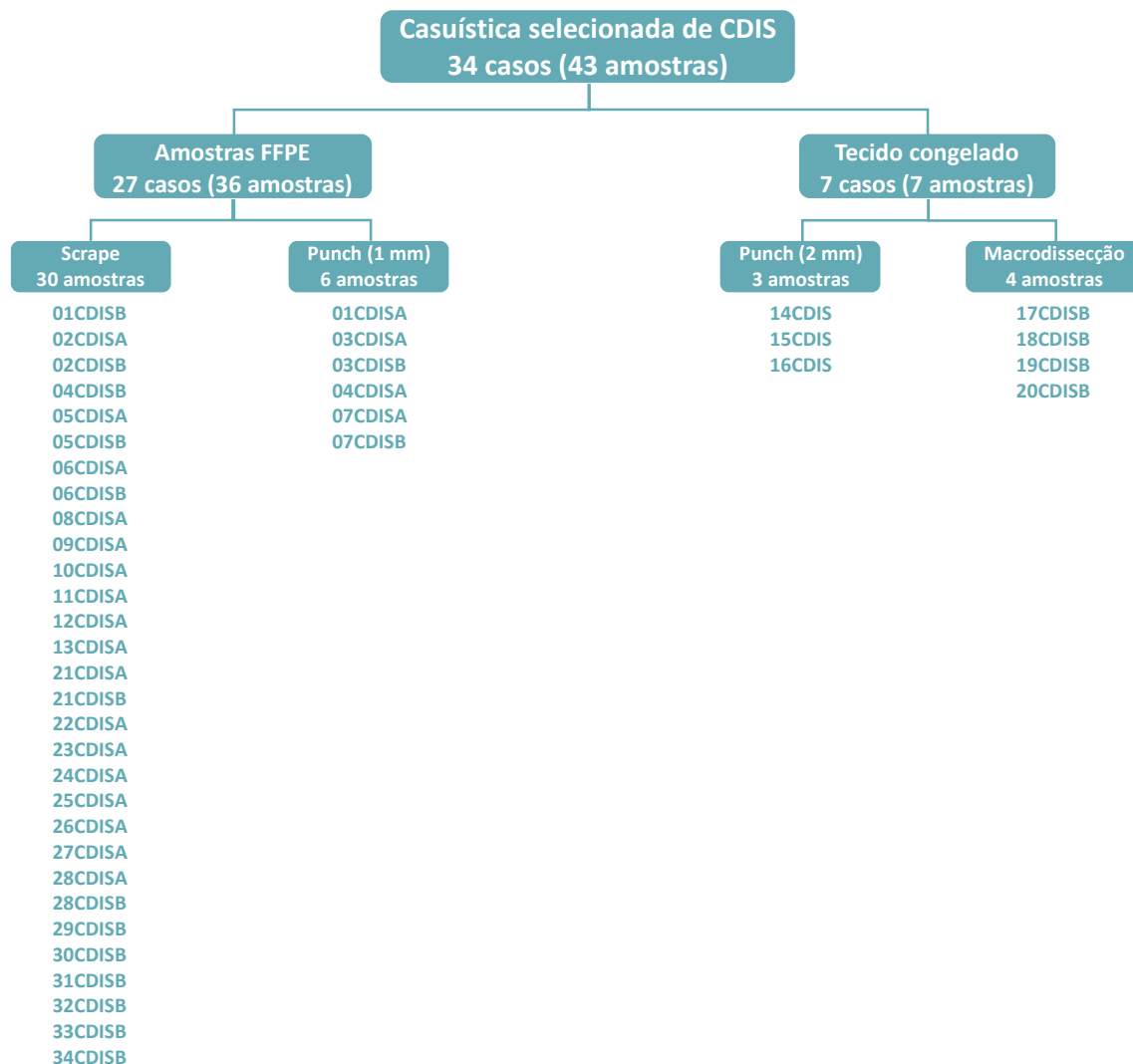
Para as amostras com pouco material disponível nos blocos, para evitar que esgotássemos o material, utilizamos o procedimento de *scrape* (Figura 5), e adotamos o protocolo do biobanco que já havia sido estabelecido para esse procedimento. Para esses pacientes os blocos de parafina foram submetidos a cortes histológicos de 5 µm de espessura, sem coloração, para posterior *scrape* das áreas CDIS previamente marcadas nas correspondentes lâminas HE. Cada caso teve de 1 a 3 blocos de parafina submetidos a cortes histológicos e cada bloco de parafina foi submetido a 8 a 12 cortes histológicos. A extração de DNA foi realizada com o uso do kit QIAamp DNA Mini Kit, seguindo as recomendações do

fabricante (Qiagen). A desparafinização deste material foi realizada com a adição de 160 μL de solução de desparafinização, agitação por vortex por 10 segundos e incubação a 56°C e em agitação a 600 rpm por 3 minutos. Depois de resfriar em temperatura ambiente por 5 minutos, foram adicionados 180 μL de tampão ATL, seguido de agitação por vortex por 10 segundos e centrifugação a 10000 rpm por 1 minuto. A digestão foi realizada com a adição de 20 μL de proteinase K na fase inferior do tubo. A reação foi incubada a 56°C e em agitação a 600 rpm por cerca de 20 horas. Para as amostras cuja digestão não havia ocorrido completamente, foram adicionados mais 20 μL de proteinase K e a reação foi incubada a 56°C e em agitação máxima por 5 minutos. Em seguida, a reação foi incubada a 90°C por 1 hora, depois de resfriar em temperatura ambiente por 5 minutos foram adicionados 2 μL de RNase A (100 mg/mL) e a reação foi incubada a temperatura ambiente por 2 minutos. Em seguida, foram adicionados 200 μL de tampão AL, seguido de mais 200 μL de etanol 100%. O lisado foi transferido para uma coluna MinElute (Qiagen), a qual foi centrifugada a 8000 rpm por 1 minuto. A purificação teve início com a adição de 500 μL de tampão AW1 e centrifugação a 8000 rpm por 1 minuto, seguido da adição de 500 μL de tampão AW2 e centrifugação a 8000 rpm por 1 minuto, seguido da adição de 500 μL de tampão álcool 70% e centrifugação a 8000 rpm por 1 minuto. Em seguida, a coluna foi centrifugada em velocidade máxima por 3 minutos. A eluição seguiu com a adição de 20 ou 30 μL de TAE para material de biópsia ou peça cirúrgica, respectivamente. A coluna foi mantida em temperatura ambiente por 5 minutos e depois centrifugada em velocidade máxima por 1 minuto. O eluato foi quantificado através de espectrofotometria por Nanodrop (Thermo Fisher Scientific) e por fluorescência por Qubit (Thermo Fisher Scientific).

Em relação aos tecidos congelados, as áreas CDIS, referentes às áreas marcadas nas lâminas HE, foram submetidas a uma dissecação manual com o uso de um bisturi estéril (Figura 5), e posteriormente submetidas à extração de DNA com o uso do QIAamp DNA Blood Midi kit (Qiagen). Quando possível, as áreas correspondentes ao tecido normal foram removidas. Todos esses procedimentos foram realizados pela equipe do biobanco do ACCCC.

Além disso, 3 pacientes (14CDIS, 15CDIS e 16CDIS) foram selecionados de um estudo prévio do grupo em que o DNA já havia sido extraído de amostras que foram submetidas a *punches* de 2 mm de diâmetro (Figura 5). O DNA destas amostras foi extraído e purificado com o uso do kit QIAamp DNA FFPE Tissue Kit (Qiagen), seguindo as recomendações do fabricante.

Em relação a papa leucocitária, o DNA de 9 pacientes (11CDIS, 13CDIS, 14CDIS, 15CDIS, 16CDIS, 18CDIS, 19CDIS, 20CDIS e 27CDIS) da casuística foi submetido à extração com o uso dos kits QIAmp DNA Blood Mini QIAcube Kit (Qiagen) e Gentra Puregene Blood Kit (Qiagen), de acordo com as recomendações do fabricante. Todos esses procedimentos foram realizados pela equipe do biobanco do ACCCC.



Legenda: As amostras FFPE e de tecido congelado foram submetidas a scrape ou punch para coleta das áreas de CDIS marcadas nas lâminas HE para posterior extração de DNA. Abreviações: FFPE, amostra fixada em formalina e embebida em parafina; HE, hematoxilina e eosina.

Figura 5 - Método de coleta das áreas CDIS.

3.5.4 Critérios de controle de qualidade de DNA

Todas as amostras foram submetidas a uma avaliação de qualidade por espectrofotometria no equipamento Nanodrop 1000 (ThermoFisher Scientific). Consideramos como critério de qualidade as amostras que apresentaram uma razão de absorbância 260/280 próxima a 1,8 e uma razão de absorbância 260/230 próxima a 2,0. As amostras com qualidade insatisfatória foram repurificadas em coluna MinElute com o uso do kit QIAamp DNA Mini Kit (Qiagen). As amostras de DNA com qualidade foram submetidas a uma avaliação de quantidade por fluorescência no equipamento *Qubit* (Thermo Fisher Scientific). Em seguida, as amostras foram submetidas a uma análise de integridade do DNA no equipamento *4200 TapeStation* (Agilent Technologies) com o uso do ensaio *ScreenTape Genomic DNA*, considerando os valores de DIN (*DNA Integrity Number*), os quais variam de 1 a 10, em que quanto maior o valor de DIN melhor é a integridade do DNA. Esse último procedimento foi realizado pela equipe do biobanco do ACCCC.

3.5.5 Construção de bibliotecas para sequenciamento NGS com painéis multigenes

Todas as bibliotecas foram construídas utilizando o mesmo kit *Ion AmpliSeq library kit 2.0* (Thermo Fisher Scientific), seguindo as recomendações do fabricante. Inicialmente, submetemos as amostras a um tratamento com 1 μ L (1U/ μ L) da enzima Uracil-DNA-glicosilase (UDG) (Thermo Fisher Scientific), incubando o DNA a 37°C por 10 minutos e em seguida a 85°C por 5 minutos. Esse procedimento é realizado para diminuir a taxa de transições C:G > T:A, as quais correspondem a artefatos comuns em amostras fixadas em formalina e embebidas em parafina. Dados do grupo mostraram de forma objetiva que o método é eficiente para diminuir os artefatos observados nos resultados de NGS decorrentes de desaminação de citosina em uracila, ou seja, da 5-metilcitosina (5-mC) em timina (Berra et al. 2019). Em seguida, foi realizada uma reação de PCR *multiplex* para amplificação dos alvos dos painéis comerciais. Para o sequenciamento de DNA por NGS utilizamos 2 painéis multigenes: o painel *Ion AmpliSeq Comprehensive Cancer Panel* (CCP) (Thermo Fisher Scientific) e o painel *Ion AmpliSeq Cancer Hotspot Panel v2* (CHP2) (Thermo Fisher Scientific). Utilizamos o painel CCP para avaliar aquelas amostras que conseguimos obter uma quantidade mínima de 40 ng de DNA e o painel CHP2 para avaliar aquelas amostras que conseguimos obter uma quantidade de DNA que variou entre 10 e 40 ng.

Ambos os painéis são baseados numa reação de PCR multiplex, a qual corresponde a primeira etapa e é distinta para os painéis. O painel CCP é um painel amplo que possibilita a avaliação de 409 genes inteiros envolvidos com câncer (Quadro 2). Ele é constituído por mais de 16 mil pares de *primers*, os quais são divididos em 4 *pools* para amplificação *multiplex* de todos os exons dos 409 genes (Quadro 2). Para cada pool a reação de PCR multiplex foi realizada com 5x Ion AmpliSeq HiFi mix, 2x Ion AmpliSeq primer pool, 10 a 20 ng de DNA por *pool* de *primers*, num volume total de 10 µL, e foi incubada a 99°C por 2 minutos, seguido de 19 ciclos a 99°C por 15 segundos e a 60°C por 8 minutos. Em seguida, as reações dos 4 *pools* foram transferidas para um mesmo tubo, e após homogeneização, 20 µL foram transferidos para um novo tubo. O painel CHP2 é um painel que possibilita a avaliação de regiões *hotspots* mutacionais de 50 genes envolvidos com câncer (Quadro 2). Ele é constituído por 1 *pool* de *primers* para amplificação *multiplex* de 207 amplicons que cobrem aproximadamente 2800 mutações reportadas no Cosmic em regiões *hotspots* dos 50 genes (Quadro 2). Todos esses genes também são representados pelo painel CCP. Na construção das bibliotecas utilizando o painel CHP2, a reação de PCR multiplex foi realizada com 5x Ion AmpliSeq HiFi mix, 2x Ion AmpliSeq primer pool, 10 a 15 ng de DNA, num volume total de 20 µL, e foi incubada a 99°C por 2 minutos, seguido de 23 ciclos a 99°C por 15 segundos e a 60°C por 4 minutos.

Quadro 2 - Lista de genes dos painéis multigenes utilizados no estudo. Lista com 409 genes cobertos pelo painel CCP.

ABL1	BAP1	CCND2	CREBBP	EPHB1	FBXW7	GPR124	IRS2	MAF	MMP2	NFKB2	PDGFRA	PRKDC	SBDS	TAF1	TRIP11
ABL2	BCL10	CCNE1	CRKL	EPHB4	FGFR1	GRM8	ITGA10	MAFB	MN1	NIN	PDGFRB	PSIP1	SDHA	TAF1L	TRRAP
ACVR2A	BCL11A	CD79A	CRTC1	EPHB6	FGFR2	GUCY1A2	ITGA9	MAGEA1	MPL	NKX2-1	PER1	PTCH1	SDHB	TAL1	TSC1
ADAMTS20	BCL11B	CD79B	CSF1R	ERBB2	FGFR3	HCAR1	ITGB2	MAGI1	MRE11A	NLRP1	PGAP3	PTEN	SDHC	TBX22	TSC2
AFF1	BCL2	CDC73	CSMD3	ERBB3	FGFR4	HIF1A	ITGB3	MALT1	MSH2	NOTCH1	PHOX2B	PTGS2	SDHD	TCF12	TSHR
AFF3	BCL2L1	CDH1	CTNNA1	ERBB4	FH	HLF	JAK1	MAML2	MSH6	NOTCH2	PIK3C2B	PTPN11	SEPT9	TCF3	UBR5
AKAP9	BCL2L2	CDH11	CTNNA1	ERCC1	FLCN	HNF1A	JAK2	MAP2K1	MTOR	NOTCH4	PIK3CA	PTPRD	SETD2	TCF7L1	UGT1A1
AKT1	BCL3	CDH2	CYLD	ERCC2	FLI1	HOOK3	JAK3	MAP2K2	MTR	NPM1	PIK3CB	PTPRT	SF3B1	TCF7L2	USP9X
AKT2	BCL6	CDH20	CYP2C19	ERCC3	FLT1	HRAS	JUN	MAP2K4	MTRR	NRAS	PIK3CD	RAD50	SGK1	TCL1A	VHL
AKT3	BCL9	CDH5	CYP2D6	ERCC4	FLT3	HSP90AA1	KAT6A	MAP3K7	MUC1	NSD1	PIK3CG	RAF1	SH2D1A	TET1	WAS
ALK	BCR	CDK12	DAXX	ERCC5	FLT4	HSP90AB1	KAT6B	MAPK1	MUTYH	NTRK1	PIK3R1	RALGDS	SMAD2	TET2	WHSC1
APC	BIRC2	CDK4	DCC	ERG	FN1	ICK	KDM5C	MAPK8	MYB	NTRK3	PIK3R2	RARA	SMAD4	TFE3	WRN
AR	BIRC3	CDK6	DDB2	ESR1	FOXL2	IDH1	KDM6A	MARK1	MYC	NUMA1	PIM1	RB1	SMARCA4	TGFBR2	WT1
ARID1A	BIRC5	CDK8	DDIT3	ETS1	FOXO1	IDH2	KDR	MARK4	MYCL1	NUP214	PKHD1	RECQL4	SMARCB1	TGM7	XPA
ARID2	BLM	CDKN2A	DDR2	ETV1	FOXO3	IGF1R	KEAP1	MBD1	MYCN	NUP98	PLAG1	REL	SMO	THBS1	XPC
ARNT	BLNK	CDKN2B	DEK	ETV4	FOXP1	IGF2	KIT	MCL1	MYD88	PAK3	PLCG1	RET	SMUG1	TIMP3	XPO1
ASXL1	BMPR1A	CDKN2C	DICER1	EXT1	FOXP4	IGF2R	KLF6	MDM2	MYH11	PALB2	PLEKHG5	RHOH	SOC31	TLR4	XRCC2
ATF1	BRAF	CEBPA	DNMT3A	EXT2	FZR1	IKBKB	KRAS	MDM4	MYH9	PARP1	PML	RNASEL	SOX11	TLX1	ZNF384
ATM	BRD3	CHEK1	DPYD	EZH2	G6PD	IKBKE	LAMP1	MEN1	NBN	PAX3	PMS1	RNF2	SOX2	TNFAIP3	ZNF521
ATR	BRIP1	CHEK2	DST	FAM123B	GATA1	IKZF1	LCK	MET	NCOA1	PAX5	PMS2	RNF213	SRC	TNFRSF14	
ATRX	BTB	CIC	EGFR	FANCA	GATA2	IL2	LIFR	MITF	NCOA2	PAX7	POT1	ROS1	SSX1	TNK2	
AURKA	BUB1B	CKS1B	EML4	FANCC	GATA3	IL21R	LPHN3	MLH1	NCOA4	PAX8	POU5F1	RPS6KA2	STK11	TOP1	
AURKB	CARD11	CMPK1	EP300	FANCD2	GDNF	IL6ST	LPP	MLL	NF1	PBRM1	PPARG	RRM1	STK36	TP53	
AURKC	CASC5	COL1A1	EP400	FANCF	GNA11	IL7R	LRP1B	MLL2	NF2	PBX1	PPP2R1A	RUNX1	SUFU	TPR	
AXL	CBL	CRBN	EPHA3	FANCG	GNAQ	ING4	LTF	MLL3	NFE2L2	PDE4DIP	PRDM1	RUNX1T1	SYK	TRIM24	
BAI3	CCND1	CREB1	EPHA7	FAS	GNAS	IRF4	LTK	MLLT10	NFKB1	PDGFB	PRKAR1A	SAMD9	SYNE1	TRIM33	

Legenda: Os genes em negrito referem-se aos genes contidos no painel CHP2 que cobrem hotspots mutacionais dos respectivos genes. O painel *Ion AmpliSeq Comprehensive Cancer Panel* (CCP) é um painel multigenes amplo que avalia 409 genes inteiros. O painel *Ion AmpliSeq Cancer Hotspot Panel v2* (CHP2) é um painel multigenes que avalia regiões *hotspots* de 50 genes, os quais estão destacados em negrito.

Em seguida, foram realizadas as etapas de digestão parcial dos amplicons, ligação de adaptadores e barcodes e purificação das reações. Estas etapas foram iguais para ambos os painéis. Os alvos amplificados por ambos os painéis foram submetidos a uma reação de digestão parcial dos amplicons com a adição de 2 μ L do reagente FuPa. A reação foi incubada a 50°C por 10 minutos, 55°C por 10 minutos, 60°C por 20 minutos. Após a digestão foi realizada uma reação de ligação de adaptadores e *barcodes* nos amplicons, com a adição de 4 μ L de *switch solution*, 2 μ L de um mix previamente feito de adaptadores e *barcodes* (2 μ L de adaptador Ion P1, 2 μ L de *Ion Xpress barcode* e 4 μ L de água), 2 μ L da enzima DNA ligase. Esta reação foi incubada a 22°C por 30 minutos, 68°C por 5 minutos, 72°C por 5 minutos. Os 30 μ L de reação foram transferidos para um novo tubo para purificação, ao qual foram adicionados 45 μ L de beads magnéticas, com a posterior homogeneização e incubação em temperatura ambiente por 5 minutos. A reação foi mantida em rack magnética por 5 minutos com posterior descarte do sobrenadante e duas lavagens com 200 μ L de álcool 70%. As *beads* foram mantidas em temperatura ambiente por 5 minutos para secagem. A eluição foi realizada com a adição de 52 μ L de PCR mix (50 μ L de Platinum PCR SuperMix HiFi e 2 μ L de *library amplification primer mix*) e incubação em temperatura ambiente por 5 minutos. A reação foi mantida em rack magnética por 5 minutos com posterior transferência do sobrenadante para um novo tubo e incubação a 98°C por 2 minutos, seguido de 9 ciclos a 98°C por 15 segundos e 60°C por 1 minuto. Uma nova purificação foi realizada com a transferência dos 52 μ L da reação para um novo tubo, ao qual foram adicionados 25 μ L de *beads* magnéticas com a posterior homogeneização e incubação em temperatura ambiente por 5 minutos. A reação foi mantida em rack magnética por 5 minutos com posterior transferência do sobrenadante para um novo tubo, ao qual foram adicionados 60 μ L de *beads* magnéticas com a posterior homogeneização e incubação em temperatura ambiente por 5 minutos. A reação foi mantida em rack magnética por 5 minutos com posterior descarte do sobrenadante e duas lavagens com álcool 70%. As *beads* foram eluídas com a adição de 30 μ L de TE e posterior incubação em temperatura ambiente por 5 minutos. O eluato foi mantido em rack magnética por 5 minutos e o sobrenadante foi transferido para um novo tubo. Por fim, as bibliotecas construídas foram quantificadas por reação de PCR em tempo real com o uso do *Ion Library Quantification kit* (Thermo Fisher Scientific).

Com a finalidade de filtrar variantes germinativas e selecionarmos apenas as variantes somáticas (presentes apenas no tumor), o DNA de leucócito de 9 pacientes da casuística (11CDIS, 13CDIS, 14CDIS, 15CDIS, 16CDIS, 18CDIS, 19CDIS, 20CDIS e 27CDIS) também foi submetido a sequenciamento por NGS. O DNA dos 9 pacientes foi dividido em 4 *pools*, cada um contendo quantidades equimolares de DNA. Esta estratégia foi adotada porque mantém a capacidade de detecção de variantes germinativas e ao mesmo tempo possibilita uma redução nos custos. Desta forma, 3 *pools* tiveram o DNA de leucócito de 2 pacientes e 1 *pool* teve o DNA de leucócito de 3 pacientes. Nos *pools* com 2 pacientes é esperado que a VAF de uma variante germinativa em heterozigose seja de aproximadamente 25%, e no *pool* com 3 pacientes de aproximadamente 17%. Submetemos 20 ng de cada *pool* para a construção das bibliotecas com o painel CCP, conforme descrito anteriormente. Foram descartadas as variantes identificadas nas amostras tumorais e que também foram identificadas no *pool* de leucócito correspondente, além das variantes detectadas em qualquer um dos 4 *pools* de leucócitos, sendo do mesmo paciente ou não.

3.5.6 Sequenciamento de DNA

Utilizamos a tecnologia de sequenciamento de nova geração (NGS – *Next Generation Sequencing*) através da plataforma Ion GeneStudio S5 Plus System e Ion Proton (Thermo Fisher Scientific). Após a construção das bibliotecas de DNA com os painéis CCP e CHP2, realizou-se uma reação de amplificação dos alvos através de PCR em emulsão (emPCR) com o uso do kit Ion 540-OT2 e Ion PI Template OT2 v3 (Thermo Fisher Scientific), em que os fragmentos de DNA são ligados à superfície de beads microscópicas. Idealmente, apenas um fragmento de DNA é capturado por *bead* e amplificado individualmente milhares de vezes durante a emPCR. Em seguida, foram selecionados apenas as *beads* carregadas com material amplificado pelo processo de enriquecimento, para que em seguida a reação fosse submetida ao sequenciamento. Para o sequenciamento foi utilizado o kit Ion 540-OT2 e Ion PI Sequencing 200 v3 (Thermo Fisher Scientific), de acordo com as recomendações do fabricante, que permite o sequenciamento de fragmentos de aproximadamente 200 pb. A reação de sequenciamento foi realizada nos chips Ion 540 e Ion Pi v3, que são capazes de gerar uma quantidade de dados de 10 a 15 Gb cada um, o que representa a geração de 60 a 80 milhões de leituras.

3.5.7 Sequenciamento de amplicon

Foram selecionadas para validação as variantes potencialmente somáticas (identificadas no DNA tumoral) que atenderam aos seguintes critérios: variantes identificadas em baixa frequência alélica na amostra tumoral; variantes cujas posições genômicas foram cobertas nas duas amostras tumorais de uma mesma paciente (biópsia e peça cirúrgica), mas que foram identificadas em apenas uma das amostras tumorais (biópsia ou peça cirúrgica); variantes identificadas no DNA tumoral com frequência alélica próxima a 50%, cujo DNA de leucócito da paciente estava disponível para ser avaliado para confirmar se a variante era somática ou germinativa; variantes em regiões *hotspots* dos genes mais frequentemente alterados em câncer de mama: *PIK3CA* e *TP53*.

O programa *Primer3* (versão 4.1.0) (Untergasser et al. 2012) foi utilizado para o desenho de *primers* capazes de amplificar as regiões genômicas dessas variantes. As reações de PCR seguiram as seguintes condições: 95°C por 5 minutos, 40 ciclos de 95°C por 30 segundos, 60-62°C por 90 segundos e 72°C por 30 segundos, 68°C por 10 minutos, num volume total de 25 µL.

Em seguida, 5 µL dos produtos das reações de PCR foram submetidos a eletroforese em gel de agarose 1% para checagem da existência de amplicons únicos e com o tamanho esperado. O restante de 20 µL dos produtos das reações de PCR foram transferidos para um novo tubo para purificação, ao qual foram adicionados 36 µL de *beads* magnéticas, com a posterior homogeneização e incubação em temperatura ambiente por 5 minutos. A reação foi mantida em rack magnética por 5 minutos com posterior descarte do sobrenadante e duas lavagens com 200 µL de álcool 70%. As *beads* foram mantidas em temperatura ambiente por 5 minutos para secagem. A eluição foi realizada com a adição de 30 µL de água.

Após quantificação por fluorescência no equipamento *Qubit* (Thermo Fisher Scientific), 10-100ng do produto de PCR purificado com *beads* magnéticas foram usados para preparo de biblioteca de DNA usando o kit *Ion Plus Fragment Library* (Thermo Fisher), seguindo as recomendações do fabricante.

3.5.8 Análise dos dados de sequenciamento de DNA para seleção das variantes potencialmente somáticas

Os critérios de qualidade de sequenciamento assumidos foram os parâmetros padrão já configurados pelo fabricante nos sequenciadores Ion Proton e Ion S5, através do software Torrent Suite (versões 5.12.3 e 5.10.1).

As amostras consideradas com qualidade satisfatória de sequenciamento foram as amostras que tiveram os painéis multigenes com percentual de leituras no alvo > 50%, ou seja, a maioria das leituras foram mapeadas nos alvos (regiões *hotspot* dos 50 genes do painel CHP2 e regiões dos exons dos 409 genes do painel CCP); e cobertura vertical média $\geq 500\times$, ou seja, a média de sequenciamento das bases alvo foi de pelo menos 500 leituras.

O mapeamento e alinhamento das sequências no genoma referência (GRCh37/hg19) foram realizadas pelo programa *Torrent Suite* (versões 5.12.3 e 5.10.1) (Thermo Fisher Scientific), seguindo os parâmetros recomendados pelo fabricante.

A chamada de variantes foi realizada pelo programa *Torrent Variant Caller* (TVC) (Thermo Fisher Scientific), um *plugin* do *Torrent Suite*, seguindo os parâmetros recomendados pelo fabricante.

A seleção das variantes foi realizada com o uso do programa VarSeq (versão 2.2.5) (Golden Helix), após a importação dos arquivos VCF (*Variant Call Format*) gerados pelo TVC.

As variantes consideradas com qualidade de chamada e potencialmente somáticas (adquiridas pelo tumor) foram selecionadas no VarSeq de acordo com os seguintes critérios: cobertura de base > 100; qualidade do genótipo identificado na chamada ≥ 100 (probabilidade em escala *Phred* do genótipo chamado); quantidade de leituras do alelo variante identificados na fita senso ≥ 3 ; quantidade de leituras do alelo variante identificados na fita antisenso ≥ 3 ; tamanho aceito de homopolímero na região da variante chamada ≤ 4 ; número de observações do alelo variante naquela região em todas as amostras < 5; frequência do alelo variante nas amostras de leucócito < 0,3% ou não identificado; frequência do alelo variante no dbSNP < 0,5% ou informação ausente; variante em sítio de *splice* (≤ -2 ; $\leq +2$); variante em região codificante do tipo não sinônima com troca de aminoácido ou perda de função proteica. Nas amostras cujos painéis tiveram uma qualidade satisfatória de sequenciamento, numa primeira abordagem selecionamos variantes com frequência do alelo variante nas amostras tumorais $\geq 5\%$. Numa segunda abordagem realizamos uma busca ativa (inspeção visual) por variantes

em oncogenes e genes supressores tumorais que são mais frequentemente alterados em tumores sólidos, principalmente em câncer de mama, e selecionamos variantes em baixa frequência alélica (frequência do alelo variante nas amostras tumorais > 1%) e variantes com baixa cobertura de sequenciamento (cobertura > 40x) em oncogenes e genes supressores tumorais. Nas amostras que não passaram pelos nossos critérios de qualidade (amostras que não tiveram percentual de leituras no alvo > 50% e cobertura vertical média $\geq 500x$), numa terceira abordagem foi realizada uma busca ativa por variantes em oncogenes e genes supressores tumorais com frequência alélica $\geq 5\%$.

3.5.9 Anotação das variantes potencialmente oncogênicas

Para selecionar as variantes potencialmente oncogênicas, os bancos de dados consultados foram: *National Center for Biotechnology Information Reference Sequence Database* (RefSeq), um banco de dados de sequência referência, incluindo sequências genômicas, transcritos e proteínas; *Catalogue of Somatic Mutations in Cancer* (Cosmic), um banco de dados de mutações somáticas em câncer; *Single Nucleotide Polymorphism Database* (dbSNP) (Sherry et al. 1999), um banco de dados que agrega informações de variantes genéticas comuns identificadas em diversos estudos populacionais; *ClinVar*, um banco de dados de variantes de sequência genômica majoritariamente germinativas; e *Varsome* (Kopanos et al. 2019), uma ferramenta de busca e agregação de dados e análise do efeito de variantes genômicas. Além disso, o preditor de patogenicidade de variantes missense *Rare Exome Variant Ensemble Learner* (REVEL) (Ioannidis et al. 2016) também foi consultado. Esta ferramenta integra os escores dos preditores MutPred, FATHMM v2.3, VEST 3.0, PolyPhen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP, and phastCons. Os escores do REVEL são baseados no projeto *Database of human non-synonymous SNPs* (dbNSFP) (Liu et al. 2016) e variam de 0 a 1, sendo que acima de 0,5 a variante é considerada como provavelmente causadora de doença.

Além disso, utilizamos a plataforma *Cancer-Related Analysis of Variants Toolkit* (CRAVAT) (Douville et al. 2013; Masica et al. 2017), a qual possibilita dentre outras funcionalidades a aplicação do algoritmo *Cancer-specific High-throughput Annotation of Somatic Mutations* (CHASM-plus) (Tokheim e Karchin 2019) para a predição de variantes somáticas missense nas categorias *driver* ou *passenger*. Utilizamos o algoritmo específico para câncer de mama.

Em seguida, a anotação seguiu com uma consulta manual na plataforma *cBioPortal for Cancer Genomics* (cBioPortal) (Cerami et al. 2012; Gao et al. 2013), uma plataforma que reúne inúmeras informações de diversos estudos mutacionais de larga escala e que possibilita identificar quais variantes do nosso estudo já foram reportadas nesses estudos. Atualmente a plataforma reúne 23 estudos mutacionais em câncer de mama, porém existe certa sobreposição de amostras avaliadas entre todos esses estudos. Então, desconsiderando esses estudos com sobreposição de amostras, as variantes do nosso estudo foram investigadas em 18 estudos mutacionais de larga escala, que somados avaliaram 7963 tumores de mama: (1) *Breast Cancer-MSK* (Razavi et al. 2018); (2) *Breast Cancer-MSK* (Razavi et al. 2020); (3) *Metastatic Breast Cancer-MSK* (Li et al. 2022); (4) *Breast Fibroepithelial Tumors-Duke-NUS* (Tan et al. 2015); (5) *Breast Cancer-METABRIC* (Curtis et al. 2012; Pereira et al. 2016); (6) *Breast Cancer-MSK* (Pareja et al. 2020); (7) *Breast Cancer-MSKCC* (Nixon et al. 2019); (8) *Breast Cancer-SMC* (Kan et al. 2018); (9) *Breast Cancer Xenografts* (Eirew et al. 2015); (10) *Breast Invasive Carcinoma* (Shah et al. 2012); (11) *Breast Invasive Carcinoma* (Banerji et al. 2012); (12) *Breast Invasive Carcinoma* (Stephens et al. 2012); (13) *Breast Invasive Carcinoma* (Ciriello et al. 2015); (14) *Juvenile Papillomatosis and Breast Cancer-MSK* (não publicado); (15) *Metastatic Breast Cancer-INSERM* (Lefebvre et al. 2016); (16) *The Metastatic Breast Cancer Project* (não publicado); (17) *Adenoid Cystic Carcinoma of the Breast-MSKCC* (Martelotto et al. 2015); (18) *Metaplastic Breast Cancer-MSK* (não publicado). Nesta plataforma também foi possível visualizar a localização da variante na sequência da proteína, e com isso identificar se a variante ocorria dentro ou fora de algum sítio proteico. Além disso, também foi possível identificar se a variante ocorria em uma região *hotspot*. As regiões consideradas pela plataforma cBioPortal como sendo regiões *hotspot* de mutações recorrentes e estatisticamente significativas em câncer são baseadas nos trabalhos de Chang e colaboradores 2016, 2018 e Gao e colaboradores 2017.

A plataforma *Precision Oncology Knowledge Base* (OncoKB) (Chakravarty et al. 2017) também foi consultada para identificar as variantes reportadas como oncogênicas e provavelmente oncogênicas, de acordo com os níveis de evidência científica da literatura. O OncoKB é uma plataforma criada e mantida pelo *Memorial Sloan Kettering Cancer Center* (MSKCC) e agrega dados mutacionais de 1066 genes, os quais são considerados pelo OncoKB como genes relacionados ao câncer, por terem sido incluídos em diversos painéis de diversos estudos de larga escala e por fazerem parte das listas *Cancer Gene Census* elaborada pelo

Wellcome Sanger Institute (Sondka et al. 2018) e Vogelstein e colaboradores (2013). Destes, o OncoKB apresenta dados mutacionais altamente curados de 5697 variantes em 685 genes de 127 tumores, e classifica como oncogênicas e provavelmente oncogênicas aquelas variantes que se tem fortes evidências e evidências sugestivas, respectivamente, de que a alteração tenha efeito oncogênico.

Em seguida, através do programa *Integrative Genomics Viewer* (IGV) (versão 5.0) (Robinson et al. 2011), foi realizada uma inspeção visual do alinhamento das sequências geradas nas amostras de leucócito para identificação e exclusão de variantes germinativas que não foram chamadas e consequentemente não foram filtradas pelo programa VarSeq. O mesmo procedimento foi realizado nas amostras tumorais com o intuito de identificar e excluir artefatos de sequenciamento.

A classificação dos genes nos quais as variantes únicas foram identificadas foi baseada nas classificações obtidas nas plataformas Cosmic e OncoKB. Foram identificados genes associados ao câncer, oncogenes, genes supressores tumorais e genes com duplo papel funcional relacionado ao câncer (oncogenes/supressores tumorais). Foram anotados todos os genes que são classificados desta forma por qualquer uma das duas plataformas.

Para evitar a possibilidade de não selecionarmos variantes verdadeiras presentes em baixa frequência alélica (VAF entre 1% e 5%) ou com uma baixa cobertura de sequenciamento mas que poderiam ter efeito oncogênico e relevância na biologia tumoral, nós fizemos uma busca dessas variantes no banco de dados Cosmic, em outros estudos que avaliaram câncer de mama, e se seriam classificadas como *driver* em câncer de mama pelo algoritmo CHASMplus que avalia variantes somáticas.

3.5.10 Análise de associação de mutações oncogênicas e provavelmente oncogênicas potencialmente somáticas com os grupos de CDIS

A análise de significância estatística da associação entre a quantidade de mutações oncogênicas e provavelmente oncogênicas potencialmente somáticas com o grupo CDIS sem progressão e CDIS com progressão foram realizados pelo programa *GraphPad Prism* (<https://www.graphpad.com/scientific-software/prism/>) (versão 9.0). O teste estatístico utilizado foi o “Teste exato de Fisher” e a significância estatística foi considerada com $P < 0,05$.

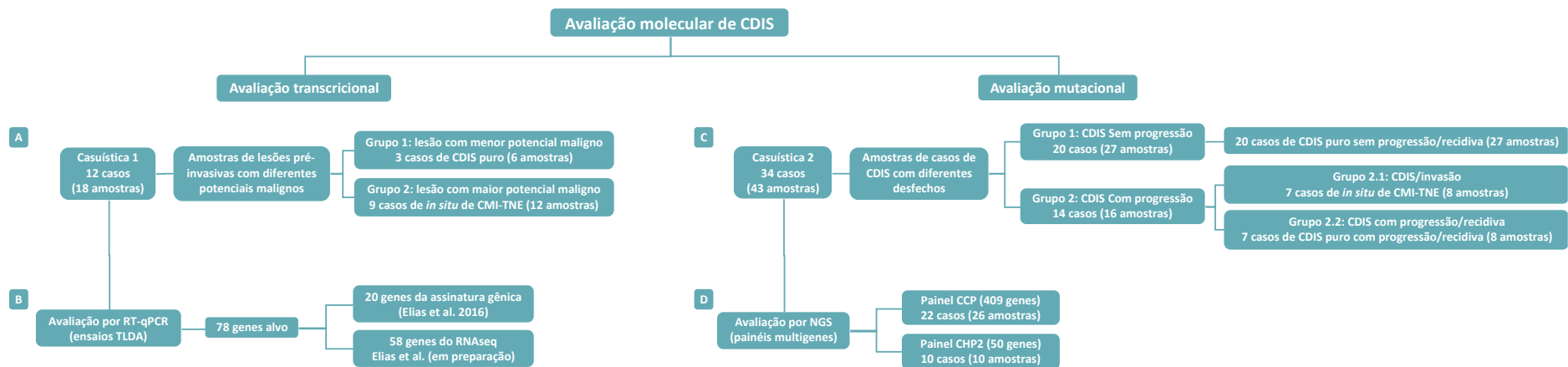
3.5.11 Análise de enriquecimento funcional

Os oncogenes, genes supressores tumorais e oncogenes/supressores acometidos por variantes foram submetidos a uma análise de enriquecimento funcional através da ferramenta STRING versão 11.5 (Szklarczyk et al. 2019) (<https://www.string-db.org/>) para avaliarmos a existência de vias bioquímicas canônicas enriquecidas. A plataforma STRING analisa interações proteicas diretas e indiretas e utiliza o banco de dados *Kyoto Encyclopedia of Genes and Genomes* (KEGG) para fazer a anotação das vias enriquecidas pelas proteínas ou genes. O critério utilizado para considerarmos enriquecimento de vias foi escore de enriquecimento $> 1,0$ e significância estatística por FDR $< 0,05$. O escore de enriquecimento é a medida que avalia o tamanho do efeito de enriquecimento e é calculado através da fórmula $\log_{10}(\text{observado/esperado})$. O parâmetro “observado” corresponde ao número de vezes que um determinado gene é descrito na rede de interação (ou via bioquímica), e “esperado” corresponde ao número de vezes que um determinado gene é descrito numa rede de interação aleatória e de mesmo tamanho (número de genes).

4. RESULTADOS

4.1 Avaliação molecular de CDIS

Foi realizada uma avaliação molecular de lesões pré-invasivas, as quais foram representadas por amostras de CDIS que apresentavam diferentes potenciais malignos. Esta avaliação foi realizada através de duas abordagens diferentes, uma análise transcricional em que se realizou uma análise de expressão gênica por RT-qPCR (ensaios TLDA) e uma análise mutacional em que se realizou o sequenciamento de DNA tumoral por NGS com o uso de 2 painéis multigenes (um painel que avaliou 409 genes inteiros e um painel que avaliou regiões *hotspots* de 50 genes). As duas abordagens foram utilizadas na avaliação de duas casuísticas distintas de CDIS. Na análise transcricional foi avaliada uma casuística composta essencialmente por lesões pré-invasivas com diferentes potenciais malignos (CDIS puro e componente *in situ* de lesão invasiva (CMI-TNE), a qual foi dividida em 2 grupos, o grupo 1 composto por amostras com menor potencial maligno (CDIS puro), e o grupo 2 composto por amostras com maior potencial maligno (componente *in situ* de CMI-TNE). Na análise mutacional foi avaliada uma casuística composta essencialmente por pacientes de CDIS puro que apresentaram diferentes desfechos. Esta casuística foi dividida em 2 grupos, o grupo 1 (denominado CDIS sem progressão) composto por CDIS puro sem evento de progressão/recidiva em pelo menos 3 anos após cirurgia, e o grupo 2 (denominado CDIS com progressão) composto por pacientes de CDIS puro na biópsia e presença de invasão na peça cirúrgica (Grupo 2.1 – CDIS/invasivo) ou CDIS na biópsia e peça cirúrgica, com posterior evento de progressão/recidiva durante o seguimento (Grupo 2.2 – CDIS/progressão/recidiva). A Figura 6 apresenta de forma resumida o delineamento experimental desta avaliação molecular de CDIS.



Legenda: A avaliação molecular de CDIS composta de avaliação transcricional e avaliação mutacional. (A) Casuística 1: 12 pacientes com lesões pré-invasivas com diferentes potenciais malignos representados por 18 amostras, sendo o grupo 1 composto por lesões pré-invasivas com menor potencial maligno (CDIS puro) com 3 pacientes de CDIS puro representados por 6 amostras (3 amostras de células epiteliais neoplásicas e 3 amostras de células mioepiteliais capturadas por microdissecção a laser) e o grupo 2 composto por lesões pré-invasivas com maior potencial maligno (componente *in situ* de CMI-TNE) com 9 pacientes de componente *in situ* de lesão invasiva CMI-TNE representados por 12 amostras (9 amostras de células epiteliais neoplásicas e 3 amostras de células mioepiteliais capturadas por microdissecção a laser). (B) Ensaio TLDA com 78 genes-alvo. (C) Casuística 2: 34 pacientes de CDIS com diferentes desfechos representados por 43 amostras. Esta casuística foi dividida em 2 grupos, o grupo 1 (denominado CDIS sem progressão) com 20 pacientes representados por 27 amostras de CDIS puro sem progressão da doença no período de seguimento, e o grupo 2 (denominado CDIS com progressão), com 14 pacientes representados por 16 amostras, o qual foi dividido em 2 subgrupos, o grupo 2.1 (CDIS/invasivo) com 7 pacientes representados por 8 amostras de componente *in situ* de lesão invasiva CMI-TNE e o grupo 2.2 (CDIS/progressão/recidiva) com 7 pacientes representados por 8 amostras de CDIS puro com progressão ou recidiva da doença no período de seguimento. (D) Painéis NGS – CCP (painel amplo que possibilita a avaliação de 409 genes inteiros) e CHP2 (painel que possibilita a avaliação de regiões *hotspots* de 50 genes).

Figura 6 - Delineamento experimental do estudo.

4.2 Avaliação transcricional de CDIS

4.2.1 Casuística selecionada

Um total de 12 pacientes, representados por 18 amostras, foram selecionadas de estudos prévios do nosso grupo (Castro et al. 2008; Pinilla-Fernández 2015; Elias et al. em preparação). Essas 12 pacientes foram representadas por lesões pré-invasivas com diferentes potenciais malignos (CDIS puro e componente *in situ* de lesão invasiva CMI-TNE). Das 12 pacientes, 3 pacientes são de CDIS puro, as quais foram representados por 6 amostras (3 amostras de células epiteliais neoplásicas e 3 amostras de células mioepiteliais capturadas por microdissecção a laser), e 9 pacientes são de componente *in situ* de CMI-TNE, as quais foram representadas por 12 amostras (9 amostras de células epiteliais neoplásicas e 3 amostras de células mioepiteliais capturadas por microdissecção a laser). A avaliação das células mioepiteliais representou uma validação de um estudo anterior do grupo (Elias et al., em preparação). As principais características clínico-patológicas desta casuística estão apresentadas na Tabela 1.

Tabela 1 - Principais características clínico-patológicas da casuística submetida a avaliação transcricional.

Grupo	Identificação do caso	Idade ao diagnóstico	Identificação da amostra	Tipo celular avaliado	Grau nuclear	ER	PR	HER2
CDIS puro	1	42	27-E	epitelial	3	+	+	-
			27-M	mioepitelial				
	2	44	44-E	epitelial	3	+	+	-
			44-M	mioepitelial				
	3	67	60-E	epitelial	3	+	+	-
			60-M	mioepitelial				
<i>in situ</i> de CMI-TNE	4	65	196-E	epitelial	3	+	+	+
			196-M	mioepitelial				
	5	42	209-E	epitelial	3	+	+	-
			209-M	mioepitelial				
	6	71	240-E	epitelial	3	+	+	-
			240-M	mioepitelial				
	7	49	8N0-E	epitelial	NA	+	+	-
	8	47	15-E	epitelial	2	+	+	+
	9	55	45-E	epitelial	3	-	-	+
	10	34	25-E	epitelial	3	+	+	+
	11	43	51-E	epitelial	3	-	-	+
	12	31	81-E	epitelial	3	-	-	+

Legenda: A nomenclatura utilizada na identificação das amostras apresenta o sufixo “E” nas amostras que tiveram as células epiteliais neoplásicas capturadas por microdissecção a laser e o sufixo “M” nas amostras que tiveram as células mioepiteliais capturadas por microdissecção a laser. Abreviações: CDIS, carcinoma ductal *in situ*; *in situ* de CMI-TNE, componente *in situ* de carcinoma mamário invasivo tipo não-especial; ER, receptor de estrógeno; PR, receptor de progesterona; HER2, receptor 2 do fator de crescimento epidérmico humano; NA, não avaliado.

4.2.2 Análise das diferenças de expressão gênica entre CDIS puro e componente *in situ* de CMI-TNE

Para a análise de expressão gênica dos dois grupos de células epiteliais neoplásicas e mioepiteliais capturadas de lesões pré-invasivas com diferentes potenciais malignos avaliamos 12 pacientes: 3 pacientes de CDIS puro e 9 pacientes de componente *in situ* de CMI-TNE. Além disso, amostras de células epiteliais neoplásicas capturadas por microdissecção a laser de outros 16 pacientes (2 pacientes de tecido de mama normal e 14 pacientes de carcinoma de mama invasivo tipo não-especial) também foram submetidas a avaliação da expressão gênica para aumentarem o grupo de amostras e contribuírem na avaliação dos

genes referência, a fim de selecionarmos de maneira mais adequada os genes candidatos a normalizadores da expressão gênica.

Foram realizados 3196 ensaios de expressão gênica, considerando a avaliação de 94 sondas em 34 amostras. Do total de 34 amostras submetidas a avaliação da expressão gênica, 3 amostras (27-M, 240-E e 240-M) foram excluídas do estudo por não terem passado nos critérios de qualidade. Desta forma, as 31 amostras que passaram pelos critérios de qualidade foram submetidas a avaliação de 6 genes referência, dos quais 3 genes (*GAPDH*, *GUSB* e *RPLP0*) foram identificados pelo geNorm como genes de expressão mais estável e foram utilizados para normalizar a expressão dos genes alvo.

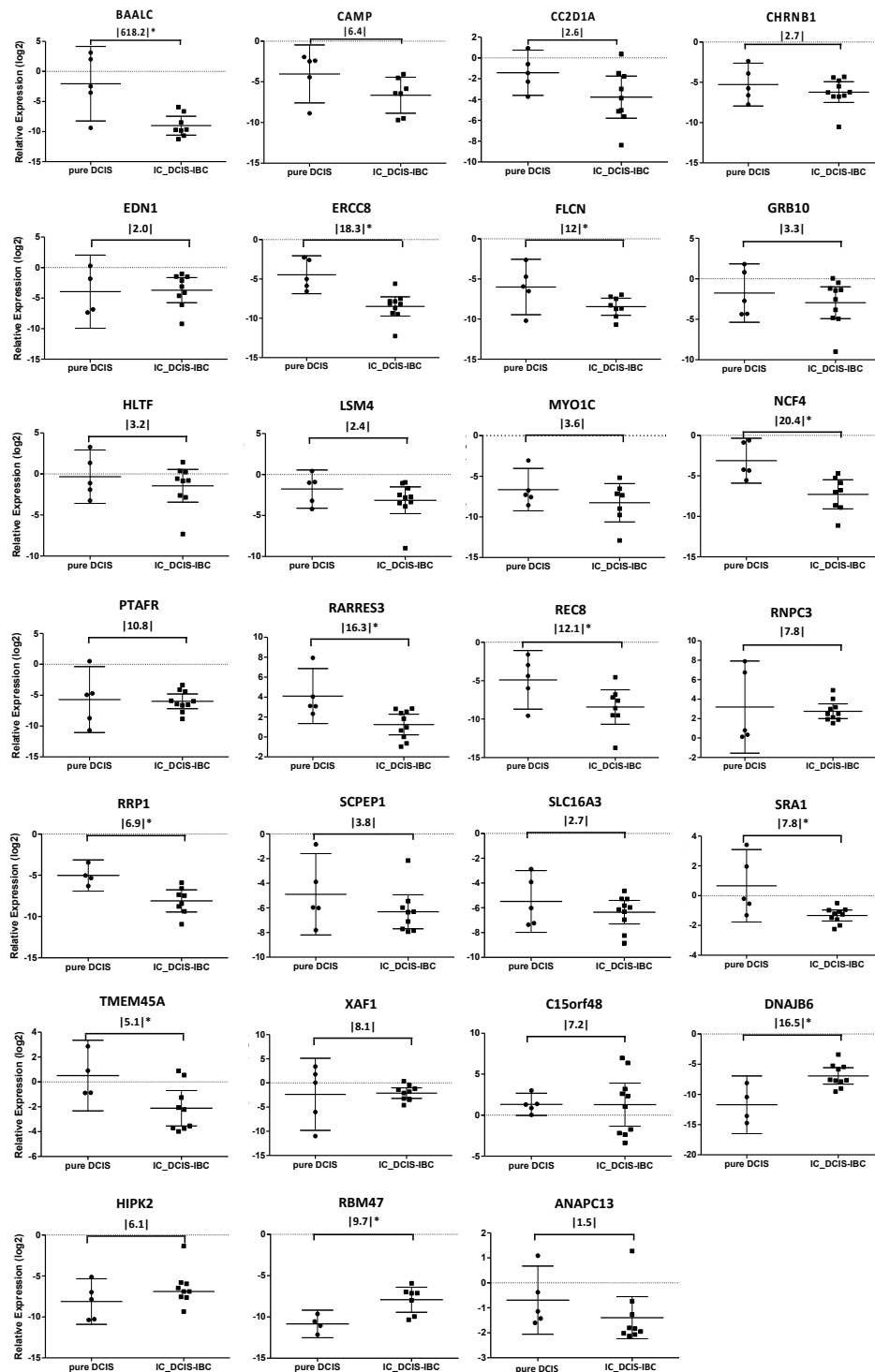
Dos 78 genes alvo avaliados, 53 genes (*ALMS1*, *ANAPC13*, *ARHGAP9*, *AZGP1*, *CAMP*, *CHRNA1*, *CPNE3*, *CTTNBP2NL*, *EDN1*, *GRB10*, *HLTF*, *INPP1*, *LSM4*, *MAPK8*, *P4HB*, *RABEPK*, *RARRES3*, *REC8*, *UTP20*, *BAALC*, *CC2D1A*, *CXCL10*, *DEGS1*, *ERCC8*, *FLCN*, *HAT1*, *LIMS1*, *MYO1C*, *NCF4*, *PLAUR*, *POLR3A*, *PTAFR*, *RNPC3*, *ROBO1*, *RRP1*, *S100A9*, *SCPEP1*, *SLC16A3*, *SRA1*, *SRRM1*, *ST6GALNAC4*, *TMEM45A*, *XAF1*, *ZBP1*, *ABCA1*, *C15orf48*, *DNAJA3*, *DNAJB6*, *HIPK2*, *MYB*, *NDFIP1*, *RBM47* e *SACM1L*) passaram pelos critérios de qualidade e puderam ser avaliados em 11 pacientes representados por 15 amostras (5 amostras de CDIS puro (27-E, 44-E, 44-M, 60-E e 60-M) e 10 amostras de componente *in situ* de CMI-TNE (8NO-E, 15-E, 25-E, 45-E, 51-E, 81-E, 196-E, 196-M, 209-E e 209-M). Do total de 795 ensaios, considerando a avaliação de 53 sondas em 15 amostras, 67 ensaios não passaram pelos critérios de qualidade e foram excluídos das análises, e desta forma, o restante de 728 ensaios (92%) que passaram pelos critérios de qualidade foram considerados nas análises de diferença de expressão gênica.

Desses 53 genes analisados, confirmamos a diferença de expressão nas células epiteliais neoplásicas e mioepiteliais capturadas dos dois grupos de amostras - CDIS puro e componente *in situ* de CMI-TNE de 49% (26/53) deles ($\text{fold change} \geq |2|$) (*BAALC*, *CAMP*, *CC2D1A*, *CHRNA1*, *C15orf48*, *DNAJB6*, *EDN1*, *ERCC8*, *FLCN*, *GRB10*, *HIPK2*, *HLTF*, *LSM4*, *MYO1C*, *NCF4*, *PTAFR*, *RARRES3*, *RBM47*, *REC8*, *RNPC3*, *RRP1*, *SCPEP1*, *SLC16A3*, *SRA1*, *TMEM45A* e *XAF1*) (Figura 7). Além disso, o gene *ANAPC13* que apresentou uma diferença de expressão ($\text{fold change} \geq |1,5|$) abaixo do critério utilizado foi selecionado para as análises posteriores de clusterização pelo fato de termos dados prévios demonstrando o papel desse gene na progressão do câncer de mama (Nakahata et al., em preparação). No gráfico de expressão do gene *ANAPC13* observa-se que no grupo de amostras de células epiteliais neoplásicas do

componente *in situ* de CMI-TNE a maior parte das amostras apresentou um nível de expressão mais baixo, e uma das amostras apresentou um nível de expressão mais alto e que deve ter interferido no valor médio de expressão, que acabou não alcançando o critério adotado (Figura 7). No entanto, dados anteriores do grupo demonstram que a diminuição no nível de expressão de *ANAPC13* deve ter um papel importante na progressão do câncer de mama.

Dos 26 genes validados, 11 genes (*BAALC*, *DNAJB6*, *ERCC8*, *FLCN*, *NCF4*, *RARRES3*, *RBM47*, *REC8*, *RRP1*, *SRA1* e *TMEM45A*) apresentaram diferença de expressão com significância estatística ($fold\ change \geq |2|$, $P < 0,05$), e foram selecionados para as análises posteriores de clusterização.

Dos 12 genes (11 genes validados e com significância estatística e o gene *ANAPC13*), 10 genes (*ANAPC13*, *RARRES3*, *BAALC*, *ERCC8*, *REC8*, *RRP1*, *SRA1*, *TMEM45A*, *NCF4* e *FLCN*) apresentaram uma diminuição da expressão no componente *in situ* de CMI-TNE quando comparado com CDIS puro, e nos grupos subsequentes da progressão (células epiteliais de lesão invasiva com e sem comprometimento linfonodal) apresentaram um aumento no nível de expressão, o qual se tornou semelhante ou um pouco mais alto do que o nível de expressão do CDIS puro. Por outro lado, 2 genes (*DNAJB6* e *RBM47*) apresentaram maior expressão no componente *in situ* do CMI-TNE quando comparado com o CDIS puro, mas permanecendo semelhante nos dois grupos de células epiteliais de lesão invasiva semelhantes ao componente *in situ* do CMI-TNE. Adicionalmente, analisamos a expressão desses 12 genes (*ANAPC13*, *BAALC*, *DNAJB6*, *ERCC8*, *FLCN*, *NCF4*, *RARRES3*, *RBM47*, *REC8*, *RRP1*, *SRA1* e *TMEM45A*) em 16 pacientes com amostras de células epiteliais capturadas por microdissecção a laser de tecido normal e de carcinoma invasivo (2 pacientes de tecido de mama normal e 14 pacientes de carcinoma de mama invasivo tipo não-especial) para termos um panorama do perfil de expressão desses 12 genes em grupos de amostras que mimetizam a progressão do câncer de mama (tecido normal, CDIS puro, componente *in situ* de CMI-TNE e CMI-TNE com e sem comprometimento linfonodal) (Apêndice 1).



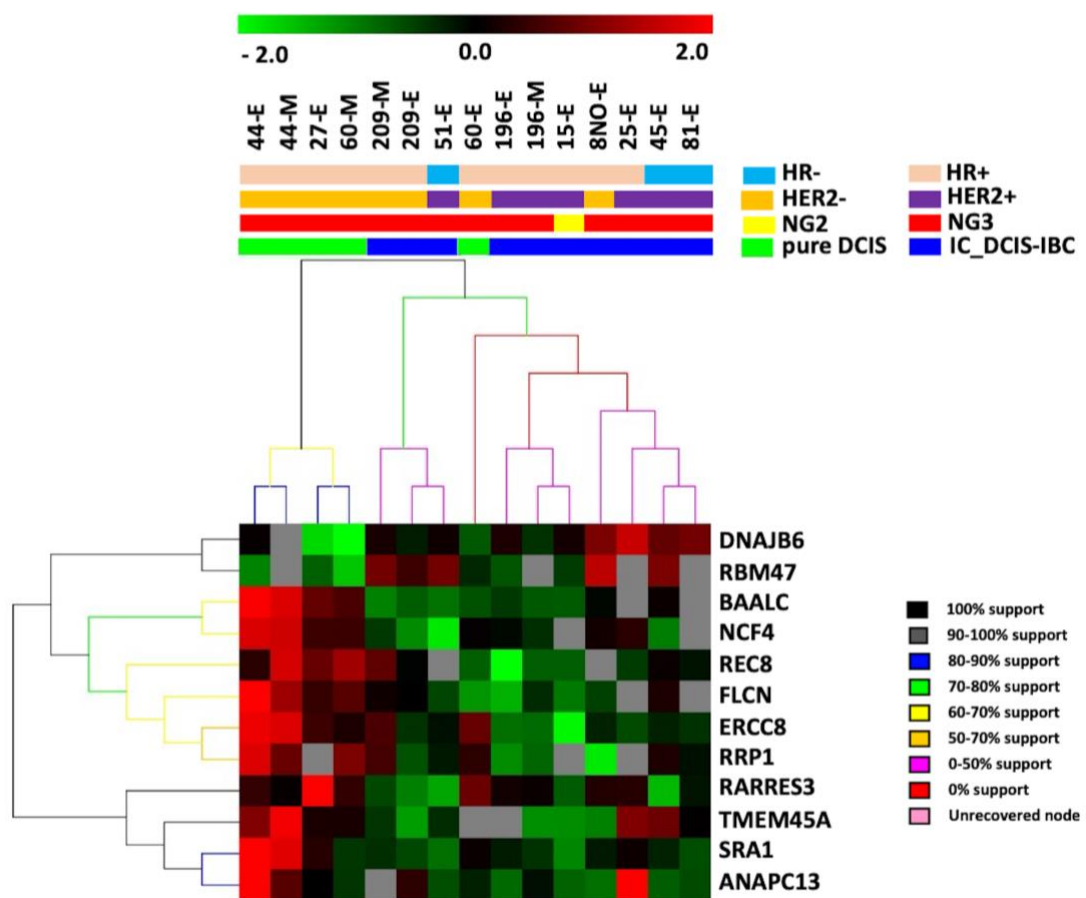
Legenda: Gráficos dotplot ilustrando os valores de expressão gênica. De acordo com nossos critérios de qualidade, 53 dos 78 genes selecionados do nosso ensaio customizado TLDA puderam ser analisados e 26 deles confirmaram nossos achados anteriores ($FC \geq |2|$). O gene ANAPC13 também está apresentado. Em colchetes estão os valores de diferença de expressão entre os grupos de amostras CDIS puro e componente *in situ* de CMI-TNE. Os asteriscos indicam valores com significância estatística ($P < 0,05$). Abreviações: *pure DCIS*, carcinoma ductal *in situ*; *IC_DCIS-IBC*, componente *in situ* de CMI-TNE.

Figura 7 - Genes diferencialmente expressos entre amostras de CDIS puro e de componente *in situ* de CMI-TNE confirmados por RT-qPCR usando Low Taqman Density Arrays (TLDA).

4.2.3 Análise de clusterização

Para avaliar se o padrão de expressão dos 12 genes (*ANAPC13*, *BAALC*, *DNAJB6*, *ERCC8*, *FLCN*, *NCF4*, *RARRES3*, *RBM47*, *REC8*, *RRP1*, *SRA1*, *TMEM45A*) selecionados era capaz de discriminar células epiteliais neoplásicas e mioepiteliais capturadas de 11 pacientes representados por 15 lesões pré-invasivas de CDIS com diferentes potenciais malignos (5 CDIS puro e 10 componente *in situ* de CMI-TNE) realizamos uma clusterização hierárquica não-supervisionada com distância Euclidiana e média das distâncias.

O padrão de expressão dos 12 genes está apresentado na Figura 8.



Legenda: Clusterização hierárquica não-supervisionada com distância Euclidiana e média das distâncias de 15 amostras, incluindo 5 amostras de CDIS puro (verde) e 10 amostras de componente *in situ* de CMI-TNE (azul). A clusterização hierárquica foi baseada em valores em escala logarítmica de base 2 obtidos da validação por RTq-PCR (ensaios TLDA). As colunas, linhas, cores vermelho e verde representam amostras, genes, aumento e diminuição de expressão, respectivamente. A escala de cor do dendrograma representa o suporte estatístico da técnica *bootstrap* (linhas pretas = 100% de suporte estatístico), a qual foi realizada com 1000 permutações. Abreviações: *pure DCIS*, carcinoma ductal *in situ* puro; *IC_DCIS-IBC*, componente *in situ* de CMI-TNE; NG, grau nuclear; HER2, fator de crescimento epidérmico; HR, receptores hormonais.

Figura 8 - Clusterização hierárquica baseada no perfil de expressão de 12 genes diferencialmente expressos entre amostras CDIS puro e componente *in situ* de CMI-TNE.

O padrão de expressão desses 12 genes foi capaz de discriminar relativamente bem os 11 pacientes. Estes 11 pacientes foram 3 pacientes de CDIS puro e 8 pacientes de componente *in situ* de CMI-TNE). Os 3 pacientes de CDIS puro são representados por 5 amostras, sendo 3 amostras de células epiteliais neoplásicas (27-E, 44-E e 60-E) e duas amostras de células mioepiteliais (44-M e 60-M). Os 8 pacientes de componente *in situ* de CMI-TNE são representados por 10 amostras, sendo 8 amostras de células epiteliais neoplásicas (8NO-E, 15-E, 25-E, 45-E, 51-E, 81-E, 196-E, 209-E) e duas amostras de células mioepiteliais (196-M e 209-M).

Considerando pelo menos uma das amostras de cada caso, 100% (3/3) dos pacientes (27, 44 e 60) de CDIS puro agruparam juntos no mesmo ramo do cluster. Considerando as duas amostras de cada caso (amostras de células epiteliais neoplásicas e mioepiteliais), 80% (4/5) das amostras de CDIS puro agruparam juntas no mesmo ramo do cluster, ou seja, uma amostra de células epiteliais neoplásicas (60-E) agrupou no ramo do cluster em que agruparam também as amostras do componente *in situ* de CMI-TNE. Em relação ao componente *in situ* de CMI-TNE, 100% dos pacientes e das amostras agruparam no mesmo ramo do cluster, demonstrando semelhança no perfil de expressão entre cada uma das amostras de células epiteliais neoplásicas e mioepiteliais. Curiosamente a amostra 60-E (epitelial) foi agrupada com as amostras de componente *in situ* do CMI-TNE, isso é, com o perfil de expressão de células com maior potencial maligno, e a amostra 60-M permaneceu no grupo das amostras CDIS. Interessantemente, este caso 60 se refere a uma paciente diagnosticada com CDIS, um caso de CDIS legítimo, ou seja, que tanto a biópsia quanto a peça cirúrgica apresentaram somente o componente *in situ*, e que apresentou progressão da doença 3 anos após tratamento convencional, pois apresentou nova lesão ipsilateral invasiva (CMI-TNE) com componente *in situ* associado. Este mesmo caso apresentou um perfil de expressão baseado em 132 GDE identificadas na abordagem de RNAseq que foi semelhante ao que detectamos neste estudo. Porém, na avaliação prévia de RNAseq, amostras 60-E e 60-M agruparam juntas com as amostras de componente *in situ* de CMI-TNE.

4.2.4 Análise de enriquecimento funcional

Inicialmente, através de uma busca no banco de dados do GeneOntology (Ashburner et al. 2000; Carbon et al. 2019), fizemos uma anotação dos processos biológicos com os quais

esses 12 genes estão envolvidos. Os genes foram organizados em 11 classes biológicas e aqueles pertencentes a mais de um processo biológico foram organizados de maneira hierárquica (Tabela 2).

Tabela 2 - Processos biológicos dos 12 genes cujo perfil de expressão foi capaz de discriminar amostras de CDIS puro de amostras de componente *in situ* de CMI-TNE.

Processo Biológico	Genes mais expressos nas células epiteliais/mioepiteliais em CDIS puro	Genes menos expressos nas células epiteliais/mioepiteliais em CDIS puro
Reparo de quebra de dupla-fita via recombinação homóloga	<i>REC8</i>	
Reparo de excisão de nucleotídeo acoplada a transcrição	<i>ERCC8</i>	
Resposta celular a estímulo estrogênico	<i>SRA1</i>	
Regulação negativa de proliferação celular	<i>RARRES3</i>	
Regulação positiva de adesão celular	<i>FLCN</i>	
Organização de matriz extracelular		<i>DNAJB6</i>
Diferenciação de células progenitoras hematopoiéticas		<i>RBM47</i>
Via de sinalização do receptor do fator de crescimento endotelial vascular	<i>NCF4</i>	
Processamento de rRNA	<i>RRP1</i>	
Divisão celular	<i>ANAPC13</i>	
Desconhecido	<i>BAALC</i> e <i>TMEM45A</i>	

Legenda: Apenas 2 (*DNAJB6* e *RBM47*) dos 12 genes não apresentaram diminuição da expressão nas amostras pré-invasivas com maior potencial maligno (componente *in situ* de CMI-TNE) quando comparado com as amostras de CDIS puro. Processos biológicos envolvidos com reparo de DNA foram representados por 2 (*ERCC8* e *REC8*) dos 12 genes.

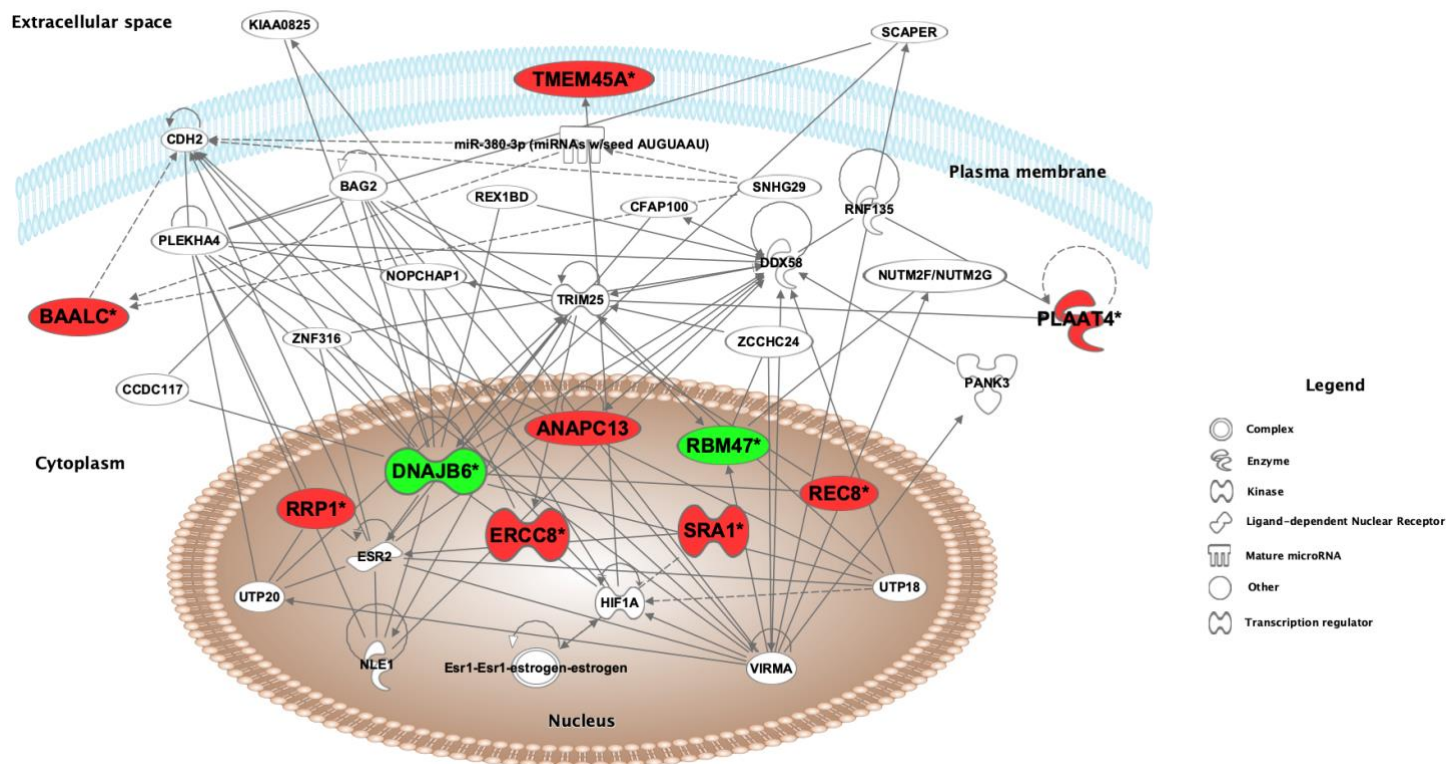
Em seguida, realizamos uma análise funcional *in silico* com os dados de expressão desses 12 genes utilizando o programa *Ingenuity Pathway Analysis* (IPA) (Qiagen). Desta forma, identificamos vias bioquímicas canônicas (Tabela 3) e redes de interação gênica enriquecidas por esses genes (Figura 9).

Tabela 3 - Vias canônicas que estão enriquecidas pelos 12 genes cujo perfil de expressão foi capaz de discriminar amostras de CDIS puro de amostras de componente *in situ* de CMI-TNE.

Vias Canônicas	P-valor
Reparo por excisão de nucleotídeo	1,86E-02
Via de sinalização de fosfolipase D	3,23E-02
Eicosanóides	3,38E-02
Cinase tipo-polo	3,49E-02
Via de sinalização MAPK	4,11E-02

Legenda: A via de reparo por excisão de nucleotídeo é uma das vias de reparo de DNA que atua de acordo com 5 etapas: reconhecimento da lesão, abertura da dupla hélice de DNA onde está localizada a lesão, dupla incisão nas extremidades da lesão, síntese do novo DNA utilizando como molde a fita não danificada, e ligação da porção 5' da nova fita sintetizada à sequência original; a via de sinalização de fosfolipase D é uma das vias de fosfolipases que atuam convertendo fosfolipídios em mediadores lipídicos que regulam múltiplos processos celulares, tais como, proliferação, migração e invasão celulares; a via de eicosanóides é uma via de metabolismo de fosfolipídios de membrana, como por exemplo, o ácido araquidônico; a via cinase tipo-polo é uma via envolvida na formação do fuso mitótico; a via de sinalização MAPK é uma via envolvida na proliferação celular, invasão e sobrevivência celulares.

Identificamos que os 12 genes formam uma rede de interação relacionada a função e desenvolvimento do sistema cardiovascular, desenvolvimento celular e movimentação celular, e apresentou um score igual a 28 ($-\log[P\text{-valor}]$). Vale ressaltar que a rede de interação gênica foi formada por 10 dos 12 genes, os quais interagem funcionalmente de maneira direta ou indireta (Figura 9).



Legenda: A análise do programa IPA mostrou que 10 dos 12 genes diferencialmente expressos entre as amostras de células epiteliais neoplásicas e mioepiteliais de CDIS puro e de componente *in situ* de CMI-TNE formam uma rede de interação (score 28). Esta rede está relacionada com movimento celular, desenvolvimento celular e desenvolvimento do sistema cardiovascular. Esta rede é baseada nos valores de diferença de expressão dos 12 genes. A rede de interação gênica é representada por nós e linhas entre os nós. As linhas contínuas e tracejadas representam interações diretas e indiretas entre as moléculas, respectivamente. Os nós vermelhos e verdes representam genes menos e mais expressos em amostras de componente *in situ* de CMI-TNE, respectivamente. Nós brancos representam outros genes automaticamente incluídos pelo IPA. Os nós são ilustrados por uma variedade de formas que representam as classes funcionais dos produtos gênicos.

Figura 9 - Rede de interação gênica.

Estes dados da avaliação transcricional de CDIS validaram os resultados de diferença de expressão entre células epiteliais neoplásicas e mioepiteliais de CDIS e componente *in situ* de lesão invasiva que foram gerados através da metodologia de RNAseq num estudo anterior do grupo. Esse trabalho demonstrou que os padrões de expressão dos genes diferencialmente expressos em ambas as células epiteliais neoplásicas e mioepiteliais foram igualmente capazes em discriminar amostras de CDIS puro e de componente *in situ* de CMI-TNE. Os dados foram reunidos e organizados no manuscrito Elias et al., (em preparação) e está em fase final de preparação para ser submetido à publicação.

4.3 Avaliação mutacional de CDIS

4.3.1 Casuística selecionada

A partir de uma lista de 560 pacientes identificados como carcinoma ductal *in situ* (CDIS) nas plataformas *Recruit* (262 pacientes com material no Arquivo da Patologia) e *Biobank* (298 pacientes com material no Biobanco) do ACCCC, foi realizada uma análise dos prontuários eletrônicos na plataforma *H2TC* da mesma instituição. Após a identificação das pacientes com diagnóstico de CDIS na biópsia e que atendiam aos critérios de inclusão do estudo, a identificação e obtenção dos blocos de parafina e dos tecidos congelados disponíveis, assim como de seus respectivos cortes histológicos com a topografia correta representando o tumor, e a revisão anatomopatológica das amostras cujas características atendiam aos critérios de inclusão do estudo, conforme descrito em *Materiais e Métodos*, conseguimos selecionar um total de 34 pacientes, representados por 43 amostras tumorais, onde 9 pacientes foram representados por duas amostras, uma do material de biópsia e outra do material de peça cirúrgica (Tabela 4).

Tabela 4 - Pacientes selecionados para sequenciamento de DNA por NGS.

Grupos	Subgrupos	Pacientes	Amostras
Grupo 1: CDIS Sem progressão	-	20	27
Grupo 2: CDIS Com progressão	Grupo 2.1 - CDIS/invasão	7	8
	Grupo 2.2 - CDIS/progressão/recidiva	7	8
Total		34	43

A casuística foi organizada em 2 grupos, um primeiro grupo nomeado CDIS sem progressão (Grupo 1) foi composto por 20 pacientes, os quais tiveram laudo anatomopatológico de carcinoma ductal *in situ* (CDIS puro) de mama tanto no material da biópsia quanto no material da peça cirúrgica e que não apresentou progressão e/ou recidiva da doença por pelo menos 3 anos. O segundo grupo foi nomeado CDIS com progressão (Grupo 2) e foi composto por 14 pacientes com laudo anatomopatológico de CDIS puro na biópsia, sendo dividido em dois subgrupos, em que cada um deles foi formado por 7 pacientes. O grupo 2.1 (CDIS/invasivo) foi formado por pacientes que tiveram laudo anatomopatológico de CDIS puro no material da biópsia e no material da cirurgia tiveram laudo anatomopatológico de carcinoma de mama invasivo tipo não especial com componente *in situ* associado (CMI-TNE + CDIS); o grupo 2.2 (CDIS/progressão/recidiva) foi formado por pacientes que tiveram laudo anatomopatológico de CDIS puro tanto no material da biópsia quanto no material da cirurgia e que em algum momento do período de seguimento a paciente evoluiu com manifestação de progressão da doença (nova lesão ipsilateral invasiva) ou com recidiva (nova lesão ipsilateral *in situ*). As características anatomopatológicas e antecedentes familiares das pacientes do grupo 2.2 estão apresentadas no Apêndice 2. Considerando o total de 560 pacientes avaliadas com diagnóstico de CDIS na biópsia, encontramos sete pacientes que tiveram progressão/recidiva, ou seja, 1,3%. Destas 7 pacientes, 4 pacientes (28CDIS, 30CDIS, 33CDIS e 34CDIS) apresentaram progressão da doença (recidiva ipsilateral de lesão invasiva) e 3 pacientes (29CDIS, 31CDIS e 32CDIS) apresentaram recidiva de CDIS (recidiva ipsilateral de lesão *in situ*).

Aproximadamente 95% da casuística foi formada por pacientes de alto grau (90% do grupo 1 (CDIS sem progressão) e 100% do grupo 2 (CDIS com progressão); 76% das pacientes com expressão positiva dos receptores hormonais (85% do grupo 1 e 64% do grupo 2) e 41% com presença de comedocarcinoma (35% do grupo 1 e 50% do grupo 2). A mediana de idade das pacientes ao diagnóstico foi de 48 anos (48 anos do grupo sem progressão e 48,5 anos do grupo com progressão). O tempo mediano de seguimento da casuística foi de 6 anos, sendo que no grupo 1 foi de 6 anos, com intervalo de 3 a 9 anos; e no grupo 2.2 (CDIS/progressão/recidiva) foi de 7 anos, com intervalo de 5 a 12 anos. No grupo 2.2 (CDIS/progressão/recidiva), o tempo mediano entre o diagnóstico do tumor primário e o diagnóstico da progressão/recidiva foi de 2 anos (intervalo de 1 a 7 anos) (Tabela 5).

Tabela 5 - Características histológicas e moleculares das 34 pacientes de CDIS selecionados (43 amostras) e quantidade e qualidade do DNA obtido para sequenciamento por NGS.

Grupos	Subgrupos	Identificação da amostra	Laudo na biópsia	Laudo na peça cirúrgica	ER	PR	HER2	Grau Nuclear	Comedocarcinoma	Idade ao diagnóstico	Procedimento cirúrgico	Tempo de seguimento (anos)	Período entre o tumor primário e a progressão/recidiva (anos)	Tipo de recorrência	Material avaliado	Tipo de material	Quantidade total de DNA extraído (ng)	Integridade do DNA (DIN)			
Grupo 1: CDIS sem progressão	-	01CDISA	CDIS	CDIS	+	+	1+	2	não	60	RS	3	-		B	FFPE	0	NA			
		01CDISB	CDIS	CDIS	+	+	1+	3	não						PC	FFPE	2	NA			
		02CDISA	CDIS	CDIS	+	+	2+	3	sim	50	M	6			B	FFPE	126	2,3			
		02CDISB	CDIS	CDIS	+	+	2+	3	sim						PC	FFPE	236	4			
		03CDISA	CDIS	CDIS	-	-	3+	3	sim	59	M	5			B	FFPE	13	NA			
		03CDISB	CDIS	CDIS	-	-	3+	3	sim						PC	FFPE	190	2,6			
		04CDISA	CDIS	CDIS	+	+	0+	3	não	49	M	6			B	FFPE	30	2			
		04CDISB	CDIS	CDIS	+	+	0+	3	não						PC	FFPE	8	NA			
		05CDISA	CDIS	CDIS	+	+	3+	3	não	48	RS	6			B	FFPE	42	4			
		05CDISB	CDIS	CDIS	+	+	3+	3	não						PC	FFPE	26	3,6			
		06CDISA	CDIS	CDIS	+	-	3+	3	não	48	RS	7			B	FFPE	141	3			
		06CDISB	CDIS	CDIS	+	-	3+	3	não						PC	FFPE	264	4			
		07CDISA	CDIS	CDIS	+	+	1+	2	não	44	RS	6			B	FFPE	42	2,7			
		07CDISB	CDIS	CDIS	+	+	1+	3	não						PC	FFPE	163	3,7			
		08CDISA	CDIS	CDIS	+	+	1+	1	não	64	RS	5			B	FFPE	6	NA			
		09CDISA	CDIS	CDIS	+	+	3+	3	sim						B	FFPE	5	NA			
		10CDISA	CDIS	CDIS	-	-	3+	3	sim	64	M	5			B	FFPE	14	1,9			
		11CDISA	CDIS	CDIS	+	+	1+	3	não						B	FFPE	13	2,1			
		12CDISA	CDIS	CDIS	+	+	2+	3	não	42	M	6			B	FFPE	35	2,3			
		13CDISA	CDIS	CDIS	+	+	1+	3	não						B	FFPE	28	1,8			
14CDISB	CDIS	CDIS	+	+	0+	3	sim	42	M	5	PC	TC			73	NA					
15CDISB	CDIS	CDIS	+	+	3+	3	não				PC	TC			314	6,4					
16CDISB	CDIS	CDIS	+	+	0+	2	não	77	RS	7	PC	TC			120	6,5					
17CDISB	CDIS	CDIS	+	+	3+	3	não				PC	TC			691	7,7					
18CDISB	CDIS	CDIS	-	+	3+	3	sim	37	M	8	PC	TC			847	7,5					
19CDISB	CDIS	CDIS	+	+	1+	3	não				PC	TC			896	7,1					
20CDISB	CDIS	CDIS	+	-	3+	3	sim	46	M	6	PC	TC			127	8					
Grupo 2: CDIS com progressão	Subgrupo 1: Foco(s) de invasão no material da cirúrgica	21CDISA	CDIS	<i>in situ</i> de CMI-TNE	+	+	3+	3	não	33	M	2			-		B	FFPE	36	1,1	
		21CDISB	CDIS	<i>in situ</i> de CMI-TNE	+	+	3+	3	não								PC	FFPE	95	4	
		22CDISA	CDIS	<i>in situ</i> de CMI-TNE	+	+	1+	3	não	67	NI	6					B	FFPE	4	NA	
		23CDISA	CDIS	<i>in situ</i> de CMI-TNE	+	+	3+	3	não								B	FFPE	33	2,6	
		24CDISA	CDIS	<i>in situ</i> de CMI-TNE	+	+	3+	3	sim	57	NI	5					B	FFPE	24	1,9	
		25CDISA	CDIS	<i>in situ</i> de CMI-TNE	-	-	3+	3	não								B	FFPE	47	NA	
		26CDISA	CDIS	<i>in situ</i> de CMI-TNE	+	+	-	3	não	52	NI	4					B	FFPE	7	NA	
		27CDISA	CDIS	<i>in situ</i> de CMI-TNE	-	-	1+	3	sim								B	FFPE	124	1,8	
		28CDISA	CDIS	CDIS	+	+	3+	3	sim	35	M	7	3	progressão			B	FFPE	15	3,7	
		28CDISB	CDIS	CDIS	+	+	3+	3	sim					PC			FFPE	99	3		
		29CDISB	CDIS	CDIS	+	+	3+	3	sim	50	RS	12	2	recidiva			PC	FFPE	27	2	
		30CDISB	CDIS	CDIS	-	-	3+	3	não					PC			FFPE	25	2,5		
		31CDISB	CDIS	CDIS	+	+	1+	3	não	48	RS	6	4	progressão			PC	FFPE	177	2,6	
		32CDISB	CDIS	CDIS	+	+	2+	3	sim					PC			FFPE	57	3,1		
	33CDISB	CDIS	CDIS	-	-	3+	3	sim	40	AM	7	1	progressão	PC			FFPE	36	NA		
	34CDISB	CDIS	CDIS	-	-	3+	3	sim					PC	FFPE			94	NA			
		Subgrupo 2: Progressão/Recidiva durante o seguimento																			

Legenda: A nomenclatura das amostras corresponde a numeração do caso com um sufixo “A” para material de biópsia e “B” para material da cirúrgica. Abreviações: AM, adenomastectomia; B, biópsia; CDIS, carcinoma ductal *in situ*; *in situ* de CMI-TNE, componente *in situ* de carcinoma mamário invasivo tipo não especial; FFPE, material fixado em formalina e embebido em parafina; M, mastectomia; NA, não avaliado; PC, peça cirúrgica; RS, ressecção segmentar; TC, tecido congelado.

4.3.2 Sequenciamento de DNA

Todas as 43 amostras da casuística foram submetidas à extração de DNA, sendo que para 36 amostras obtivemos quantidade de DNA suficiente para o sequenciamento por NGS usando os painéis multigenes CHP2 e CCP. As 7 amostras (01CDISA, 01CDISB, 04CDISB, 08CDISA, 09CDISA, 22CDISA e 26CDISA) que não apresentaram quantidade suficiente de DNA foram excluídas do estudo.

Em relação as 36 amostras tumorais com quantidade de DNA suficiente para o sequenciamento por NGS, a integridade do DNA (DIN) variou de 1,1 a 7,7 e a quantidade variou de 13 ng a 896 ng (Tabela 5). Dessas, 16 amostras apresentaram valor de $DIN \geq 3$, que é considerado um valor de integridade adequado para sucesso no sequenciamento NGS (02CDISB, 05CDISA, 05CDISB, 06CDISA, 06CDISB, 07CDISB, 15CDISB, 16CDISB, 17CDISB, 18CDISB, 19CDISB, 20CDISB, 21CDISB, 28CDISA, 28CDISB e 32CDISB) e 15 amostras apresentaram $DIN < 3$ (02CDISA, 03CDISB, 04CDISA, 07CDISA, 10CDISA, 11CDISA, 12CDISA, 13CDISA, 21CDISA, 23CDISA, 24CDISA, 27CDISA, 29CDISB, 30CDISB e 31CDISB). Para 5 amostras (03CDISA, 14CDISB, 25CDISA, 33CDISB e 34CDISB) a avaliação do DIN falhou (Tabela 5). Na experiência do nosso grupo com dados de amostras FFPE, o DNA com valor de $DIN > 3$ resulta em um sequenciamento de DNA com melhor qualidade. Entretanto, devido a dificuldade de obtenção desse tipo de amostra, as amostras com DNA com valor de $DIN < 3$ também foram incluídas para o sequenciamento de DNA por NGS.

Das 36 amostras tumorais sequenciadas, 26 amostras foram sequenciadas utilizando o painel de 409 genes, o qual exige uma quantidade mínima de 40 ng de DNA, sendo necessário 10 ng para cada *pool* de *primers*; e 10 amostras com o painel de 50 genes, uma vez que para esse teste é usado um único *pool* de *primers* que exige uma quantidade mínima de 10 ng de DNA. Além disso, 4 *pools* com DNA de leucócitos foram sequenciados pelo painel 409 genes para filtrar variantes germinativas e com isso selecionarmos as variantes com maior chance de serem somáticas, ou seja, variantes adquiridas durante o processo tumorigênico. As variantes identificadas no DNA do tumor foram avaliadas no DNA de leucócitos de 9 pacientes cujos DNAs de leucócitos estavam disponíveis no nosso biobanco (11CDIS, 13CDIS, 14CDIS, 15CDIS, 16CDIS, 18CDIS, 19CDIS, 20CDIS e 27CDIS). Esses DNAs foram organizados em 4 *pools* (3 *pools* com duas amostras e 1 *pool* com 3 amostras). A análise pareada (tumor-leucócito do mesmo caso) foi possível para 8 (11CDIS, 14CDIS, 15CDIS, 16CDIS, 18CDIS,

19CDIS, 20CDIS e 27CDIS) das 9 pacientes, pois a biblioteca construída com o DNA tumoral do caso 13CDIS não atingiu qualidade mínima para ser avaliada (Tabela 6).

No total, das 36 amostras tumorais submetidas ao sequenciamento de DNA, 9 amostras (04CDISA, 05CDISA, 05CDISB, 13CDISA, 24CDISA, 25CDISA, 28CDISB, 29CDISB e 33CDISB) tiveram dados que não atenderam os critérios mínimos de qualidade. Entretanto, incluímos as amostras 05CDISA e 28CDISB no estudo, pois através de uma busca ativa (inspeção visual) identificamos nessas amostras variantes potencialmente somáticas em oncogene e gene supressor tumoral (GST). Desta forma, um total de 29 amostras permaneceram no estudo (Tabela 6).

Estas 29 amostras representaram 23 pacientes: 15 pacientes do grupo 1 e 8 pacientes do grupo 2 (3 pacientes do grupo 2.1 e 5 pacientes do grupo 2.2). Das 29 amostras analisadas, 24 amostras foram analisadas com o painel de 409 genes (painel CCP) e 5 amostras com o painel de *hotspots* de 50 genes (painel CHP2).

Tabela 6 - Dados do sequenciamento de DNA das 29 amostras tumorais e dos 4 pools de leucócitos.

Grupo	Subgrupo	Identificação da amostra	Material avaliado	Tipo de material	Conteúdo tumoral	Painel	Quantidade de DNA (ng) na construção da biblioteca	Integridade do DNA tumoral (DIN)	DNA de leucócito disponível	Amostras que atenderam aos critérios de qualidade	Bases	Bases com qualidade de sequenciamento	Leituras	Leituras mapeadas	Tamanho médio das leituras	Uniformidade	Ts/Tv	Cobertura				Cobertura média	Leituras no alvo (%)
Grupo 1: CDIS sem progressão		02CDISA	B	FFPE	80%	CCP	60	2,3	não	sim	1.903.128.779	1.753.872.188	23.398.630	23.059.113	81 bp	32%	2,8	98%	74%	43%	21%	1.009	92%
		02CDISB	PC	FFPE	70%	CCP	60	4	não	sim	1.435.881.152	1.281.769.088	15.213.940	15.039.453	94 bp	54%	2,4	98%	85%	63%	33%	798	98%
		03CDISA	B	FFPE	60%	CHP2	10	NA	não	sim	251.184.237	230.338.868	2.595.920	2.583.511	96 bp	68%	2,7	100%	100%	96%	92%	10.373	99%
		03CDISB	PC	FFPE	70%	CCP	60	2,6	não	sim	2.020.341.404	1.846.317.324	20.355.127	20.135.521	99 bp	71%	2,4	99%	95%	83%	53%	1.123	99%
		05CDISA*	B	FFPE	60%	CCP	32	4	não	não	331.185.151	301.444.645	3.671.534	3.491.879	90 bp	39%	2,9	96%	45%	21%	7%	136	72%
		06CDISA	B	FFPE	70%	CCP	60	3	não	sim	962.625.061	873.113.363	8.729.064	8.693.488	110 bp	62%	2,4	95%	76%	62%	30%	502	99%
		06CDISB	PC	FFPE	60%	CCP	60	4	não	sim	2.316.016.835	2.017.959.572	20.362.152	20.205.894	113 bp	86%	2,1	99%	97%	93%	73%	1.216	99%
		07CDISA	B	FFPE	70%	CCP	38	2,7	não	sim	4.908.545.728	4.464.389.510	47.222.832	46.758.971	103 bp	34%	2,5	90%	68%	50%	35%	2.745	99%
		07CDISB	PC	FFPE	70%	CCP	60	3,7	não	sim	2.044.931.908	1.842.495.267	18.750.134	18.559.316	109 bp	75%	2,3	99%	94%	84%	60%	1.104	99%
		10CDISA	B	FFPE	90%	CHP2	10	1,9	não	sim	165.340.003	151.844.629	2.175.656	2.143.676	75 bp	50%	3,0	100%	98%	90%	61%	4.092	58%
		11CDISA	B	FFPE	90%	CHP2	10	2,1	sim	sim	73.675.910	68.669.914	713.386	711.486	103 bp	86%	2,2	100%	100%	100%	90%	3.048	100%
		12CDISA	B	FFPE	70%	CCP	36	2,3	não	sim	1.314.793.787	1.202.943.184	15.528.113	15.139.948	84 bp	28%	2,4	94%	54%	32%	16%	718	95%
		14CDISB	PC	TC	70%	CCP	40	NA	sim	sim	2.115.364.311	1.943.565.468	22.807.178	22.637.808	92 bp	20%	1,0	93%	39%	24%	17%	1.196	99%
		15CDISB	PC	TC	20%	CCP	40	6,4	sim	sim	3.784.256.552	3.358.423.831	29.132.779	29.029.169	129 bp	91%	2,3	100%	99%	97%	87%	1.873	100%
		16CDISB	PC	TC	80%	CCP	40	6,5	sim	sim	2.841.627.914	2.560.578.922	24.693.400	24.550.128	115 bp	54%	2,4	98%	85%	68%	48%	1.527	100%
		17CDISB	PC	TC	30%	CCP	60	7,7	não	sim	1.740.015.064	1.521.037.306	14.077.398	14.006.099	123 bp	84%	2,1	99%	97%	90%	60%	876	99%
		18CDISB	PC	TC	30%	CCP	60	7,5	sim	sim	1.749.055.110	1.542.918.149	14.564.018	14.491.635	120 bp	83%	2,2	99%	97%	89%	59%	902	99%
		19CDISB	PC	TC	60%	CCP	60	7,1	sim	sim	4.988.436.679	4.452.455.742	39.602.211	39.442.353	125 bp	89%	2,5	100%	99%	97%	89%	2.500	99%
		20CDISB	PC	TC	40%	CCP	60	8	sim	sim	5.900.680.115	5.239.305.759	47.108.469	46.904.106	125 bp	91%	2,4	100%	99%	98%	93%	2.976	99%
Grupo 2: CDIS com progressão	Subgrupo 1: Foco (s) de invasão no material da cirúrgica	21CDISA	B	FFPE	40%	CCP	36	1,1	não	sim	1.897.249.211	1.721.646.517	17.373.781	17.167.010	109 bp	70%	2,3	99%	92%	79%	55%	1.027	99%
		21CDISB	PC	FFPE	60%	CCP	80	4	não	sim	2.754.563.954	2.484.142.522	23.136.297	22.926.053	119 bp	86%	2,3	100%	98%	94%	75%	1.417	99%
		23CDISA	B	FFPE	70%	CCP	40	2,6	não	sim	1.563.198.501	1.425.593.961	16.783.209	16.598.187	93 bp	31%	2,7	94%	60%	38%	21%	880	99%
		27CDISA	B	FFPE	70%	CCP	60	1,8	sim	sim	2.379.046.824	2.175.686.069	23.919.670	23.714.744	99 bp	63%	2,3	99%	92%	79%	51%	1.331	99%
	Subgrupo 2: Progressão/recidiva durante o seguimento	28CDISA	B	FFPE	40%	CHP2	10	3,7	não	sim	109.093.976	99.798.659	1.163.570	1.155.940	93 bp	58%	3,7	100%	99%	90%	68%	4.472	98%
		28CDISB*	PC	FFPE	70%	CCP	60	3	não	não	1.069.360.400	979.779.525	13.191.985	12.889.522	81 bp	39%	2,6	97%	63%	33%	12%	348	58%
		30CDISB	PC	FFPE	90%	CHP2	15	2,5	não	sim	696.333.468	637.183.598	6.571.020	6.540.344	105 bp	86%	3,0	100%	100%	100%	100%	28.608	98%
		31CDISB	PC	FFPE	90%	CCP	80	2,6	não	sim	1.203.342.882	1.085.432.129	14.136.333	13.965.145	85 bp	45%	2,5	97%	77%	50%	22%	655	94%
		32CDISB	PC	FFPE	90%	CCP	50	3,1	não	sim	2.844.475.449	2.538.013.496	29.435.299	29.220.229	96 bp	54%	2,6	99%	90%	73%	46%	1.594	99%
		34CDISB	PC	FFPE	90%	CCP	80	NA	não	sim	2.637.338.962	2.317.445.566	24.657.049	24.540.721	106 bp	50%	2,4	98%	86%	66%	42%	1.446	99%
Pool de leucócitos		Pool 1 de leucócitos	-	-	-	CCP	20	-	-	sim	3.375.695.254	2.965.354.528	28.119.613	27.962.612	120 bp	93%	2,3	99%	99%	98%	91%	1.747	99%
		Pool 2 de leucócitos	-	-	-	CCP	20	-	-	sim	3.462.620.435	3.019.720.550	28.183.345	28.051.300	122 bp	95%	2,3	100%	99%	98%	92%	1.758	99%
		Pool 3 de leucócitos	-	-	-	CCP	20	-	-	sim	3.042.124.846	2.703.214.882	26.296.583	26.119.082	115 bp	93%	2,3	100%	99%	98%	89%	1.593	99%
		Pool 4 de leucócitos	-	-	-	CCP	20	-	-	sim	3.079.311.486	2.755.151.156	28.820.117	28.573.796	106 bp	90%	2,3	100%	99%	96%	85%	1.717	99%

Legenda: Dados de sequenciamento das 29 amostras tumorais e dos 4 *pools* de DNA de leucócitos que passaram pelos nossos critérios de qualidade e que foram analisados. *Pool* 1 de DNA de leucócitos: pacientes 13CDIS e 18CDIS; *Pool* 2 de DNA de leucócitos: pacientes 11CDIS e 19CDIS; *Pool* 3 de DNA de leucócitos: pacientes 20CDIS e 27CDIS; *Pool* 4 de DNA de leucócitos: pacientes 14CDIS, 15CDIS e 16CDIS. (*) As amostras 05CDISA e 28CDISB não passaram pelos critérios de qualidade de sequenciamento, mas foram incluídas no estudo devido as variantes identificadas por busca ativa (inspeção visual). Abreviações: B, biópsia; CCP, painel de 409 genes; CHP2, painel de 50 genes; DIN, número de integridade do DNA; FFPE, tecido fixado em formalina e embebido em parafina; NA, não avaliado; PC, peça cirúrgica; TC, tecido congelado; Ts/Tv, razão transição/transversão.

4.3.3 Identificação de variantes potencialmente somáticas em CDIS

Inicialmente analisamos as 27 amostras que passaram pelos critérios de qualidade de sequenciamento. Foi detectado um total de 221 variantes em 138 genes. Das 221 variantes, excluimos um total de 117 (50 variantes com frequência alélica > 0,5% no dbSNP; 58 variantes germinativas presentes nos *pools* de 6 pacientes com DNA de leucócitos (15CDIS, 16CDIS, 18CDIS, 19CDIS, 20CDIS e 27CDIS); 6 variantes que na inspeção visual do alinhamento mostraram ser artefatos de sequenciamento; e 3 variantes com VAF próximo a 50% que são classificadas no ClinVar como benignas), totalizando 104 variantes potencialmente somáticas.

Em seguida, realizamos uma busca ativa em oncogenes e GST sabidamente envolvidos com tumores sólidos, especialmente em câncer de mama, usando como critérios VAF > 1% e cobertura de sequenciamento > 40x com VAF > 5%. Nesse processo detectamos 4 variantes em oncogenes/GST (*HNF1A* c.526+1G>A, p. (spl?), VAF 3%; *PIK3CA* c.278G>A, p.Arg93Gln, VAF 3%; *PPP2R1A* c.773G>A, p.Arg258His, VAF 4%; *RB1* c.1654C>T, p.Arg552Ter, VAF 45%), totalizando 108 variantes potencialmente somáticas em 84 genes. Além disso fizemos uma busca de variantes em oncogenes e GST na chamada de variantes gerada pelo TVC em amostras que não preencheram os critérios de qualidade de sequenciamento e detectamos uma variante no gene *SMARCA4* (c.1777G>T, p.Glu593Ter, VAF 82%) no caso 05CDIS cuja análise visual demonstrou ser real. Desta forma, identificamos um total de 109 variantes potencialmente somáticas em 85 genes, detectadas em 23 pacientes de CDIS.

Detectamos variantes potencialmente somáticas em oncogenes ou GST em 69,5% (16/23) (02CDIS, 03CDIS, 05CDIS, 06CDIS, 07CDIS, 12CDIS, 15CDIS, 17CDIS, 21CDIS, 23CDIS, 27CDIS, 28CDIS, 30CDIS, 31CDIS, 32CDIS e 34CDIS) das pacientes (Tabela 7).

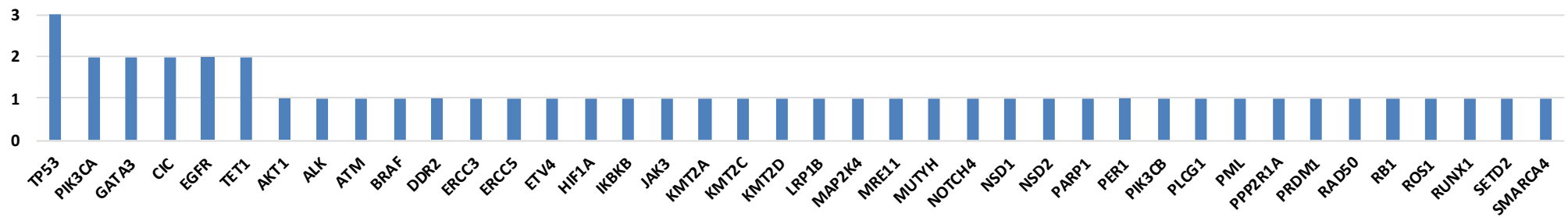
Tabela 7 - Variantes detectadas em oncogenes e genes supressores de tumor.

		Painel		Conteúdo tumoral (%)		Biopsia		Peça cirúrgica		Validação		Classe do gene	Tipo de variante (origem)	Classificação oncogênica da variante (OncoKB)	Efeito oncogênico da variante (OncoKB)	Identificador da variante no Cosmic	Identificador da variante no dbSNP (v.155)	Classificação patogênica da variante (ClinVar)			
Grupo	Subgrupo	Caso	Biopsia	Peça cirúrgica	Biopsia	Peça cirúrgica	Gene	HGVs c.	HGVs p.	Cobertura	VAF (%)								Cobertura	VAF (%)	Cobertura
Grupo 1: CDIS sem progressão	02CDIS	CCP	CCP	80	70	GATA3	c.1050>T>C	p. (spl7)	22982	38	7597	52	NA	onco/GST	PS	provavelmente oncogênica	provável perda de função	NR	NR	NR	
						KMT2A	c.10327G>A	p.Ala3443Thr	178	69	1290	41		GST	PS	ND	ND	1235525	rs201345346	NR	
						MAP2K4	c.339_340delGA	p.Lys114Argfs*28	137	31	668	63		GST	PS	provavelmente oncogênica	provável perda de função	NR	NR	NR	
						PRDM1	c.230T>C	p.Val777Ala	44	43	111	37		GST	PS	ND	ND	NR	rs147363823	NR	
	06CDIS	CCP	CCP	70	60	MUTYH	c.190_191delGC	p.Ala64Profs*28	741	52	1164	53	NA	GST	PS	provavelmente oncogênica	provável perda de função	NR	NR	NR	
						ERCC3	c.1201A>G	p.Ile401Val	1042	40	1605	42		GST	PS	ND	ND	NR	rs115176021	NR	
						SETD2	c.2546G>T	p.Cys849Phe	7	43	1026	61		GST	PS	ND	ND	NR	rs776511280	VUS	
						KMT2C	c.2353C>A	p.Pro785Thr	467	47	821	53		GST	PS	ND	ND	NR	rs777757726	NR	
	07CDIS	CCP	CCP	70	70	NRAS	c.355G>T	p.Asp119Tyr	166	35	1541	0	18692	0	onco	PS	ND	ND	NR	NR	NR
						AKT1	c.49G>A	p.Glu17Lys	10508	46	4587	57	NA	onco	PS	oncogênica	ganho de função	33765	rs121434592	VUS	
						PML	c.766G>A	p.Ala256Thr	568	15	133	2		GST	PS	ND	ND	NR	NR	NR	
						EGFR	c.2293G>A	p.Val765Met	876	6	523	0	175938	1	onco	PS	ND	ND	28603	rs374873413	VUS
						LRP1B	c.7366G>A	p.Val2456Ile	93	47	573	48	NA	GST	PS	ND	ND	NR	rs150174082	NR	
						PPP2R1A	c.773G>A	p.Arg258His	2304	4	1884	0		GST	PS	inconclusivo	ND	144147	rs863225094	patogênica	
						BRAF	c.87delC	p.Gly30Alafs*24	312	95	165	81		onco	PS	ND	ND	NR	NR	NR	
						PLCG1	c.3824G>A	p.Arg1275Gln	0	0	923	54		NA	onco	PS	ND	ND	1026808	rs372257814	NR
	03CDIS	CHP2	CCP	60	70	HNFI1A	c.526+1G>A	p. (spl7)	0	0	522	3	100962	0	GST	PS	provavelmente oncogênica	provável perda de função	21487	rs1364708195	PP
						PIK3CA	c.278G>A	p.Arg93Gln	0	0	901	3	254565	0	onco	PS	provavelmente oncogênica	provável ganho de função	86041	rs1064793663	patogênica
	05CDIS**	CCP	NA	60	NA	SMARCA4	c.1777G>T	p.Glu593Ter	590	82			NA	NA	GST	PS	provavelmente oncogênica	provável perda de função	NR	NR	NR
	12CDIS	CCP	NA	70	NA	TP53	c.743G>C	p.Arg248Pro	288	29			NA	onco/GST	PS	provavelmente oncogênica	provável perda de função	11491	rs11540652	patogênica	
						PIK3CB	c.827G>A	p.Arg276Lys	156	41	NA	onco		PS	ND	ND	NR	rs368178860	NR		
						NSD1	c.7500G>C	p.Met2500Ile	301	42		onco/GST		PS	ND	ND	NR	rs770120153	conflitante		
						ROS1	c.5419T>C	p.Phe1807Leu	199	37	onco	PS		ND	ND	NR	rs376765085	NR			
	15CDIS*	NA	CCP	NA	20	CIC	c.5768C>T	p.Ala1923Val		NA	4764	49	NA	GST	PS	ND	ND	NR	rs780624177	NR	
	17CDIS	NA	CCP	NA	30	HIF1A	c.1163T>A	p.Val388Asp			760	35	NA	onco	PS	ND	ND	NR	rs771469186	NR	
PER1						c.887C>G	p.Ser296Cys		NA	121	44	GST		PS	ND	ND	NR	rs140853060	NR		
					JAK3	c.905G>A	p.Ser302Asn			594	50		onco	PS	ND	ND	NR	rs751055857	NR		
21CDIS	CCP	CCP	40	60	ETV4	c.515dupA	p.His172Glnfs*11	4230	41	3225	22	NA	onco	PS	ND	ND	NR	NR	NR		
					IKBKB	c.1336C>T	p.Arg446Trp	419	26	8026	11		onco	PS	ND	ND	1210456	rs202136671	VUS		
					GATA3	c.614dupT	p.Met205Ilefs*99	968	10	2155	52		onco/GST	PS	provavelmente oncogênica	provável perda de função	NR	NR	NR		
					TP53	c.822_823dupTT	p.Cys275Phefs*71	1808	16	1761	42		58731	58	onco/GST	PS	provavelmente oncogênica	provável perda de função	NR	NR	NR
23CDIS	CCP	NA	70	NA	PARP1	c.1103C>T	p.Thr368Met	396	59			NA	GST	PS	ND	ND	NR	rs749502012	NR		
					ALK	c.1652G>A	p.Arg551Gln	2825	51	NA	onco		PS	ND	ND	1182536	rs374136388	VUS			
27CDIS*	CCP	NA	70	NA	DDR2	c.1384C>G	p.Pro462Ala	1981	15				NA	onco	PS	ND	ND	NR	NR	NR	
					TP53	c.524G>A	p.Arg175His	686	74	NA	onco/GST	somática		oncogênica	perda de função	10648	rs28934578	patogênica			
					RB1	c.1654C>T	p.Arg552Ter	87	45		GST	PS		provavelmente oncogênica	provável perda de função	1300225	rs121913303	patogênica			
					NOTCH4	c.4465C>T	p.Arg1489Ter	635	25	onco/GST	PS	provavelmente oncogênica		provável perda de função	NR	rs1231739465	NR				
Grupo 2: CDIS com progressão	28CDIS**	CHP2	CCP	40	70	PIK3CA	c.3140A>G	p.His1047Arg	2471	31	42	40	7076	31	onco	PS	oncogênica	ganho de função	94986	rs121913279	patogênica
	30CDIS	NA	CHP2	NA	90	EGFR	c.2293G>A	p.Val765Met		NA	2791	3	81702	1	onco	PS	ND	ND	28603	rs374873413	VUS
	31CDIS	NA	CCP	NA	90	MRE11	c.1726C>T	p.Arg576Ter			169	42	NA	GST	PS	provavelmente oncogênica	provável perda de função	NR	rs774277300	patogênica	
						KMT2D	c.15345delT	p.Phe5115Leufs*32		NA	444	56		onco/GST	PS	provavelmente oncogênica	provável perda de função	NR	NR	NR	
						ERCC5	c.3056A>G	p.Asp1019Gly			254	41		GST	PS	ND	ND	NR	rs370456073	NR	
						TET1	c.6016A>G	p.Ile2006Val			154	50		onco/GST	PS	ND	ND	NR	rs775056557	NR	
	32CDIS	NA	CCP	NA	90	ATM	c.8560C>T	p.Arg2854Cys			151	44	NA	GST	PS	ND	ND	NR	rs201958469	conflitante	
						CIC	c.6573_6607del	p.Leu2192Profs*49		NA	1338	61		onco/GST	PS	provavelmente oncogênica	provável perda de função	NR	NR	NR	
						RUNX1	c.1073C>T	p.Thr358Met			451	35		onco/GST	PS	ND	ND	NR	rs962671280	VUS	
						NSD2	c.3634C>G	p.Leu1212Val			206	52		onco	PS	ND	ND	NR	rs945268423	NR	
	34CDIS	NA	CCP	NA	90	RAD50	c.508A>G	p.Lys170Glu			691	56	NA	GST	PS	ND	ND	NR	rs1750464972	VUS	
						TET1	c.3866C>T	p.Thr1289Met			369	48		onco/GST	PS	ND	ND	4015294	rs575341146	NR	
						PIK3CA	c.3140A>G	p.His1047Arg		NA	135	16		17467	14	onco	PS	oncogênica	ganho de função	94986	rs121913279

Legenda: CCP, painel de 409 genes; CHP2, painel de 50 genes; NA, não avaliado; ND, não determinada(o); onco, oncogene; GST, gene supressor tumoral; PS, potencialmente somática; VUS, variante de significado clínico incerto; (*) As pacientes 15CDIS e 27CDIS tiveram o DNA de leucócito avaliado em *pools* e as variantes têm maior chance de serem somáticas. No caso 27CDIS, a variante c.524G>A no gene TP53 (VAF 74%) foi avaliada no DNA do leucócito do caso 27CDIS individualmente e confirmada como somática; (**) A amostra de biópsia do caso 05CDIS e a amostra de peça cirúrgica do caso 28CDIS não passaram pelos critérios de qualidade de sequenciamento e suas variantes foram identificadas por busca ativa (inspeção visual).

Realizamos por sequenciamento NGS de amplicon a validação de 6 variantes (Tabela 7) no DNA tumoral. Das 6 variantes, 3 não foram confirmadas (*HNF1A* c.526+1G>A, p. (spl?), VAF 3%; *PIK3CA* c.278G>A, p.Arg93Gln, VAF 3%; *NRAS* c.355G>T, p.Asp119Tyr, VAF 35%) (Tabela 7), justamente aquelas que apresentaram VAF ou cobertura de sequenciamento baixa, e foram excluídas das análises posteriores, restando 106 variantes potencialmente somáticas detectadas em 83 genes em 23 pacientes. Dessas, 45 variantes estavam presentes em 40 genes descritos no Cosmic ou OncoKB como sendo oncogenes/GST (14 oncogenes, 19 GST e 7 oncogenes/GST). O efeito oncogênico das 45 variantes foi investigado na plataforma OncoKB, das quais identificamos 15 variantes classificadas como oncogênicas ou provavelmente oncogênicas em 12 genes, uma variante classificada como inconclusiva e 29 variantes não classificadas (Tabela 7).

Dos 40 genes classificados como oncogenes/GST, 85% (34/40) deles apresentaram variantes potencialmente somáticas em apenas um caso e 15% (6/40) em mais de um caso (Figura 10). O gene *TP53* foi o que apresentou variantes no maior número de pacientes (13%; 3/23); seguido pelos genes *CIC*, *EGFR*, *GATA3*, *PIK3CA* e *TET1*, que apresentaram variantes em 8,7% (2/23) das pacientes, e pelos genes *AKT1*, *KMT2C*, *MAP2K4*, *RB1* e *RUNX1* que apresentaram variantes em 4,3% (1/23) das pacientes (Tabela 7; Figura 10).



Legenda: A frequência absoluta corresponde a quantidade de pacientes que apresentaram alguma variante potencialmente somática no respectivo gene.

Figura 10 - Frequência de pacientes com variantes potencialmente somáticas.

O gene *TP53* apresentou 3 variantes classificadas como oncogênica ou provavelmente oncogênica em 3 pacientes - c.743G>C, p.Arg248Pro (VAF 29%) no caso 12CDIS; c.822_823dupTT, p.Cys275Phefs*71 (VAF 58%) no caso 21CDIS, e a c.524G>A, p.Arg175His (VAF 74%) no caso 27CDIS, esta última confirmada como somática pela avaliação no DNA de leucócito da amostra correspondente. O gene *PIK3CA* apresentou uma variante oncogênica em duas pacientes - c.3140A>G, p.His1047Arg nas pacientes 28CDIS (VAF 31%) e 34CDIS (VAF 14%), a qual foi validada no DNA do tumor de ambas as pacientes. O gene *GATA3* apresentou duas variantes provavelmente oncogênicas em duas pacientes - c.1050+2T>C, p. (spl?) (VAF 38%) no caso 02CDIS e c.614dupT, p.Met205Ilefs*99 (VAF 52%) no caso 21CDIS. O gene *AKT1* apresentou a variante oncogênica c.49G>A, p.Glu17Lys (VAF 46%) no caso 07CDIS; o gene *KMT2C* apresentou a variante com efeito oncogênico não determinado c.2353C>A, p.Pro785Thr (VAF 53%) no caso 06CDIS; o gene *MAP2K4* apresentou a variante provavelmente oncogênica c.339_340delGA, p.Lys114Argfs*28 (VAF 63%) no caso 02CDIS; o gene *RB1* apresentou a variante provavelmente oncogênica c.1654C>T, p.Arg552Ter (VAF 45%) no caso 27CDIS; e o gene *RUNX1* apresentou a variante com efeito oncogênico não determinado c.1073C>T, p.Pro785Thr (VAF 35%) no caso 32CDIS (Tabela 7).

Em relação as variantes que foram identificadas com VAF próximo a 50% nas amostras cujo DNA de leucócito não estava disponível para ser avaliado, não podemos descartar a possibilidade de serem variantes de origem germinativa. Nos chama atenção as 6 pacientes (02CDIS, 03CDIS, 06CDIS, 07CDIS, 21CDIS e 28CDIS) que avaliamos a biópsia e a peça cirúrgica. Nessas 6 pacientes detectamos 4 variantes em duas pacientes do grupo CDIS sem progressão. No caso 06CDIS a variante provavelmente oncogênica c.190_191delGC, p.Ala64Profs*28 no gene *MUTYH* (VAF 53%) e a variante c.2353C>A, p.Pro785Thr no gene *KMT2C* (VAF 53%). E no caso 07CDIS a variante oncogênica c.49G>A, p.Glu17Lys no gene *AKT1* (VAF 46%) e a variante c.7366G>A, p.Val2456Ile no gene *LRP1B* (VAF 48%), com VAF próximo a 50% e com valores parecidos nas duas amostras, sinalizando uma maior possibilidade de serem germinativas, pois fragmentos diferentes do mesmo tumor apresentaram VAFs bastante semelhantes (Tabela 7).

Adicionalmente, avaliamos os genes de predisposição ao câncer de mama e com VAF próximo a 50% para detectarmos variantes classificadas como patogênica/provavelmente patogênica (P/PP) e variantes de significado clínico incerto (VUS, do inglês *variant of uncertain significance*). Identificamos duas variantes no caso 32CDIS do grupo CDIS com

progressão/recidiva. Uma delas é classificada no ClinVar como VUS no GST *RAD50* (c.508A>G, p.Lys170Glu, VAF 56%). O algoritmo REVEL avalia essa variante com escore 0,2, o algoritmo CHASMplus considera a variante como *passenger* e a plataforma Varsome considera a variante como VUS. A outra variante é classificada como conflitante no GST *ATM* (c.8560C>T, p.Arg2854Cys, VAF 44%), a qual foi reportada uma vez como provavelmente benigna e 11 vezes como VUS no ClinVar. O algoritmo REVEL avalia essa variante com escore 0,9, o algoritmo CHASMplus considera a variante como *driver* e a plataforma Varsome considera essa variante como sendo patogênica. O escore REVEL próximo a 1 (escore que varia de 0 a 1) indica maior possibilidade da variante apresentar patogenicidade. Na revisão do prontuário deste caso, foi identificado que a paciente apresenta antecedentes familiares de câncer de mama e mutação no gene *BRCA2*.

Em relação as variantes com VAF próximo de 50% que foram identificadas nas pacientes 15CDIS e 27CDIS cujo DNA de leucócito foi avaliado em *pools*, tais como, a variante c.5768C>T, p.Ala1923Val no gene *CIC* (VAF 49%) e a variante c.1654C>T, p.Arg552Ter no gene *RB1* (45%) (Tabela 7), a possibilidade de serem variantes germinativas é menor, uma vez que os respectivos DNA de leucócitos estavam presentes em *pools* e a qualidade do sequenciamento dessas amostras foi considerada boa, com pelo menos 85% das leituras apresentando cobertura de pelo menos 500x (Tabela 6).

Em relação as variantes identificadas nas 4 pacientes (02CDIS, 06CDIS, 07CDIS e 21CDIS) em que conseguimos avaliar os dois materiais (biópsia e peça cirúrgica) utilizando o mesmo painel, de 19 variantes apenas 3 delas (c.355G>T, p.Asp119Tyr no gene *NRAS*; c.2293G>A, p.Val765Met no gene *EGFR*; c.773G>A, p.Arg258His no gene *PPP2R1A*) foram identificadas em uma amostra, mas não foram identificadas na outra amostra do mesmo caso. Esse resultado demonstrou que as biópsias e peças cirúrgicas da nossa casuística apresentaram uma grande semelhança no repertório mutacional.

4.3.3.1 Variantes potencialmente somáticas classificadas como oncogênicas e provavelmente oncogênicas detectadas em oncogenes e GST em CDIS

Um total de 15 variantes oncogênicas/provavelmente oncogênicas foram identificadas em 12 genes (Apêndice 3) em 11 pacientes, sendo 6 variantes em 5 pacientes do grupo 1 (2CDIS, 5CDIS, 6CDIS, 7CDIS e 12CDIS), e 9 variantes em 6 pacientes do grupo 2 (21CDIS, 27CDIS, 28CDIS, 31CDIS, 32CDIS e 34CDIS). A variante oncogênica no gene *PIK3CA*

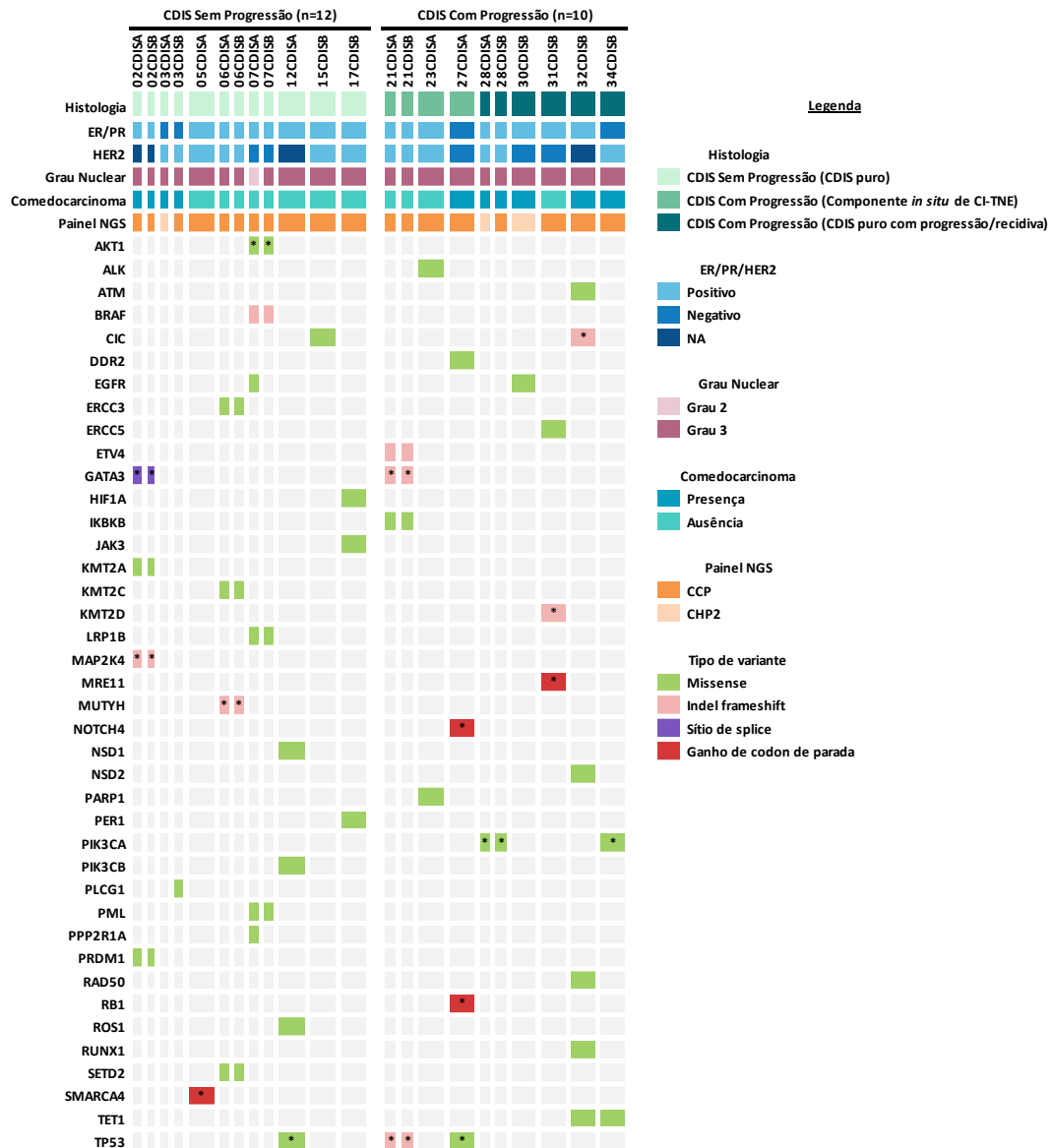
(c.3140A>G), conhecida como uma mutação ativadora em região *hotspot*, foi detectada em duas pacientes do grupo 2.2 (CDIS/progressão/recidiva) (28CDIS (VAF 31%) e 34CDIS (VAF 16%)) (Tabela 7).

No grupo 1, das 6 variantes potencialmente somáticas classificadas como oncogênicas/provavelmente oncogênicas, identificamos a variante c.743G>C, p.Arg248Pro (VAF 29%) no gene *TP53* no caso 12CDIS com maior possibilidade de ser somática, pois embora não teve sua origem confirmada devido a indisponibilidade do DNA de leucócito desse caso, ela foi detectada numa frequência alélica menor do que 50%.

No grupo 2, das 9 variantes potencialmente somáticas classificadas como oncogênicas/provavelmente oncogênicas, detectamos a variante somática c.524G>A, p.Arg175His (VAF 74%) no gene *TP53* no caso 27CDIS, cujo DNA de leucócito foi avaliado individualmente; e outras 3 variantes com maior possibilidade de serem somáticas, as variantes que resultam em códon de parada prematuro nos genes *NOTCH4* (c.4465C>T, p.Arg1489Ter, VAF 25%) e *RB1* (c.1654C>T, p.Arg552Ter, VAF 45%), ambas identificadas no caso 27CDIS cujo DNA de leucócito foi avaliado em *pool*, e a variante missense no gene *PIK3CA* (c.3140A>G, p.His1047Arg) nas pacientes 28CDIS (VAF 31%) e 34CDIS (VAF 14%), ambas validadas no DNA do tumor.

De forma geral, comparando o número de pacientes com variantes oncogênicas/provavelmente oncogênicas nos dois grupos, o grupo 1 (CDIS sem progressão) apresentou 33% (5/15) das pacientes avaliados e o grupo 2 (CDIS com progressão) apresentou 75% (6/8) das pacientes avaliados com variantes oncogênicas/provavelmente oncogênicas (Figura 11).

Não houve associação estatisticamente significativa entre a quantidade de mutações potencialmente somáticas nos genes *TP53* e *PIK3CA* e os grupos de CDIS. Mutações potencialmente somáticas foram observadas no gene *TP53* em 6,6% (1/15) do grupo 1 e em 25% (2/8) do grupo 2 (P-valor = 0,2909). No gene *PIK3CA*, mutações potencialmente somáticas ativadoras não foram observadas no grupo 1 e foram observadas em 25% (2/8) do grupo 2 (P-valor = 0,1212).



Legenda: Variantes não sinônimas em oncogenes/genes supressores tumorais identificados nos grupos CDIS sem progressão e CDIS com progressão pelos painéis multigenes CCP e CHP2. As colunas representam os tumores e as linhas representam os genes. As características clínico-patológicas das pacientes e os painéis utilizados estão no topo. Os tipos de variantes estão codificados por cores de acordo com a legenda. O asterisco (*) indica variantes oncogênicas ou provavelmente oncogênicas. A amostra de biópsia do caso 05CDIS e a amostra de peça cirúrgica do caso 28CDIS não passaram pelo controle de qualidade, suas variantes foram identificadas por busca ativa (inspeção visual).

Figura 11 - Repertório de variantes em CDIS.

4.3.3.2 Variantes potencialmente somáticas de significado clínico incerto (VUS) detectadas em oncogenes e GST em CDIS

Das 29 variantes não classificadas no OncoKB, identificamos 3 variantes classificadas como VUS nos oncogenes *ALK* (c.1652G>A, p.Arg551Gln, VAF 51%), *EGFR* (c.2293G>A, p.Val765Met, VAF 1%) e *IKBKB* (c.1336C>T, p.Arg446Trp, VAF 11%), todas reportadas no Cosmic. A variante em *EGFR* foi validada no DNA tumoral das pacientes 07CDIS (VAF 1%) e 30CDIS (VAF 1%).

4.3.4 Análise de enriquecimento funcional

Com o intuito de identificarmos as vias bioquímicas mais representadas pelo conjunto de 40 oncogenes/GST acometidos pelas 45 variantes identificadas em nosso estudo, realizamos uma análise de enriquecimento no programa STRING, o qual avalia o banco de dados KEGG. Nesta análise, identificamos 113 vias enriquecidas, das quais a via de reparo de DNA por união de extremidades não homólogas foi a que apresentou o maior escore de enriquecimento. Das 45 variantes, 7 (15,5%) foram encontradas em 7 genes que participam de vias metabólicas de reparo de DNA: união de extremidades não homólogas (*non-homologous end joining* (NHEJ)), recombinação homóloga (*homologous recombination* (HR)), reparo por excisão de base (*base excision repair* (BER)) e reparo por excisão de nucleotídeo (*nucleotide excision repair* (NER)) (Tabela 8).

Tabela 8 - Vias de reparo de DNA enriquecidas.

Vias canônicas	Genes na rede	Escore de enriquecimento	FDR
União de extremidades não homólogas	MRE11A, RAD50	1,89	0,0013
Recombinação homóloga	ATM, MRE11A, RAD50	1,55	0,0003
Reparo por excisão de base	MUTYH, PARP1	1,45	0,0076
Reparo por excisão de nucleotídeo	ERCC3, ERCC5	1,31	0,0137

Além disso, identificamos 90 vias enriquecidas (Apêndice 4) pelos 12 genes alterados (Apêndice 3) pelas 15 variantes oncogênicas/provavelmente oncogênicas.

5. DISCUSSÃO

Neste estudo realizamos uma avaliação molecular (transcricional e mutacional) de uma casuística composta por pacientes com CDIS que apresentaram diferentes desfechos no tempo de seguimento. Esta casuística é de difícil seleção e as amostras são de difícil processamento, principalmente as biopsias emblocadas em parafina. Além disso, utilizamos critérios rigorosos na análise dos resultados observados. Estes fatores são pontos fortes do nosso estudo.

Em relação a avaliação transcricional, a qual foi realizada em uma casuística composta por amostras de células epiteliais neoplásicas e mioepiteliais capturadas de lesões pré-invasivas que apresentam diferentes potenciais malignos (CDIS puro e componente *in situ* de CMI-TNE), nossos resultados corroboraram os resultados observados nos estudos prévios do grupo (Castro et al. 2008; Sens-Abuázar et al. 2012; Elias et al. 2016) em relação ao padrão de diferença de expressão gênica. Neste estudo confirmamos que há um padrão geral de pequenas diferenças de expressão entre os grupos de amostras e que a maioria dos genes avaliados apresenta uma menor expressão nas amostras do componente *in situ* de CMI-TNE em relação ao CDIS puro (Figura 7), reforçando nossa observação anterior de que as etapas iniciais da progressão de CDIS parecem ocorrer preferencialmente por uma diminuição de expressão gênica (Castro et al. 2008, Elias et al. 2016).

Em relação a análise de expressão gênica das amostras de células epiteliais neoplásicas e mioepiteliais de CDIS puro e de componente *in situ* de CMI-TNE, pudemos confirmar a diferença de expressão em 49% (26/53) dos genes analisados. Contudo, uma limitação desta etapa do nosso estudo foi a impossibilidade de incluir novas amostras e precisarmos trabalhar com amostras de estudos anteriores do grupo, ou seja, com amostras não-independentes (Castro et al. 2008, Pinilla-Fernández 2015, Elias et al., em preparação). E nem todas as amostras que já haviam sido processadas e avaliadas nos estudos anteriores do nosso grupo tinham cDNA disponível e com qualidade para sua utilização neste estudo. Assim, a quantidade de amostras que passaram pelos critérios de qualidade rigorosos que adotamos foi extremamente pequeno. Isso ressalta a importância de uma validação desses dados numa casuística maior e com amostras independentes.

Contudo, neste estudo conseguimos identificar um grupo menor de genes e aparentemente mais robustos em relação aos estudos anteriores do nosso grupo, pois selecionamos 78 genes, dos quais 20 genes (Elias et al. 2016) haviam sido identificados por *microarray* e confirmados por RT-qPCR (ensaios TLDA) como diferencialmente expressos em amostras de células epiteliais neoplásicas capturadas de CDIS puro e de componente *in situ* de CMI-TNE, e 58 genes (Elias et al., em preparação) que haviam sido identificados por RNAseq como diferencialmente expressos em amostras de células epiteliais neoplásicas e mioepiteliais capturadas de CDIS puro e de componente *in situ* de CMI-TNE, e conseguimos chegar numa assinatura composta por apenas 12 genes cujo perfil de expressão foi capaz de distinguir amostras de células epiteliais neoplásicas e mioepiteliais de lesões pré-invasivas com diferentes potenciais malignos (CDIS puro e componente *in situ* de CMI-TNE). Além disso, nosso estudo foi uma validação do estudo anterior de RNAseq (Elias et al., em preparação) com amostras semi-independentes, pois 9 das 15 amostras avaliadas eram diferentes em relação ao estudo anterior.

Nosso estudo também confirmou os resultados observados no estudo de RNAseq (Elias et al., em preparação) quanto ao padrão de agrupamento das amostras de células epiteliais neoplásicas e mioepiteliais capturadas das lesões pré-invasivas que apresentam diferentes potenciais malignos (CDIS puro e amostras de componente *in situ* de CMI-TNE). Elias e colaboradores (em preparação), ao avaliarem por RNAseq a porção 3' dos transcritos, analisaram a diferença de expressão das células epiteliais neoplásicas e mioepiteliais capturadas a laser dos dois tipos de lesões pré-invasivas (CDIS puro e componente *in situ* de CMI-TNE). Os resultados mostraram que o padrão de expressão dos GDE entre os dois tipos de lesões pré-invasivas (CDIS puro e componente *in situ* de CMI-TNE) identificados tanto na avaliação das células epiteliais neoplásicas, como mioepiteliais, assim como os GDE comuns dos dois tipos de células, foram igualmente capazes de distinguir amostras de CDIS puro das amostras de componente *in situ* de CMI-TNE, assim como, a amostra de CDIS puro do caso que manifestou progressão da doença. Em nosso estudo, nós fomos capazes de reproduzir resultados semelhantes utilizando uma assinatura gênica baseada em somente 12 genes.

O perfil de expressão dos 12 genes foi capaz de distinguir amostras de CDIS puro das amostras de componente *in situ* de CMI-TNE, e também foi capaz de discriminar um caso de CDIS referente a uma paciente que apresentou progressão da doença 3 anos após o tratamento convencional. A amostra de células epiteliais neoplásicas 60-E agrupou com as

amostras de componente *in situ* de CMI-TNE, as quais apresentam um maior potencial maligno. Contudo, nessa avaliação por RNAseq as amostras de células epiteliais neoplásicas e mioepiteliais (60-E e 60-M) agruparam juntas. Já em nosso estudo, tais amostras não foram agrupadas juntas, ao invés disso, a amostra de células epiteliais neoplásicas 60-E agrupou com as amostras de componente *in situ* de CMI-TNE e a amostra de mioepiteliais 60-M agrupou com as amostras de CDIS puro. Já é sabido que as células mioepiteliais têm um papel de supressor de invasão tumoral (Polyak and Hu 2005; Mitchell et al. 2020). E os resultados baseados no perfil de expressão dos 12 genes sugerem que a programação molecular identificada nas amostras de células epiteliais neoplásicas pode ter potencial de prever a progressão para doença invasiva, o que não parece ocorrer nas amostras enriquecidas com células mioepiteliais. Entretanto, é fundamental destacar que o estudo atual realizou uma avaliação transcricional de uma casuística bastante pequena. Então, é importante considerar que os resultados observados devem ser interpretados com bastante cautela, pois podem ter sido consequência de um fenômeno aleatório.

Os 12 genes que compõem nossa assinatura estão relacionados a processos biológicos relevantes à tumorigênese, tais como, reparo de quebra de dupla-fita de DNA via recombinação homóloga (*REC8*), reparo por excisão de nucleotídeo acoplado à transcrição (*ERCC8*), resposta celular à estímulo estrogênico (*SRA1*), regulação negativa da proliferação celular (*RARRES3*), regulação positiva da adesão celular (*FLCN*), organização da matriz extracelular (*DNAJB6*), diferenciação de células progenitoras hematopoiéticas (*RBM47*), via de sinalização do receptor do fator de crescimento endotelial vascular (*NCF4*), processamento de rRNA (*RRP1*), divisão celular (*ANAPC13*), e ainda, processos biológicos desconhecidos (*BAALC* e *TMEM45A*).

Dos 12 genes, *DNAJB6* e *RBM47* foram os únicos que apresentaram expressão diminuída em CDIS puro em relação ao componente *in situ* de CMI-TNE, assim como, apresentaram uma expressão diminuída em CDIS puro em relação ao tecido normal (Apêndice 1). E no grupo de carcinomas invasivos (CMI-TNE), a expressão de ambos os genes foi encontrada como aumentada em relação ao componente *in situ* de CMI-TNE, evidenciando um aumento de suas expressões ao longo do processo de progressão de CDIS. Esses resultados não são corroborados pelos dados da literatura que avaliaram a expressão dos genes *DNAJB6* e *RBM47*. Contudo, os trabalhos que avaliaram a expressão desses 2 genes, avaliaram pacientes e linhagens celulares de câncer de mama de subtipos agressivos e metastáticos, e a

maioria da casuística submetida a avaliação transcricional em nosso estudo foi composta por pacientes do subtipo luminal.

O gene *DnaJ heat shock protein family (Hsp40) member B6 (DNAJB6)* codifica um membro bastante conservado da família DnaJ/Hsp40 de proteínas co-chaperonas, as quais atuam em conjunto com as proteínas Hsp70 para atuarem na conformação tridimensional correta de outras proteínas (Mitra et al. 2008). Existem duas isoformas de DNAJB6, a isoforma maior é encontrada principalmente no núcleo celular (DNAJB6a) e a isoforma menor é encontrada principalmente no citoplasma (DNAJB6b). DNAJB6a é capaz de reverter o processo de transição epitélio-mesênquima, pois é capaz de aumentar a expressão de DKK1, um inibidor da via de sinalização Wnt, sugerindo que DNAJB6a atua na manutenção de características morfológicas epiteliais (Meng et al. 2016). Mitra e colaboradores (2008) avaliaram por RT-qPCR o nível de expressão do gene *DNAJB6a* nas linhagens celulares MDA-MB-231 e MDA-MB-435 transfectadas com plasmídeos construídos para expressar *DNAJB6a*, encontrando baixos níveis de expressão, o qual foi confirmado no nível proteico por *immunoblotting*. E através de *tissue microarray* (TMA) encontraram a expressão proteica de DNAJB6a progressivamente diminuída nas pacientes de câncer de mama progressivamente mais avançados. Nos ensaios *in vitro* os autores encontraram que a expressão de *DNAJB6a* esteve associada com menor migração e invasão das linhagens celulares.

O gene *RNA binding motif protein 47 (RBM47)* codifica uma proteína que é um membro de uma família proteica responsável por se ligar a moléculas de RNA, as proteínas ligantes de RNA (*RNA binding protein – RBP*), as quais são as principais responsáveis por uma regulação gênica pós-transcricional (Shivalingappa et al. 2021). Vanharanta e colaboradores (2014) sugeriram que *RBM47* seja um gene supressor tumoral e que a inativação da sua atividade está associada com a progressão do câncer de mama. Os autores realizaram uma análise integrada combinando dados de expressão gênica de linhagens celulares de câncer de mama triplo negativo metastático obtidos dos trabalhos de Minn e colaboradores (2005) e Bos e colaboradores (2009), com dados clínicos de uma coorte de 368 pacientes de câncer de mama dos trabalhos de Minn e colaboradores (2005) e Wang e colaboradores (2005), com dados mutacionais de pacientes de câncer de mama basaloide com metástase cerebral e pulmonar do trabalho de Ding e colaboradores (2010). Os autores encontraram o gene *RBM47* com baixo nível de expressão, com mutações que geram uma proteína truncada e associado com recidiva em cérebro e pulmões.

Os demais 10 genes da nossa assinatura tiveram sua expressão aumentada em CDIS puro em relação ao componente *in situ* de CMI-TNE, assim como, apresentaram uma expressão aumentada em CDIS puro em relação ao tecido normal (Apêndice 1). Esses dados são corroborados por outros estudos da literatura que avaliaram a expressão desses genes.

O gene *anaphase promoting complex subunit 13* (*ANAPC13*) codifica um componente do complexo promotor de anáfase que controla a progressão do ciclo celular. Ele não passou pelos nossos critérios de validação ($\text{fold change} \geq |2|$) devido a apenas um único valor aberrante de expressão em uma única amostra. Caso desconsiderássemos esse valor aberrante, ele teria sido validado (fold change : 2,53) com significância estatística (P-valor: 0,01). Nosso grupo tem investigado o papel de *ANAPC13* no processo de progressão de CDIS. Nesses estudos foi identificado por *cDNA microarray* e RT-qPCR sua expressão gênica diminuída em amostras pré-invasivas com maior potencial maligno, ou seja, em amostras de componente *in situ* de CMI-TNE em relação a amostras de CDIS puro (Castro et al. 2008; Sens-Abuázar et al. 2012; Elias et al. 2016). Além disso, também foi identificado por imunohistoquímica que a expressão proteica de *ANAPC13* também está diminuída em amostras de componente *in situ* de CMI-TNE em relação a amostras de CDIS puro. Inclusive, também foram observadas maiores taxas de sobrevida global e sobrevida livre de doença em pacientes de carcinoma ductal invasivo com expressão proteica positiva de *ANAPC13* (Sens-Abuázar et al. 2012). O nível de expressão de *ANAPC13* também está sendo estudado em ensaios com culturas celulares com resultados promissores que comprovam que células com a expressão de *ANAPC13* diminuída demonstra maior capacidade de invasão (manuscrito em preparação).

O gene *BAALC binder of MAP3K1 and KLF4* (*BAALC*) foi identificado por estudos de expressão gênica em pacientes com leucemia mielóide aguda (*acute myeloid leukaemia* – AML) e é considerado como tendo um papel de oncogene em AML através da ativação da via de sinalização ERK (Morita et al. 2015). Weber e colaboradores 2016 avaliaram por PCR quantitativo um total de 632 amostras de medula óssea e de sangue periférico de 142 pacientes com leucemia mielóide aguda citogeneticamente normais (*cytogenetically normal acute myeloid leukaemia* – CN-AML) e encontraram uma associação entre a expressão aumentada de *BAALC* e uma menor sobrevida livre de doença e menor sobrevida global. Birgersson e colaboradores (2021) realizaram uma avaliação *in silico* da expressão gênica de

BAALC com os dados do TCGA de 960 pacientes com carcinoma de mama invasivo, e realizaram ensaios *in vitro* de proliferação celular, migração e invasão tumoral para avaliar o efeito funcional da expressão de *BAALC* em câncer de mama. Os autores encontraram que a expressão de *BAALC* é progressivamente aumentada em tumores primários e metástases quando comparado com tecidos normais de mama; sua expressão gênica aumentada está associada com uma redução na sobrevida livre de metástase a distância; e que sua expressão proteica aumentada em linhagens celulares MCF-7 aumenta a proliferação celular, migração e invasão tumoral. Os autores ainda mostraram que a inativação por RNA de interferência (siRNA) da expressão de *BAALC* resulta em diminuição da proliferação celular, migração e invasão tumoral em linhagens celulares Hs578T.

O gene *ERCC excision repair 8, CSA ubiquitin ligase complex subunit (ERCC8)* é um membro da via de reparo do DNA por excisão de nucleotídeo (*nucleotide excision repair – NER*). Uma meta-análise realizada por Moslehi e colaboradores (2020), que avaliou dados de *cDNA microarray*, encontrou uma expressão aumentada de *ERCC8* em câncer de mama invasivo em comparação com tecido normal de mama.

O gene supressor tumoral *folliculin (FLCN)* expressa uma proteína que é ligante de AMPK e ativadora de RagC/D-GTPase envolvida na via de sinalização mTOR. El-Houjeiri e colaboradores (2021) desenvolveram diversas linhagens celulares de câncer de mama luminal que tinham o gene *FLCN* depletado. Os autores encontraram uma associação entre a inibição da sua expressão e um estímulo ao crescimento tumoral através da indução de diversas vias de sinalização, tais como, autofagia, biogênese lisossomal, glicólise e angiogênese.

O gene *neutrophil cytosolic factor 4 (NCF4)* expressa a proteína p40phox, que é um componente regulatório do complexo NADPH oxidase, um sistema enzimático multicomponente cuja principal função é imunológica. Ao nosso conhecimento, evidências sobre a alteração de sua expressão em câncer de mama ainda não foram relatadas na literatura.

O gene *Retinoic acid receptor responder 3 (RARRES3)*, também conhecido como *phospholipase A and acyltransferase 4 (PLAAT4)*, expressa uma proteína que possui atividades que regulam o crescimento celular, diferenciação, apoptose, através da inibição da via Ras/MAPK (Tsai et al. 2007). Hsu e colaboradores (2015) avaliaram as linhagens celulares MCF-7 e MDA-MB231 de câncer de mama transfectadas com vetores para aumentarem ou inibirem a expressão de *RARRES3*. Os autores encontraram que o aumento da sua expressão diminuiu

a proliferação celular em ambas as linhagens celulares. Ao contrário, a inibição da sua expressão aumentou a proliferação das linhagens celulares. Além disso, o grupo também demonstrou que a expressão aumentada de RARRES3 é capaz de inibir a fosforilação de EGFR e as transduções de sinais mediadas por EGFR, tais como, a fosforilação de Raf/ERK1/2, PI3K/Akt; enquanto a inibição da atividade de RARRES3 aumenta a atividade dessas vias de sinalização.

O gene *REC8 meiotic recombination protein (REC8)* codifica um membro da família proteica SMC de manutenção da estrutura cromossômica. Diferentes estudos que avaliaram diferentes tipos de tumor têm encontrado resultados conflitantes a respeito do papel deste gene no câncer. Embora ele já tenha sido encontrado como uma atividade promotora de migração, invasão e angiogênese em carcinoma hepatocelular (Han et al. 2021), ele também já foi encontrado como tendo um papel de supressor de angiogênese em linhagens celulares de câncer gástrico (Liu et al. 2020). E ainda, Liu e colaboradores (2015) encontraram que *REC8* é metilado pela via de sinalização PI3K em câncer de tireoide, sugerindo-o como um novo gene supressor tumoral.

O gene *ribosomal RNA processing 1 (RRP1)* codifica uma proteína que atua no processo de biogênese de ribossomos. O seu homólogo RRP1B já foi demonstrado como tendo papel no processo de progressão tumoral e metástase em câncer de mama. Crawford e colaboradores (2007) transfectaram as linhagens celulares Mvt-1 e 4T1 com vetor de expressão contendo o gene RRP1B para obterem uma expressão aumentada de RRP1B. Ao transfectarem essas células em camundongos, eles identificaram uma associação de expressão aumentada de RRP1B com progressão tumoral e metástase.

O gene *steroid receptor RNA activator 1 (SRA1)* codifica RNAs codificadores e não codificador de proteínas, os quais são capazes de interagir com diversas proteínas ligantes de RNA e com fatores de transcrição através de estruturas secundárias (Lanz et al. 1999). Estes complexos proteicos incluem múltiplos receptores nucleares e proteínas envolvidas em silenciamento gênico, sendo considerado um importante regulador transcricional e potencialmente envolvido em diversos mecanismos, tais como, tumorigênese e progressão tumoral (Leygue et al. 1999). Yan e colaboradores (2015) avaliaram por RNAseq e RT-qPCR a expressão gênica de linhagens celulares HeLa (câncer de colo de útero) e MDA-MB-231 (câncer de mama) submetidas a duas condições, transfectadas com vetores para aumentar ou inibir a expressão de SRA1. Os autores encontraram que o movimento celular é

potencialmente regulado por SRA1. E através de ensaios de motilidade celular, observaram que a inibição e aumento da expressão gênica de SRA1 resulta em diminuição e aumento da motilidade das células tumorais, respectivamente.

O gene *transmembrane protein 45A* (*TMEM45A*) pertence a uma grande família de genes que codificam proteínas preditas como proteínas transmembrana. Diversos estudos têm reportado alterações na expressão de *TMEM45A* em diferentes tipos de tumores. A expressão aumentada de *TMEM45A* tem sido mostrada como um evento que favorece a resistência a quimioterapia em linhagens celulares de câncer de mama e fígado em condições de hipóxia (Flamant et al. 2012). E esse aumento de expressão tem sido associado com um pior prognóstico em pacientes com câncer de mama (Flamant et al. 2012), bexiga (Urquidí et al. 2012) e ovário (Crijns et al. 2009). Guo e colaboradores (2015) silenciaram o gene *TMEM45A* em duas linhagens celulares (HO-8910 e A-2780) de câncer de ovário e encontraram que a inibição da sua expressão está associada a uma diminuição da proliferação celular, indução de uma parada do ciclo celular e diminuição da invasão tumoral. Além do gene *ANAPC13*, o gene *TMEM45A* também teve o seu papel avaliado em CDIS. Lee e colaboradores (2012) capturaram por microdissecção a laser as células epiteliais neoplásicas de 53 tumores de CDIS (44 amostras de CDIS puro e 9 amostras de componente *in situ* de CMI-TNE) e de 51 tumores de CMI-TNE. Ao avaliarem a expressão gênica desses tumores por *cDNA microarray* compararam as diferenças entre os grupos de amostras de lesões pré-invasivas (CDIS e componente *in situ* de CMI-TNE) e invasivas (CMI-TNE) e identificaram um grupo de 74 genes diferencialmente expressos, cujo perfil de expressão foi capaz de discriminar as amostras de CDIS das amostras de CMI-TNE. Um total de 10 genes (validados por RT-qPCR) foram selecionados para serem submetidos a uma avaliação do efeito da sua expressão. Para isso os autores utilizaram linhagens celulares de CDIS (linhagens CDIS.COM e SUM225) para criar um modelo animal de xenoenxerto com camundongos que receberam essas linhagens celulares. Ao avaliarem a progressão de CDIS para CMI-TNE nesse ensaio *in vivo*, encontraram que a progressão foi significativamente aumentada quando a expressão do gene *TMEM45A* foi inibida pela transfecção com vetores virais contendo RNA de interferência.

Outro achado interessante em relação aos 12 genes, é que 10 (*ANAPC13*, *BAALC*, *ERCC8*, *FLCN*, *NCF4*, *RARRES3*, *REC8*, *RRP1*, *SRA1* e *TMEM45A*) deles apresentaram uma expressão diminuída nas amostras com maior potencial maligno (componente *in situ* de CMI-TNE). Estes genes participam de processos biológicos relacionados com a tumorigênese, tais

como, reparo de DNA, regulação negativa da proliferação celular, regulação positiva da adesão celular, organização da matriz extracelular, divisão celular, entre outros. Baseando-se no padrão de expressão desses genes, estes resultados sugerem que tais processos biológicos podem estar sendo menos ativados em nível transcricional nas amostras com maior potencial maligno. Desta forma, as células do componente in situ de CMI-TNE poderiam apresentar uma menor atividade de reparo de DNA, maior atividade proliferativa, menor atividade de adesão celular, menor organização da matriz extracelular e menor controle da divisão celular.

Além disso, um outro subgrupo de 10 (*ANAPC13*, *BAALC*, *DNAJB6*, *ERCC8*, *RARRES3* (*PLAAT4*), *RBM47*, *REC8*, *RRP1*, *SRA1* e *TMEM45A*) dos 12 genes formam uma rede de interação. Esta rede está relacionada com as funções de movimento celular, desenvolvimento celular, e desenvolvimento do sistema cardiovascular.

E ao investigarmos possíveis vias bioquímicas enriquecidas pelo pequeno grupo de 12 genes, identificamos 5 vias (via de reparo por excisão de nucleotídeo, via de sinalização de fosfolipase D, via cinase tipo-polo, via de eicosanoides e via de sinalização MAPK) enriquecidas por eles.

A via de reparo por excisão de nucleotídeo é uma das vias de reparo de DNA que protegem as células contra danos genômicos e sua perda de capacidade de reparo de DNA resulta em instabilidade genômica. No estudo de Latimer e colaboradores (2010), os autores encontraram que a capacidade de reparo de DNA desta via estava diminuída em câncer de mama esporádico em estágio inicial em relação ao tecido de mama normal.

A via de sinalização de fosfolipase D é uma das vias de fosfolipases que mediam sinalizações intracelulares e extracelulares, convertendo fosfolipídios em mediadores lipídicos que regulam múltiplos processos celulares, tais como, proliferação, migração, invasão e angiogênese. Esta via já foi encontrada como ativada em culturas celulares de câncer de mama altamente proliferativas (Ye et al. 2013).

As cinases tipo-polo (*polo-like kinase* - PLK) são uma família conservada de 5 membros, cujas proteínas estão envolvidas em vias de transdução de sinal relacionadas a formação e manutenção do fuso mitótico, e com isso estão envolvidas no controle da progressão do ciclo celular, e consequentemente, com o crescimento e proliferação celular. Além disso, evidências sugerem sua participação também em processos de reparo do DNA, autofagia, apoptose e transição epitélio-mesênquima (Iliaki et al. 2021). Karlin e colaboradores (2014) avaliaram os dados do TCGA de 773 tumores primários e de 107 tecidos normais de mama, e

identificaram que a cinase tipo 1 (PLK1) tem sua expressão aumentada em câncer de mama subtipo triplo-negativo.

Os eicosanoides consistem em lipídios bioativos com 20 átomos de carbono derivados do metabolismo de ácidos graxos poli-insaturados, tais como o ácido araquidônico, o qual pode ser metabolizado através de 3 vias bioquímicas, a via da ciclo-oxigenase, a via da lipo-oxigenase e a via epoxigenase P450. Os eicosanoides, tais como os prostanoides (prostaglandinas e tromboxanos) podem modular diversos processos biológicos, como por exemplo, a proliferação celular, adesão celular, migração celular, angiogênese, entre outros (Gomes et al. 2018). Watkins e colaboradores (2005) avaliaram por RT-qPCR a expressão gênica do eicosanoide tromboxano A2 em 120 carcinomas mamários e 32 tecidos mamários normais e encontraram uma associação significativa entre os níveis elevados do mRNA de tromboxano A2 e tumores com alto grau e menor sobrevida livre de doença.

A via de sinalização MAPK (*mitogen activated protein kinase*) é uma via de sinalização que promove a proliferação tumoral, sobrevivência, invasão, processo de transição epitélio-mesênquima e metástase, podendo estar ativada em diversos tipos de tumores em estágios iniciais e avançados (Najafi et al. 2019). A via de sinalização MAPK tem um papel central no estado proliferativo da célula tumoral, sendo a via canônica composta por 4 vias que são classificadas de acordo com o membro da família de MAPK presente ao final da cascata de fosforilação, são elas: ERK 1 e 2 (*extracellular signal-related Kinases*), JNK 1, 2 e 3 (*c-JUN N-terminal kinase*), p-38-MAPK e ERK5. Fatores de crescimento, tais como o EGFR, são considerados os principais reguladores da via de ERK 1 e 2, e o estresse celular é o principal indutor das vias JNK e p38-MAPK, enquanto estes dois fatores iniciadores são capazes de ativar a cascata de ERK5. Essas cascatas de sinalização, no núcleo, ativam fatores de transcrição como c-Jun c-Fos e c-Myc que promovem a transcrição de genes que regulam o ciclo celular, modulando assim, a proliferação e a sobrevivência celular (Pritchard e Hayward 2013). A família do receptor do fator de crescimento epidermal (*epidermal growth factor receptor* - EGFR) desencadeia sinais proliferativos e antiapoptóticos, e a cascata de transdução de sinal frequentemente ativada em câncer de mama é iniciada pela família EGFR e seus 3 membros homólogos da família (ERBB2, ERBB3 e ERBB4), caracterizados por um sítio de ligação extracelular, uma região transmembrana e um domínio tirosino-quinase citoplasmático. A ativação do EGFR é decorrente da ligação de fatores relacionados ao EGF (*epidermal growth factor*), que induzem a homo ou heterodimerização do receptor e a

estimulação de vias tirosino-quinase intrínsecas. Esse estímulo leva à auto-fosforilação dos resíduos de tirosina no domínio citoplasmático do receptor, criando sítios de ligação para proteínas adaptadoras e enzimas das cascatas da MAPK e PI3K, ligadas à proliferação celular, angiogênese, invasão, metástases e inibição do processo de apoptose (Najafi et al. 2019).

Em relação a avaliação mutacional, uma casuística composta por 34 pacientes de CDIS com diferentes desfechos, representados por 43 amostras, foram submetidos a extração de DNA. Das 43 amostras, 84% (36/43) foram amostras fixadas em formalina e embebidas em parafina (FFPE) (Figura 5). E embora a extração de DNA com qualidade de amostras FFPE seja menos desafiador do que a extração de RNA, também tivemos dificuldade na obtenção de DNA com qualidade para esse tipo de análise. O fato de não termos conseguido obter quantidade suficiente de DNA de 7 amostras para avaliação por painéis multigenes ocorreu pelo fato de que o tecido mamário apresentou grande quantidade de tecido adiposo, o qual já se tem conhecimento de que atrapalha bastante a extração de DNA. Além disso, CDIS é uma lesão tipicamente pequena, frequentemente diagnosticada precocemente, e na nossa casuística grande parte do material disponível para análise molecular era oriundo de biópsias pequenas. Além disso, o DNA de amostras FFPE apresentam maior nível de degradação, dificultando ainda mais a obtenção de resultados moleculares de alta qualidade. Dentre as 7 amostras cujo DNA extraído foi insuficiente para ser avaliado pelos painéis multigenes todas eram amostras FFPE, e 5 delas eram material de biópsia.

Contudo, conseguimos analisar 80% (29/36) das amostras submetidas ao sequenciamento de DNA, e destas, 83% (24/29) foram avaliadas com o painel amplo de 409 genes inteiros (painel CCP). Consequentemente, como a grande maioria das amostras foi avaliada por um painel amplo de genes, não conseguimos obter uma alta cobertura de sequenciamento para todas as amostras. Inclusive, 70% (17/24) dessas amostras eram FFPE. Diferentemente, as 5 amostras avaliadas pelo painel de 50 genes (CHP2) apresentaram uma alta cobertura de sequenciamento para todas as amostras.

Embora não conseguimos obter uma alta cobertura de sequenciamento em todas as amostras avaliadas pelo painel CCP, a grande maioria das variantes detectadas foram identificadas nas amostras avaliadas por este painel. Das 45 variantes que foram identificadas nos 40 oncogenes/GST, apenas duas variantes (c.3140A>G, p.His1047Arg no gene PIK3CA (VAF 31%) no caso 28CDIS e c.2293G>, p.Val765Met no gene EGFR (VAF 3%) no caso 30CDIS) foram identificadas pelo painel CHP2. Inclusive, as posições genômicas de ambas as variantes

também são cobertas pelo painel CCP, o qual possibilitou a identificação dessas duas variantes em outras duas pacientes, a variante c.3140A>G, p.His1047Arg no gene PIK3CA (VAF 14%) no caso 34CDIS e a variante c.2293G>, p.Val765Met no gene EGFR (VAF 6%) no caso 07CDIS. Esses resultados demonstraram que a utilização de um painel multigenes amplo, como o painel de 409 genes que utilizamos para avaliar a grande maioria das amostras em nosso estudo, possibilitou uma avaliação muito mais adequada do perfil mutacional de CDIS. Cosimo e colaboradores (2020) também realizaram uma avaliação do perfil mutacional de CDIS, avaliando uma casuística com 27 pacientes de CDIS puro, das quais 3 pacientes eram de CDIS puro com progressão da doença. Os autores utilizaram os mesmos 2 painéis multigenes que utilizamos em nosso estudo. Inicialmente utilizaram o painel CHP2, e posteriormente o painel CCP para reavaliar as pacientes previamente avaliados pelo painel CHP2 em que não havia sido identificada nenhuma variante somática.

Em nosso estudo foram identificados 53% (8/15) das pacientes do grupo CDIS puro sem progressão com pelo menos uma variante potencialmente somática em oncogenes ou genes supressores de tumor. Este percentual é semelhante ao que foi relatado no trabalho de Cosimo e colaboradores (2020), que avaliaram 24 pacientes de CDIS puro sem progressão e identificaram variantes somáticas em 50% (12/24) das pacientes avaliadas. Vale mencionar que uma limitação do nosso estudo neste contexto foi o período curto de seguimento deste grupo de pacientes. Em relação ao grupo de CDIS puro que apresentou progressão ou recidiva da doença (grupo 2.2), identificamos 100% (5/5) das pacientes com pelo menos uma variante potencialmente somática. Este percentual foi superior ao encontrado por Cosimo e colaboradores (2020), que avaliaram 3 pacientes de CDIS puro que apresentaram progressão da doença e encontraram variantes somáticas em 67% (2/3) das pacientes.

Vale mencionar que embora conseguimos avaliar apenas 5 pacientes de CDIS que manifestaram progressão ou recidiva, foram investigados 560 prontuários eletrônicos de pacientes do ACCCC para conseguirmos selecionar 7 pacientes, dos quais conseguimos analisar os dados de sequenciamento de DNA com qualidade de 5 pacientes. Esta pequena proporção de 1,3% (7/560) de pacientes de CDIS que manifestaram progressão ou recidiva corrobora os achados do estudo de Cagnacci-Neto (2014). Em sua dissertação de mestrado, em que também avaliou retrospectivamente pacientes do ACCCC, avaliou um número muito semelhante de pacientes e encontrou um percentual de pacientes com progressão/recidiva praticamente igual ao encontrado em nosso estudo. Cagnacci-Neto (2014) observou uma

proporção de 1,4% (8/565) pacientes de CDIS que manifestaram progressão ou recidiva da doença.

Em relação aos genes mais frequentemente alterados em câncer de mama, identificamos nas pacientes do grupo CDIS sem progressão os genes *PIK3CA* e *TP53* tendo pelo menos uma variante potencialmente somática em 0% (0/15) e 6,6% (1/15) das pacientes, respectivamente; enquanto Cosimo e colaboradores (2020) identificaram variantes somáticas nos genes *PIK3CA* e *TP53* em 25% (6/24) e 8,3% (2/24) das pacientes, respectivamente. Pareja e colaboradores (2020) realizaram o sequenciamento do exoma de 7 pacientes de CDIS puro sem progressão, 25 amostras de componente *in situ* de CMI-TNE e 24 amostras de CMI-TNE. Em relação às pacientes de CDIS puro sem progressão, os autores identificaram variantes somáticas nos genes *PIK3CA* e *TP53* em 0% (0/7) e 14% (1/7) das pacientes, respectivamente. Nesta comparação pudemos observar que nossos resultados foram mais semelhantes aos resultados obtidos por Pareja et al. (2020). Interessantemente, embora os percentuais de mutação nos genes *PIK3CA* e *TP53* nas pacientes CDIS puro sem progressão não tenham sido semelhantes entre os 3 estudos (o nosso estudo, o de Cosimo et al. (2020) e o de (Pareja et al. 2020)), as casuísticas dos 3 estudos são semelhantes em relação a expressão dos receptores hormonais e ao grau nuclear, em que a grande maioria das pacientes dos 3 estudos são luminais e de intermediário e alto grau nuclear.

Em relação ao grupo de CDIS puro que apresentou progressão ou recidiva da doença, identificamos variantes potencialmente somáticas nos genes *PIK3CA* e *TP53* em 40% (2/5) e 0% (0/5) das pacientes, respectivamente; enquanto Cosimo e colaboradores (2020) identificaram variantes somáticas nos genes *PIK3CA* e *TP53* em 33% (1/3) das pacientes cada um. Já em relação ao grupo de CDIS que apresentou componente invasivo no material da peça cirúrgica, identificamos variantes potencialmente somáticas nos genes *PIK3CA* e *TP53* em 0% (0/3) e 67% (2/3) das pacientes, respectivamente; enquanto Pareja e colaboradores (2020) identificaram variantes somáticas nos genes *PIK3CA* e *TP53* em 41% (11/27) e 52% (14/27) das pacientes, respectivamente.

A partir dos resultados obtidos em seu estudo, Pareja e colaboradores (2020) sugeriram que o CDIS puro que não manifesta progressão é geneticamente menos avançado do que o componente *in situ* de CMI-TNE, pois não apresentam mutações somáticas em genes chaves que são comumente alterados em CMI-TNE, tais como, o gene *PIK3CA*. Os resultados obtidos em nosso estudo corroboram esta ideia. Enquanto identificamos variantes

potencialmente somáticas nos genes *PIK3CA* e *TP53* em 0% (0/15) e 6,6% (1/15) das pacientes do grupo CDIS puro sem progressão, respectivamente, identificamos 25% (2/8) das pacientes do grupo CDIS com progressão com variantes potencialmente somáticas em cada um dos genes *PIK3CA* e *TP53*. É importante ressaltar que este grupo foi composto por 2 subgrupos, um deles com pacientes de CDIS puro que manifestaram progressão ou recidiva da doença e o outro com pacientes que tiveram laudo anatomopatológico de CDIS na biópsia e que apresentaram invasão na peça cirúrgica, ou seja, após a biópsia constatou-se que se tratava de amostras de componente *in situ* de CMI-TNE, uma lesão pré-invasiva com maior potencial maligno.

O racional que utilizamos para agrupar esses dois tipos de lesões (CDIS puro que manifestou progressão/recidiva e o CDIS que na peça cirúrgica constatou-se que se tratava de componente *in situ* de CMI-TNE) num mesmo grupo, nomeado CDIS com progressão, e separados do outro grupo de CDIS puro, nomeado CDIS sem progressão, foi baseado no fato que uma lesão CDIS com focos de invasão, teria uma conduta clínica diferenciada e mais agressiva do que uma lesão indolente de CDIS puro. Além disso, outro aspecto que apoiou esse desenho experimental foram nossos resultados prévios, em que observamos uma semelhança em nível transcricional entre as células epiteliais neoplásicas do componente *in situ* de CMI-TNE e as células epiteliais neoplásicas do componente invasivo de CMI-TNE (Castro et al. 2008). E a partir de outro resultado prévio do grupo obtido com a utilização da metodologia de RNAseq para avaliar as diferenças de expressão gênica entre as células epiteliais neoplásicas e mioepiteliais de CDIS puro e de componente *in situ* de CMI-TNE, foi identificado o agrupamento num mesmo ramo do cluster de amostras de CDIS puro que tiveram progressão e as amostras de componente *in situ* de CMI-TNE (Elias et al., em preparação).

No grupo CDIS com progressão, subgrupo com invasão na peça cirúrgica, identificamos o caso 27CDIS com a maior quantidade de variantes oncogênicas ou provavelmente oncogênicas de toda a casuística. Este caso apresentou a variante oncogênica c.524G>A, p.Arg175His no gene *TP53* (VAF 74%) e as variantes provavelmente oncogênicas c.1654C>T, p.Arg552Ter no gene *RB1* (VAF 45%) e c.4465C>T, p.Arg1489Ter no gene *NOTCH4* (VAF 25%). A variante no gene *TP53* foi avaliada no DNA de leucócito do mesmo caso e confirmada como somática. As variantes nos genes *RB1* e *NOTCH4* foram avaliadas num *pool* de DNA que continha o DNA de leucócitos do mesmo caso, sendo consideradas potencialmente somáticas.

A variante no gene *RB1* deverá ser submetida a validação devido a sua VAF ser próxima de 50% e porque em nosso processo de busca ativa ela foi identificada com baixa cobertura de sequenciamento. É importante destacar que o caso 27CDIS é do subtipo triplo negativo, sendo o subtipo que apresenta a maior proporção de mutações nos genes *TP53* e *RB1* (Koboldt et al. 2012).

Vale destacar que as posições genômicas de diversas regiões *hotspot* dos genes *PIK3CA*, tais como R93, E542, N345, E545, H1047; e *TP53*, tais como Y163, R175, R248, F270, R273, C275, R282 foram devidamente cobertas no sequenciamento de DNA das 23 pacientes que tiveram os dados com qualidade e puderam ser analisados.

Em relação as variantes identificadas nas 4 pacientes (02CDIS, 06CDIS, 07CDIS e 21CDIS) em que conseguimos avaliar os dois materiais (biópsia e peça cirúrgica) utilizando o mesmo painel, de 19 variantes apenas 3 delas (c.355G>T, p.Asp119Tyr no gene *NRAS*; c.2293G>A, p.Val765Met no gene *EGFR*; c.773G>A, p.Arg258His no gene *PPP2R1A*) foram identificadas em uma amostra, mas não foram identificadas na outra amostra do mesmo caso. A variante c.2293G>A, p.Val765Met no gene *EGFR* foi identificada na biópsia do caso 7CDIS com VAF 6%, mas não foi identificada na peça cirúrgica do mesmo caso. Na validação foi confirmada na biópsia com VAF 1%, mas não foi confirmada na peça cirúrgica. Nos chama a atenção o fato dela ter sido confirmada na validação com um VAF discrepante em relação ao VAF em que foi identificada, e que foi identificada na biópsia, mas não foi identificada na peça cirúrgica. Contudo, ela também foi identificada na peça cirúrgica do caso 30CDIS com VAF 3% e na validação foi confirmada com VAF 1%. Esta variante está localizada no domínio cinase da proteína. Embora ela ainda não tenha sido reportada em estudos que avaliaram câncer de mama, ela está próxima à variante H773L, a qual é classificada pelo OncoKB como oncogênica e já foi identificada em câncer de mama invasivo no estudo METABRIC (Pereira et al. 2016). Embora ela ainda não tenha sido relatada em outros estudos que avaliaram câncer de mama, ela é reportada no Cosmic em tumores de pulmão, de pele e em linfoma; e é predita como *driver* em câncer de mama. A variante c.355G>T, p.Asp119Tyr no gene *NRAS* foi identificada na biópsia com VAF 35%, mas não foi confirmada na validação. E a variante c.773G>A, p.Arg258His no gene *PPP2R1A*, foi identificada na biópsia do caso 07CDIS com VAF 4%, mas não foi identificada na peça cirúrgica. Esta variante tem um efeito oncogênico classificado pelo OncoKB como inconclusivo, está reportada no Cosmic e ainda deverá ser submetida a validação para ser confirmada.

Em relação as variantes com efeito oncogênico não classificado no OncoKB, presentes em genes de predisposição ao câncer de mama, identificamos duas variantes (c.8560C>T, p.Arg2854Cys no gene *ATM*; c.508A>G, p.Lys170Glu no gene *RAD50*) que aparentam ter relevância. Ambas as variantes foram identificadas no caso 32CDIS, o qual é do grupo CDIS com progressão (grupo 2.2), tendo apresentado recidiva de lesão in situ 7 anos após o tratamento, tem antecedentes familiares de câncer de mama e mutação no gene *BRCA2* (Apêndice 2). A variante c.8560C>T, p.Arg2854Cys no gene *ATM* (VAF 44%) é classificada como conflitante no ClinVar (uma interpretação como benigna e 11 como VUS), apresenta escore REVEL 0,9, não é reportada no Cosmic, está em região de domínio da proteína codificada pelo gene, e já foi identificada em estudos que avaliaram variantes patogênicas em indivíduos com câncer familiar de mama (Paglia et al. 2010), e outros, como câncer de próstata (Paulo et al. 2018), colorretal (Yurgelun et al. 2017) e de pâncreas (Earl et al. 2020). E a variante c.508A>G, p.Lys170Glu no gene *RAD50* (VAF 56%) é classificada como VUS no ClinVar, não está catalogada no Cosmic, e está em região de domínio da proteína codificada pelo gene, porém ainda não foi identificada em estudos que avaliaram variantes patogênicas em indivíduos com câncer familiar de mama.

Uma variante que também nos chama a atenção é a variante c.190_191delGC, p.Ala64Profs*28 no gene *MUTYH* (VAF 53%) identificada no caso 06CDIS. Esta variante deve ser de origem germinativa e estar presente em heterozigose, devido ao fato de que foi identificada com VAF próximo a 50% tanto na biópsia quanto na peça cirúrgica. Embora esta variante não esteja reportada no ClinVar e os estudos da literatura indicam que ainda não há um consenso sobre o real efeito de variantes que resultam em perda de função monoalélica no gene *MUTYH* (Rennert et al. 2012; Win et al. 2016; Fulk et al. 2019) na predisposição ao câncer de mama, esta variante deverá ser submetida a validação.

Um outro resultado a ser mencionado da nossa avaliação mutacional de CDIS foi que identificamos diferentes vias bioquímicas de reparo de DNA enriquecidas pelos 40 oncogenes e GST acometidos pelas 45 variantes potencialmente somáticas. Interessantemente, no estudo de Pareja e colaboradores (2020), ao realizarem uma caracterização do perfil mutacional somático, através da avaliação do exoma completo de 7 pacientes de CDIS puro, os autores identificaram que uma das assinaturas mutacionais dominantes nesses tumores era a assinatura mutacional 3 (Nik-Zainal et al. 2016), uma assinatura mutacional de deficiência de reparo por recombinação homóloga.

Desta forma, em relação a avaliação molecular de CDIS realizada neste estudo, nossos achados da avaliação transcricional corroboram as evidências anteriores do nosso grupo de que as mudanças na expressão gênica ocorrem no CDIS puro antes da manifestação do fenótipo invasivo, e revelam uma assinatura baseada em 12 genes que aparentemente tem um potencial de prever a progressão de CDIS para a doença invasiva. E os achados da avaliação mutacional, os quais foram gerados pelo primeiro estudo do nosso grupo que avaliou o perfil mutacional de CDIS, evidenciam que mutações nos oncogenes e genes supressores de tumor mais frequentemente alterados em câncer de mama estão em maior frequência nas pacientes de CDIS puro que sofreram progressão/recidiva e nas lesões de CDIS em que se identificou invasão tumoral na peça cirúrgica. Essas evidências sugerem a possibilidade do CDIS puro que manifesta progressão/recidiva e o componente *in situ* de CMI-TNE apresentarem mais alterações mutacionais ou transcricionais do que o CDIS puro que não manifesta progressão/recidiva. Interessantemente, ao realizarmos uma análise comparativa das vias enriquecidas nas avaliações mutacional e transcricional, identificamos duas vias bioquímicas comuns as duas abordagens. Ao compararmos as 5 vias enriquecidas pelos 12 genes da assinatura de expressão com as 90 vias enriquecidas pelos 12 genes alterados por variantes oncogênicas/provavelmente oncogênicas, identificamos as vias de sinalização de fosfolipase D e de sinalização MAPK enriquecidas tanto na avaliação transcricional quanto na avaliação mutacional de CDIS. Contudo, é necessário avaliar um número maior de pacientes de CDIS que manifestaram diferentes desfechos para estimar o potencial desses achados em serem biomarcadores de risco de progressão do CDIS. Biomarcadores com essa finalidade são essenciais para possibilitar que pacientes com diagnóstico de CDIS puro sem risco estimado para progredir recebam tratamentos mais brandos garantindo as excelentes taxas de sobrevida.

6. CONCLUSÃO

Um total de 11 pacientes representados por 15 amostras de células epiteliais neoplásicas e mioepiteliais de lesões pré-invasivas com diferentes potenciais malignos (CDIS puro e componente *in situ* de CMI-TNE) foram avaliadas por RT-qPCR (ensaios TLDA) e apresentaram dados qualificados para análise de 53 genes. Os resultados confirmam a diferença de expressão de 49% (26/53) deles.

Identificamos uma assinatura composta por 12 genes cujo perfil de expressão foi capaz de distinguir os dois tipos de lesões pré-invasivas com diferentes potenciais malignos (CDIS puro e componente *in situ* de CMI-TNE). Nossos resultados sugerem que a assinatura gênica baseada nos 12 genes tem potencial de predizer a progressão de CDIS para a doença invasiva, corroborando as evidências anteriores do nosso grupo de que as mudanças na expressão gênica ocorrem nas células das lesões de CDIS puro antes da manifestação do fenótipo invasivo.

Um total de 23 pacientes com laudo anatomopatológico de CDIS na biópsia tiveram o perfil mutacional avaliado por NGS com painéis multigenes. O repertório mutacional da nossa casuística mostra a presença de variantes potencialmente somáticas em genes frequentemente alterados em câncer de mama, com *TP53* em 13% (3/23) dos pacientes, *PIK3CA* e *GATA3* em 8,7% (2/23) das pacientes, e *AKT1*, *KMT2C*, *MAP2K4*, *RB1* e *RUNX1* em 4,3% (1/23) das pacientes.

Observamos uma maior proporção de mutações nos genes *TP53* e *PIK3CA* no grupo CDIS com progressão. O gene *TP53* apresentou variantes em 6,6% (1/15) e 25% (2/8) dos grupos CDIS sem e com progressão, respectivamente; e *PIK3CA* em 0% (0/15) e 25% (2/8) dos grupos CDIS sem e com progressão. Porém, não identificamos uma associação estatisticamente significativa.

7. REFERÊNCIAS BIBLIOGRÁFICAS

Allred DC. Ductal carcinoma in situ: Terminology, classification, and natural history. *Journal of the National Cancer Institute - Monographs*. 2010;(41):134–8.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. *Nature Genetics* [Internet]. 2000 May;25(1):25–9. Available from: http://www.nature.com/articles/ng0500_25

Badve S, A'Hern RP, Ward AM, Millis RR, Pinder SE, Ellis IO, et al. Prediction of local recurrence of ductal carcinoma in situ of the breast using five histological classifications: A comparative study with long follow-up. *Human Pathology*. 1998;29(9):915–23.

Badve SS, Gökmen-Polar Y. Ductal carcinoma in situ of breast: update 2019. *Pathology*. 2019;51(6):563–9.

Baldan-Pineda PH. Identificação de genes supostamente envolvidos com o processo de invasão tumoral em câncer de mama através da técnica RaSH [Dissertação]. [São Paulo]: Fundação Antônio Prudente; 2007.

Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* [Internet]. 2012 Jun 21;486(7403):405–9. Available from: <http://www.nature.com/articles/nature11154>

Berra CM, Torrezan GT, de Paula CA, Hsieh R, Lourenço S v., Carraro DM. Use of uracil-DNA glycosylase enzyme to reduce DNA-related artifacts from formalin-fixed and paraffin-embedded tissues in diagnostic routine. *Applied Cancer Research* [Internet]. 2019 Dec 10;39(1):7. Available from: <https://appliedcr.biomedcentral.com/articles/10.1186/s41241-019-0075-2>

Birgersson M, Chi M, Miller C, Brzozowski JS, Brown J, Schofield L, et al. A Novel Role for Brain and Acute Leukemia Cytoplasmic (BAALC) in Human Breast Cancer Metastasis. *Frontiers in Oncology* [Internet]. 2021 Apr 22;11. Available from: <https://www.frontiersin.org/articles/10.3389/fonc.2021.656120/full>

Bombonati e Sgroi. The Molecular Pathology of Breast Cancer Progression. *J Pathol*. 2011;223(2):307–17.

Bos PD, Zhang XHF, Nadal C, Shu W, Gomis RR, Nguyen DX, et al. Genes that mediate breast cancer metastasis to the brain. *Nature* [Internet]. 2009 Jun 6;459(7249):1005–9. Available from: <http://www.nature.com/articles/nature08021>

Bremer T, Whitworth PW, Patel R, Savala J, Barry T, Lyle S, et al. A biological signature for breast ductal carcinoma in situ to predict radiotherapy benefit and assess recurrence risk.

Clinical Cancer Research [Internet]. 2018 Dec 1;24(23):5895–901. Available from: <http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-18-0842>

Brennan ME, Robin Turner FM, Stefano Ciatto Mb, Luke Marinovich M, James French MR, Petra Macaskill F, et al. Ductal Carcinoma in Situ at Core-Needle Biopsy: Meta-Analysis of Underestimation and Predictors of Invasive Breast Cancer 1. *Radiology*. 2011;260.

Cagnacci-Neto R. Carcinoma ductal in situ de mama: avaliação de potenciais fatores prognósticos [Dissertação]. [São Paulo]: Fundação Antônio Prudente; 2014.

Carbon S, Douglass E, Dunn N, Good B, Harris NL, Lewis SE, et al. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research* [Internet]. 2019 Jan 8;47(D1):D330–8. Available from: <https://academic.oup.com/nar/article/47/D1/D330/5160994>

Cardoso F, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rubio IT, et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology* [Internet]. 2019;30(8):1194–220. Available from: <https://doi.org/10.1093/annonc/mdz173>

Carraro DM, Elias E v., Andrade VP. Ductal carcinoma in situ of the breast: Morphological and molecular features implicated in progression. *Bioscience Reports*. 2014;34(1):19–28.

Carraro DM, Ferreira EN, Molina G de C, Puga RD, Abrantes EF, Trapé AP, et al. Poly (A)+ Transcriptome Assessment of ERBB2-Induced Alterations in Breast Cell Lines. Ho PL, editor. *PLoS ONE* [Internet]. 2011 Jun 22;6(6):e21022. Available from: <https://dx.plos.org/10.1371/journal.pone.0021022>

Castro NP, Osório CA, Torres C, Bastos EP, Mourão-Neto M, Soares FA, et al. Evidence that molecular changes in cells occur before morphological alterations during the progression of breast ductal carcinoma. *Breast Cancer Research* [Internet]. 2008 Oct 17;10(5):R87. Available from: <http://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr2157>

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discovery* [Internet]. 2012 May;2(5):401–4. Available from: <http://cancerdiscovery.aacrjournals.org/lookup/doi/10.1158/2159-8290.CD-12-0095>

Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: A Precision Oncology Knowledge Base. *JCO Precision Oncology* [Internet]. 2017 Nov;(1):1–16. Available from: <https://ascopubs.org/doi/10.1200/PO.17.00011>

Chang MT, Asthana S, Gao SP, Lee BH, Chapman JS, Kandoth C, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nature Biotechnology* [Internet]. 2016 Feb 30;34(2):155–63. Available from: <http://www.nature.com/articles/nbt.3391>

Chang MT, Bhattarai TS, Schram AM, Bielski CM, Donoghue MTA, Jonsson P, et al. Accelerating Discovery of Functional Mutant Alleles in Cancer. *Cancer Discovery* [Internet]. 2018 Feb 1;8(2):174–83. Available from: <http://cancerdiscovery.aacrjournals.org/lookup/doi/10.1158/2159-8290.CD-17-0321>

Cho SY, Park SY, Bae YK, Kim JY, Kim EK, Kim WG, et al. Standardized pathology report for breast cancer. *Journal of Pathology and Translational Medicine*. 2021;55(1):1–15.

Ciriello G, Gatza ML, Beck AH, Wilkerson MD, Rhie SK, Pastore A, et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* [Internet]. 2015 Oct 8;163(2):506–19. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867415011952>

di Cosimo S, Appierto V, Silvestri M, Ortolan E, de Cecco L, Veneroni S, et al. Primary tumor somatic mutations in the blood of women with ductal carcinoma in situ of the breast. *Annals of Oncology* [Internet]. 2020 Mar 1;31(3):435–7. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0923753419432286>

Crawford NPS, Qian X, Ziogas A, Papageorge AG, Boersma BJ, Walker RC, et al. Rrp1b, a New Candidate Susceptibility Gene for Breast Cancer Progression and Metastasis. Collier HA, editor. *PLoS Genetics* [Internet]. 2007 Nov 30;3(11):e214. Available from: <https://dx.plos.org/10.1371/journal.pgen.0030214>

Crijns APG, Fehrmann RSN, de Jong S, Gerbens F, Meersma GJ, Klip HG, et al. Survival-related profile, pathways, and transcription factors in ovarian cancer. Narod S, editor. *PLoS Medicine* [Internet]. 2009 Feb 3;6(2):e1000024. Available from: <https://dx.plos.org/10.1371/journal.pmed.1000024>

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* [Internet]. 2012 Jun 18;486(7403):346–52. Available from: <http://www.nature.com/articles/nature10983>

DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, et al. Breast cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*. 2019;69(6):438–51.

DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA: A Cancer Journal for Clinicians*. 2017;67(6):439–48.

Ding L, Ellis MJ, Li S, Larson DE, Chen K, Wallis JW, et al. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* [Internet]. 2010 Apr 15;464(7291):999–1005. Available from: <http://www.nature.com/articles/nature08989>

Douville C, Carter H, Kim R, Niknafs N, Diekhans M, Stenson PD, et al. CRAVAT: cancer-related analysis of variants toolkit. *Bioinformatics* [Internet]. 2013 Mar 1;29(5):647–8. Available from: <https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btt017>

Earl J, Galindo-Pumariño C, Encinas J, Barreto E, Castillo ME, Pachón V, et al. A comprehensive analysis of candidate genes in familial pancreatic cancer families reveals a high frequency of potentially pathogenic germline variants. *EBioMedicine* [Internet]. 2020 Mar 1;53:102675. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2352396420300505>

Eirew P, Steif A, Khattra J, Ha G, Yap D, Farahani H, et al. Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* [Internet]. 2015 Feb 26;518(7539):422–6. Available from: <http://www.nature.com/articles/nature13952>

El-Houjeiri L, Biondini M, Paquette M, Kuasne H, Pacis A, Park M, et al. Folliculin impairs breast tumor growth by repressing TFE3-dependent induction of the Warburg effect and angiogenesis. *Journal of Clinical Investigation* [Internet]. 2021 Nov 15;131(22). Available from: <https://www.jci.org/articles/view/144871>

Elias EV, de Castro NP, Pineda PHB, Abuázar CS, Osorio CABT, Pinilla MG, et al. Epithelial cells captured from ductal carcinoma in situ reveal a gene expression signature associated with progression to invasive breast cancer. *Oncotarget*. 2016;7(46).

Elshof LE, Tryfonidis K, Slaets L, van Leeuwen-Stok AE, Skinner VP, Dif N, et al. Feasibility of a prospective, randomised, open-label, international multicentre, phase III, non-inferiority trial to assess the safety of active surveillance for low risk ductal carcinoma in situ - The LORD study. *European Journal of Cancer* [Internet]. 2015;51(12):1497–510. Available from: <http://dx.doi.org/10.1016/j.ejca.2015.05.008>

Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, et al. Laser Capture Microdissection. *Science* (1979) [Internet]. 1996 Nov 8;274(5289):998–1001. Available from: <https://www.science.org/doi/10.1126/science.274.5289.998>

Ferreira EN, de Campos Molina G, Puga RD, Nagai MA, Campos AHJFM, Guimarães GC, et al. Linear mRNA amplification approach for RNAseq from limited amount of RNA. *Gene* [Internet]. 2015 Jun 15;564(2):220–7. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0378111915003649>

Flamant L, Roegiers E, Pierre M, Hayez A, Sterpin C, de Backer O, et al. TMEM45A is essential for hypoxia-induced chemoresistance in breast and liver cancer cells. *BMC Cancer* [Internet]. 2012 Dec 6;12(1):391. Available from: <https://bmccancer.biomedcentral.com/articles/10.1186/1471-2407-12-391>

Francis A, Fallowfield L, Rea D. The LORIS Trial: Addressing overtreatment of ductal carcinoma in situ. *Clinical Oncology*. 2015;27(1):6–8.

Freitas R, Rodrigues DCN, da Silveira Corrêa R, Peixoto JE, de Oliveira HVCG, Rahal RMS. Contribution of the unified health care system to mammography screening in Brazil, 2013. *Radiologia Brasileira*. 2016;49(5):305–10.

Fulk K, LaDuca H, Black MH, Qian D, Tian Y, Yussuf A, et al. Monoallelic MUTYH carrier status is not associated with increased breast cancer risk in a multigene panel cohort. *Familial Cancer*

[Internet]. 2019 Apr 23;18(2):197–201. Available from: <http://link.springer.com/10.1007/s10689-018-00114-4>

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling* [Internet]. 2013 Apr 2;6(269). Available from: <https://www.science.org/doi/10.1126/scisignal.2004088>

Gao J, Chang MT, Johnsen HC, Gao SP, Sylvester BE, Sumer SO, et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Medicine* [Internet]. 2017 Dec 23;9(1):4. Available from: <https://genomemedicine.biomedcentral.com/articles/10.1186/s13073-016-0393-x>

Gebrim LH, Shida JY, Hegg R, Topis T, Mattar A. Avaliação do tempo de início do tratamento, estadiamento histopatológico e positividade dos biomarcadores (RE, RP, HER-2) em 3.566 pacientes tratadas pelo SUS no período de 2012 a 2014, no Hospital Pérola Byington. *Revista Brasileira de Mastologia* [Internet]. 2014 [cited 2022 Apr 9];24(3):65–9. Available from: https://www.mastology.org/wp-content/uploads/2015/06/MAS_v24n3_65-69.pdf

Gomes R, Costa S, Colquhoun A. Eicosanoids and cancer. *Clinics* [Internet]. 2018 Oct 9;73(Suppl 1). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6096979/?report=classic>

Grimm LJ, Rahbar H, Abdelmalak M, Hall AH, Ryser MD. Ductal Carcinoma in Situ: State-of-the-Art Review. *Radiology* [Internet]. 2022 Jan;302(2):246–55. Available from: <http://pubs.rsna.org/doi/10.1148/radiol.211839>

Guo J, Chen L, Luo N, Yang W, Qu X, Cheng Z. Inhibition of TMEM45A suppresses proliferation, induces cell cycle arrest and reduces cell invasion in human ovarian cancer cells. *Oncology Reports* [Internet]. 2015 Jun 1;33(6):3124–30. Available from: <https://www.spandidos-publications.com/10.3892/or.2015.3902>

Han J, Bai Y, Wang J, Xie X li, Li A di, Ding Q, et al. REC8 promotes tumor migration, invasion and angiogenesis by targeting the PKA pathway in hepatocellular carcinoma. *Clinical and Experimental Medicine* [Internet]. 2021 Aug 7;21(3):479–92. Available from: <https://link.springer.com/10.1007/s10238-021-00698-9>

Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast cancer. Vol. 5, *Nature Reviews Disease Primers*. 2019.

Helleman J, Mortier G, de Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* [Internet]. 2007 Feb 9;8(2):R19. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17291332>

Hernandez L, Wilkerson PM, Lambros MB, Campion-Flora A, Rodrigues DN, Gauthier A, et al. Genomic and mutational profiling of ductal carcinomas in situ and matched adjacent invasive

breast cancers reveals intra-tumour genetic heterogeneity and clonal selection. *The Journal of Pathology* [Internet]. 2012 May;227(1):42–52. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/path.3990>

Hsu TH, Jiang SY, Chan WL, Eckert RL, Scharadin TM, Chang TC. Involvement of RARRES3 in the regulation of Wnt proteins acylation and signaling activities in human breast cancer cells. *Cell Death & Differentiation* [Internet]. 2015 May 7;22(5):801–14. Available from: <http://www.nature.com/articles/cdd2014175>

Hu Z, Fan C, Oh DS, Marron JS, He X, Qaqish BF, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*. 2006 Apr 27;7.

Hwang ES, Hyslop T, Lynch T, Frank E, Pinto D, Basila D, et al. The COMET (Comparison of Operative versus Monitoring and Endocrine Therapy) trial: a phase III randomised controlled clinical trial for low-risk ductal carcinoma in situ (DCIS). *BMJ Open*. 2019;9(3):e026797.

Iliaki S, Beyaert R, Afonina IS. Polo-like kinase 1 (PLK1) signaling in cancer and beyond. *Biochemical Pharmacology* [Internet]. 2021 Nov 1;193:114747. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0006295221003634>

INCA. Instituto Nacional de Câncer [Internet]. Ministério da Saúde. 2020 [cited 2022 Apr 11]. Available from: <https://www.inca.gov.br>

Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *The American Journal of Human Genetics* [Internet]. 2016 Oct 6;99(4):877–85. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0002929716303706>

Jiang H, Kang D chul, Alexandre D, Fisher PB. RaSH, a rapid subtraction hybridization approach for identifying and cloning differentially expressed genes. *Proceedings of the National Academy of Sciences* [Internet]. 2000 Nov 7;97(23):12684–9. Available from: <https://pnas.org/doi/full/10.1073/pnas.220431297>

Kan Z, Ding Y, Kim J, Jung HH, Chung W, Lal S, et al. Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. *Nature Communications* [Internet]. 2018 Dec 30;9(1):1725. Available from: <http://www.nature.com/articles/s41467-018-04129-4>

Karlin KL, Mondal G, Hartman JK, Tyagi S, Kurley SJ, Bland CS, et al. The Oncogenic STP Axis Promotes Triple-Negative Breast Cancer via Degradation of the REST Tumor Suppressor. *Cell Reports* [Internet]. 2014 Nov 20;9(4):1318–32. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S2211124714008651>

Kim J, Kim JY, Lee HB, Lee YJ, Seong MK, Paik N, et al. Characteristics and prognosis of 17 special histologic subtypes of invasive breast cancers according to World Health Organization classification: comparative analysis to invasive carcinoma of no special type. *Breast Cancer*

Research and Treatment [Internet]. 2020;184(2):527–42. Available from: <https://doi.org/10.1007/s10549-020-05861-6>

Knudsen ES, Ertel A, Davicioni E, Kline J, Schwartz GF, Witkiewicz AK. Progression of ductal carcinoma in situ to invasive breast cancer is associated with gene expression programs of EMT and myoepithelia. *Breast Cancer Research and Treatment* [Internet]. 2012 Jun 2;133(3):1009–24. Available from: <http://link.springer.com/10.1007/s10549-011-1894-3>

Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, et al. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61–70.

Kopanos C, Tsiolkas V, Kouris A, Chapple CE, Albarca Aguilera M, Meyer R, et al. VarSome: the human genomic variant search engine. Wren J, editor. *Bioinformatics* [Internet]. 2019 Jun 1;35(11):1978–80. Available from: <https://academic.oup.com/bioinformatics/article/35/11/1978/5146783>

Lakahani S, Ellis I, Schnitt S, Tan P, van de Vijver M. WHO Classification of tumours of the breast. 4th ed. International Agency for Research on Cancer, editor. Vol. 4. Lyon: World Health Organization; 2012.

Lakhani SR, O'Hare MJ. The mammary myoepithelial cell - Cinderella or ugly sister? *Breast Cancer Research* [Internet]. 2001 Feb 1;3(1):1. Available from: <http://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr260>

Langerød A, Zhao H, Borgan Ø, Nesland JM, Bukholm IR, Ik Dahl T, et al. TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. *Breast Cancer Research* [Internet]. 2007 Jun 15;9(3):R30. Available from: <http://breast-cancer-research.biomedcentral.com/articles/10.1186/bcr1675>

Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, et al. A Steroid Receptor Coactivator, SRA, Functions as an RNA and Is Present in an SRC-1 Complex. *Cell* [Internet]. 1999 Apr;97(1):17–27. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0092867400807114>

Latimer JJ, Johnson JM, Kelly CM, Miles TD, Beaudry-Rodgers KA, Lallanne NA, et al. Nucleotide excision repair deficiency is intrinsic in sporadic stage I breast cancer. *Proceedings of the National Academy of Sciences* [Internet]. 2010 Dec 14;107(50):21725–30. Available from: <https://pnas.org/doi/full/10.1073/pnas.0914772107>

Lee S, Stewart S, Nagtegaal I, Luo J, Wu Y, Colditz G, et al. Differentially expressed genes regulating the progression of ductal carcinoma in situ to invasive breast cancer. *Cancer Research* [Internet]. 2012 Sep 1;72(17):4574–86. Available from: <http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-12-0636>

Lefebvre C, Bachelot T, Filleron T, Pedrero M, Campone M, Soria JC, et al. Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. Mardis ER, editor. *PLOS Medicine*

[Internet]. 2016 Dec 27;13(12):e1002201. Available from: <https://dx.plos.org/10.1371/journal.pmed.1002201>

Lemos NA de F, Junior RF, Moreira MAR, Rodrigues FFO. Natural history of ductal carcinoma in situ. *Mastology*. 2018;28(2):114–8.

Leygue E, Dotzlaw H, Watson PH, Murphy LC. Expression of the steroid receptor RNA activator in human breast tumors. *Cancer Research* [Internet]. 1999 [cited 2022 Apr 12];59(17):4190–3. Available from: <https://aacrjournals.org/cancerres/article/59/17/4190/505542/Expression-of-the-Steroid-Receptor-RNA-Activator>

Li Q, Jiang B, Guo J, Shao H, del Priore IS, Chang Q, et al. INK4 Tumor Suppressor Proteins Mediate Resistance to CDK4/6 Kinase Inhibitors. *Cancer Discovery* [Internet]. 2022 Feb;12(2):356–71. Available from: <http://cancerdiscovery.aacrjournals.org/lookup/doi/10.1158/2159-8290.CD-20-1726>

Lin CY, Vennam S, Purington N, Lin E, Varma S, Han S, et al. Genomic landscape of ductal carcinoma in situ and association with progression. *Breast Cancer Research and Treatment* [Internet]. 2019 Nov 17;178(2):307–16. Available from: <http://link.springer.com/10.1007/s10549-019-05401-x>

Liu D, Shen X, Zhu G, Xing M. REC8 is a novel tumor suppressor gene epigenetically robustly targeted by the PI3K pathway in thyroid cancer. *Oncotarget* [Internet]. 2015 Nov 17;6(36):39211–24. Available from: <https://www.oncotarget.com/lookup/doi/10.18632/oncotarget.5391>

Liu M, Xu W, Su M, Fan P. REC8 suppresses tumor angiogenesis by inhibition of NF- κ B-mediated vascular endothelial growth factor expression in gastric cancer cells. *Biological Research* [Internet]. 2020 Dec 21;53(1):41. Available from: <https://biolres.biomedcentral.com/articles/10.1186/s40659-020-00307-1>

Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Human Mutation* [Internet]. 2016 Mar 1;37(3):235–41. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/humu.22932>

Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, et al. Gene expression profiles of human breast cancer progression. *Proceedings of the National Academy of Sciences* [Internet]. 2003 May 13;100(10):5974–9. Available from: <https://pnas.org/doi/full/10.1073/pnas.0931261100>

Martelotto LG, de Filippo MR, Ng CK, Natrajan R, Fuhrmann L, Cyrta J, et al. Genomic landscape of adenoid cystic carcinoma of the breast. *The Journal of Pathology* [Internet]. 2015 Oct 1;237(2):179–89. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/path.4573>

Masica DL, Douville C, Tokheim C, Bhattacharya R, Kim R, Moad K, et al. CRAVAT 4: Cancer-Related Analysis of Variants Toolkit. *Cancer Research* [Internet]. 2017 Nov 1;77(21):e35–8. Available from: <http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-17-0338>

Meng E, Shevde LA, Samant RS. Emerging roles and underlying molecular mechanisms of DNAJB6 in cancer. *Oncotarget* [Internet]. 2016 Aug 16;7(33):53984–96. Available from: <https://www.oncotarget.com/lookup/doi/10.18632/oncotarget.9803>

Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, et al. Genes that mediate breast cancer metastasis to lung. *Nature* [Internet]. 2005 Jul 28;436(7050):518–24. Available from: <http://www.nature.com/articles/nature03799>

Mitchell E, Jindal S, Chan T, Narasimhan J, Sivagnanam S, Gray E, et al. Loss of myoepithelial calponin-1 characterizes high-risk ductal carcinoma in situ cases, which are further stratified by T cell composition. *Molecular Carcinogenesis* [Internet]. 2020 Jul 5;59(7):701–12. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/mc.23171>

Morita K, Masamoto Y, Kataoka K, Koya J, Kagoya Y, Yashiroda H, et al. BAALC potentiates oncogenic ERK pathway through interactions with MEKK1 and KLF4. *Leukemia* [Internet]. 2015 Nov 8;29(11):2248–56. Available from: <https://www.nature.com/articles/leu2015137>

Morrow M, van Zee KJ, Solin LJ, Houssami N, Chavez-MacGregor M, Harris JR, et al. Society of Surgical Oncology-American Society for Radiation Oncology-American Society of Clinical Oncology consensus guideline on margins for breast-conserving surgery with whole-breast irradiation in ductal carcinoma in situ. *Journal of Clinical Oncology*. 2016;34(33):4040–6.

Moslehi R, Tsao HS, Zeinomar N, Stagnar C, Fitzpatrick S, Dzutsev A. Integrative genomic analysis implicates ERCC6 and its interaction with ERCC8 in susceptibility to breast cancer. *Scientific Reports* [Internet]. 2020 Dec 4;10(1):21276. Available from: <https://www.nature.com/articles/s41598-020-77037-7>

Najafi M, Ahmadi A, Mortezaee K. Extracellular-signal-regulated kinase/mitogen-activated protein kinase signaling as a target for cancer therapy: an updated review. *Cell Biology International* [Internet]. 2019 Nov 11;43(11):1206–22. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/cbin.11187>

Naume B, Zhao X, Synnestvedt M, Borgen E, Russnes HG, Lingjaerde OC, et al. Presence of bone marrow micrometastasis is associated with different recurrence risk within molecular subtypes of breast cancer. *Molecular Oncology* [Internet]. 2007 Sep;1(2):160–71. Available from: <http://doi.wiley.com/10.1016/j.molonc.2007.03.004>

Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature*. 2016;534(7605):47–54.

Nixon MJ, Formisano L, Mayer IA, Estrada MV, González-Ericsson PI, Isakoff SJ, et al. PIK3CA and MAP3K1 alterations imply luminal A status and are associated with clinical benefit from

pan-PI3K inhibitor buparlisib and letrozole in ER+ metastatic breast cancer. *npj Breast Cancer* [Internet]. 2019 Dec 23;5(1):31. Available from: <http://www.nature.com/articles/s41523-019-0126-6>

Osório CABT, Júnior MAC, Soares FA. Avaliação de resposta patológica em câncer de mama após quimioterapia neoadjuvante: Padronização de protocolo adaptado. *Jornal Brasileiro de Patologia e Medicina Laboratorial*. 2012;48(6):447–53.

Page DL, Dupont WD, Rogers LW, Jensen RA, Schuyler PA. Continued local recurrence of carcinoma 15–25 years after a diagnosis of low grade ductal carcinoma in situ of the breast treated only by biopsy. *Cancer*. 1995;76(7):1197–200.

Page DL, DUPONT WD, ROGERS LW, MARIE LANDENBERGER. Intraductal Carcinoma of the Breast: Follow-up After Biopsy Only. 1982;

Paglia L la, Laugé A, Weber J, Champ J, Cavaciuti E, Russo A, et al. ATM germline mutations in women with familial breast cancer and a relative with haematological malignancy. *Breast Cancer Research and Treatment* [Internet]. 2010 Jan 29;119(2):443–52. Available from: <http://link.springer.com/10.1007/s10549-009-0396-z>

Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, et al. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *New England Journal of Medicine* [Internet]. 2004 Dec 30;351(27):2817–26. Available from: <http://www.nejm.org/doi/abs/10.1056/NEJMoa041588>

Pang JMB, Savas P, Fellowes AP, Mir Arnau G, Kader T, Vedururu R, et al. Breast ductal carcinoma in situ carry mutational driver events representative of invasive breast cancer. *Modern Pathology* [Internet]. 2017 Jul 24;30(7):952–63. Available from: <http://www.nature.com/articles/modpathol201721>

Pareja F, Brown DN, Lee JY, da Cruz Paula A, Selenica P, Bi R, et al. Whole-exome sequencing analysis of the progression from non-low-grade ductal carcinoma in situ to invasive ductal carcinoma. *Clinical Cancer Research* [Internet]. 2020 Jul 15;26(14):3682–93. Available from: <http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-19-2563>

Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *Journal of Clinical Oncology* [Internet]. 2009 Mar 10;27(8):1160–7. Available from: <http://ascopubs.org/doi/10.1200/JCO.2008.18.1370>

Paulo P, Maia S, Pinto C, Pinto P, Monteiro A, Peixoto A, et al. Targeted next generation sequencing identifies functionally deleterious germline mutations in novel genes in early-onset/familial prostate cancer. Eng C, editor. *PLOS Genetics* [Internet]. 2018 Apr 16;14(4):e1007355. Available from: <https://dx.plos.org/10.1371/journal.pgen.1007355>

Pereira B, Chin SF, Rueda OM, Vollan HKM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes.

Nature Communications [Internet]. 2016 Sep 1;7(1):11479. Available from: <http://www.nature.com/articles/ncomms11479>

Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature* [Internet]. 2000 Aug;406(6797):747–52. Available from: <http://www.nature.com/articles/35021093>

Pinilla-Fernández M. Avaliação da expressão de genes codificadores e não codificadores em células epiteliais do carcinoma ductal invasivo de mama de pacientes com e sem linfonodo acometido [Tese]. [São Paulo]: Fundação Antônio Prudente; 2015.

Polyak K, Hu M. Do Myoepithelial Cells Hold the Key for Breast Tumor Progression? *Journal of Mammary Gland Biology and Neoplasia* [Internet]. 2005 Jul 28;10(3):231–47. Available from: <http://link.springer.com/10.1007/s10911-005-9584-6>

Pritchard AL, Hayward NK. Molecular Pathways: Mitogen-Activated Protein Kinase Pathway Mutations and Drug Resistance. *Clinical Cancer Research* [Internet]. 2013 May 1;19(9):2301–9. Available from: <http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-12-0383>

Rakovitch E, Nofech-Mozes S, Hanna W, Baehner FL, Saskin R, Butler SM, et al. A population-based validation study of the DCIS Score predicting recurrence risk in individuals treated by breast-conserving surgery alone. *Breast Cancer Research and Treatment* [Internet]. 2015 Jul 29;152(2):389–98. Available from: <http://link.springer.com/10.1007/s10549-015-3464-6>

Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasan N, et al. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell* [Internet]. 2018 Sep 10;34(3):427–438.e6. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1535610818303684>

Razavi P, Dickler MN, Shah PD, Toy W, Brown DN, Won HH, et al. Alterations in PTEN and ESR1 promote clinical resistance to alpelisib plus aromatase inhibitors. *Nature Cancer* [Internet]. 2020 Apr 23;1(4):382–93. Available from: <http://www.nature.com/articles/s43018-020-0047-1>

Rennert G, Lejbkowitz F, Cohen I, Pinchev M, Rennert HS, Barnett-Griness O. MutYH mutation carriers have increased breast cancer risk. *Cancer* [Internet]. 2012 Apr 15;118(8):1989–93. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/cncr.26506>

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nature Biotechnology* [Internet]. 2011 Jan 1;29(1):24–6. Available from: <http://www.nature.com/articles/nbt.1754>

Rosner Dutz. Noninvasive breast carcinoma- results of a national survey by the American College of Surgeons. *Annals of Surgery*. 1980;242(5):629–30.

Rudloff U, Jacks LM, Goldberg JI, Wynveen CA, Brogi E, Patil S, et al. Nomogram for predicting the risk of local recurrence after breast-conserving surgery for ductal carcinoma in situ.

Journal of Clinical Oncology [Internet]. 2010 Aug 10;28(23):3762–9. Available from: <http://ascopubs.org/doi/10.1200/JCO.2009.26.8847>

Russell TD, Jindal S, Agunbiade S, Gao D, Troxell M, Borges VF, et al. Myoepithelial cell differentiation markers in ductal carcinoma in situ progression. The American Journal of Pathology [Internet]. 2015 Nov 1;185(11):3076–89. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0002944015004307>

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: A Free, Open-Source System for Microarray Data Management and Analysis. Biotechniques [Internet]. 2003 Feb;34(2):374–8. Available from: <https://www.future-science.com/doi/10.2144/03342mt01>

Sanders ME, Schuyler PA, Dupont WD, Page DL. The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. Cancer. 2005;103(12):2481–4.

Sanders ME, Schuyler PA, Simpson JF, Page DL, Dupont WD. Continued observation of the natural history of low-grade ductal carcinoma in situ reaffirms proclivity for local recurrence even after more than 30 years of follow up. Mod Pathol. 2015;28(5):662–9.

Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (1979) [Internet]. 1995 Oct 20;270(5235):467–70. Available from: <https://www.science.org/doi/10.1126/science.270.5235.467>

Schuetz CS, Bonin M, Clare SE, Nieselt K, Sotlar K, Walter M, et al. Progression-Specific Genes Identified by Expression Profiling of Matched Ductal Carcinomas *In situ* and Invasive Breast Tumors, Combining Laser Capture Microdissection and Oligonucleotide Microarray Analysis. Cancer Research [Internet]. 2006 May 15;66(10):5278–86. Available from: <http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-05-4610>

Sens-Abuázar C, Napolitano e Ferreira E, Osório CABT, Krepschi ACV, Ricca TI, Castro NP, et al. Down-regulation of ANAPC13 and CLTCL1: Early Events in the Progression of Preinvasive Ductal Carcinoma of the Breast. Translational Oncology [Internet]. 2012 Apr;5(2):113-IN8. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1936523312800325>

Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature [Internet]. 2012 Jun 4;486(7403):395–9. Available from: <http://www.nature.com/articles/nature10933>

Sherry ST, Ward M, Sirotkin K. dbSNP-database for single nucleotide polymorphisms and other classes of minor genetic variation. Genome Res [Internet]. 1999 Aug;9(8):677–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10447503>

Shivalingappa PKM, Sharma V, Shiras A, Bapat SA. RNA binding motif 47 (RBM47): emerging roles in vertebrate development, RNA editing and cancer. Molecular and Cellular

Biochemistry [Internet]. 2021 Dec 9;476(12):4493–505. Available from: <https://link.springer.com/10.1007/s11010-021-04256-5>

Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin* [Internet]. 2022;72(1):7–33. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/35020204>

Silverstein MJ. The University of Southern California/Van Nuys prognostic index for ductal carcinoma in situ of the breast. *The American Journal of Surgery* [Internet]. 2003 Oct;186(4):337–43. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0002961003002654>

Solin LJ, Gray R, Baehner FL, Butler SM, Hughes LL, Yoshizawa C, et al. A multigene expression assay to predict local recurrence risk for ductal carcinoma in situ of the breast. *JNCI Journal of the National Cancer Institute* [Internet]. 2013 May 15;105(10):701–10. Available from: <https://academic.oup.com/jnci/article-lookup/doi/10.1093/jnci/djt067>

Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nature Reviews Cancer* [Internet]. 2018 Nov 6;18(11):696–705. Available from: <http://www.nature.com/articles/s41568-018-0060-1>

Sørli T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences* [Internet]. 2001 Sep 11;98(19):10869–74. Available from: <https://pnas.org/doi/full/10.1073/pnas.191367098>

Sørli T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences* [Internet]. 2003 Jul 8;100(14):8418–23. Available from: <https://pnas.org/doi/full/10.1073/pnas.0932692100>

Stephens PJ, Tarpey PS, Davies H, van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* [Internet]. 2012 Jun 16;486(7403):400–4. Available from: <http://www.nature.com/articles/nature11017>

Stuart KE, Houssami N, Taylor R, Hayen A, Boyages J. Long-term outcomes of ductal carcinoma in situ of the breast: A systematic review, meta-analysis and meta-regression analysis. *BMC Cancer* [Internet]. 2015;15(1). Available from: <http://dx.doi.org/10.1186/s12885-015-1904-7>

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*. 2021;71(3):209–49.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* [Internet]. 2019 Jan

- 8;47(D1):D607–13. Available from: <https://academic.oup.com/nar/article/47/D1/D607/5198476>
- Tan J, Ong CK, Lim WK, Ng CCY, Thike AA, Ng LM, et al. Genomic landscapes of breast fibroepithelial tumors. *Nature Genetics* [Internet]. 2015 Nov 5;47(11):1341–5. Available from: <http://www.nature.com/articles/ng.3409>
- Tokheim C, Karchin R. CHASMap Reveals the Scope of Somatic Missense Mutations Driving Human Cancers. *Cell Syst* [Internet]. 2019 Jul 24;9(1):9–23.e8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31202631>
- Tsai FM, Shyu RY, Jiang SY. RIG1 suppresses Ras activation and induces cellular apoptosis at the Golgi apparatus. *Cellular Signalling* [Internet]. 2007 May;19(5):989–99. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0898656806003019>
- Tsunoda AT, Nunes JS, Watanabe APHU, Santos Junior LA, Mauad EC, Brentani RR. Controle de qualidade em rastreamento mamográfico no Brasil: experiência do Hospital de Câncer de Barretos TT - Quality control in mammographic breast cancer screening in Brazil: the Barretos Cancer Hospital experience. *Rev bras mastologia* [Internet]. 2013;23(1):12–8. Available from: http://www.rbmastologia.com.br/wp-content/uploads/2015/06/MAS-v23n1%7B%5C_%7D12-18.pdf
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Research* [Internet]. 2012 Aug;40(15):e115–e115. Available from: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gks596>
- Urquidí V, Goodison S, Cai Y, Sun Y, Rosser CJ. A candidate molecular biomarker panel for the detection of bladder cancer. *Cancer Epidemiology Biomarkers & Prevention* [Internet]. 2012 Dec;21(12):2149–58. Available from: <https://cebp.aacrjournals.org/lookup/doi/10.1158/1055-9965.EPI-12-0428>
- Vanharanta S, Marney CB, Shu W, Valiente M, Zou Y, Mele A, et al. Loss of the multifunctional RNA-binding protein RBM47 as a source of selectable metastatic traits in breast cancer. *Elife* [Internet]. 2014 Jun 4;3. Available from: <https://elifesciences.org/articles/02734>
- Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer Genome Landscapes. *Science* (1979) [Internet]. 2013 Mar 29;339(6127):1546–58. Available from: <https://www.sciencemag.org/lookup/doi/10.1126/science.1235122>
- Wallis MG, Clements K, Kearins O, Ball G, MacArtney J, Lawrence GM. The effect of DCIS grade on rate, type and time to recurrence after 15 years of follow-up of screen-detected DCIS. *British Journal of Cancer* [Internet]. 2012;106(10):1611–7. Available from: <http://dx.doi.org/10.1038/bjc.2012.151>
- Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *The Lancet*

[Internet]. 2005 Feb;365(9460):671–9. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0140673605179471>

Watkins G, Douglas-Jones A, Mansel RE, Jiang WG. Expression of thromboxane synthase, TBXAS1 and the thromboxane A2 receptor, TBXA2R, in human breast cancer. *International Seminars in Surgical Oncology* [Internet]. 2005 Dec 26;2(1):23. Available from: <https://issoonline.biomedcentral.com/articles/10.1186/1477-7800-2-23>

Weber S, Haferlach T, Alpermann T, Perglerová K, Schnittger S, Haferlach C, et al. Feasibility of *BAALC* gene expression for detection of minimal residual disease and risk stratification in normal karyotype acute myeloid leukaemia. *British Journal of Haematology* [Internet]. 2016 Dec 1;175(5):904–16. Available from: <https://onlinelibrary.wiley.com/doi/10.1111/bjh.14343>

Weinmann S, Leo MC, Francisco M, Jenkins CL, Barry T, Leesman G, et al. Validation of a Ductal Carcinoma *In Situ* Biomarker Profile for Risk of Recurrence after Breast-Conserving Surgery with and without Radiotherapy. *Clinical Cancer Research* [Internet]. 2020 Aug 1;26(15):4054–63. Available from: <http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-19-1152>

Wellings SR, Jensen HM. On the Origin and Progression of Ductal Carcinoma in the Human Breast2. *JNCI: Journal of the National Cancer Institute*. 1973 May;50(5):1111–8.

WHO. WHO Classification of Tumours: Breast Tumours. 5th ed. International Agency for Research on Cancer, editor. Lyon; 2019.

Win AK, Reece JC, Dowty JG, Buchanan DD, Clendenning M, Rosty C, et al. Risk of extracolonic cancers for people with biallelic and monoallelic mutations in *MUTYH*. *International Journal of Cancer* [Internet]. 2016 Oct 1;139(7):1557–63. Available from: <https://onlinelibrary.wiley.com/doi/10.1002/ijc.30197>

Yan Y, Cooper C, Hamedani MK, Guppy B, Xu W, Tsuyuki D, et al. The steroid receptor RNA activator protein (SRAP) controls cancer cell migration/motility. *FEBS Letters* [Internet]. 2015 Dec 21;589(24PartB):4010–8. Available from: <http://doi.wiley.com/10.1016/j.febslet.2015.11.007>

Ye Q, Kantonen S, Henkels KM, Gomez-Cambronero J. A new signaling pathway (JAK-Fes-phospholipase D) that is enhanced in highly proliferative breast cancer cells. *Journal of Biological Chemistry* [Internet]. 2013 Apr 5;288(14):9881–91. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0021925820673495>

Yurgelun MB, Kulke MH, Fuchs CS, Allen BA, Uno H, Hornick JL, et al. Cancer susceptibility gene mutations in individuals with colorectal cancer. *Journal of Clinical Oncology* [Internet]. 2017 Apr 1;35(10):1086–95. Available from: <https://ascopubs.org/doi/10.1200/JCO.2016.71.0012>

Anexo 1 - Carta de aprovação do estudo pelo Comitê de Ética em Pesquisa.



A.C. Camargo Cancer Center

Centro Integrado de Diagnóstico, Tratamento, Ensino e Pesquisa

**COMITÊ DE ÉTICA
EM PESQUISA - CEP**

São Paulo, 27 de dezembro de 2017.

À

Dra. Dirce Carraro

Aluno: Paulo Henrique Baldan Pineda (Doutorado)

Ref.: Projeto de Pesquisa nº. 1746/13F

"AVALIAÇÃO TRANSCRICIONAL DE LESÕES PRÉ-INVASIVAS DE CARCINOMA DUCTAL DE MAMA: VALIDAÇÃO DE UMA ASSINATURA GÊNICA COM POTENCIAL DE PREDIÇÃO DE PROGRESSÃO DE CARCINOMA DUCTAL IN SITU."

Os membros do Comitê de Ética em Pesquisa em Seres Humanos da Fundação Antonio Prudente – A.C. Camargo Cancer Center, em sua última reunião de **05/12/2017**, **tomaram conhecimento e aprovaram** o seguinte documento:

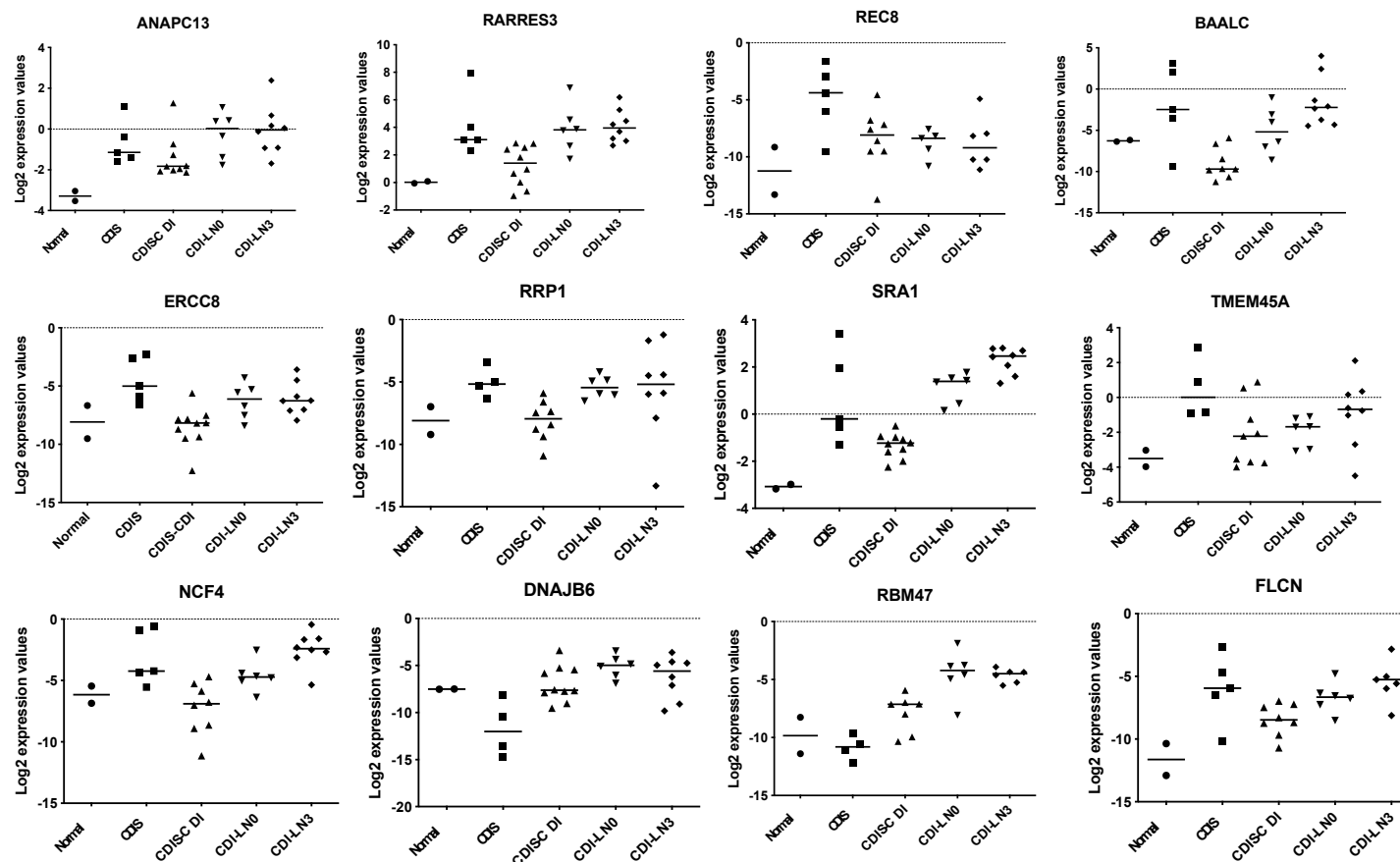
- Solicitação de dispensa da submissão da documentação obrigatória e análise ética do projeto acima mencionado por se tratar de um projeto afiliado ao temático intitulado: "Aspectos moleculares envolvidos no desenvolvimento e progressão do carcinoma ductal de mama: investigação da progressão do carcinoma in situ e do papel da mutação em BRCA1 no tumor triplo negativo" registrado neste CEP sob nº 1746/13. O projeto afiliado em referência será um Tese de Doutorado do aluno Paulo Henrique Baldan Pineda.

Atenciosamente,

Dr. Jefferson Luiz Gross

1º Vice-Coordenador do Comitê de Ética em Pesquisa

Apêndice 1 - Valores de expressão dos 12 genes nas etapas da progressão de CDIS.



Legenda: Valores de expressão dos 12 genes nas etapas da progressão de CDIS. Valores de expressão dos 12 genes cujo perfil de expressão foi capaz de discriminar as amostras de células epiteliais e mioepiteliais de CDIS puro de componente in situ de CMI-TNE em amostras que representam as etapas do processo de progressão de CDIS. Abreviações: CDIS, carcinoma ductal in situ; CDIS-CDI, componente in situ de CMI-TNE; CDI-LN0, carcinoma invasivo sem comprometimento linfonodal; CDI-LN3, carcinoma invasivo com até 3 linfonodos comprometidos.

Apêndice 2 - Características anatomopatológicas e antecedentes familiares do grupo CDIS com progressão ou recidiva.

Caso	Idade ao diagnóstico	Lateralidade dos tumores primário e progressão/recidiva	Progressão/recidiva	Tempo entre o primário e a progressão/recidiva (anos)	Tempo de seguimento	Procedimento cirúrgico	Características anatomopatológicas do tumor primário (peça cirúrgica)	Radioterapia no tumor primário	Antecedentes pessoais (AP)	Antecedente familiar (AF)
28CDIS	35	Direita	Progressão (CMI-TNE + CDIS)	3	7	Mastectomia	CDIS tipos sólido e cribriforme com comedonecrose, GN3, com cancerização de lóbulos. Medida estimada da neoplasia: 6,5cm. Presença de necrose, microcalcificações, multifocalidade, multicentricidade. Margens cirúrgicas livres, porém a margem profunda foi exigua (1,5mm).	Sim	Painel 37 genes + MLPA de BRCA1 e BRCA2. Resultado negativo para variantes patogênicas.	Mãe com câncer de mama aos 59 anos e negativa para variantes patogênicas nos genes BRCA1/2. Avó materna com câncer de mama aos 80 anos. Bisavó materna com câncer de mama aos 60 anos. Tia materna com câncer de endométrio aos 57 anos.
29CDIS	50	Esquerda	Recidiva (CDIS)	2	8	Ressecção segmentar	CDIS tipos sólido e cribriforme com comedonecrose, GN3. Medida estimada da neoplasia: 2,0cm. Margens cirúrgicas livres, porém exiguas.	Não informado no prontuário	-	Tia materna com câncer de mama aos 85 anos, tia materna com câncer de mama aos 70 anos, tia materna com câncer de ovário aos 45 anos, tia materna com câncer de ovário aos 50 anos e prima materna com câncer de mama aos 36 anos. Tem origem Judia Ashkenazi.
30CDIS	49	Direita	Progressão (CMI-TNE + CDIS)	1	11	Ressecção segmentar	CDIS tipos sólido e cribriforme com comedonecrose, GN3. Medida estimada da neoplasia: 2,0cm. Presença de microcalcificações, multifocalidade. Margens cirúrgicas livres. Imunoistoquímica de P63+ (sem microinvasão).	Sim	-	Mãe com câncer de pulmão (não fumante), avó materna com câncer colorretal, tia materna com câncer de útero e fígado, prima materna com câncer de estômago (falecida aos 40 anos) e 3 tios maternos com câncer de próstata.
31CDIS	48	Esquerda	Recidiva (CDIS)	4	6	Ressecção segmentar	CDIS tipo sólido com comedonecrose, GN3, com cistificação de ductos. Medida estimada da neoplasia: 3,0cm. Presença de necrose e microcalcificações. Margens cirúrgicas livres, porém a profunda foi exigua (0,4mm). Na ampliação de margens não houve neoplasia residual. Imunoistoquímica de P63+ (sem microinvasão).	Sim	-	Tia materna com câncer de mama aos 85 anos.
32CDIS	37	Esquerda	Recidiva (CDIS)	7	8	Ressecção segmentar	CDIS tipos sólido e micropapilar com comedonecrose, GN3, com extensão lobular. Presença de microcalcificações. Margens cirúrgicas livres, porém a medial foi exigua (0.1mm). A ampliação de margem possibilitou a remoção do CDIS residual com margens cirúrgicas livres.	Sim	Painel 37 genes + MLPA de BRCA1 e BRCA2. Resultado positivo para variante patogênica c.156_157insAlu no éxon 3 do gene BRCA2 em HT (confirmado no MLPA). Encontrada 2 VUS: p.Arg2854Cys em ATM e variante intrônica em PMS2: ENST0000026584	Mãe com câncer de mama aos 54 anos, tia materna com câncer de mama aos 50 anos, avó materna com câncer de mama. Antecedente pessoal: nega câncer de intestino e ovário.
33CDIS	40	Direita	Progressão (CMI-TNE + CDIS)	1	7	Adenomastectomia	CDIS tipos cribriforme e micropapilar com comedonecrose, GN3, com extensão aos lóbulos. Medida estimada da neoplasia: 3,5 cm. Presença de multifocalidade. Linfonodo sentinela pesquisado e negativo. Margens cirúrgicas livres.	Não	-	Tia com câncer de mama aos 70 anos, tia com câncer de útero.
34CDIS	51	Esquerda	Progressão (CMI-TNE + CDIS)	2	4	Ressecção segmentar	CDIS tipos sólido e cribriforme com comedonecrose, GN3, com colonização de lóbulos. Medida estimada da cirurgia: 3,0cm. Presença de microcalcificações. Margens cirúrgicas livres.	Sim	-	Avó materna com câncer colorretal, pai com câncer de bexiga e próstata.

Apêndice 3 - Relação dos 12 genes com mutações em CDIS.

Gene	Classe do gene	Caso	HGVS c.	HGVS p.	Classificação oncogênica da variante (OncoKB)	Efeito oncogênico da variante (OncoKB)
AKT1	onco	07CDIS	c.49G>A	p.Glu17Lys	oncogênica	ganho de função
CIC	onco/GST	32CDIS	c.6573_6607del	p.Leu2192Profs*49	provavelmente oncogênica	provável perda de função
GATA3	onco/GST	02CDIS	c.1050+2T>C	p. (spl?)	provavelmente oncogênica	provável perda de função
		21CDIS	c.614dupT	p.Met205Ilefs*99	provavelmente oncogênica	provável perda de função
KMT2D	onco/GST	31CDIS	c.15345delT	p.Phe5115Leufs*32	provavelmente oncogênica	provável perda de função
MAP2K4	GST	02CDIS	c.339_340delGA	p.Lys114Argfs*28	provavelmente oncogênica	provável perda de função
MRE11	GST	31CDIS	c.1726C>T	p.Arg576Ter	provavelmente oncogênica	provável perda de função
MUTYH	GST	06CDIS	c.190_191delGC	p.Ala64Profs*28	provavelmente oncogênica	provável perda de função
NOTCH4	onco/GST	27CDIS	c.4465C>T	p.Arg1489Ter	provavelmente oncogênica	provável perda de função
PIK3CA	onco	28CDIS 34CDIS	c.3140A>G	p.His1047Arg	oncogênica	ganho de função
RB1	GST	27CDIS	c.1654C>T	p.Arg552Ter	provavelmente oncogênica	provável perda de função
SMARCA4	GST	05CDIS	c.1777G>T	p.Glu593Ter	provavelmente oncogênica	provável perda de função
TP53	onco/GST	12CDIS	c.743G>C	p.Arg248Pro	provavelmente oncogênica	provável perda de função
		21CDIS	c.822_823dupTT	p.Cys275Phefs*71	provavelmente oncogênica	provável perda de função
		27CDIS	c.524G>A	p.Arg175His	oncogênica	perda de função

Legenda: Relação dos 12 genes com mutações em CDIS. Um total de 12 genes foram acometidos por 15 variantes oncogênicas ou provavelmente oncogênicas em CDIS.

Apêndice 4 - Vias bioquímicas enriquecidas pelos 12 genes com mutações em CDIS.

ID da via no KEGG	Nome da via no KEGG	Quantidade de genes na via	Escore	FDR
hsa05223	Non-small cell lung cancer	4 de 68	1,98	0,00000361
hsa05218	Melanoma	4 de 72	1,96	0,00000378
hsa05214	Glioma	4 de 72	1,96	0,00000378
hsa05212	Pancreatic cancer	4 de 73	1,95	0,00000378
hsa05220	Chronic myeloid leukemia	4 de 75	1,94	0,00000378
hsa01522	Endocrine resistance	5 de 95	1,93	0,000000813
hsa05213	Endometrial cancer	3 de 57	1,93	0,0000944
hsa05219	Bladder cancer	2 de 41	1,9	0,0022
hsa04664	Fc epsilon RI signaling pathway	3 de 66	1,87	0,00013
hsa04973	Carbohydrate digestion and absorption	2 de 44	1,87	0,0025
hsa05222	Small cell lung cancer	4 de 92	1,85	0,00000626
hsa05230	Central carbon metabolism in cancer	3 de 69	1,85	0,00014
hsa01524	Platinum drug resistance	3 de 70	1,84	0,00014
hsa05215	Prostate cancer	4 de 96	1,83	0,00000688
hsa05210	Colorectal cancer	3 de 82	1,78	0,00021
hsa04923	Regulation of lipolysis in adipocytes	2 de 54	1,78	0,0036
hsa04012	ErbB signaling pathway	3 de 83	1,77	0,00021
hsa04370	VEGF signaling pathway	2 de 57	1,76	0,004
hsa05224	Breast cancer	5 de 145	1,75	0,00000315
hsa04211	Longevity regulating pathway	3 de 87	1,75	0,00023
hsa04218	Cellular senescence	5 de 150	1,74	0,00000315
hsa04919	Thyroid hormone signaling pathway	4 de 119	1,74	0,0000148
hsa04213	Longevity regulating pathway - multiple species	2 de 61	1,73	0,0044
hsa05225	Hepatocellular carcinoma	5 de 160	1,71	0,00000315
hsa05161	Hepatitis B	5 de 159	1,71	0,00000315
hsa04929	GnRH secretion	2 de 63	1,71	0,0045
hsa05418	Fluid shear stress and atherosclerosis	4 de 130	1,7	0,0000188
hsa05142	Chagas disease	3 de 99	1,69	0,00033
hsa04620	Toll-like receptor signaling pathway	3 de 101	1,69	0,00034
hsa05221	Acute myeloid leukemia	2 de 66	1,69	0,0049
hsa05211	Renal cell carcinoma	2 de 66	1,69	0,0049
hsa04917	Prolactin signaling pathway	2 de 69	1,67	0,0051
hsa05226	Gastric cancer	4 de 144	1,66	0,0000262
hsa05167	Kaposi sarcoma-associated herpesvirus infection	5 de 187	1,64	0,00000361
hsa04668	TNF signaling pathway	3 de 112	1,64	0,00044
hsa05169	Epstein-Barr virus infection	5 de 193	1,63	0,00000361
hsa04722	Neurotrophin signaling pathway	3 de 114	1,63	0,00045
hsa05160	Hepatitis C	4 de 156	1,62	0,000034
hsa04071	Sphingolipid signaling pathway	3 de 116	1,62	0,00046
hsa04935	Growth hormone synthesis, secretion and action	3 de 118	1,62	0,00047
hsa04662	B cell receptor signaling pathway	2 de 78	1,62	0,0064
hsa01521	EGFR tyrosine kinase inhibitor resistance	2 de 78	1,62	0,0064
hsa05166	Human T-cell leukemia virus 1 infection	5 de 211	1,59	0,00000378
hsa05135	Yersinia infection	3 de 125	1,59	0,00054
hsa04926	Relaxin signaling pathway	3 de 128	1,58	0,00056
hsa04210	Apoptosis	3 de 132	1,57	0,0006
hsa05235	PD-L1 expression and PD-1 checkpoint pathway in cancer	2 de 88	1,57	0,0076
hsa05017	Spinocerebellar ataxia	3 de 135	1,56	0,00062
hsa04666	Fc gamma R-mediated phagocytosis	2 de 90	1,56	0,0078
hsa05162	Measles	3 de 138	1,55	0,00064
hsa04914	Progesterone-mediated oocyte maturation	2 de 94	1,54	0,0084
hsa05231	Choline metabolism in cancer	2 de 96	1,53	0,0086
hsa04933	AGE-RAGE signaling pathway in diabetic complications	2 de 98	1,52	0,0088
hsa04660	T cell receptor signaling pathway	2 de 101	1,51	0,0092
hsa04625	C-type lectin receptor signaling pathway	2 de 102	1,5	0,0092
hsa05206	MicroRNAs in cancer	3 de 160	1,49	0,00097
hsa04066	HIF-1 signaling pathway	2 de 106	1,49	0,0098
hsa05163	Human cytomegalovirus infection	4 de 218	1,48	0,00011
hsa04931	Insulin resistance	2 de 107	1,48	0,0098
hsa04725	Cholinergic synapse	2 de 110	1,47	0,0102
hsa05203	Viral carcinogenesis	3 de 182	1,43	0,0014
hsa04611	Platelet activation	2 de 122	1,43	0,0119
hsa04380	Osteoclast differentiation	2 de 122	1,43	0,0119
hsa04152	AMPK signaling pathway	2 de 120	1,43	0,0119
hsa04110	Cell cycle	2 de 120	1,43	0,0119
hsa04068	FoxO signaling pathway	2 de 127	1,41	0,0125
hsa05165	Human papillomavirus infection	5 de 325	1,4	0,0000188
hsa05205	Proteoglycans in cancer	3 de 196	1,4	0,0017
hsa04140	Autophagy - animal	2 de 130	1,4	0,0128
hsa04915	Estrogen signaling pathway	2 de 133	1,39	0,0132
hsa04910	Insulin signaling pathway	2 de 133	1,39	0,0132
hsa05132	Salmonella infection	3 de 209	1,37	0,002
hsa04550	Signaling pathways regulating pluripotency of stem cells	2 de 140	1,37	0,0142
hsa05131	Shigellosis	3 de 218	1,35	0,0022
hsa04072	Phospholipase D signaling pathway	2 de 147	1,35	0,0154
hsa04932	Non-alcoholic fatty liver disease	2 de 148	1,34	0,0154
hsa04150	mTOR signaling pathway	2 de 151	1,33	0,0158
hsa04934	Cushing syndrome	2 de 153	1,33	0,016
hsa04630	JAK-STAT signaling pathway	2 de 160	1,31	0,0172
hsa05164	Influenza A	2 de 165	1,3	0,0181
hsa04062	Chemokine signaling pathway	2 de 186	1,24	0,0225
hsa04010	MAPK signaling pathway	3 de 288	1,23	0,0044
hsa04510	Focal adhesion	2 de 198	1,22	0,025
hsa04015	Rap1 signaling pathway	2 de 202	1,21	0,0257
hsa05200	Pathways in cancer	5 de 517	1,2	0,00013
hsa05170	Human immunodeficiency virus 1 infection	2 de 204	1,2	0,0259
hsa04024	cAMP signaling pathway	2 de 208	1,2	0,0266
hsa04014	Ras signaling pathway	2 de 226	1,16	0,0308
hsa04151	PI3K-Akt signaling pathway	3 de 350	1,15	0,0066
hsa05168	Herpes simplex virus 1 infection	3 de 479	1,01	0,0126

Legenda: Vias bioquímicas enriquecidas pelos 12 genes com mutações em CDIS. Um total de 90 vias bioquímicas estão enriquecidas pelos 12 genes que apresentaram 15 variantes oncogênicas ou provavelmente oncogênicas em CDIS.