

**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA
Programa de Pós-graduação em Ciência Biológicas
(Bioquímica)**

PAOLA DE AVELAR CARPINETTI OLIVEIRA

**Identificação e estudo de biomarcadores personalizados
para avaliação e seguimento de pacientes com câncer de
reto tratados com quimioradioterapia neoadjuvante**

**VERSÃO CORRIGIDA DA TESE
Conforme Resolução CoPGr 5890**

**A versão original encontra-se disponível na Secretaria de Pós Graduação do
IQ-USP**

São Paulo
Data do depósito
27/11/2014

PAOLA DE AVELAR CARPINETTI OLIVEIRA

**Identificação e estudo de biomarcadores personalizados para
avaliação e seguimento de pacientes com câncer de reto
tratados com quimiorradioterapia neoadjuvante**

*Tese apresentada ao Instituto de Química da
Universidade de São para obtenção do Título Doutor
em Ciências (Bioquímica)*

*Orientadora: Dr^a. Anamaria Aranha Camargo
Co-orientador: Dr. Raphael Bessa Parmigiani*

**São Paulo
2015**

Ficha Catalográfica

Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP.

C298i Carpinetti-Oliveira, Paola de Avelar
Identificação e estudo de biomarcadores personalizados
para avaliação e seguimento de pacientes com câncer de reto
tratados com quimiorradioterapia neoadjuvante / Paola de
Avelar Carpinetti-Oliveira. -- São Paulo, 2015.
133p.

Tese (doutorado) - Instituto de Química da Universidade de
São Paulo. Departamento de Bioquímica.

Orientador: Camargo, Anamaria Aranha

Co-orientador: Parmigiani, Raphael Bessa

1. Genoma : Biologia molecular 2. Neoplasias colorretais
I. T. II. Camargo, Anamaria Aranha, orientador III. Parmigiani,
Raphael Bessa, co-orientador

574.88 CDD

A Deus e à minha família.

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ser essencial em minha vida, meu guia, socorro presente nas horas de angústia e companheiro nos momentos de glória.

À minha orientadora, Dra. Anamaria Aranha Camargo, pela paciência na orientação, pelo grande aprendizado e incentivo que tornaram possível a conclusão deste trabalho.

Ao meu co-orientador, Dr. Raphael Bessa Parmigiani, pela orientação, disponibilidade e conhecimentos transmitidos. Sua ajuda foi essencial em todas as etapas.

À Dra. Angelita Habr-Gama, ao Dr. Rodrigo Perez e ao Dr. Jorge Sabbaga pela colaboração fundamental para o desenvolvimento desta tese e pelos ensinamentos clínicos.

A todos do Centro de Oncologia Molecular – HSL, que de alguma forma contribuíram para o desenvolvimento deste trabalho, seja com as discussões enriquecedoras ou com conselhos e incentivos valiosos. Vocês tornaram o dia a dia no laboratório mais agradável.

Às minhas amigas Bruna, Camila, Gabriela, Mariana, Mayra e Nayara. E aos meus amigos Andrei, Fábio e Luiz Felipe. Vocês foram ingredientes mais que especiais nessa minha jornada, e os levarei para a vida toda.

À Camila Ramos e à Gabriela Barnabé, que desde o início foram meus pilares, transmitindo força e conhecimentos. Vocês são meus exemplos tanto pessoal quanto profissional.

À Paula Asprino, Fabiana Bettoni e Fernanda Koyama pelas constantes discussões e pelos experimentos de sequenciamento.

À Natália Felício, Bruna Quevedo e Bruna Hessel pela ajuda com as amostras e pela dedicação ao banco de tumores.

Aos meus grandes amigos "da BQI 2006" presentes em todas as fases da minha vida. Principalmente à Fernanda, Zaira e ao Thales que me ouviram e me aconselharam durante todo o doutorado.

À minha segunda família Iolanda, Getúlio e Tássia pela torcida, incentivo e orações.

À minha família, por sua capacidade de acreditar e investir em mim. Especialmente, aos meus pais, Rosa e Jaür, que fizeram tudo ser possível, e aos meus irmãos, Pamela, Ian e Iuri, pelos incentivos e torcida.

Ao Lucas, meu esposo, que contribuiu ativamente neste trabalho, me auxiliando de inúmeras maneiras, inclusive na bancada! Além de transmitir, de uma forma especial, força e coragem. Seu carinho, paciência e capacidade de me trazer paz na correria entre os experimentos fizeram essa caminhada menos árdua.

Às agências financiadora Fundação de Amparo à Pesquisa do Estado de São Paulo, Conselho Nacional do Desenvolvimento Científico e Tecnológico, e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pelo auxílios financeiros e concessão de bolsa de doutorado. Além do Instituto Ludwig de Pesquisa sobre o Câncer, Instituto de Ensino e Pesquisa do Hospital Sírio-Libanês e Instituto de Química da USP pela estrutura fornecida e auxílios financeiros.

Combati o bom combate, completei a carreira, guardei a fé.

(II Timóteo 4:7)

RESUMO

CARPINETTI-OLIVEIRA, P. A. **Identificação e estudo de biomarcadores personalizados para avaliação e seguimento de pacientes com câncer de reto tratados com quimiorradioterapia neoadjuvante.** 2015. 133 p. Tese – Programa de Pós-graduação em Ciência Biológicas (Bioquímica), Instituto de Química da Universidade de São Paulo, São Paulo.

O tratamento padrão para pacientes com câncer de reto localmente avançado consiste no uso de quimiorradioterapia neoadjuvante (QRTn), seguida por cirurgia. Uma fração significativa dos pacientes responde completamente ao tratamento e no momento da reavaliação não apresenta evidência clínica nem radiológica de doença. Uma abordagem alternativa, *Watch and Wait*, propõe não operar imediatamente esses pacientes e submetê-los a um protocolo de observação frequente, a fim de evitar as morbidades associadas à cirurgia. No entanto, a avaliação da resposta ao tratamento ainda é um desafio, devido à subjetividade da avaliação clínica e a ausência de exames radiológicos suficientemente sensíveis e específicos para garantir a ausência de células tumorais residuais ou capazes de detectar a recorrência precoce da doença. *DNA* circulante contendo alterações genéticas específicas do tumor (*ctDNA*) pode ser encontrado na fração livre de células do sangue e tem sido utilizado para monitorar a dinâmica tumoral em tumores sólidos. Avanços recentes das tecnologias de sequenciamento permitem a identificação eficiente e rápida e a um custo relativamente baixo de alterações genéticas em tumores individuais, superando o problema imposto pela ausência de alterações genéticas recorrentes nesses tumores. Essas alterações podem ser utilizadas como biomarcadores personalizados para monitorar a resposta ao tratamento, detectar doença residual e a recidiva precoce do tumor. O objetivo deste trabalho foi identificar e estudar biomarcadores personalizados em pacientes com câncer de reto localmente avançado tratados com QRTn e avaliar a capacidade desses biomarcadores para monitorar a dinâmica tumoral, e auxiliar na definição da conduta cirúrgica e na detecção da recidiva precoce da doença. Biópsias de seis pacientes com adenocarcinoma de reto distal (cT2-3N0-1M0), foram coletadas prospectivamente pré-tratamento. O *DNA* genômico extraído a partir das biópsias foi usado para construir bibliotecas tipo *mate-pair* para o sequenciamento do genoma completo, utilizando a plataforma SOLiD. Rearranjos inter e intracromossômicos foram identificados utilizando programas computacionais desenvolvidos pelo nosso grupo de pesquisa e em seguida foram validados utilizando *PCR* e sequenciamento Sanger. Foram validadas, pelo menos, três variações estruturais para cada paciente. Amostras de plasma foram coletadas no momento do diagnóstico, depois da QRTn e durante o seguimento. *DNA* circulante total foi extraído a partir das amostras

de plasma e ensaios personalizados foram desenvolvidos para monitorar a presença de variações estruturais através de *PCR* Digital. *ctDNA* foi detectado em todas amostras de plasma pré-tratamento de pacientes com tumores T3. A detecção desses biomarcadores apresentou boa correlação com a resposta ao tratamento, no entanto, esta abordagem não foi sensível o suficiente para detectar doença residual. Para dois pacientes que desenvolveram doença metastática foi verificado um aumento nos níveis de *ctDNA* com pelo menos 36 semanas antes do diagnóstico clínico de doença metastática, sendo possível correlacionar os níveis de *ctDNA* detectados em coletas subsequentes com a resposta ao tratamento sistêmico de segunda linha. Este estudo, embora de caráter exploratório, gerou dados relevantes e suficientes para justificar a realização de estudos adicionais para avaliar a aplicação dos biomarcadores personalizados na definição da conduta cirúrgica e no acompanhamento de pacientes com câncer de reto tratados com QRTn.

Palavras-chaves: câncer, biomarcadores, *DNA* circulante, rearranjo cromossômico e *PCR* Digital

ABSTRACT

CARPINETTI-OLIVEIRA, P. A. **Identification and study of personalized biomarkers for assessment and follow-up of patients with rectal cancer treated with neoadjuvant chemoradiotherapy.** 2015. 133 p. PhD Thesis – Graduate Program in Biochemistry. Chemistry Institute, Universidade de São Paulo, São Paulo.

The standard treatment for patients with locally advanced rectal cancer comprises in neoadjuvant chemo radiotherapy (nCRT), followed by surgery. A significant fraction of these patients show complete response to the treatment and at the time of reassessment, there are no clinical and nor radiological evidence of residual tumor. An alternative approach, Watch and Wait, proposes not to immediately operate these patients, but to submit them to a protocol of frequent observation in order to avoid the morbidities associated with radical surgery. However, assessment of treatment response remains a significant challenge due to the subjectivity of the clinical examination and to the lack of sufficiently sensitive tools to ensure the absence of tumor cells or to detect early disease recurrence. Circulating DNA carrying tumor-specific genetic alterations (circulating tumor DNA – ctDNA) can be found in the cell-free fraction of the blood and has been successfully used to monitor the tumor dynamics in solid tumors. Recent advances in sequencing technologies have enabled the rapid and cost effective identification of genetic alterations in individual tumors, overcoming the problem imposed by the absence of recurrent genetic alterations in these tumors. These alterations can be used as personalized biomarkers to monitor treatment response, detect residual disease and early tumor recurrence. The purpose of this work was to identify and validate the use of personalized biomarkers for patients with locally advanced rectal cancer treated with nCRT and to evaluate the ability of these biomarkers to monitor the tumor dynamics, to define surgical approach and to detect early recurrence of the disease. Pre-treatment biopsies from 6 patients with cT2-3N0-1M0 distal rectal adenocarcinoma were prospectively collected. Genomic DNA extracted from the biopsies was used to construct mate-pair libraries for whole genome sequencing using SOLiD platform. Inter and intrachromosomal rearrangements were identified using an in-house bioinformatics pipeline and validated using PCR amplification and Sanger sequencing. At least three structural variations were validated for each patient. Plasma samples were collect at diagnosis, after nCRT and follow-up. Circulating DNA was obtained from the plasma samples and personalized assays were designed to monitor the presence of structural variations using Droplet Digital PCR. ctDNA was detected in all pre-treatment plasma samples for patients with T3 tumors. The detection of these biomarkers showed a good correlation with the treatment response, nonetheless, the approach was not sensitive enough to detect residual disease.

In two patients who developed metastatic disease, an increase in ctDNA levels was observed at least 36 weeks before clinical detection of metastatic disease, and it was possible to correlate the level of ctDNA in subsequent plasma samples with response to the second-line treatment. This study, although exploratory, generated relevant and sufficient data to support additional studies to evaluate the use of personalized biomarkers in the surgical management and follow-up of rectal cancer patients treated with nCRT.

Keywords: cancer, biomarkers, circulating DNA, chromosomal rearrangement and Digital PCR

LISTA DE ABREVIATURAS E SIGLAS

°C – graus Celsius

5-FU – 5-fluorouracila

μg – micrograma

μL – microlitro

μM – micromolar

AJCC – American Joint Committee on Cancer

BLAST – *Basic Local Alignment Search Tool*

BLAT – *BLAST-Like Alignment Tool*

CA 15-3 – Antígeno carcinogênico 15-3, do inglês *cancer antigen 15-3*

CCR – câncer colorretal

CEA – Antígeno carcinoembrinário, do inglês *carcinoembryonic antigen*

cfDNA – DNA livre circulante, do inglês *cell-free DNA*

CT – Tomografia Computadorizada, do inglês *Computed Tomography*

CTC – Célula tumoral circulante

ctDNA – DNA tumoral circulante, do inglês *circulating tumor DNA*

ddPCR – PCR Digital em gotas, do inglês *Droplet Digital PCR*

DNA – ácido desoxirribonucleico, do inglês *deoxyribonucleic acid*

dsDNA – DNA dupla fita, do inglês *double-stranded DNA*

EDTA – ácido etilendiamino tetra-acético, do inglês *ethylenediamine tetraacetic acid*

FAP – Polipose Adenomatosa Familiar, do inglês *Familial Adenomatous Polyposis*

gDNA – DNA genômico

HAOC – Hospital Alemão Oswaldo Cruz

HNPCC – Câncer coloretal hereditário sem polipose, do inglês *Hereditary Nonpolyposis Colorectal Cancer*

M – molar

MGB – *Minor Groove Binding*

min – minutos

mL – mililitros

ND – Não Detectável

ng – nanograma

nm – nanômetro

nt – nucleotídeo

PARE – *Personalized Analysis of Rearranged Ends*

pb – pares de base

PCR – reação em cadeia da polimerase, do inglês *Polimerase Chain Reaction*

PET – Tomografia por Emissão de Pósitrons, do inglês *Positron Emission Tomography*

QRTn – quimiorradioterapia neoadjuvante

RCc – resposta completa clínica

RCp – resposta completa patológica

RNA – ácido ribonucleico, do inglês *ribonucleic acid*

RNase – Ribonuclease

RT – radioterapia

SDS – Dodecil sulfato de sódio, do inglês *sodium dodecyl sulfate*

TE – *Tris-EDTA Buffer*

TRG – Grau de resposta do tumor, do inglês *tumor regression grade*

VE – Variação estrutural cromossômica

LISTA DE FIGURAS

Figura 1 – Anatomia do reto.	20
Figura 2 – Modelo de evolução do câncer colorretal a partir de um pólipó.	22
Figura 3 – Mecanismos de liberação de <i>DNA</i> circulante no sangue.	33
Figura 4 – Estratégia de identificação de variações estruturais pela análise de sequências pareadas.	38
Figura 5 – Utilização da <i>PARE</i> para a detecção de <i>ctDNA</i>	40
Figura 6 – Esquema de coleta de sangue.	48
Figura 7 – Construção da biblioteca do tipo <i>mate-pair</i>	53
Figura 8 – Sequenciamento na plataforma SOLiD.	56
Figura 9 – Resumo dos principais filtros usados para a detecção de variações estruturais.	59
Figura 10 – Estratégia para o desenho dos iniciadores utilizados nas reações de <i>PCR</i> para a validação das variações estruturais.	61
Figura 11 – Validação das variações estruturais através da análise de sequências utilizando a ferramenta <i>BLAT</i>	62
Figura 12 – Etapas do <i>PCR Digital</i> em gotas.	65
Figura 13 – Estratégia de <i>Nested PCR</i> , utilizando <i>PCR Digital</i>	67

Figura 14 – Gráficos com a distribuição das distâncias entre as sequências pareadas.....	73
Figura 15 – Representação gráfica dos candidatos a variações estruturais	76
Figura 16 – Representação da validação dos candidatos a VE por <i>PCR</i> e sequenciamento pelo método de Sanger.....	78
Figura 17 – Estabilidade das variações estruturais em amostras de tumor residual e metástase.....	82
Figura 18 – Esquema com os momentos das coletas de sangue e principais desfechos clínicos de cada paciente.	83
Figura 19 – Comparação de detecção do <i>ctDNA</i> usando <i>ddPCR</i> sem e com pré-amplificação.	87
Figura 20 – Detecção dos biomarcadores personalizados por <i>PCR</i> digital: paciente com resposta incompleta.	91
Figura 21 – Detecção dos biomarcadores personalizados por <i>PCR</i> digital: paciente com resposta patológica completa.	92
Figura 22 – Detecção dos biomarcadores personalizados por <i>PCR</i> digital: paciente com resposta clínica completa seguida de doença metastática.....	96
Figura 23 – Detecção dos biomarcadores personalizados por <i>PCR</i> digital: paciente com resposta incompleta seguida de metástase.	97

LISTA DE QUADROS

Quadro 1 – Dados clínicos patológicos e de seguimentos dos pacientes	71
Quadro 2 – Relação dos biomarcadores testados no plasma	89
Quadro 3 – Sequências dos oligonucleotídeos utilizados para a validação dos candidatos a VE.	122
Quadro 4 – Sequências dos oligonucleotídeos utilizados para a detecção do <i>DNA</i> tumoral circulante (<i>ctDNA</i>) no plasma dos pacientes.....	124
Quadro 5 – Relação de todos os candidatos a VE e detalhamento das validações	125

LISTA DE TABELAS

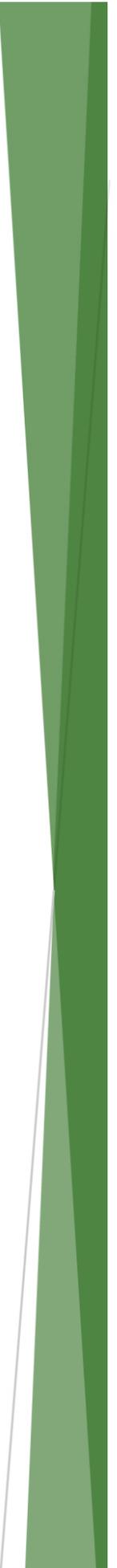
Tabela 1 – Dados de sequenciamento, alinhamento e cobertura das sequências pareadas geradas para os diferentes tumores	74
Tabela 2 – Resultado da análise para detecção de variações estruturais em cada paciente	75
Tabela 3 – Resumos das validações	80
Tabela 4 – Quantificação do <i>DNA</i> circulante nas amostras de plasma dos pacientes	85
Tabela 5 – Resultados dos ensaios de detecção do <i>ctDNA</i> para todos os ensaios e amostras	130

SUMÁRIO

1	INTRODUÇÃO	20
1.1	Câncer de reto.....	20
1.2	Conduta de observação constante (<i>Watch and Wait Approach</i>).....	27
1.3	<i>DNA</i> circulante no plasma.....	30
1.4	Biomarcadores personalizados.....	36
2	OBJETIVOS	44
2.1	Objetivo geral.....	44
2.2	Objetivos específicos	44
3	MATERIAIS E MÉTODOS	46
3.1	Pacientes e amostras biológicas	46
3.2	Coleta e processamento do sangue periférico	48
3.3	Extração de <i>DNA</i>	49
3.4	Amplificação do genoma inteiro.....	51
3.5	Construção da biblioteca de <i>DNA</i> genômico do tipo <i>mate-pair</i>	52
3.6	<i>PCR</i> em emulsão e sequenciamento na plataforma SOLiD™	54
3.7	Análise de bioinformática	57
3.8	Validação das Variações Estruturais	60
3.9	Extração de <i>DNA</i> circulante do plasma.....	63
3.10	Quantificação do <i>DNA</i> circulante total do plasma	63
3.11	Deteção do <i>DNA</i> tumoral circulante no plasma	65

4	RESULTADOS	70
4.1	Pacientes incluídos e seguimento	70
4.2	Sequenciamento e identificação das variações estruturais	72
4.3	Validação das Variações Estruturais	77
4.4	Estabilidade das variações estruturais em amostras de tumor residual e metástase	81
4.5	Quantificação do <i>DNA</i> circulante total	83
4.6	Detecção do <i>DNA</i> tumoral circulante no plasma	86
5	DISCUSSÃO	99
6	CONCLUSÃO	110
	REFERÊNCIAS	113
	APÊNDICE A – Sequências dos oligonucleotídeos (iniciadores e sondas)	122
	APÊNDICE B – Detalhamento das validações para todos os pacientes	125
	APÊNDICE C – Resultados dos ensaios de detecção do <i>ctDNA</i> para todos os ensaios e amostras	130
	LISTA DE ANEXOS	133
	ANEXO A – <i>Abstract</i> do Manuscrito em Fase Final de Preparação	134
	ANEXO B – Manuscrito Aceito para Publicação no Periódico <i>Genomics</i>	135
	ANEXO C – <i>Termo de Consentimento Livre e Esclarecido</i>	136
	ANEXO D – <i>Aprovação do Projeto no Comitê de Ética</i>	137
	ANEXO E – <i>Súmula Curricular</i>	138

1 INTRODUÇÃO



1 INTRODUÇÃO

1.1 CÂNCER DE RETO

O intestino grosso é a porção final do tubo digestivo, responsável pela absorção de água, sais minerais e a formação do bolo fecal. Medindo cerca de 1,5m de comprimento o intestino grosso é dividido em três segmentos de acordo com a anatomia e função de cada porção, a saber são elas: ceco, cólon e reto. Dentre esses segmentos o reto encontra-se na porção final do referido intestino, mais especificamente localizado entre a alça sigmoide e o canal anal, medindo entre 12 a 15 cm de comprimento. Por sua vez o reto é ainda subdividido em três porções, com base na sua localização anatômica: alto, médio e baixo (Figura 1) (Devita, Lawrence *et al.*, 2008).

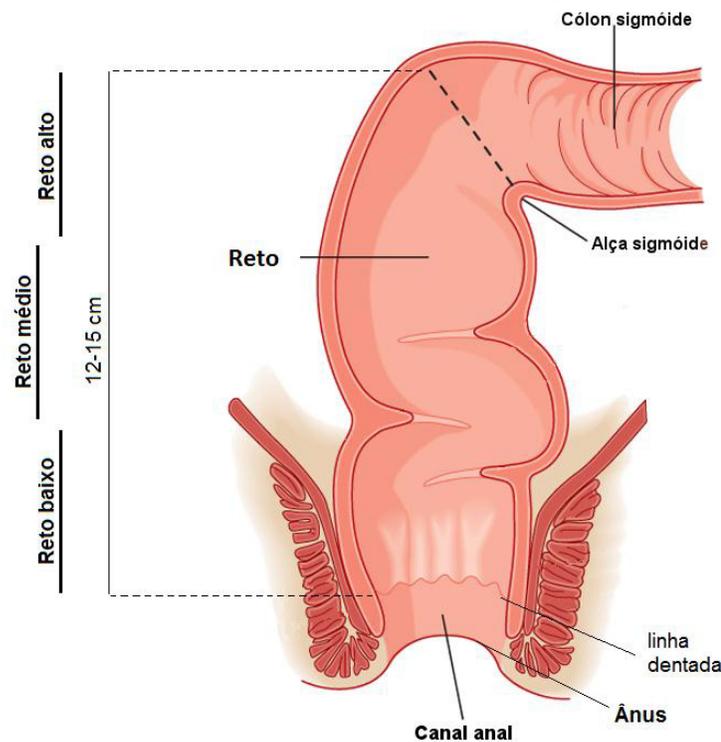


Figura 1 – Anatomia do reto. (Adaptada de Koeppen & Stanton: *Berne and Levy Physiology* – 6ª Edição, 2010)

Dentre as anomalias que acometem esta região do tubo digestório destaca-se o câncer colorretal (CCR), ele inicia-se na camada superficial do revestimento intestinal e tem sua progressão caracterizada pela invasão das camadas celulares mais profundas. Na maioria dos casos, esses tumores são originados da transformação de células glandulares epiteliais (adenocarcinomas), evoluindo segundo um modelo sequencial a partir de um pólipó benigno (adenoma), destaca-se o fato deste modelo ser bem caracterizado para o CCR. Nesse modelo, a progressão gradual e lenta do tecido normal para um carcinoma está intrinsecamente associada ao acúmulo de alterações genéticas e epigenéticas em genes como *APC*, *KRAS*, *PIK3CA*, *SMAD4* e *TP53*, os quais atribuem às células mutadas vantagens adaptativas essenciais para o desenvolvimento do fenótipo maligno (Figura 2) (Meyerson, Gabriel *et al.*, 2010; Vogelstein, Papadopoulos *et al.*, 2013).

Um outro fator com importante papel na progressão do CCR é a instabilidade cromossômica cujas consequências incluem desde alterações no número de cópias (duplicações, amplificações e deleções) a alterações estruturais (inserções, inversões, deleções e translocações). Tais alterações ocorrem nas fases iniciais da tumorigênese persistindo ao longo do desenvolvimento do tumor. (Lengauer, Kinzler *et al.*, 1998; Negrini, Gorgoulis *et al.*, 2010).

As alterações genéticas determinantes para a obtenção do fenótipo maligno podem estar presentes nas linhagens germinativas, tais como ocorre nas síndromes de câncer colorretal hereditário: Polipose Adenomatosa Familiar (*FAP*), associada às mutações no

gene *APC*, câncer colorretal hereditário sem polipose (*HNPCC*), síndrome associada a mutações em genes de reparo do *DNA* (como o *hMSH2*, *hMLH1* e *hMSH6*). Entretanto, a maioria dessas alterações são adquiridas somaticamente e acabam por ser responsáveis pelos casos esporádicos de neoplasias (Lynch e De La Chapelle, 2003).

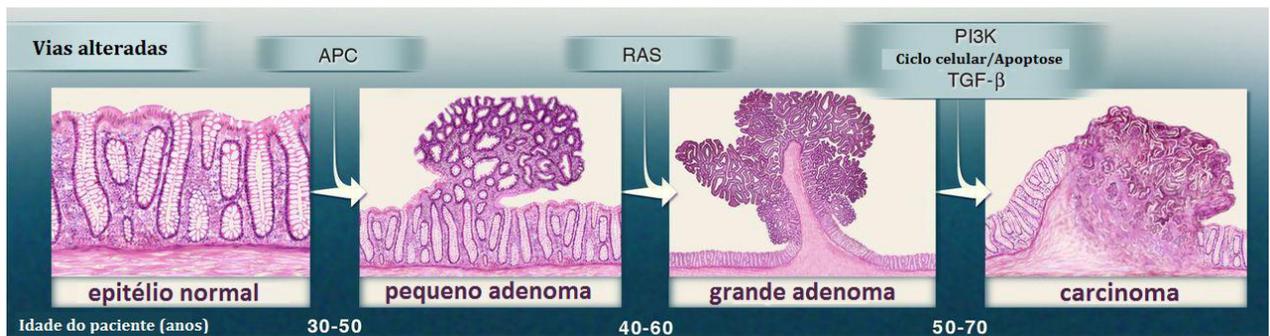


Figura 2 – Modelo de evolução do câncer colorretal a partir de um pólipo. As principais vias de sinalização que levam à tumorigênese são mostradas nas transições entre cada estágio da progressão tumoral. Um dos muitos genes que codificam componentes das vias citadas poderá estar alterado. As idades do paciente mostram os intervalos de tempo durante os quais os genes *drivers* serão geralmente mutados. (Adaptada de Vogelstein, *B. et. al* 2013)

No mundo, o câncer colorretal (cólon e reto) é o terceiro tipo de neoplasia mais comum em homens (746 mil casos), e o segundo em mulheres (614 mil casos) e é responsável por aproximadamente 695.000 mortes anualmente (Ferlay, Soerjomataram *et al.*, 2014). Nos Estados Unidos, é estimado que mais de um quarto (28%) dos casos incidentes de CCR ocorram no reto (Cheng, Chen *et al.*, 2001). No Brasil, as taxas de incidência e mortalidade por câncer de cólon e reto são crescentes e essa doença atinge ambos os sexos, sendo sua incidência maior na população acima de 50 anos (INCA, 2014). Em 2012 foram reportados 14.270 casos de morte de pacientes com esta doença, sendo

o câncer de reto responsável por 4.014 (28%) dos óbitos e para 2014 foram estimados 32.600 novos casos de câncer colorretal. Para o desenvolvimento do câncer colorretal sabe-se que fatores hereditários são responsáveis apenas por uma pequena porção dos casos, em torno de 5 a 10%, ou seja, o tipo esporádico da doença é predominante (Lynch e De La Chapelle, 2003). Outros fatores de risco são: histórico pessoal de pólipos, doenças inflamatórias intestinais, sedentarismo, hábitos alimentares (como dieta rica em carne vermelha), tabagismo e alcoolismo (Lotfi, Spencer *et al.*, 1986; Wu, Paganini-Hill *et al.*, 1987; Willett, Stampfer *et al.*, 1990; Liang, Chen *et al.*, 2009).

O rastreamento do câncer de cólon e reto é feito com a realização de exames periódicos como sangue oculto nas fezes, sigmoidoscopia flexível e colonoscopia. Dentre esses o padrão ouro para rastreamento do CCR é a colonoscopia, na qual é possível visualizar todo o cólon além da possibilidade de obtenção de biópsias e ressecção de pólipos. No entanto, tal exame possui as seguintes desvantagens: precisa de preparo prévio do intestino, sedação do paciente e a existência de um pequeno risco de ocorrência de hemorragia e desenvolvimento de infecção. Para indivíduos sem histórico familiar de CCR, os exames devem ser feitos após os 50 anos de idade e devem ser repetidos a cada 5-10 anos; indivíduos com histórico familiar, esses devem iniciar os exames 10 anos antes da idade com a qual o parente foi diagnosticado com CCR (Walsh e Terdiman, 2003). O câncer de reto quando diagnosticado precocemente apresenta taxas de remissão de 74%. Essas taxas decaem com a progressão da doença, variando entre 52 e 32% para tumores

de classificação T3N0 e T4N0, respectivamente, conforme publicado na 7ª edição do manual de estadiamento AJCC (*American Joint Committee on Cancer*) em 2010 (American Cancer Society, 2014).

O estadiamento desses tumores é feito como base no Sistema de Classificação TNM, o qual reflete a extensão anatômica da doença com base na avaliação de três componentes: extensão do tumor através das camadas do trato intestinal (T), comprometimento de linfonodos (N) e metástase à distância (M), conforme recomendação da AJCC (*American Joint Committee on Cancer*) 2002 (Sobin, Gospodarowicz *et al.*, 2009).

O tratamento cirúrgico é o principal tratamento curativo para o câncer de reto, sendo que a extensão da cirurgia acaba por variar proporcionalmente com o estadiamento da doença. Para tumores iniciais (T1N0 e T2N0) é possível uma abordagem cirúrgica mais conservadora, com a excisão local do tumor, enquanto que nos casos mais avançados da doença (T3 e T4 ou naqueles com linfonodos comprometidos) emprega-se uma modalidade cirúrgica mais radical, com a total excisão do mesorreto, podendo ocorrer a perda da função esfínteriana (Kosinski, Habr-Gama *et al.*, 2012).

Diferentes estudos demonstraram os benefícios do uso combinado da abordagem cirúrgica seguida da adoção de tratamento complementar (adjuvante) com rádio e quimioterapia em pacientes com tumores T3-4. Essa abordagem combinada contribuiu para uma importante redução dos índices de recidiva local e controle sistêmico da doença

(Thomas e Lindblad, 1988; Krook, Moertel *et al.*, 1991). Com os resultados significativos obtidos com a terapia combinada fez-se a utilização também no período pré-operatório (neoadjuvante), com objetivo primário o controle local da doença. Esses estudos sugeriram que além de ser mais eficiente no controle local da doença, a quimiorradioterapia neoadjuvante (QRTn) pode promover redução significativa do tumor primário, melhora as condições de ressecção do tumor, aumenta as taxas de preservação da função esfinteriana, e é menos tóxica (Sauer, Becker *et al.*, 2004; Moser, Ritz *et al.*, 2008; Roh, Colangelo *et al.*, 2009).

A QRTn tem sido indicada como rotina para tumores cT3-4 e/ou cN1-2 localizados no reto baixo e médio. No tratamento combinado neoadjuvante mais utilizado (também conhecido como “esquema de curso longo”), o paciente recebe a radioterapia (RT) com dose total variando entre 4500 e 5040 cGy, comumente fracionada em doses diárias de 180 cGy. Concomitante à RT é administrada a quimioterapia, na qual frequentemente é utilizada a associação do 5-fluorouracila (350mg/m²/dia) com Leucovorina (20mg/m²/dia) por cinco dias na primeira e quinta semana da RT. A resposta ao tratamento é avaliada após, pelo menos, 6-8 semanas no término da QRTn, esse período entre o tratamento e a reavaliação clínica é chamado de intervalo de descanso (Collette, Bosset *et al.*, 2007; Tulchinsky, Shmueli *et al.*, 2008).

A resposta à QRTn varia significativamente entre os pacientes a ela submetida, mesmo quando tratados dentro de uma mesma instituição, e apenas o grupo de pacientes

que respondem bem ao tratamento apresenta benefícios comprovados quanto ao aumento da sobrevida global. Feito o tratamento, podem ser observados três possíveis tipos de desfecho no quadro do paciente: resposta completa, resposta incompleta e a não resposta ao tratamento.

Para se classificar as respostas fazem-se uma série de análises, definidas a partir de critérios clínicos e exames de imagem. A resposta completa, denominada Resposta Completa Clínica (RCc), consiste na observância de resultado negativo no exame do toque retal (ausência de cicatrizes ou úlceras), resultados negativos na ressonância magnética ou ultrassonografia endorretal e níveis normais de antígeno carcinoembrionário (CEA) no sangue. Em outro cenário, se após o intervalo de descanso for detectada a presença de úlceras ou cicatrizes no local do tumor primário, e depois da cirurgia o exame histopatológico revelar ausência de células tumorais (pTON0), essa resposta recebe o nome de Completa Patológica (RCp). Por fim, no quadro de Resposta Incompleta o paciente apresenta uma regressão no tamanho do tumor, mas tanto os exames clínicos, de imagem e histopatológico da peça cirúrgica, revelam a presença de células tumorais viáveis (Therasse, Arbuck *et al.*, 2000; Kosinski, Habr-Gama *et al.*, 2012).

Adite-se que mesmo com os avanços obtidos no tratamento do câncer de reto, complicações pós-operatórias e uma taxa significativa de mortalidade (2-3%) são frequentemente reportadas (Emmertsen e Laurberg, 2013). Quando as complicações pós-operatórias são consideradas em conjunto, elas podem acometer mais de 40% dos

pacientes até um ano após a cirurgia (Paun, Cassie *et al.*, 2010; Emmertsen e Laurberg, 2013). A intervenção cirúrgica também está associada a uma elevada taxa de consequências funcionais adversas. Dentre essas as mais frequentemente apresentadas pelos pacientes com tumores de reto estão: disfunção erétil, incontinência fecal e distúrbios urinários. Nos casos nos quais não é possível a preservação da função esfinteriana do paciente e se faz necessária a colostomia definitiva, o impacto na qualidade de vida do indivíduo em questão pode ser ainda maior (Vironen, Kairaluoma *et al.*, 2006).

1.2 CONDUTA DE OBSERVAÇÃO CONSTANTE (*WATCH AND WAIT APPROACH*)

A implementação da QRTn, como tratamento padrão de tumores de reto localizados, revelou que a parcela dos pacientes que respondiam completamente ao tratamento, e mesmo assim eram submetidos à cirurgia, era de 15-27% (Maas, Nelemans *et al.*, 2010). Essa observação levou a comunidade médica a questionar a necessidade da indicação de cirurgia para esse grupo específico de pacientes.

Nesse cenário, o grupo de médicos liderados pela Dra. Angelita Habr-Gama é pioneiro no Brasil e no mundo em aperfeiçoar os protocolos de QRTn para o tratamento de pacientes com tumor de reto e em propor uma conduta de observação frequente, sem intervenção cirúrgica imediata (*Watch and Wait Approach*), para os pacientes que apresentam resposta completa clínica após o tratamento neoadjuvante, visto que os

benefícios para tal grupo de pacientes é questionável devido aos seus riscos e morbidades associados à cirurgia aqui supracitados.

Em um trabalho publicado em 2004, o grupo da Dra. Habr-Gama apresentou sua experiência com a conduta *Watch and Wait*. Foram acompanhados 265 pacientes com tumor de reto distal cirurgicamente ressecável e tratados com QRTn (5-FU+Leucovorin+5040cGy). Dos 265 pacientes, 71 (26.8%) apresentaram RCc, foram poupados de cirurgia imediata, e submetidos a consultas constantes de acompanhamento. Os 194 pacientes restantes, apresentaram resposta clínica incompleta ou não responderam, após 8 semanas do término do tratamento, e foram submetidos à cirurgia radical com excisão total do mesorreto e análise histopatológica do espécime cirúrgico; desses pacientes operados, 22 (11,3%) apresentaram RCp (ausência de células tumorais viáveis no espécime cirúrgico) e poderiam ter sido poupados da intervenção cirúrgica. A fim de avaliar os riscos da estratégia *Watch and Wait*, os autores avaliaram se havia algum impacto na sobrevida entre os pacientes com resposta completa submetidos à cirurgia radical, pacientes com RCp, e o grupo que foi submetido ao protocolo de observação constante (RCc). Não foram observadas diferenças estatisticamente significativas em termos de sobrevida livre de doença (83 versus 92% para RCc e RCp respectivamente, $p=0.09$) e uma pequena diferença foi observada na sobrevida global (100 versus 88% para RCc e RCp respectivamente, $p=0.01$). O número de recidivas sistêmicas foi igual nos dois grupos e dois pacientes do grupo de resposta clínica completa apresentaram recidiva local

da doença. Esses pacientes que recidivaram foram submetidos à cirurgia de resgate e não foi constatado impacto na sobrevida global desses pacientes (Habr-Gama, Perez *et al.*, 2004).

Em outro estudo publicado recentemente pelo grupo, no qual houve a inclusão de 70 pacientes, demonstrou-se que com alterações no esquema de QRTn foi possível aumentar as taxas de resposta completa ao tratamento neoadjuvante. No grupo de pacientes avaliados, 39 (57%) apresentaram resposta completa sustentada com mediana de seguimento de 56 meses. Desses pacientes, 4 (10%) desenvolveram recorrência local tardia (>12 meses de seguimento) (Habr-Gama, Sabbaga *et al.*, 2013).

Em conjunto, esses trabalhos indicam que a abordagem cirúrgica imediata não traz benefícios para a sobrevida livre de doença nesse grupo pacientes e que a estratégia de observação frequente sem intervenção cirúrgica, demonstra ser uma abordagem atraente, uma vez que elimina os riscos e morbidades associadas à intervenção cirúrgica radical.

Por outro lado, o sucesso da abordagem *Watch and Wait* depende de uma série de fatores que incluem desde uma equipe preparada de cirurgiões e oncologistas, bem treinados e experientes para a avaliação desses pacientes, a existência de um sistema de atendimento altamente capaz de imediatamente identificar e viabilizar um seguimento frequente e contínuo dos pacientes, no caso de eventual recorrência local ou sistêmica da doença (Kosinski, Habr-Gama *et al.*, 2012).

Dada essa conjuntura, fica evidente a necessidade do desenvolvimento de marcadores específicos e sensíveis para avaliar a resposta ao tratamento e a presença de doença residual, bem como para realizar o seguimento dos pacientes submetidos a esse protocolo. Especificamente, esses marcadores seriam úteis para: i) adoção de um tratamento personalizado para o câncer de reto, auxiliando na escolha da conduta cirúrgica pós QRTn, o que beneficiaria principalmente os pacientes com RCp que, apesar de não apresentarem células tumorais viáveis após o tratamento neoadjuvante, são hoje submetidos à intervenção cirúrgica, mesmo dentro da conduta *Watch and Wait*, devido à baixa especificidade dos exames clínicos e de imagem; ii) permitir uma maior difusão da conduta *Watch and Wait*, uma vez que traria mais segurança e menor subjetividade na avaliação da resposta a QRTn; iii) reduzir os custos do seguimento, e exposição do paciente à radiação, visto que a tomografia computadorizada e *PET/CT* (Tomografia por Emissão de Pósitrons acoplada a uma tomografia computadorizada) são exames comumente utilizados para essa finalidade; e iv) permitir a detecção precoce de recidivas, para as quais seria realizada a cirurgia de resgate, quando ocorrer recorrência local, ou quimioterapia sistêmica, quando ocorrer o desenvolvimento de metástases, aumentando assim as chances de sobrevida global dos pacientes.

1.3 *DNA* CIRCULANTE NO PLASMA

DNAs circulantes são fragmentos de ácido desoxirribonucleico presentes em fluidos corpóreos, como o sangue, que pode ser encontrado livre de qualquer compartimento ou

dentro de pequenas vesículas e micropartículas extracelulares. A existência de ácidos nucléicos livre de células na circulação de indivíduos sadios foi demonstrada pela primeira vez em 1948, mas a importância desses achados ainda não era compreendida (Mandel e Metais, 1948). Trinta anos mais tarde, Leon e colaboradores propuseram sua utilidade como biomarcador capaz de auxiliar no diagnóstico de pacientes com câncer, uma vez que seus estudos evidenciaram níveis elevados de *DNA* circulante no soro de pacientes com câncer quando comparados a indivíduos sadios e que possivelmente parte desse *DNA* tinha origem tumoral, devido a correlação negativa entre os níveis de *DNA* circulante e a terapia efetiva anticâncer (Leon, Shapiro *et al.*, 1977). No entanto, em estudos posteriores, foi verificado que a elevação do *DNA* circulante não é exclusiva de condições malignas e os níveis de *DNA* no plasma estão também aumentados em outras condições fisiológicas e patológicas, como por exemplo, gravidez, trauma, infarto do miocárdio e em doenças pró-inflamatórias, invalidando, a princípio, sua utilização como marcador molecular para o diagnóstico de câncer (Lo, Corbetta *et al.*, 1997; Lo, Rainer *et al.*, 2000; Chang, Chia *et al.*, 2003; Galeazzi, Morozzi *et al.*, 2003; Gahan, 2008).

O primeiro trabalho a comprovar que o tumor contribui com a liberação de *DNA* circulante foi publicado em 1994, quando Sorenson e colaboradores detectaram mutações pontuais tumor-específica em um gene comumente mutado em câncer (*KRAS*) no *DNA* circulante de pacientes com carcinoma de pâncreas (Sorenson, Pribish *et al.*, 1994). Estudos utilizando xenotransplantes de tumores humanos em camundongos

imunodeficientes também foram importantes para confirmar a origem tumoral de parte do *DNA* circulante total e caracterizar as propriedades do *ctDNA*. Esses estudos foram capazes de correlacionar o aumento da massa tumoral com a maior liberação de *DNA* tumoral no plasma, bem como o decréscimo nas concentrações de *ctDNA* nos camundongos, tanto após a ressecção do tumor quanto após a intervenção terapêutica citotóxica de sucesso. A partir desses estudos também foi possível caracterizar que o *DNA* circulante de origem tumoral, *ctDNA*, apresenta uma fragmentação muito maior que *DNA* circulante liberados por células normais (Rago, Huso *et al.*, 2007; Thierry, Mouliere *et al.*, 2010; Mouliere, Robert *et al.*, 2011).

Os mecanismos de liberação dessas moléculas de *DNA* na circulação ainda não estão completamente compreendidos, todavia, a principal hipótese é que o *DNA* circulante seja originado a partir de células apoptóticas e/ou necróticas de diferentes tecidos. Em condições fisiológicas normais, macrófagos e outras células especializadas são responsáveis por fagocitar células que morrem por necrose ou apoptose, assim, os níveis de *DNA* circulante encontrados em indivíduos sadios são baixos. No entanto, em condições de dano ao tecido por exercício exaustivo, inflamação e crescimento tumoral, a remoção dessas células por fagocitose não ocorre eficientemente e pode resultar no acúmulo de detritos celulares, como fragmentos de *DNA* (Figura 3) (Schwarzenbach, Hoon *et al.*, 2011; Crowley, Di Nicolantonio *et al.*, 2013).

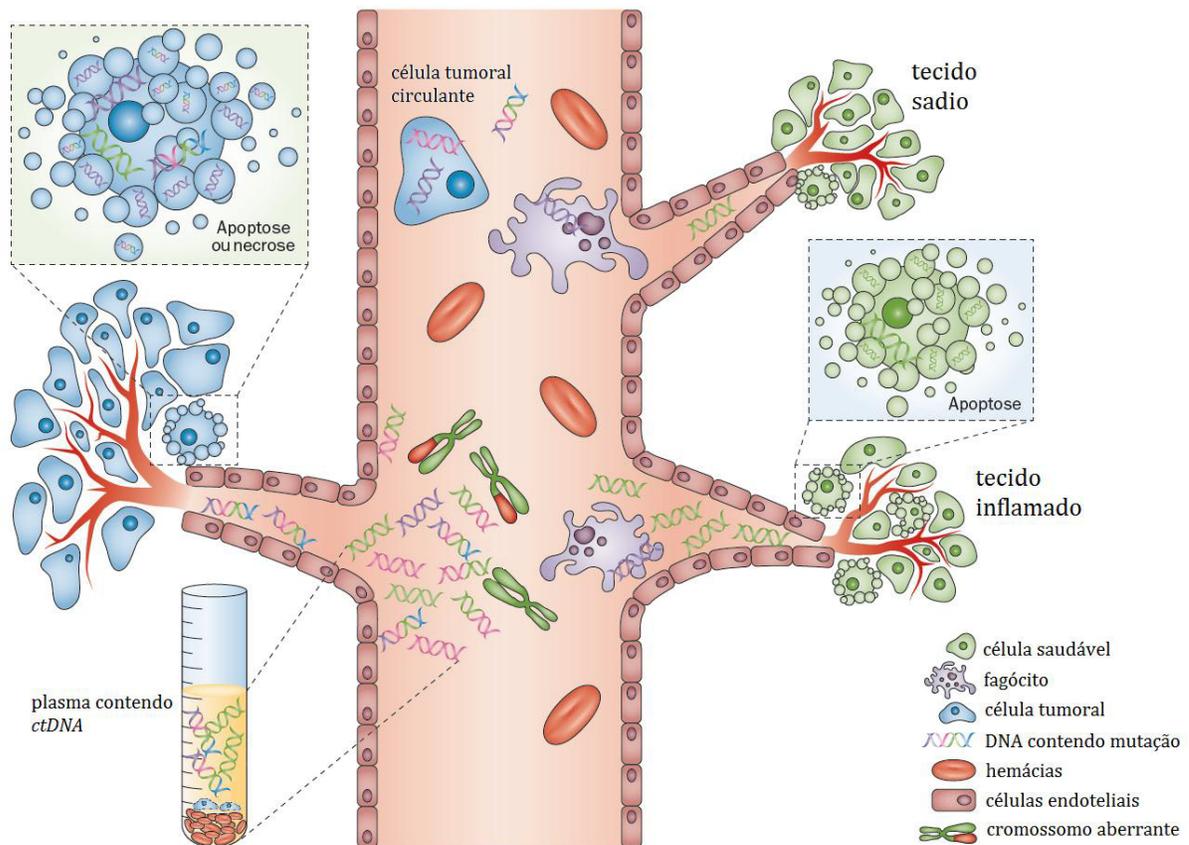


Figura 3 – Mecanismos de liberação de DNA circulante no sangue. O DNA circulante é resultado da liberação do DNA de células de tecido sadio que são renovadas, e em maior proporção por células do tecido inflamado e/ou tumoral, no qual mecanismos de morte como apoptose e necrose são frequentemente observados. Através de amostras de sangue é possível a extração do DNA circulante total e a identificação da fração que provém de células tumorais (*ctDNA*). O *ctDNA* pode ser quantificado através de ensaios que detectam alterações genéticas específicas de células tumorais, como mutações pontuais (na figura representadas por hélices com fitas coloridas em roxo, rosa verde e azul), alterações do número de cópias (porção vermelha dos cromossomos) e rearranjos cromossômicos (hélices rosa e verde). (Adaptada de Crowley, E. *et al.* 2013)

Inicialmente, foi proposto que o *ctDNA* era proveniente de células tumorais circulantes que, quando rompidas, liberavam DNA livre na circulação. No entanto, estudos recentes que avaliaram, simultaneamente, níveis de *ctDNA* e CTC em amostras de sangue de pacientes com câncer de mama e colorretal e concluíram que as CTCs não são a provável fonte de *ctDNA*. Essa conclusão foi gerada devido ao fato de que, em ambos estudos, não foi encontrada correlação entre os níveis de *ctDNA* e CTCs e quantidades

significativas de *ctDNA* puderam ser isoladas de amostras em que não foi possível o isolamento de CTC (Dawson, Tsui *et al.*, 2013; Bettegowda, Sausen *et al.*, 2014).

Após a elucidação das bases genéticas do câncer e os recentes avanços nos estudos de genômica, várias abordagens para detectar tumores de forma não invasiva, utilizando *ctDNA* têm sido propostas. Esses estudos basearam-se em características genéticas (instabilidade de microssatélite, mutação pontuais em oncogenes e genes supressores de tumor e instabilidade cromossômica) e epigenéticas (metilação de regiões promotora de genes supressores de tumor) comumente encontradas em tumores, para detecção indireta dos mesmos, através da identificação de *ctDNA* no sangue, fezes e urina de paciente com diferentes tipos de tumores (Nawroz, Koch *et al.*, 1996; Diehl, Schmidt, Durkee *et al.*, 2008; Devos, Tetzner *et al.*, 2009; Esposito, Bardelli *et al.*, 2014).

Estudos que quantificaram *ctDNA* em pacientes com câncer colorretal mostraram que o número de moléculas de *DNA* tumoral na circulação é progressivamente aumentado em tumores invasivos, correlacionando, assim, a carga tumoral com o estadiamento do tumor (Diehl, Schmidt, Choti *et al.*, 2008; Bettegowda, Sausen *et al.*, 2014). Consistentemente, um estudo que comparou os níveis de *ctDNA* dos pacientes com doença metastática com os dos pacientes com doença localizada, especificamente para os tumores colorretais, detectou *ctDNA* em 100% e 73% das amostras, respectivamente. O *ctDNA* também mostrou ser uma ferramenta de alto valor prognóstico para pacientes com doença avançada, pois foi encontrada uma correlação negativa entre a concentração

de *ctDNA* e a sobrevida, ou seja, pacientes com CCR metastático com níveis de *ctDNA* relativamente baixos viviam significativamente mais tempo do que os pacientes com níveis mais elevados (Bettegowda, Sausen *et al.*, 2014).

Um desses estudos também comparou a sensibilidade de detecção de *ctDNA* contra a de *CEA*, biomarcador utilizado rotineiramente na clínica para seguimento de indivíduos com CCR, para monitorar a doença antes e depois da ressecção cirúrgica. A diferença de sensibilidade foi estatisticamente significativa em ambos os momentos: a detecção de níveis elevados de *ctDNA* e *CEA* no período pré-operatório foi de 100% contra 56%, e a redução média encontrada após a cirurgia foi de 32,5% e 99%, respectivamente. Esses resultados refletem algumas das limitações do marcador padrão atual, que além de estar em níveis normais no momento do diagnóstico em uma parcela significativa dos pacientes, para os pacientes *CEA* positivos, após a ressecção do tumor, quase 70% não apresenta uma redução nos níveis desse marcador, como seria esperado (Diehl, Schmidt, Choti *et al.*, 2008).

As principais vantagens em se estudar *DNA* tumoral livre na circulação são: a obtenção através de métodos não-invasivos, estabilidade (estável por vários anos nas amostras de plasma ou soro armazenadas adequadamente) e a relativa simplicidade de uso (pode ser extraído usando procedimentos simples de purificação) (Alix-Panabieres, Schwarzenbach *et al.*, 2012). Somado a essas vantagens tem-se que os biomarcadores baseados em *ctDNA* vem mostrando-se mais informativos, sensíveis e específicos quando

comparados a marcadores proteicos, como CA-15-3 e CEA para pacientes com tumor de mama e colorretal, respectivamente (Diehl, Schmidt, Choti *et al.*, 2008; Dawson, Tsui *et al.*, 2013).

1.4 BIOMARCADORES PERSONALIZADOS

Atualmente é possível a detecção de mutações pontuais e variações estruturais por métodos precisos e sensíveis. No entanto, devido a heterogeneidade genética característica dos tumores sólidos, surge o desafio da utilização na prática clínica de marcadores de *ctDNA* baseados na análise de mutações em genes específicos (Tcga, 2012; Drier, Lawrence *et al.*, 2013; Vogelstein, Papadopoulos *et al.*, 2013). Estudos utilizando a detecção de *ctDNA*, através da identificação de mutações em genes pontuais comumente envolvidos no processo de tumorigênese, não foram suficientemente abrangentes para a avaliação de todos os pacientes. Para minimizar esse obstáculo estão sendo desenvolvidos painéis de mutações frequentemente encontradas em determinados tipos de tumores e, mais recentemente, a busca por marcadores personalizados (Garcia-Aguilar, Chen *et al.*, 2011; Esposito, Bardelli *et al.*, 2014; Lim, Becker *et al.*, 2014).

Em vista dos recentes avanços nas tecnologias de sequenciamento, chamada de Sequenciamento de Nova Geração (*Next Generation Sequencing*), e aumento da capacidade de processamento das sequências geradas, surgiu a possibilidade de se caracterizar as alterações genéticas de um determinado tumor de forma individualizada, e desta forma superar parte das limitações anteriormente abordadas (Ten Bosch e Grody,

2008; Roukos, 2010). Uma característica quase universal de câncer humano é o rearranjo generalizado de cromossomos como resultado da instabilidade cromossômica (Lengauer, Kinzler *et al.*, 1998; Albertson, Collins *et al.*, 2003; Negrini, Gorgoulis *et al.*, 2010). Tais alterações estruturais começam a ocorrer nas fases iniciais da tumorigênese e, quando ocorrem nas etapas iniciais do processo, têm grandes chances de persistirem ao longo do desenvolvimento do tumor (Vogelstein, Papadopoulos *et al.*, 2013). Tais alterações não estão presentes em células normais e devem ser perfeitamente específicas.

Com isto, uma nova geração de biomarcadores tornou-se disponível, na qual a detecção do *ctDNA* é baseada nas alterações genéticas específicas de cada tumor. Em um trabalho, publicado no início de 2010 por pesquisadores da Universidade Johns Hopkins, o genoma completo de 10 tumores (6 tumores colorretais e 4 tumores de mama) foi sequenciado e alterações cromossômicas de forma personalizada foram identificadas. Para isso os pesquisadores utilizaram-se de uma metodologia, denominada *PARE* (*Personalized Analysis of Rearranged Ends*), que se baseia na produção de sequências pareadas (*mate-pair reads*) as quais, uma vez mapeadas na sequência de referência do genoma humano, permitem identificar variações na estrutura dos cromossomos presentes nas células tumorais (Figura 4). Essas variações estruturais, inter e intracromossômicas foram identificadas em todos os tumores com uma média de 9 variações por tumor. Posteriormente, as alterações identificadas através dessa metodologia foram validadas por *PCR* (*Polymerase Chain Reaction*) e utilizadas na

detecção indireta de células tumorais no sangue dos pacientes em diferentes pontos do tratamento (Leary, Kinde *et al.*, 2010).

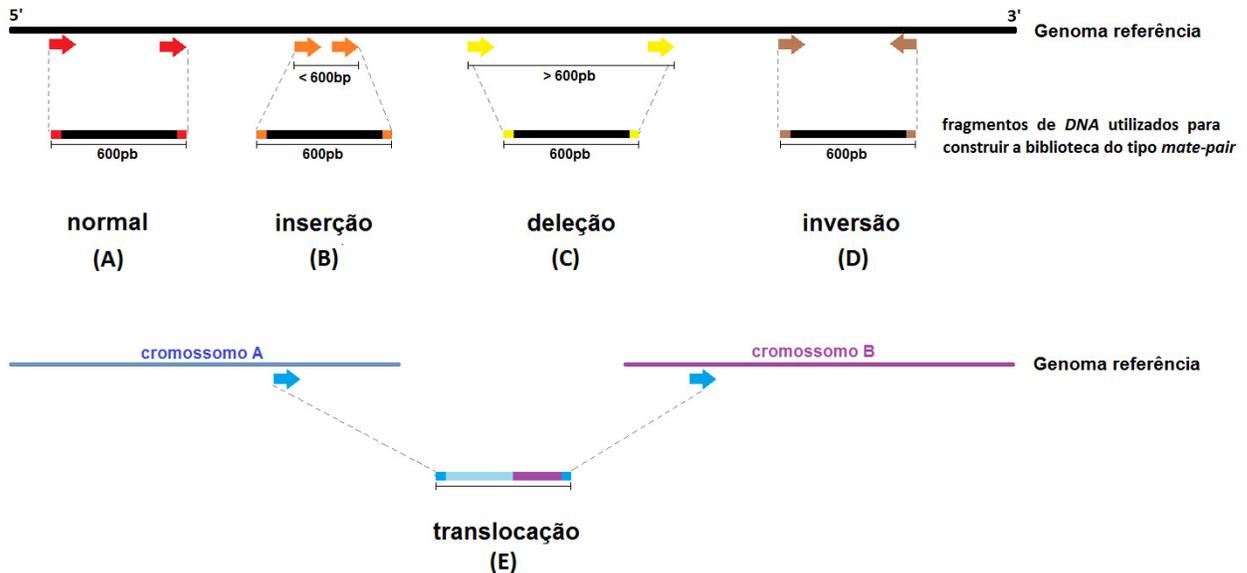


Figura 4 – Estratégia de identificação de variações estruturais pela análise de sequências pareadas. Através do sequenciamento de bibliotecas do tipo *mate-pair* é possível identificar variações estruturais quando realizado o alinhamento dessas sequências pareadas contra a sequência do genoma humano de referência. Basicamente, com essa metodologia, fragmentos de DNA de tamanho conhecido (por exemplo 600 pb) têm somente suas extremidades sequenciadas (~50 bp). Devido a construção da biblioteca e ao sequenciamento serem direcionados, é possível obter a informação quanto à orientação dessas sequências pareadas no genoma que originou a biblioteca (qual região está à montante ou à jusante). **(A)** Quando a região do genoma de interesse (genoma tumoral) não possui nenhuma alteração estrutural, espera-se que as sequências das extremidades do fragmento correspondente a essa região, quando comparadas com o genoma de referência, distem ~600pb uma da outra, devido a essa faixa de tamanho ter sido escolhida durante a construção da biblioteca. **(B)** Quando as sequências pareadas alinham-se ao genoma de referência com distância menor do que o esperado (<600 pb) indica que no genoma tumoral ocorreu uma inserção naquela região. **(C)** Se a distância for maior (> 600pb) há evidência de que parte daquela região foi deletada no genoma tumoral. **(D)** É possível também identificar quando ocorreu uma inversão, uma vez que nesse contexto as sequências pareadas serão alinhadas em fitas opostas do genoma referência, como indicado pela orientação das setas, onde a seta (\rightarrow) corresponde a fita mais (5' \rightarrow 3') e a seta (\leftarrow) representa a fita menos (3' \leftarrow 5'). **(E)** Quando houver uma translocação, após o alinhamento com o genoma referência, as sequências pareadas serão alinhadas em cromossomos distintos, no entanto, no genoma tumoral essas regiões estão unidas fisicamente, pois fazem parte do mesmo fragmento de DNA.

A estratégia utilizando a metodologia *PARE* foi capaz de avaliar a resposta dos pacientes ao tratamento, com alta especificidade e sensibilidade, pois as alterações analisadas não estavam presentes em células normais e foi possível detectar um genoma

tumoral em meio a 400.000 normais. Outra vantagem da metodologia *PARE* é a redução dos custos de sequenciamento, uma vez que essa estratégia mostrou ser capaz de identificar variações estruturais no genoma tumoral sem a necessidade do sequenciamento do genoma normal do mesmo paciente.

Um outro grupo de pesquisadores, através de uma estratégia semelhante, foi capaz de encontrar rearranjos cromossômicos em 2 pacientes com câncer de mama e 1 paciente com osteossarcoma, detectando-os de maneira sensível e específica no plasma/soro destes pacientes (Mcbride, Orpana *et al.*, 2010).

Esses estudos demonstraram que essa abordagem apresenta uma reduzida taxa de resultados falso-positivos, pois a amplificação de fusões aberrantes de sequências de *DNA*, que envolvem centenas e até milhares de pares de bases, não ocorre no *DNA* de células normais. Assim os ensaios para a detecção de alterações estruturais apresentam vantagens práticas sobre mutações pontuais para monitorar a doença residual (Mcbride, Orpana *et al.*, 2010; Dawson, Tsui *et al.*, 2013; Bettegowda, Sausen *et al.*, 2014).

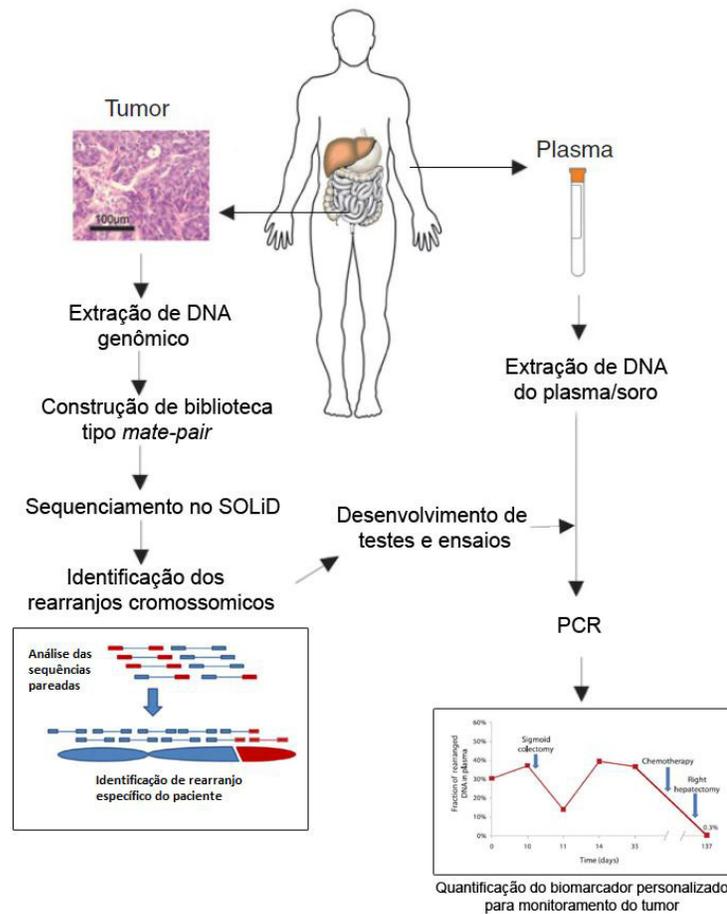


Figura 5 – Utilização da PARE para a detecção de *ctDNA*. (Adaptada de Leary, R. J. *et al.* 2010)

Mais recentemente, um artigo publicado no *The New England Journal of Medicine*, mostrou que o *ctDNA* representa uma ferramenta mais efetiva no monitoramento de câncer de mama metastático. Esse trabalho monitorou trinta mulheres com câncer de mama metastático e comparou os resultados dos exames de imagem de acompanhamento do tumor com os ensaios para detecção de *ctDNA*, CTCs e CA 15-3 (marcador utilizado atualmente na clínica), e concluiu que o *DNA* tumoral circulante

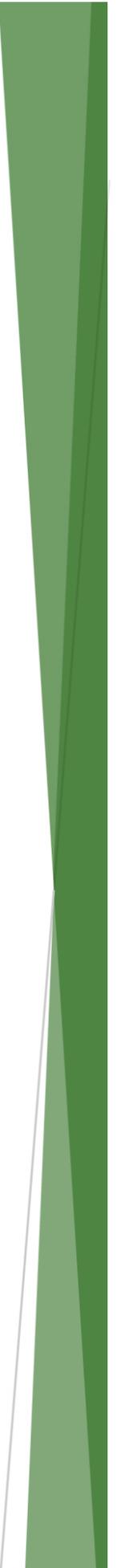
representa um biomarcador sensível, específico e informativo para câncer de mama metastático (Dawson, Tsui *et al.*, 2013).

Até 2014, os estudos utilizando biomarcadores personalizados baseados em *DNA* tumoral circulante eram restritos a tipos específicos de tumores, na maioria das vezes incluíram poucos pacientes e utilizaram diferentes metodologia para a detecção desses biomarcadores. Por isso, com a finalidade de avaliar a utilidade clínica desses biomarcadores personalizados de maneira padronizada e envolvendo vários tipos câncer, foi realizado um estudo, publicado em 2014 na *Science Translational Medicine*, que avaliou a capacidade do *ctDNA* para detectar tumores em 640 pacientes. Esse trabalho concluiu que esse tipo de marcador apresenta capacidades distintas de detecção da doença entre os diferentes tipos tumorais. O *ctDNA* foi detectável em mais de 75% dos doentes com câncer avançado de pâncreas, ovários, colorretal, de bexiga, gastresofágico, mama, melanoma, hepatocelular, e cabeça e pescoço avançado, mas em menos do que 50% dos pacientes com tumores cerebrais primários, renais, de próstata, ou câncer de tireoide. Em pacientes com tumores localizados, o *ctDNA* foi detectado em 73, 57, 48 e 50% dos pacientes com câncer colorretal, câncer gastro-esofágico, câncer de pâncreas, e adenocarcinoma de mama, respectivamente (Bettegowda, Sausen *et al.*, 2014).

Associando os recentes avanços na medicina personalizada e a real necessidade de biomarcadores mais sensíveis para o monitoramento de pacientes com tumores de reto tratados com QRTn, de modo a evitar cirurgias desnecessárias e reduzir as morbidades

associadas a esse procedimento, o objetivo desse estudo exploratório foi avaliar o potencial dos biomarcadores personalizados, baseados em *DNA* tumoral circulante, e sua utilidade no manejo da doença.

2 OBJETIVOS



2 OBJETIVOS

2.1 OBJETIVO GERAL

Identificar biomarcadores personalizados baseados em *ctDNA* e estudar sua utilidade para avaliar a resposta de pacientes com câncer de reto (T2/T3) submetidos ao tratamento com radio e quimioterapia neoadjuvante e a capacidade desses marcadores em detectar recidiva precoce da doença durante o seguimento desses pacientes.

2.2 OBJETIVOS ESPECÍFICOS

- Utilizar a metodologia *PARE* para identificar variações cromossômicas estruturais presentes no genoma tumoral de cada paciente com câncer de reto.
- Validar a ocorrência dessas variações estruturais no genoma tumoral por meio de *PCR* e sequenciamento Sanger e desenhar ensaios otimizados para a detecção dos mesmos em amostras de *DNA* extraído a partir do plasma desses pacientes.
- Avaliar a presença dessas alterações em amostras de plasma dos pacientes por *PCR* Digital em diferentes momentos do tratamento e acompanhamento, mais especificamente nos momentos de: diagnóstico inicial, após o término do tratamento neoadjuvante, avaliação clínica e radiológica e visitas de seguimento após o término do tratamento.

3 MATERIAIS E MÉTODOS



3 MATERIAIS E MÉTODOS

3.1 PACIENTES E AMOSTRAS BIOLÓGICAS

Foram selecionados 6 pacientes com câncer de reto distal recrutados entre 17/06/2010 e 31/03/2011, sendo 2 (CCR1 e CCR3) com estadio T2 e 4 (CCR2, PCR1, IR1 e IR2) com estadio T3. Todos foram submetidos ao protocolo de tratamento neoadjuvante com radioterapia (5040-5400cGy) e quimioterapia baseada em 5-FU e Leucovorin, e reavaliados, pelo menos, 12 semanas após o término do tratamento (Habr-Gama *et al.*, 2004 e Habr-Gama *et al.* 2009). Durante a reavaliação da resposta foram realizados exames clínicos, como o toque retal e nível sérico de CEA, e exames endoscópicos e radiológicos.

Os pacientes com evidência clínica ou radiológica da doença foram encaminhados para a cirurgia radical imediata, incluindo total excisão do mesorreto. Aqueles sem evidência de doença residual não foram operados e foram seguidos um acompanhamento rigoroso (estratégia *Watch and Wait*). Os pacientes com resposta incompleta foram seguidos após a cirurgia radical a cada três meses para os primeiros 2 anos e cada 6 meses depois do terceiro ano. Os pacientes com resposta completa clínica foram acompanhados a cada 1-2 meses para o primeiro ano, a cada 3 meses para o segundo ano e a cada 6 meses a partir do terceiro ano.

Dos 6 pacientes em questão, 3 deles apresentaram resposta completa clínica (CCR1, CCR2 e CCR3), 1 paciente teve resposta patológica completa (PCR1) e 2 deles tiveram

resposta incompleta (IR1 e IR2). Dos pacientes com RCc, 1 (CCR1) apresentou recidiva local da doença 51 semanas após o término do tratamento, e o CCR2 foi diagnosticado com metástase hepática após 62 semanas do término do tratamento. No grupo de resposta incompleta, o IR1 desenvolveu metástase no fígado após 36 semanas do término do tratamento e no arco costal após 82 semanas.

Além das biópsias do tumor primário, amostra do tumor residual e metástase; também foram coletadas amostras de sangue de todos os pacientes em diferentes etapas do tratamento: i) no diagnóstico inicial; ii) após o término do tratamento neoadjuvante; iii) no momento da avaliação clínica e radiológica da resposta ao tratamento; iv) nas visitas de seguimento, conforme ilustrado na Figura 6.

Os tecidos tumorais de todos os pacientes foram coletados através de biópsia endoscópica antes do início do tratamento, o qual foi congelado e macro dissecado ao microscópio óptico para que, pelo menos, 80% do tecido fosse composto por células tumorais. Todas as amostras foram coletadas somente após o consentimento informado do paciente mediante a assinatura do termo de consentimento livre e esclarecido do doador ao Banco de Tecido e Seroteca do Hospital Alemão Oswaldo Cruz (BTS-HAOC), conforme modelo no ANEXO C. Este projeto foi submetido e aprovado pelo comitê de ética em pesquisa do Hospital Alemão Oswaldo Cruz, sob protocolo de pesquisa número 22/08 (ANEXO D).

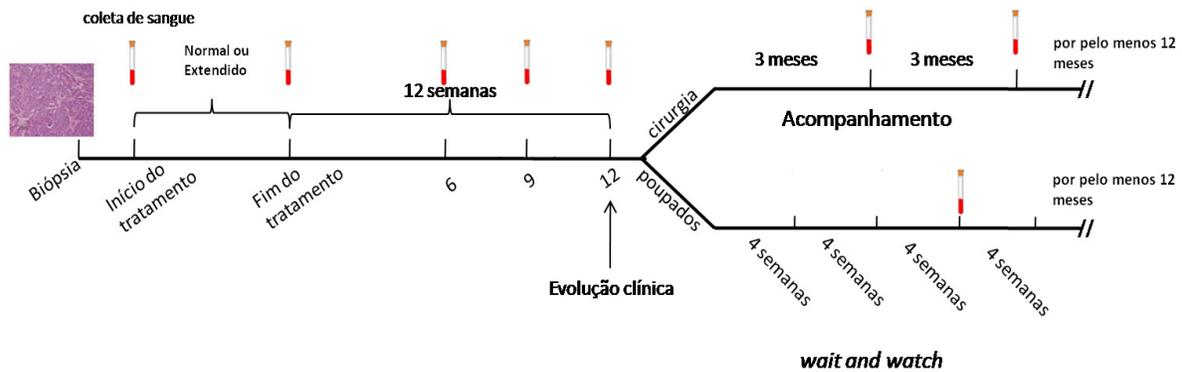


Figura 6 – Esquema de coleta de sangue.

3.2 COLETA E PROCESSAMENTO DO SANGUE PERIFÉRICO

Amostras de 9 mL de sangue periférico foram coletadas em 3 tubos (BD Vacutainer®) com solução anticoagulante (K_3EDTA) de cada paciente em vários momentos do tratamento. A fim de garantir a preservação e integridade do *DNA* livre no sangue, tais amostras foram transportadas em gelo e processadas em um período máximo de 2 horas. O processamento consistiu em, primeiramente, centrifugar os tubos a $800 \times g$, $4^\circ C$ por 10 minutos, para a concentração das células e coleta do plasma, seguido de uma segunda centrifugação a $11.000 \times g$ a $4^\circ C$, para a precipitação dos detritos celulares presentes no plasma e por fim foram separadas em alíquotas de 1 mL de plasma e armazenadas a $-80^\circ C$.

Como amostra de tecido normal de cada paciente fora coletada a camada de células leucoplaquetária do sangue. Essas células foram obtidas quando o sangue foi centrifugado para a separação do plasma. Após a centrifugação, o anel esbranquiçado formado entre a camada superior (plasma) e a camada inferior (hemácias), fez-se a retirada do plasma e

então anel foi recolhido com o auxílio de uma pipeta. Como a camada objetivada de células encontra-se em contato direto com a camada de hemácias, foi necessário remover as células vermelhas contaminantes, para não prejudicar a qualidade do *DNA*. Desta forma, as células foram inicialmente incubadas em 1 mL do tampão *RBLB* (*Red Blood Lysis Buffer*: NH_4Cl 155mM, KHCO_3 10mM e *EDTA* 0,1mM pH 8) a temperatura ambiente por 10 min, para o rompimento das células vermelhas. Em seguida as células foram sedimentadas por centrifugação a 400 x *g*, a temperatura ambiente por 10 min, feito isso descartou-se e o sobrenadante. A etapa de incubação com o tampão *RBLB* foi realizada duas vezes. Na sequência as células restantes da etapa anterior foram lavadas duas vezes com *PBS* (*Phosphate-Buffered Saline*: 137mM NaCl, 2,7mM KCl, 10mM Na_2HPO_4 , 2mM KH_2PO_4) e sedimentadas com as mesmas condições de centrifugação da etapa anterior, congeladas em gelo seco e armazenadas a -80°C .

3.3 EXTRAÇÃO DE *DNA*

Extração simultânea *RNA/DNA* com Trizol®

O *DNA* genômico (*gDNA*) das biópsias foi obtido através do método de extração simultânea de *RNA* e *DNA* utilizando o reagente Trizol (Invitrogen), para que o *RNA* dessas amostras fosse utilizado em outros projetos posteriormente. Para isso procedeu-se da seguinte forma, os fragmentos de tecidos congelados foram pulverizados em nitrogênio líquido e tratados com Trizol (Invitrogen) para a extração do *RNA* total. Após a coleta da

fração aquosa contendo o *RNA* total, à fase orgânica formada foi adicionado 1mL de *BEB* (*Back Extraction Buffer*: 4M de tiocianato de guanidina, 50mM de citrato de sódio e 1M de Tris-base pH 8,0) na proporção de 1:2 de Trizol. Após homogeneização, a amostra foi centrifugada a 12.000 x g por 30 min a 4°C. Recuperou-se a fase aquosa, na qual foi adicionado clorofórmio (500µL para cada 1mL de Trizol) e novamente foi realizada a centrifugação. A fase aquosa resultante foi precipitada com isopropanol e o *gDNA* foi solubilizado em tampão TE (Tris-*EDTA*: 10mM Tris, 1mM *EDTA*) e armazenado a 4°C. A qualidade do *DNA* foi avaliada por eletroforese em gel de agarose (UltraPure Life Technologies) 0,8% e também pelas razões 260/280 e 260/230 nm, obtidas no espectrofotômetro NanoDrop (Thermo).

Extração com fenol:clorofórmio

As amostras de tecido normal (células brancas do sangue), tumor residual e metástase tiveram o *gDNA* extraído através do método convencional de fenol:clorofórmio. Inicialmente as células foram lisadas com proteinase K (concentração final de 200 µg/mL) (Promega) em tampão TES (10mM Tris-HCl pH 8,0; 50mM *EDTA* pH 8; *SDS* 0,5%), a temperatura de 55°C por 16h. Ao tecido digerido foi acrescentada RNase A (concentração final de 40µg/mL) e a mistura foi incubada a 37°C por 1 hora. Após a digestão das proteínas e *RNA*, foi adicionado fenol (Invitrogen) na proporção 1:1 de material. Após a homogeneização por 5 min, foi realizada uma centrifugação a 1.800 x g, 4°C por 15 min.

O precipitado foi descartado, e ao sobrenadante foi adicionado igual volume de fenol:clorofórmio:álcool isoamílico (25:24:1). O material foi homogeneizado e centrifugado conforme etapa anterior. A fase aquosa, contendo o *DNA*, foi coletada e transferida para um novo tubo, no qual foi adicionado clorofórmio:álcool isoamílico (24:1), que foi homogeneizado e submetido à centrifugação. O *DNA*, presente na fase aquosa, foi precipitado com 2,5 volumes de etanol absoluto (Merck) e 1/10 volume de acetato de sódio 3 M, à -20°C por 16 horas. Após a incubação, o material foi centrifugado à 17.000 x *g*, 4°C por 30 min. O *DNA* foi lavado duas vezes com etanol 70% e o *pellet* foi seco à temperatura ambiente. Para solubilizar o *DNA* foi utilizado tampão TE e foi realizada uma incubação a 65°C por 30 min. Para finalizar a solubilização do *DNA*, o material foi incubado a 37°C quanto necessário. O *DNA* foi armazenado à 4°C.

3.4 AMPLIFICAÇÃO DO GENOMA INTEIRO

Para a construção das bibliotecas de *DNA* genômico do tipo *mate-pair* foram necessários aproximadamente 8 µg de *gDNA*, uma grande quantidade considerando-se a quantidade de material obtido a partir da biópsia. Desta forma, utilizou-se o sistema *GenomiPhi V2* (GE) para a amplificação de 50 ng do *gDNA*, em uma reação de amplificação isotérmica (30°C) utilizando-se a *DNA* polimerase *phi29* e pequenos iniciadores randômicos (6pb). Em seguida essa reação foi incubada por 3 horas e produziu até 7µg de *DNA* genômico. Através dessa abordagem, foram geradas quantidades suficientes de

material para a construção das bibliotecas tipo *mate-pair* bem como para as posteriores ampliações e validações dos rearranjos cromossômicos por *PCR* e sequenciamento.

3.5 CONSTRUÇÃO DA BIBLIOTECA DE *DNA* GENÔMICO DO TIPO *MATE-PAIR*

Na construção das bibliotecas do tipo *mate-pair*, foi utilizado o SOLiD™ *Long Mate Paired Library Construction Kit* (Life Technologies). Ao utilizar-se este tipo de biblioteca, é possível sequenciar as duas extremidades de fragmentos de *DNA* (~50pb), de maneira que a direção e orientação dessas sequências pareadas sejam mantidas conforme encontradas no fragmento de origem. Resumidamente, o *gDNA* obtido a partir da biópsia do tumor primário foi fragmentado, utilizando o sonicador Covaris S2 System em condições específicas de tempo e frequência (modo: *Frequency Sweeping*; *Duty cycle*: 2%; intensidade 7 por 12 ciclos, utilizando tubo 13x55mm), que foram padronizadas para garantir uma maior concentração de fragmentos na faixa de interesse (600-1000 pares de bases). Após fragmentação por sonicação, as extremidades dos fragmentos de *gDNA* foram reparadas de maneira a permitir a ligação de um adaptador às extremidades dos mesmos. Com os adaptadores ligados aos fragmentos, foi então realizado fracionamento do *DNA* em gel de agarose *Low Melting* 1% e os fragmentos de tamanho entre 600 a 1000 pares de base foram cortados e purificados do gel. O *DNA* purificado foi tratado com enzimas específicas, circularizado e digerido com diferentes enzimas de restrição e modificação de acordo com o protocolo do fabricante, conforme ilustrado na Figura 7.

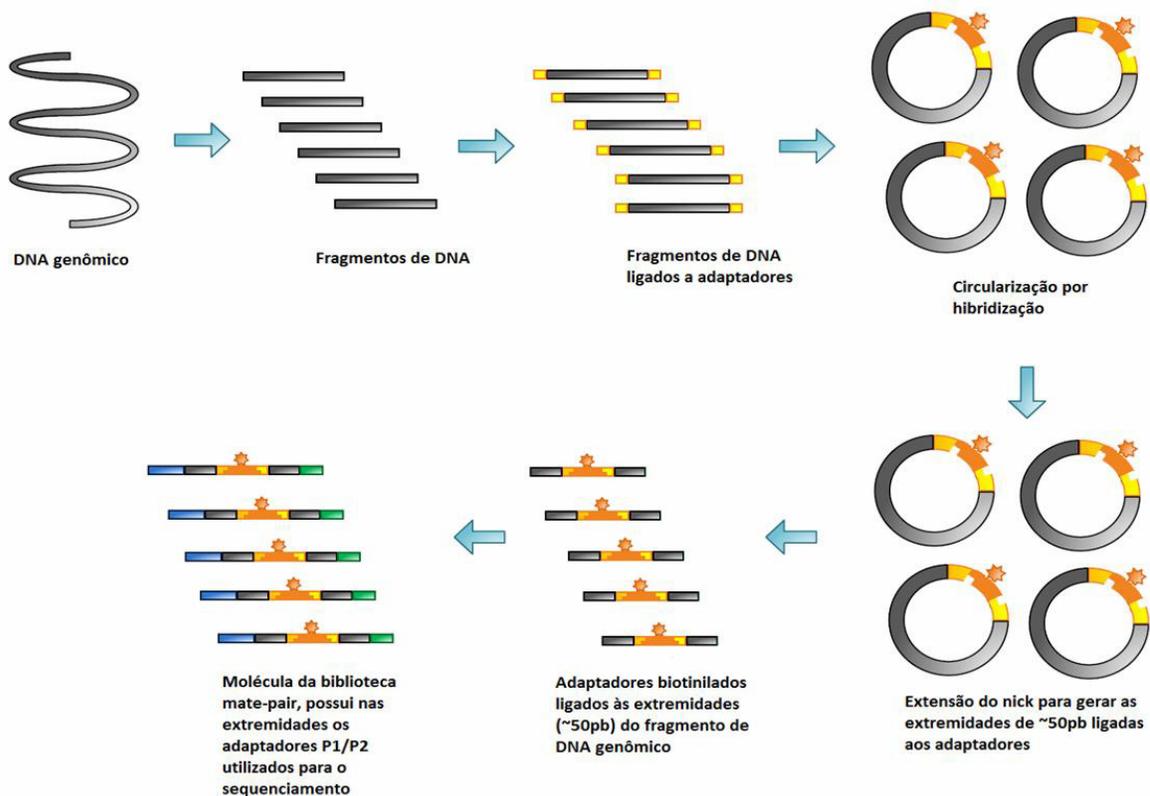


Figura 7 – Construção da biblioteca do tipo *mate-pair*. O DNA genômico foi fragmentado com o uso de sonicador e os fragmentos gerados foram tratados com enzimas (*End Polishing 1 e 2*) que converteram as extremidades dos fragmentos em extremidades cegas e fosfatadas no 5', para garantir a subsequente ligação dos adaptadores *LMP CAP* pela enzima *T4 DNA ligase*. Com os adaptadores ligados, os fragmentos foram fracionados por eletroforese em gel de agarose *Low Melting 1%*. Os fragmentos entre 600-1000 pb foram purificados do gel. Para a circularização, foi adicionado um adaptador interno biotinilado, e assim o DNA circulariza através de hibridização intramolecular entre os adaptadores *LMP* e o adaptador interno. No entanto, os adaptadores *LMP* tem um fosfato ausente na sua extremidade 5', o que promove uma falha na sequência de DNA (*nick*). Esse *nick* é utilizado pela *DNA polimerase I* para sintetizar aproximadamente 50-60 pb, utilizando a fita que não possui o *nick* como molde. Posteriormente, esse DNA circular foi digerido com endo e exonucleases originando fragmentos lineares, nos quais os adaptadores *LMP* e interno estão no interior, flanqueado pelas extremidades dos fragmentos de DNA de estudo. Como etapa final, foram ligados os adaptadores externos P1 e P2, necessários no *PCR* em emulsão e sequenciamento. Através desses adaptadores, a biblioteca foi também amplificada de modo a gerar um número maior de moléculas para serem sequenciadas. (Adaptada de Applied Biosystems SOLiD 4 System: Library Preparation Guide, pág.43, Edição de 2010)

Como última etapa, a biblioteca foi amplificada com iniciadores específicos para os adaptadores (10-12 ciclos) através de *PCR* convencional. Os fragmentos amplificados foram purificados utilizando *beads* Agencourt AMPure XP (Beckman Coulter), eluídos em 25 µL de TE, e o material foi quantificado em fluorômetro Qubit 2.0 (Invitrogen), utilizando

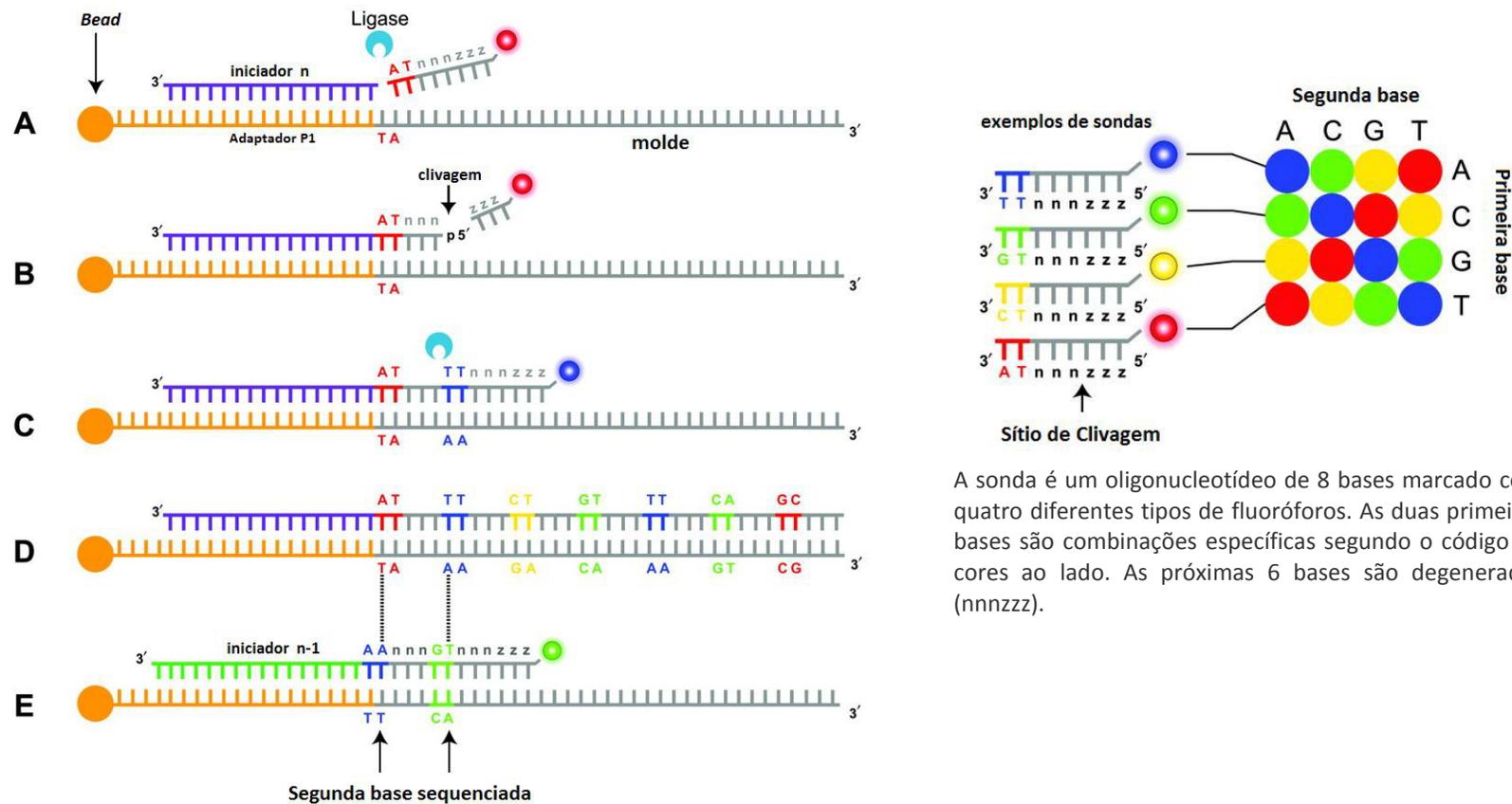
o ensaio *dsDNA High Sensitivity* (Invitrogen). Ao final do protocolo, foi obtida uma biblioteca de *DNA* na qual somente as extremidades (~50 pb) dos fragmentos isolados inicialmente estão ligadas a adaptadores específicos e necessários para o *PCR* em emulsão e sequenciamento na plataforma SOLiD™.

Para garantir a qualidade das bibliotecas, ou seja, uma maior concentração dos fragmentos entre 275 e 325 pb (faixa esperada quando somando o tamanho dos adaptadores e duas sequências de ~50 pb), elas foram purificadas utilizando-se o *e-Gel System 2%* (Invitrogen) como indicado pelo fabricante, e eluídas em 20 µL de TE. Posteriormente determinou-se a faixa de tamanho dos fragmentos purificados por Bioanalyzer (Agilent) utilizando chip *High sensitivity DNA assay*.

3.6 *PCR* EM EMULSÃO E SEQUENCIAMENTO NA PLATAFORMA SOLiD™

A plataforma SOLiD (*Supported Oligonucleotide Ligation and Detection*) é uma tecnologia de sequenciamento em larga escala que baseia-se na ligação sequencial de oligonucleotídeos marcados pela ação da enzima *DNA* ligase. Essa tecnologia utiliza uma biblioteca de fragmentos de *DNA* ligados a adaptadores de oligonucleotídeos (descrita no item anterior), que são por sua vez ligados à *beads* de sequenciamento e amplificados clonalmente por *PCR* em emulsão. Para o *PCR* em emulsão foi preparada uma solução contendo todos os componentes necessários para uma *PCR*, e a biblioteca de fragmentos *mate-pair* foi utilizada como molde para a *PCR*. Após homogeneizadas com *beads*

recobertas por sequências complementares a um dos adaptadores da biblioteca, essa fase aquosa foi misturada com uma solução oleosa e submetida à agitação constante. Ao final, foi obtida uma emulsão com micro gotas (microrreatores) cada qual, teoricamente contendo uma *bead*, uma molécula molde e os reagentes necessários para uma reação de *PCR* convencional. A emulsão foi então submetida à amplificação em termociclador, de maneira que em cada microrreator ocorreu amplificação clonal de uma única molécula. Após a amplificação, as *beads* então recobertas por milhares de cópias de uma mesma molécula foram recuperadas e enriquecidas, conforme protocolo do fabricante. Posteriormente ao enriquecimento as *beads* foram imobilizadas em uma superfície e encaminhadas para o sequenciamento. As sequências de 50-60 bp foram obtidas após vários ciclos de hibridização, ligação e clivagem, nos quais foram utilizados iniciadores complementadores aos adaptadores (P1 e P2) utilizados na construção das bibliotecas de *DNA* e oligonucleotídeos marcados com fluoróforos. A Figura 8 ilustra o princípio básico deste sequenciamento.



3.7 ANÁLISE DE BIOINFORMÁTICA

A análise de bioinformática foi realizada em parceria com a recém doutora por esse mesmo programa, Elisa Rennó Donnard Moreira, que foi responsável pelo desenvolvimento do *pipeline* para a identificação de variações estruturais no genoma tumoral, visando a sua utilização como biomarcadores personalizados. Essas análises compuseram parte da tese de doutorado defendida pela aluna este ano (Moreira, 2014). Essa ferramenta foi desenvolvida para a identificação eficiente dessas alterações genéticas sem a necessidade do sequenciamento do tecido normal de cada paciente, e de alta cobertura de sequenciamento. Deste modo, há uma significativa redução de custo associados a essa abordagem, facilitando sua aplicação clínica. A artigo científico que descreve essa abordagem já encontra-se disponível na versão online (Donnard, Carpinetti *et al.*, 2015).

Em resumo, as sequências geradas na plataforma SOLiD (Life Technologies) foram mapeadas contra a sequência de referência do genoma humano (GRCh37/hg19) com o programa BioScope (Life Technologies). Utilizando-se as sequências em que ambos os pares foram mapeados com qualidade maior ou igual a 20 ($Q \geq 20$), foram aplicados filtros que avaliam a distância esperada, e também a orientação do alinhamento desses pares de sequências. Foram mantidas na análise aquelas sequências pareadas que apresentaram distância e/ou orientações não esperadas quando alinhadas contra a

sequência de referência do genoma humano, indicando a existência de uma variação estrutural. Foram investigadas variações estruturais inter e intracromossômicas, sendo que para as intracromossômicas foram estudadas as variações envolvendo mais que quatro mil pares de bases. Nas sequências restantes, foram aplicados ainda filtros para remover pares de sequências que alinharam em regiões teloméricas e centroméricas, sequências pareadas duplicadas ou que alinharam no mesmo cromossomo em montagens alternativas do genoma humano entre outros, como apresentado na Figura 9.

Por fim, as sequências restantes foram agrupadas de modo que os pares de sequências que envolviam a mesma região foram alocados em um mesmo grupo (*cluster*). Os grupos que apresentaram pelo menos 3 pares de sequências suportando a mesma variação estrutural, envolvendo as mesmas regiões genômicas, com distâncias e orientações compatíveis, foram considerados candidatos.

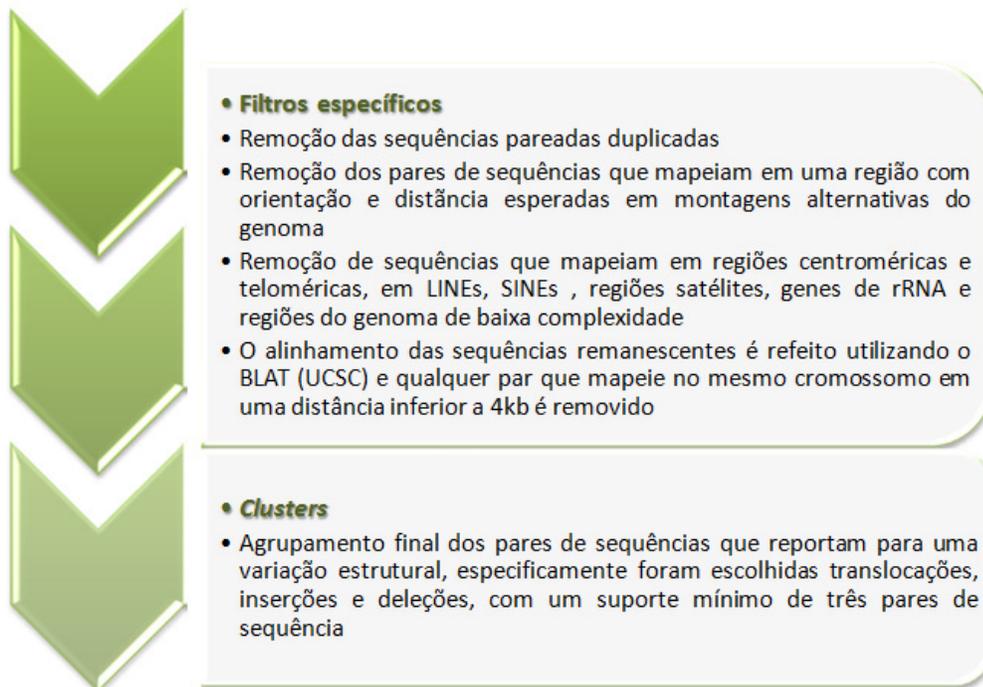
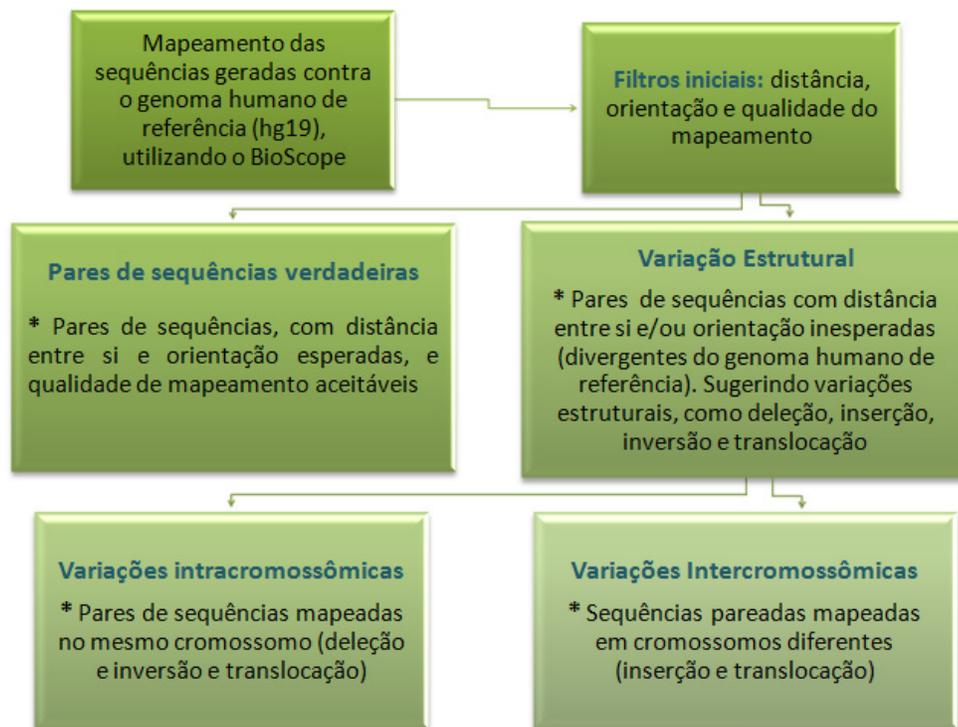


Figura 9 – Resumo dos principais filtros usados para a detecção de variações estruturais.

3.8 VALIDAÇÃO DAS VARIAÇÕES ESTRUTURAIS

Após a identificação computacional das possíveis variações estruturais presentes nas amostras dos tumores primários, as mesmas foram validadas experimentalmente através de *PCR* convencional com iniciadores flanqueando o ponto de quebra (Quadro 3 no APÊNDICE A). Para isto, foram utilizadas as regiões do genoma referência compreendida entre as sequências pareadas localizadas nos extremos da região a montante e também à jusante ao ponto de quebra da variação estrutural, conforme ilustrado na Figura 10. Essas duas regiões foram utilizadas para desenho dos iniciadores 1 e 2 de cada candidato, com o auxílio das ferramentas *online Primer3 4.0* (<http://primer3.ut.ee/>) e *IDT Oligo Analyzer 3.1* (<https://www.idtDNA.com/>). Preferencialmente, foram escolhidos para a validação os iniciadores mais próximos ao ponto de quebra, e que apresentavam tamanho entre 18 e 24 pares de base, temperatura de anelamento entre 57 e 62 °C, conteúdo de GC entre 40-70%, um máximo de três citidinas ou guanosinas consecutivas e que contivessem uma citidina ou guanosina na extremidade 3' do oligonucleotídeo.

O último critério avaliado para a escolha dos iniciadores foi a possibilidade do par selecionado também ter o potencial de amplificar outras regiões do genoma, e para essa finalidade foi utilizada a ferramenta *Primer BLAST – NCBI* (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>), a qual simula uma reação de *PCR* (*PCR in silico*) com o par de iniciadores de interesse, com base na sequência do genoma referência. Como as variações estruturais são específicas do genoma tumoral, quando o

resultado do *PCR in silico* reportava a possibilidade de amplificação de alguma região do genoma referência, o par de iniciadores foi descartado.

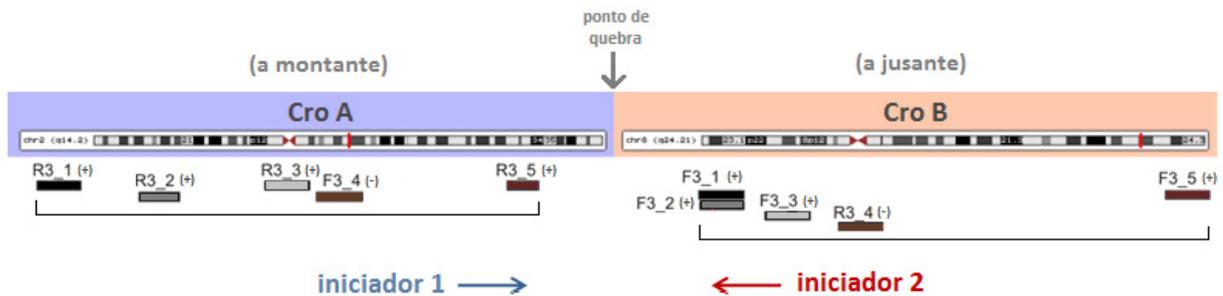


Figura 10 – Estratégia para o desenho dos iniciadores utilizados nas reações de *PCR* para a validação das variações estruturais. A figura representa uma variação estrutural intercromossômica, uma translocação, entre o cromossomo A (destacado em azul) e o cromossomo B (destacado em vermelho), o ponto de quebra, indicado pela seta cinza, é a região em que os dois cromossomos estão unidos. Os retângulos menores, localizados abaixo dos cromossomos, representam os alinhamentos das sequências pareadas (os pares estão representados com a mesma cor) que reportaram essa VE. As setas azul e vermelha, representam a região preferencial que os iniciadores 1 e 2, respectivamente, foram desenhados.

Nas reações de *PCR*, o *DNA* genômico extraído de células do sangue periférico do próprio paciente foi utilizado como controle negativo, para garantir que essas variações eram exclusivas do tecido tumoral. Os produtos de *PCR* obtidos foram fracionados por eletroforese em gel de agarose 1% e para as reações de *PCR* em que ocorreu amplificação específica somente no tecido tumoral, foram realizados sequenciamentos independentes com os dois iniciadores de cada candidato, utilizando o sequenciador 3130 Genetic Analyzer (Applied Biosystems). As sequências obtidas foram alinhadas contra a sequência do genoma humano de referência (GRCh37/hg19), utilizando a ferramenta *online* BLAT-UCSC Genome Bioinformatics (<https://genome.ucsc.edu>) (Figura 11).

Assim, as coordenadas genômicas e o ponto de quebra foram confirmados e detalhados com resolução de uma base. Foram consideradas validadas as variações em

que foi possível obter sequências que reportavam a aproximação de duas regiões de um mesmo cromossomo (deleções e inversões) ou a união de regiões de dois cromossomos distintos (translocação e inserção), e quando esses resultados estavam de acordo com as informações obtidas na análise de bioinformática.

Human BLAT Results											
BLAT Search Results											
ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	VE_FW	264	17	290	501	100.0%	7	+	127317953	127318226	274
browser details	VE_FW	205	290	500	501	98.6%	14	+	53067742	53067951	210
browser details	VE_FW	30	189	233	501	94.2%	3	-	174709150	174709197	48
browser details	VE_FW	28	304	341	501	93.6%	5	+	119711218	119711256	39
browser details	VE_FW	21	466	486	501	100.0%	X	-	41900361	41900381	21
browser details	VE_FW	21	336	356	501	100.0%	8	+	8588911	8588931	21
browser details	VE_FW	21	462	482	501	100.0%	6	+	22933476	22933496	21
browser details	VE_FW	21	133	155	501	95.7%	4	+	111496826	111496848	23
browser details	VE_FW	20	270	289	501	100.0%	4	-	15029989	15030008	20
browser details	VE_RV	303	181	483	497	100.0%	7	-	127317924	127318226	303
browser details	VE_RV	148	26	181	497	100.0%	14	-	53067742	53067897	156
browser details	VE_RV	29	100	130	497	100.0%	2	+	104912513	104912816	304
browser details	VE_RV	24	255	287	497	77.8%	3	-	97829123	97829151	29
browser details	VE_RV	22	435	458	497	87.0%	4	+	106965959	106965981	23
browser details	VE_RV	21	442	463	497	100.0%	8	-	50183935	50183957	23
browser details	VE_RV	21	115	135	497	100.0%	8	-	8588911	8588931	21
browser details	VE_RV	21	316	338	497	95.7%	4	-	111496826	111496848	23
browser details	VE_RV	21	441	461	497	100.0%	X	+	78934086	78934106	21
browser details	VE_RV	20	182	201	497	100.0%	4	+	15029989	15030008	20
browser details	VE_RV	20	254	275	497	95.5%	3	+	182635745	182635766	22
browser details	VE_iniciador_1	20	1	20	20	100.0%	7	+	127317911	127317930	20
browser details	VE_iniciador_2	20	1	20	20	100.0%	14	-	53067932	53067951	20

Figura 11 – Validação das variações estruturais através da análise de sequências utilizando a ferramenta BLAT. A figura representa o resultado da análise das sequências para um candidato a validação. Essa ferramenta verifica com quais regiões do genoma humano de referência, as sequências de interesse têm identidade, ou seja, partir de que regiões genômicas elas devem ter sido originadas. E para cada região que uma identidade é encontrada (valor dado em porcentagem), a ferramenta fornece uma pontuação (*score*) que leva em consideração a quantidade de nucleotídeos alinhados corretamente (*matches*) versus incorretamente (*mismatches*), em relação à sequência referência. Na figura, VE_Fw é a sequência gerada utilizando o iniciador 1 e VE_Rv, com o iniciador 2. Para a validação de um candidato a variação estrutural, é esperado que as sequências geradas pelos iniciadores alinhem em mais de uma região do genoma, e para as translocações, em cromossomos distintos (retângulo vermelho). A partir disso, foram conferidas as coordenadas cromossômicas, reportadas pela análise de bioinformática, através dos campos *CHRO* (cromossomo), *STRAND* (fita), *START* (início da sequência) e *END* (final da sequência). Uma vez confirmada essas coordenadas, o ponto de quebra foi determinado, grifado em vermelho.

3.9 EXTRAÇÃO DE *DNA* CIRCULANTE DO PLASMA

O *DNA* circulante foi extraído a partir de alíquotas de 3 mL de plasma, utilizando o *QIAamp MinElute Virus Vacuum Kit* (QIAGEN). A primeira etapa de extração de ácidos nucleicos do plasma foi a digestão das proteínas, para isso foram adicionados ao plasma 450 µL da solução de proteases, 3 mL da solução AL com o *RNA* carreador (6,7 µg/mL) e foi seguida a incubação a 56°C por 15 min. Após a digestão, foi adicionado 3,6 mL de etanol absoluto e o material foi mantido a temperatura ambiente por 5 min. O material foi aplicado na coluna com o auxílio de um extensor. A extração a vácuo foi feita com o sistema *QIAvac 24 plus* e *Vacuum Pump System* (QIAGEN). Após a sucção do material, a coluna foi lavada duas vezes com a solução AW1, duas vezes com a AW2 e uma vez com etanol absoluto. Depois das lavagens, o excesso de etanol foi retirado com centrifugação a 20.000 x *g* por 3 min, seguida de incubação em banho seco a 56°C por 3min. A eluição dos ácidos nucleicos foi feita com 125 µL de *Low TE*, aplicados em três etapas. O *DNA* circulante foi armazenado a -80 °C.

3.10 QUANTIFICAÇÃO DO *DNA* CIRCULANTE TOTAL DO PLASMA

Para a quantificação do *DNA* circulante total nas amostras de plasma, foi utilizada a metodologia do *PCR* Digital em gotas (*Droplet Digital PCR – ddPCR*). Para quantificação foi utilizado o ensaio *RNaseP Copy Number Reference* (Life Technologies), no qual iniciadores e sonda (oligonucleotídeo ligado ao fluoróforo *VIC*) foram especificamente desenhados para detectar o gene de cópia única do componente H1 da Ribonuclease P (H1RNA),

localizado no cromossomo 14 (Chr14:20811565 na construção 37 do NCBI), e que o *amplicon* possui 87 pb e está dentro de um único éxon. Foram utilizados dois volumes de *DNA* circulante, 2 e 4 μ L, que permitiram inferir sobre a presença de inibidores da *PCR* e calcular o volume máximo de *DNA* circulante utilizado por reação, na etapa de detecção do *ctDNA*.

Sucintamente, foram misturados todos os componentes necessários para uma *PCR* (*ddPCR Supermix™*, iniciadores, sonda e *DNA* circulante), que foram transferidos para o *QX200™ Droplet Generator* (Bio-Rad), equipamento responsável por particionar os 20 μ L da reação em 20.000 gotas imersas em óleo. Essas gotas foram levadas para um termociclador convencional e foram amplificadas por 45 ciclos. Depois de amplificar as moléculas de *DNA*, as gotas foram analisadas pelo *QX200™ Droplet Reader* (Bio-Rad). Esse equipamento analisa a integridade, o volume das gotas e quantifica a fluorescência de cada gota de forma individualizada. Assim, é possível determinar o número total de gotas analisadas, em quantas gotas ocorreu a amplificação (que continham o *DNA* de interesse) e o número de gotas negativas (com fluorescência basal). Esses números, quando analisados no *software* QuantaSoft (Bio-Rad), são corrigidos de acordo com a distribuição de Poisson e assim são estimadas as quantidades de cópias da sequência alvo por microlitro da reação. Para garantir a acurácia dos resultados, são consideradas na

quantificação somente as gotas que apresentam volume e integridade compatíveis com o esperado, além de ser exigido um mínimo de 10.000 gotas aceitáveis para os cálculos de quantificação. As etapas do PCR digital em gotas estão ilustradas na Figura 12.

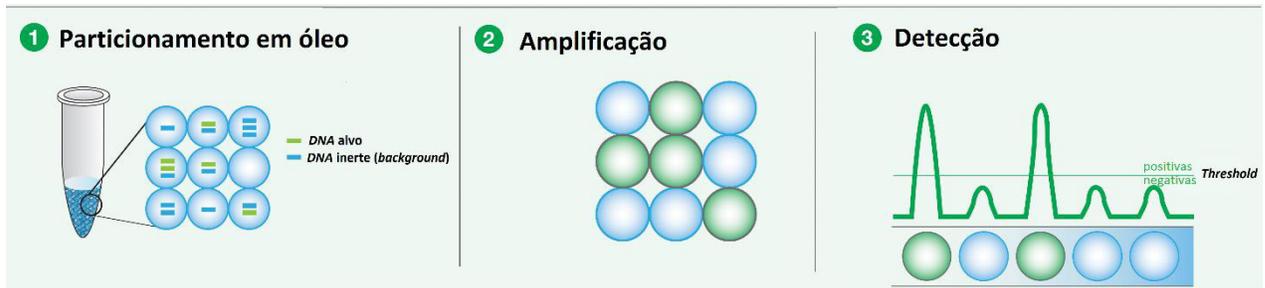


Figura 12 – Etapas do PCR Digital em gotas. (1) Na primeira etapa, a reação contendo todos os componentes necessários para a PCR foram particionadas em milhares de nano gotas, contendo um nanolitro em cada. Com esse particionamento, algumas gotas não continham DNA (fragmentos azuis e verdes) e outras continham uma ou mais fragmentos. (2) Após os ciclos de amplificação, utilizando termociclador convencional, nas gotas que continham o DNA alvo (fragmento verde) ocorreu a amplificação e conseqüente liberação do fluoróforo em condições saturantes (representada pelas gotas verdes). (3) A fluorescência das gotas foi lida individualmente, e os números de gotas positivas e negativas foram utilizados para calcular o número de cópias do DNA alvo na reação. (Adaptada do material de divulgação do QX200™ Droplet Digital™ PCR System – Bio-Rad)

3.11 DETECÇÃO DO DNA TUMORAL CIRCULANTE NO PLASMA

Para a detecção das variações estruturais, conseqüentemente o *ctDNA*, foram customizados iniciadores flanqueando o ponto de quebra e sondas *MGB (Minor Groove Binding - TaqMan)* ligadas ao fluoróforo 6-FAM (Life Technologies) localizadas exatamente no ponto de quebra da variação estrutural (Quadro 4 do APÊNDICE A). As sondas foram desenhadas utilizando o software *PrimerExpress* (Applied Biosystems), seguindo os parâmetros sugeridos no QX200™ Droplet Digital™ PCR System Brochure (Bio-Rad).

Devido ao baixo número de cópias de *ctDNA* no plasma, foi realizada uma etapa prévia de 20 ciclos de amplificação (sem a utilização da sonda). Na pré-amplificação do *DNA* circulante foi realizada uma reação *multiplex* utilizando SsoFast™ Probes Supermix (Bio-Rad), na qual foram adicionados, na mesma reação, iniciadores externos (Quadro 4) para duas variações estruturais específicas do paciente e um par de iniciadores para uma região controle (componente H1 da Ribonuclease P – RNase P), em um volume total de 30µL. Foram analisados 75% do volume (~93µL) de *DNA* circulante obtido para cada amostra de plasma em reações independentes, sendo que o volume utilizado por reação de pré-amplificação variou entre 4-7µL. Depois da amplificação, foram utilizados 3µL de cada produto de *PCR* para a detecção do *ctDNA* no sistema QX200™ Droplet Digital™ *PCR*. Para garantir a especificidade da reação, os iniciadores utilizados na primeira amplificação não foram os mesmos utilizados no *ddPCR*, e também foram utilizadas sondas específicas (Quadro 4) para o ponto de quebra de cada variação estrutural, seguindo assim a estratégia de *Nested PCR*, como ilustrado na Figura 13. Na *ddPCR*, foram feitos ensaios individuais, para a detecção das duas variações estruturais e também da região controle (RNase P).

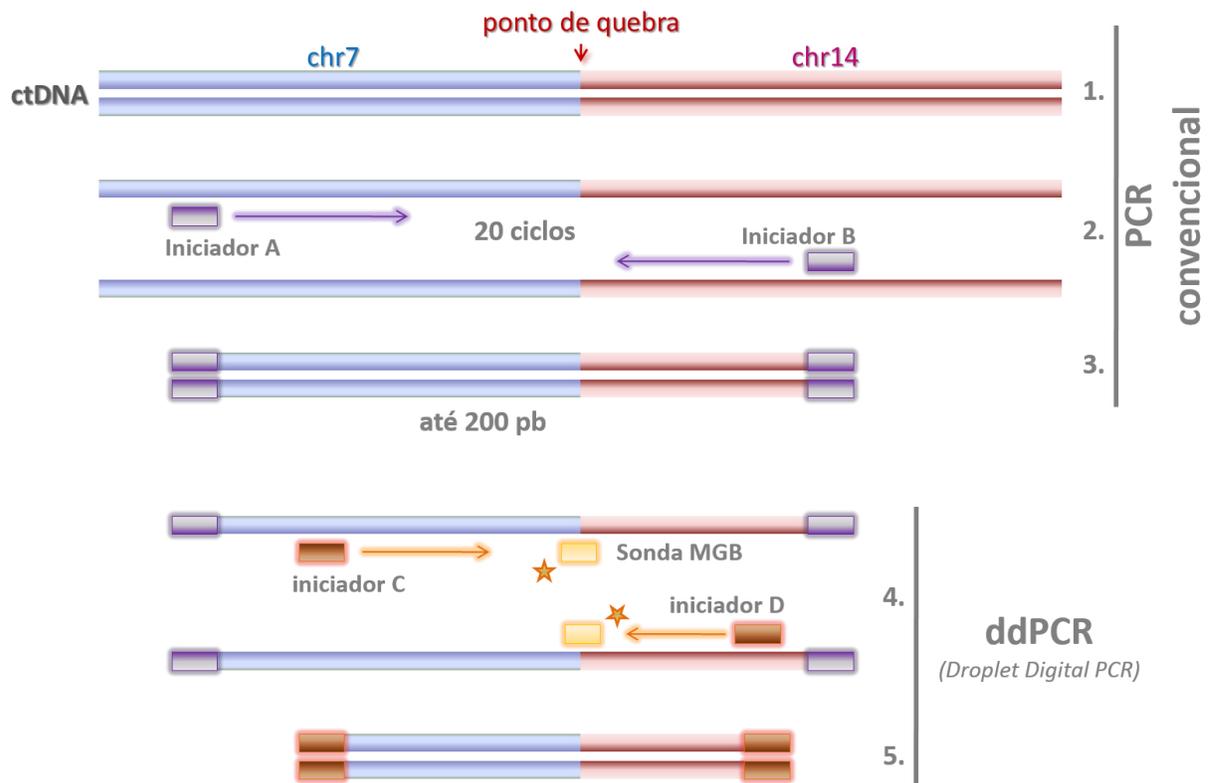


Figura 13 – Estratégia de *Nested PCR*, utilizando *PCR Digital*. Na primeira etapa, foi realizado um *PCR multiplex* convencional para aumentar o número de moléculas alvo na reação. Foram utilizados, em uma mesma reação, iniciadores específicos flanqueando o ponto de quebra das variações estruturais (translocação) do tumor do paciente e iniciadores para o gene da RNase P. Na segunda etapa foi feito um *ddPCR* para quantificar o número de cópias do *DNA* tumoral presentes na primeira reação. Os iniciadores desenhados para a segunda etapa estão localizados em uma região mais interna dos fragmentos produzidos na primeira etapa e garantem a especificidade da técnica juntamente com as sondas *MGB* ligadas a fluoróforos (6-*FAM* para a sonda específica para a variação estrutural e *VIC* para o gene da RNase P). Desta forma, somente as gotas que apresentam a molécula alvo (*ctDNA*) terão a fluorescência liberada e serão detectadas como gotas positivas.

Análise do ddPCR

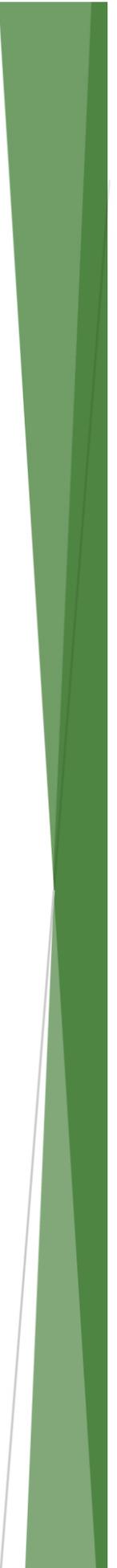
Para análise das reações de *ddPCR*, os resultados obtidos após a leitura das gotas no *QX200™ Droplet Reader™* foram analisados utilizando o *software* QuantaSoft™ (Bio-Rad). Primeiramente, foi realizado o cálculo do valor de fluorescência limite (*threshold*), que separa a população de gotas positivas e negativas. Para esse cálculo, foram utilizados os valores de fluorescência obtidos nas reações controles, ou seja, as reações utilizando o

DNA da biópsia (controle positivo), o *DNA* normal (controle negativo) e as que não continham *DNA* (sem amostra). Primeiramente, essas reações controles foram analisadas utilizando a opção *Auto Analyze*; em seguida foi obtido o valor médio de fluorescência das populações negativas e positivas para cada ensaio. Através desses valores foi possível calcular a amplitude entre a média de fluorescência das gotas negativas e positivas. O *threshold* foi estabelecido, para cada ensaio separadamente, como sendo 30% da amplitude somado à média de fluorescência da população negativa.

A partir do cálculo do *threshold* foi possível realizar a análise das reações para detectar o *ctDNA*. Todas as reações de uma mesma amostra para o mesmo ensaio foram agrupadas, utilizando a função *merged*. Para a identificação da população de gotas positivas foi aplicado o valor do *threshold* específico de cada ensaio. Desta forma, foi possível obter o número de cópias do alvo de cada ensaio para cada amostra analisada, assumindo a distribuição de Poisson. Esses valores foram, posteriormente, normalizados pelo volume de plasma, volume de eluição e volume de *DNA* circulante utilizado na reação de pré-amplificação.

Para o cálculo da taxa de falso positivo foram realizadas 16 reações para cada ensaio, utilizando como amostra o *DNA* do tecido normal pareado. A partir da análise dessas reações foi estabelecido o limite de detecção (*Limit of Detection – LOD*), com intervalo de confiança de 95%, para cada ensaio. Esse limite foi utilizado como critério para determinar quando ocorreu amplificação específica do alvo.

4 RESULTADOS



4 RESULTADOS

4.1 PACIENTES INCLUÍDOS E SEGUIMENTO

Nesse estudo, foram selecionados seis pacientes com câncer de reto, sendo dois pacientes com tumores de estadio T2N1 (CCR1 e CCR3) e quatro pacientes com T3N0-1 (CCR2, PCR1, IR1 e IR2). O critério utilizado para a escolha dos pacientes foi obter um conjunto de pacientes com diferentes respostas ao tratamento neoadjuvante do câncer retal localmente avançado, representando o leque de resposta observado na prática clínica. Todos pacientes foram submetidos ao protocolo de tratamento neoadjuvante com radioterapia (5040-5400cGy) e quimioterapia baseada em 5-FU, e foram reavaliados, pelo menos, 12 semanas após o término do tratamento. O tempo mediano de seguimento desses pacientes foi de 101 semanas (variando entre 53 e 186 semanas) contadas após o término do tratamento. Os dados clínicos, patológicos e de seguimento dos pacientes incluídos no estudo estão listados no Quadro 1.

Os três pacientes (CCR1, CCR2 e CCR3) com resposta completa clínica foram submetidos ao protocolo de observação constante (*Wait and Watch Approach*) e foram poupados da cirurgia imediata. Desse grupo de pacientes, um (CCR1) apresentou recidiva local 51 semanas ao término da QRTn e foi submetido a cirurgia de resgate para remoção do tumor residual. Outro paciente (CCR2) teve diagnosticada doença metastática 62 semanas após o término do tratamento e foi submetido à quimioterapia sistêmica (semana 73), no entanto, ao final da primeira linha de tratamento (semana 99) foi

verificada a progressão da doença, através dos exames de imagem, e uma segunda linha de tratamento foi realizada (semana 138 até 166), a qual e também não teve sucesso. Com isso adotou-se uma terceira linha de tratamento sistêmico (semana 173), mas infelizmente o paciente faleceu algumas semanas depois (semana 186).

Quadro 1 – Dados clínicos patológicos e de seguimentos dos pacientes

Paciente	Sexo	Estadio do tumor*		Local do tumor	Resposta	GRT [§]	Seguimento
		cT	cN				
CCR1	F	2	1	baixo	Clínica Completa	---	Recorrência local ypT2N0 (sem 51) [#]
CCR2	M	3	0	baixo	Clínica Completa	---	Metástase no fígado (sem 62) [#]
CCR3	M	2	1	baixo	Clínica Completa	---	SED
PCR1	F	3	1	baixo	Patológica Completa	4	SED
IR1	M	3	1	médio/distal	Incompleta	3	Metástase no fígado (sem 36) [#]
IR2	F	3	1	baixo	Incompleta	2	SED

* Estadiamento do tumor com base no Sistema de Classificação TNM

§ Grau de regressão tumoral, obtido a partir da análise anatomopatológica do tumor ressecado, como descrito em Dworak *et. al*, 1997.

Número de semanas após o término do tratamento que ocorreu a recidiva da doença

O paciente PCR1 foi submetido à cirurgia após o intervalo de descanso, uma vez que a presença de úlcera e fibrose no local do tumor primário impossibilitou descartar a doença residual e somente o exame patológico da peça cirúrgica demonstrou a ausência de células tumorais viáveis, o que caracteriza a resposta completa patológica.

O grupo de pacientes com resposta incompleta (IR1 e IR2) foi submetido à cirurgia, após a avaliação clínica, para a remoção do tumor residual e exame patológico. Um desses pacientes (IR1) desenvolveu metástase no fígado 36 semanas após o tratamento e foi submetido à hepatectomia, posteriormente foi diagnosticado também com metástase no arco costal (semana 82).

4.2 SEQUENCIAMENTO E IDENTIFICAÇÃO DAS VARIAÇÕES ESTRUTURAIIS

Os tecidos tumorais obtidos através de biópsias endoscópicas, antes do início do tratamento, foram macrodissecados para garantir que mais de 80% sejam de células tumorais. Deles foi extraído *DNA*, o qual fora utilizado na construção de bibliotecas do tipo *mate-pair* para sequenciamento em larga escala e posterior identificação de variações estruturais exclusivas do tecido tumoral de cada paciente.

Em média, 400 milhões de sequências pareadas foram obtidas através do sequenciamento em larga escala utilizando a plataforma SOLiD. As sequências geradas foram alinhadas contra a sequência de referência do genoma humano (GRCh37/hg19) resultando em aproximadamente 80% de sequências mapeadas para cada amostra. Com a finalidade de aumentar a confiabilidade das análises, foram selecionadas sequências com qualidade de alinhamento maior ou igual a 20, o que atribui 99% de chance do alinhamento estar correto. As sequências com alinhamento único foram em média 60% das sequências geradas, indicando a alta qualidade do sequenciamento. Esses dados estão detalhados na Tabela 1.

Como esperado, o tamanho médio dos fragmentos sequenciados nas diferentes bibliotecas foi de aproximadamente 700 pb (variando entre 631 e 826 pb), como ilustrado na Figura 14. A cobertura física foi calculada como base nesses tamanhos médios e variou entre 13 e 25 vezes o genoma humano, com exceção da amostra IR2 (60 vezes), conforme Tabela 1. A cobertura real média também foi calculada e teve média de 4,7 vezes o genoma para as amostras, exceto para IR2, que apresentou cobertura de 8,7 vezes.

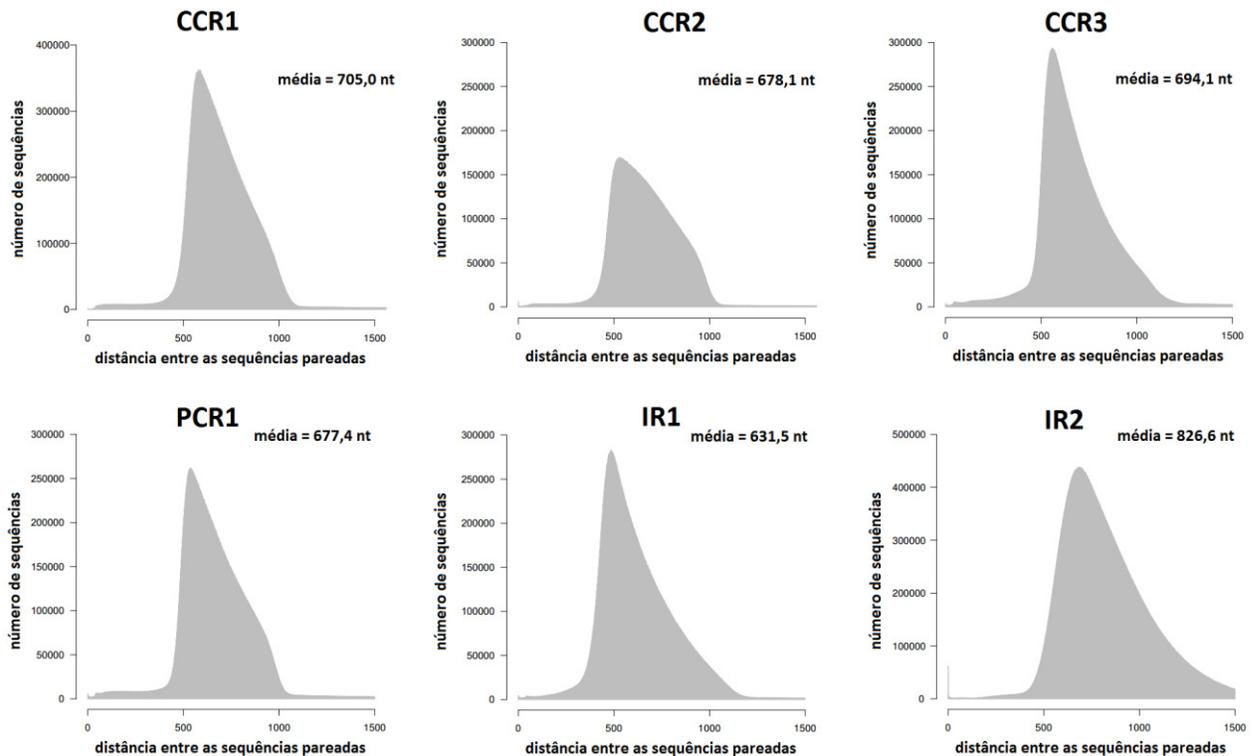


Figura 14 – Gráficos com a distribuição das distâncias entre as seqüências pareadas. As distâncias entre as seqüências pareadas foram calculadas após o alinhamento dessas seqüências contra a seqüência do genoma referência. O tamanho médio dos fragmentos sequenciados, nucleotídeos (nt), estão informados nos gráficos de cada amostra.

Tabela 1 – Dados de sequenciamento, alinhamento e cobertura das sequências pareadas geradas para os diferentes tumores

Paciente	Sequências geradas	Nucleotídeos gerados	Sequências mapeadas	Sequências mapeadas Q \geq 20	Cobertura real	Cobertura física*
CCR1	991,625,036	49,581,251,800	628,192,760 (63%)	459,743,724 (46%)	6.67x	25x
CCR2	425,460,416	29,782,229,120	350,515,565 (82%)	247,758,837 (58%)	4.83x	13x
CCR3	398,436,826	19,921,841,300	323,562,661 (81%)	256,206,045 (64%)	4.07x	20x
PCR1	385,789,584	19,289,479,200	305,832,709 (79%)	242,963,348 (63%)	3.70x	18x
IR1	393,756,912	19,687,845,600	322,877,464 (82%)	256,652,253 (65%)	4.07x	19x
IR2	1,035,604,016	51,780,200,800	788,793,614 (76%)	560,355,119 (54%)	8.65x	60x

Os resultados de alinhamento foram analisados por um conjunto de programas computacionais, desenvolvido em colaboração com o grupo de bioinformática do Centro de Oncologia Molecular, que permite a identificação de rearranjos estruturais no genoma tumoral utilizando baixa cobertura de sequenciamento (~4 vezes) e sem a necessidade do sequenciamento do genoma normal pareado dos pacientes em análise, reduzindo assim o custo do sequenciamento e viabilizando a sua aplicação na prática clínica (Donnard, Carpinetti et al., 2015).

Para a identificação dos rearranjos foram implementados uma série de filtros que eliminam artefatos de alinhamento causados por sequências de baixa qualidade, regiões repetitivas e variações estruturais presentes no genoma humano. As etapas da análise estão descritas em Materiais e Métodos e detalhadas em Donnard *et al.*2014. Ao final da análise, obteve-se uma lista com os possíveis candidatos a variação estrutural para cada paciente e desses alguns foram selecionados aleatoriamente para a validação experimental através de amplificação por *PCR* seguida de sequenciamento Sanger (Quadro 5 do Apêndice B). A Tabela 2 sumariza o número de variações estruturais, intercromossômicas (translocação/inserção) e intracromossômicas (deleção/inversões maiores que 4kb), encontradas para cada paciente. A Figura 15 se trata da representação gráfica das variações, na qual é possível observar a distribuição dos candidatos no genoma (cromossomos) de cada paciente.

Tabela 2 – Resultado da análise para detecção de variações estruturais em cada paciente

Amostra	nº variações intercromossômicas	nº variações Intracromossômicas
CCR1	5	10
CCR2	3	3
CCR3	66	15
PCR1	4	11
IR1	13	5
IR2	4	11

* Todos os candidatos possuem 3 ou mais sequencias pareadas suportando a variação estrutural

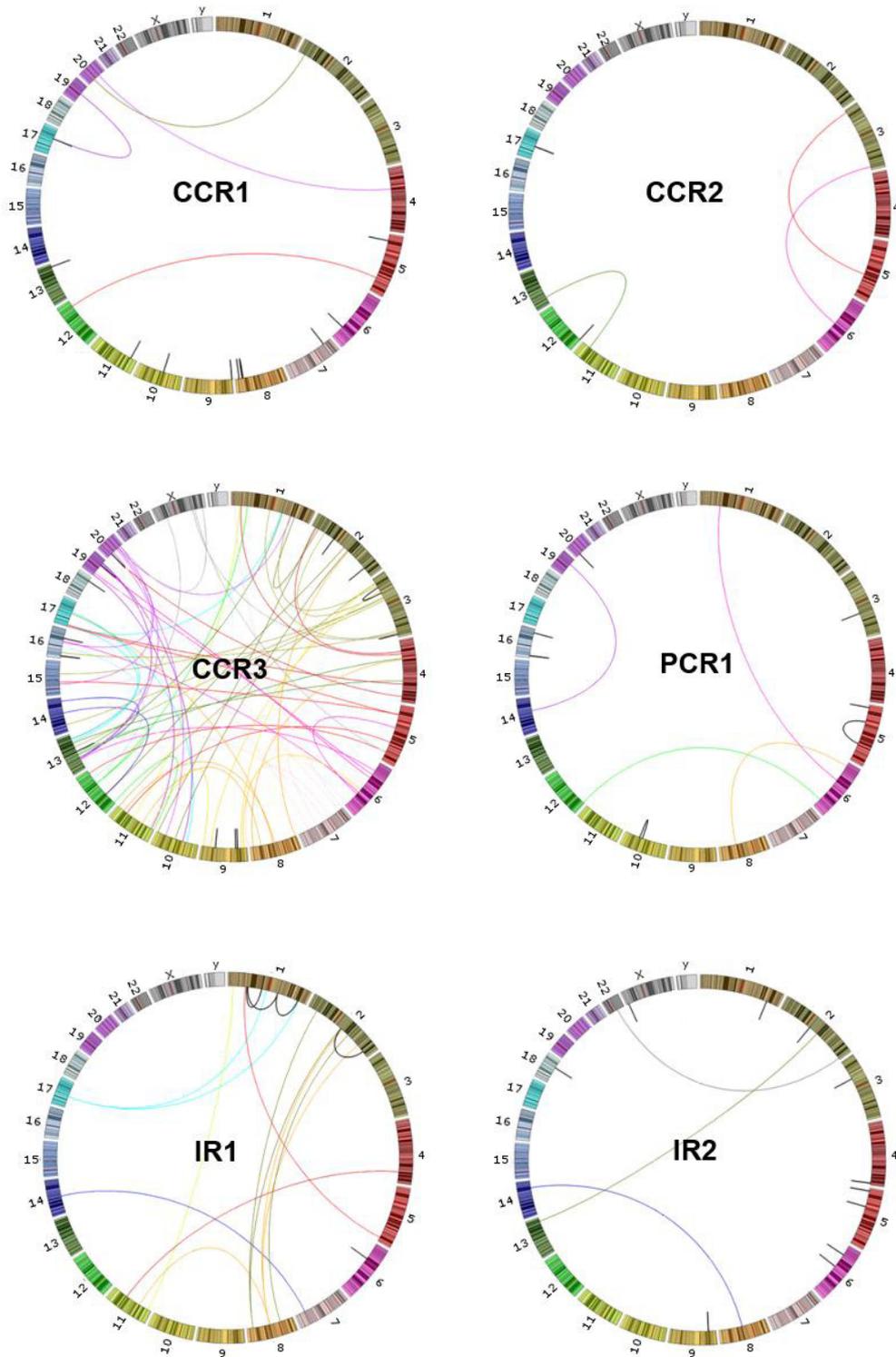


Figura 15 – Representação gráfica dos candidatos a variações estruturais. Nessa representação (*circus plot*), todos os cromossomos estão organizados em sentido horário formando um círculo. Os rearranjos intercromossômicos são representados pelas linhas longas coloridas que ligam dois cromossomos. E os intracromossômicos são as linhas curtas pretas que ligam duas regiões de um mesmo cromossomo.

O número de candidatos entre os pacientes variou significativamente, sendo os extremos as amostras dos pacientes CCR3 e CCR2 (81 e 6 candidatos, respectivamente). No entanto, essas diferenças já eram esperadas, uma vez que para tumores sólidos não é possível correlacionar o número de rearranjos cromossômicos com a localização/tipo da neoplasia nem com o estadio da doença (Tcga, 2012; Drier, Lawrence *et al.*, 2013). Outra evidência importante, a qual também está de acordo com a literatura, é a ausência de rearranjos recorrentes nas amostras tumorais dos pacientes, reforçando assim a necessidade da utilização da abordagem personalizada para a identificação dos biomarcadores.

4.3 VALIDAÇÃO DAS VARIAÇÕES ESTRUTURAS

Para a validação das possíveis variações estruturais presentes nas amostras de tumor primário, foram selecionados aleatoriamente, pelo menos cinco candidatos de cada amostra para os quais foram confeccionados iniciadores específicos flanqueando o ponto de quebra do rearranjo e que foram utilizados para a amplificação e posterior sequenciamento dos fragmentos amplificados. Nas reações de amplificação foram utilizados como molde o *DNA* do tumor primário e como controle de especificidade utilizamos o *DNA* de células do sangue periférico (linfócitos) do mesmo indivíduo, além da reação sem *DNA*. Os resultados das reações de *PCR* foram avaliados por eletroforese em gel de agarose e os produtos obtidos, nos quais observou-se ocorrência de amplificação tumor-específica, foram sequenciados pelo método de Sanger (Figura 16).

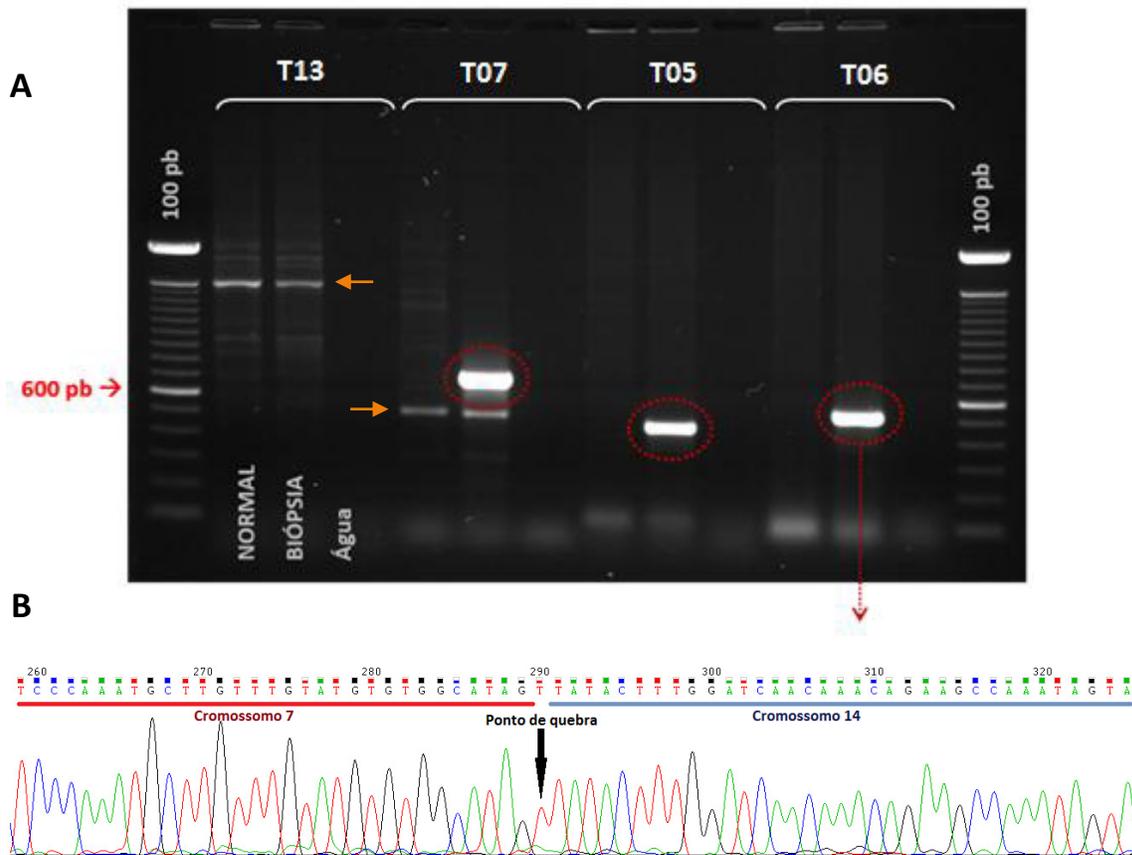


Figura 16 – Representação da validação dos candidatos a VE por PCR e sequenciamento pelo método de Sanger. A- Gel de agarose dos produtos das reações de PCR. Estão apresentadas as validações para quatro candidatos a translocação (T) do paciente IR1. Os **círculos vermelhos** destacam as bandas que apresentaram amplificação tumor-específica e foram sequenciadas pelo método de Sanger. As **setas amarelas** apontam para amplificações que ocorreram tanto no tecido normal como tumoral e foram descartadas na validação. B – O sequenciamento pelo método de Sanger proporcionou a localização exata do ponto de quebra das variações estruturais, como exemplificado no cromatograma do candidato T06 do paciente IR1.

Primeiramente, buscou-se validar as alterações envolvendo dois cromossomos, dado que esse tipo de evento ocorre quase que exclusivamente em células neoplásicas, esta abordagem aumentaria as chances desse candidato ser um bom biomarcador da presença da doença. Nos pacientes nos quais não fora possível validar pelo menos três variações estruturais intercromossômicas, optou-se pela validação das alterações

envolvendo o mesmo cromossomo, o que foi então realizado para nas amostras dos pacientes CCR1, CCR2 e CCR3.

Foram considerados candidatos validados aqueles cujo sequenciamento Sanger permitiu a obtenção de uma sequência de bases que, ao ser alinhada contra a sequência de referência do genoma humano, reportou a união de duas regiões localizadas em cromossomos distintos no genoma de referência, no caso das variações intercromossômicas. E para as variações intracromossômicas, quando esse alinhamento reportou a aproximação de duas regiões que, no genoma de referência, deveriam distar mais de 4 mil bases no mesmo cromossomo. O Apêndice B contém informações detalhadas sobre as supracitadas validações (regiões cromossômicas envolvidas, coordenadas do ponto de quebra e tamanho do fragmento amplificado). O resultado das validações está resumido na Tabela 3.

Devido a faixa de tamanho dos fragmentos de *DNA* escolhida para a construção da biblioteca tipo *mate-pair* (600-1000 pb), a validação dos candidatos foi realizada utilizando apenas um par de iniciadores, com o qual foi possível amplificar, sequenciar e identificar o ponto de quebra de cada VE. Isso porque, no sequenciamento pelo método de Sanger é possível obter, com qualidade, uma sequência de aproximadamente 600 pb em cada direção. Desta forma, quando o sequenciamento é feito com os dois iniciadores, ou seja, em ambos sentidos, é possível obter a sequência completa de um fragmento de

~1200 pb. Com isso, foi possível reduzir o tempo e os custos associados a confecção de um número maior de iniciadores e mais reações de sequenciamento, como seria necessário caso a faixa de tamanho utilizada para a construção da biblioteca tipo *mate-pair* fosse maior.

Tabela 3 – Resumos das validações

Amostra	Variações intercromossômicas		Variações intracromossômicas		total de variações validadas
	nº candidatos testados	nº validados	nº candidatos testados	nº validados	
CCR1	5	1	10	2	3
CCR2	3	1	3	2	3
CCR3	18	1	10	5	6
PCR1	3	3	---	---	3
IR1	10	9	---	---	9
IR2	3	3	7	4	7

--- nenhum candidato dessa categoria foi testado

Para todos os pacientes foi possível identificar e validar pelo menos três variações estruturais, o que representa o sucesso dessa abordagem e também das análises de bioinformática. A partir das variações estruturais validadas, foram escolhidas três que estavam mais representadas na amostra do tumor primário, para serem otimizadas e utilizadas nos ensaios de *Nested PCR*. Devido a limitações técnicas, desses ensaios otimizados, foram escolhidos somente dois que apresentaram melhor eficiência de amplificação e quantificação, para análise das amostras de plasma. Pois, uma vez que o *ctDNA* encontra-se altamente fragmentado no plasma, foram desenhados novos

iniciadores e também sondas *MGB TaqMan* (*Life Technologies*), de forma que os fragmentos a serem amplificados possuíssem no máximo 180pb, quando possível (Mouliere, Robert *et al.*, 2011).

4.4 ESTABILIDADE DAS VARIAÇÕES ESTRUTURAIS EM AMOSTRAS DE TUMOR RESIDUAL E METÁSTASE

Para as variações estruturais presentes no tumor primário serem utilizadas como biomarcadores, é essencial que elas sejam estáveis e mantidas durante a progressão tumoral, desde o tumor primário aos focos metastáticos. Para testar a estabilidade das variações estruturais, foram utilizadas amostras do tumor residual removido cirurgicamente após a QRTn dos pacientes CCR1 e IR1 e amostra do tecido metastático do paciente IR1. Então, foram realizadas reações de *PCR* para os mesmos ensaios validados com amostras do tumor primário, e os resultados estão apresentados na Figura 17.

Para a amostra do tumor residual do paciente CCR1, foram testadas as três variações estruturais validadas na biópsia pré-tratamento (T01, I01 e I02) e duas delas (I01 e I02) estavam presentes. Nas amostras do tumor residual e metástase hepática do paciente IR1, estavam presentes sete das nove VE validadas na biópsia pré-tratamento, sendo as VE T07 e T11 ausentes em ambas amostras. Esses resultados demonstram que, para as amostras testadas, a maior parte das VE presentes no tumor primário também estão no tumor residual e nódulos metastáticos. O fato de algumas alterações genéticas serem selecionadas positiva ou negativamente representa o resultado de uma seleção clonal ocorrida devido ao tratamento QRTn ou ao processo tumorigênico/metastático. Este

evento de expansão clonal é bastante conhecido e reflete a heterogeneidade tumoral existente na lesão primária (Meacham e Morrison, 2013). Para contornar esse obstáculo e utilizar as VE como biomarcadores, devem ser escolhidas mais de uma alteração para ser acompanhada no plasma, diminuindo a probabilidade de todas as VE utilizadas sofrerem seleção negativa.

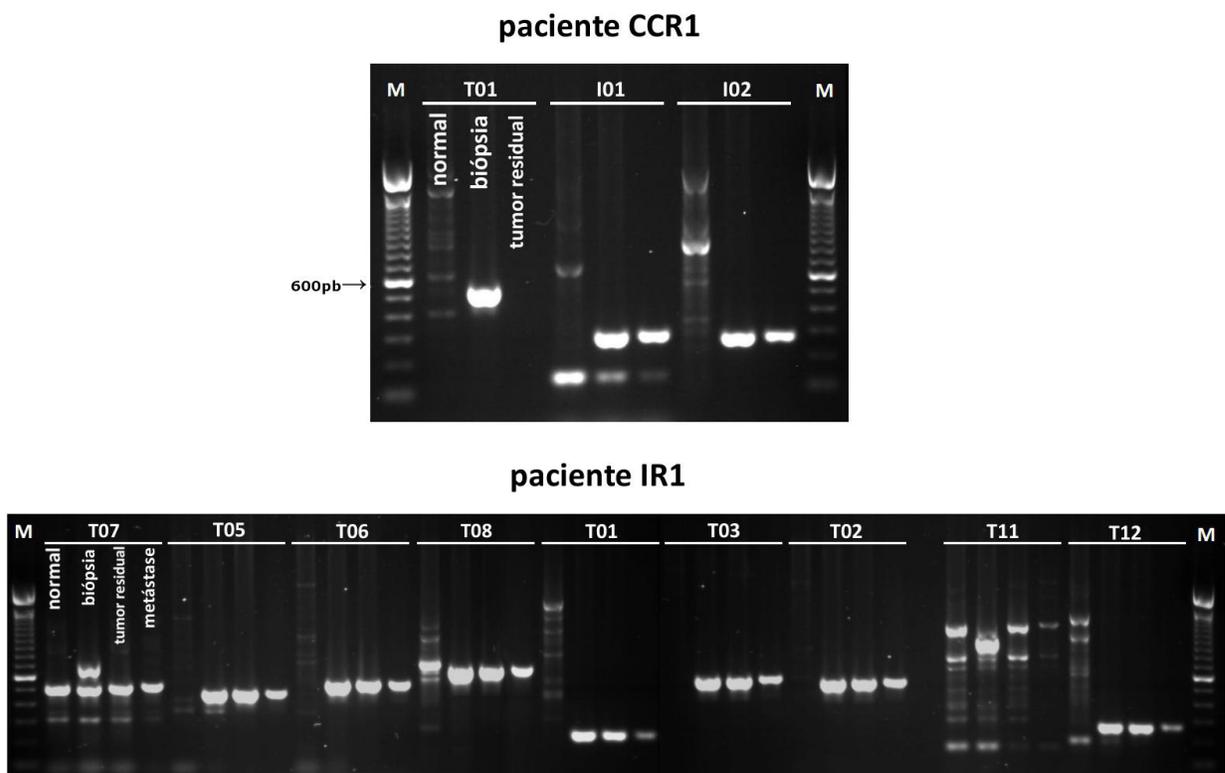


Figura 17 – Estabilidade das variações estruturais em amostras de tumor residual e metástase. Todas as variações estruturais validadas para os pacientes CCR1 e IR1 foram também avaliadas por *PCR* nas amostras de tumor residual, obtidas após a ressecção do tumor e, para o paciente IR1, também na amostra do nódulo metastático removido na hepatectomia. A figura apresenta o resultado da eletroforese em gel de agarose dos produtos de *PCR*. A identificação das variações estruturais para cada paciente está na parte superior de cada gel. (**M**) são as canaletas com o marcador de peso molecular 100pb.

4.5 QUANTIFICAÇÃO DO *DNA* CIRCULANTE TOTAL

Amostras de sangue foram coletadas dos seis pacientes em diferentes momentos: no diagnóstico (biópsia), após o término do tratamento, no momento da reavaliação clínica e durante o seguimento. No total foram realizadas 47 coletas, das quais 39 foram processadas para obtenção do *DNA* circulante total e submetidas a análise de detecção do *ctDNA* (Figura 18).

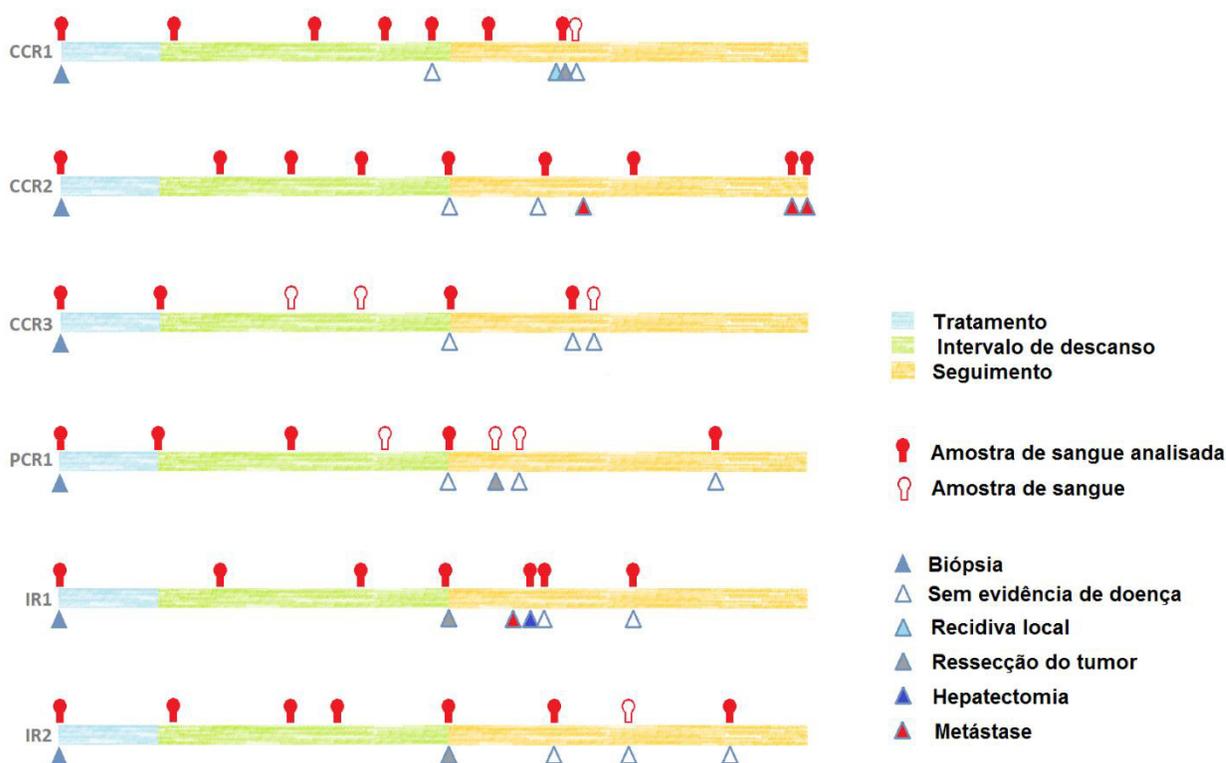


Figura 18 – Esquema com os momentos das coletas de sangue e principais desfechos clínicos de cada paciente.

Primeiramente, o *DNA* circulante de todas as amostras foi quantificado por *ddPCR* com um ensaio que amplifica um gene de cópia única do genoma humano (RNase P),

conforme descrito em Materiais e Métodos. Essa quantificação forneceu o número de cópias amplificáveis presente em cada amostra, e quando comparamos a quantificação de dois volumes crescentes (2 e 4 μ L) da mesma amostra, foi possível também estimar a presença de inibidores de *PCR*. Esses resultados foram importantes para mostrar que o processamento do plasma, a extração e a qualidade do *DNA* circulante estavam adequados. Além disso, eles também serviram para estimar o volume de *DNA* circulante utilizado na reação de pré-amplificação para a detecção do *ctDNA* de cada amostra. Na Tabela 4 estão apresentados: o volume de plasma analisado, o volume com o qual o *DNA* circulante foi eluído e o número de cópias amplificáveis para cada amostra.

Na Tabela 4 é possível observar que a quantificação do *DNA* circulante varia significativamente entre as amostras, ainda que do mesmo paciente. Esse fato pode ser decorrente das condições fisiológicas e patológicas de cada indivíduo, que interferem na quantidade de *DNA* liberado na circulação, como também pode refletir a qualidade do *DNA* circulante, que mesmo estando presente na amostra, tem dificuldade de ser amplificado por *PCR* devido à presença de inibidores ou danos na estrutura que impossibilitam a ação da *DNA* polimerase.

Tabela 4 – Quantificação do DNA circulante nas amostras de plasma dos pacientes

Paciente	Coleta	Volume plasma (mL)	Volume eluição (µL)	Ensaio	nº gotas analisadas	cópias/mL plasma*
CCR1	diagnóstico	3	125	RNase P	16240	906
CCR1	0sem	3	125	RNase P	10759	552
CCR1	6sem	3	125	RNase P	17121	1438
CCR1	9sem	3	125	RNase P	13747	417
CCR1	12sem	3	125	RNase P	15852	1031
CCR1	22sem	3	125	RNase P	16850	3042
CCR1	53sem	3	125	RNase P	15959	1792
CCR2	diagnóstico	3	125	RNase P	16714	500
CCR2	sem 3	3	125	RNase P	10562	333
CCR2	sem 6	3	125	RNase P	17976	500
CCR2	sem 9	3	125	RNase P	17090	667
CCR2	sem 13	3	125	RNase P	9963	167
CCR2	sem 46	3	125	RNase P	15517	917
CCR2	sem 84	3	125	RNase P	11399	771
CCR2	sem 158	3	125	RNase P	17239	729
CCR2	sem 170	3	125	RNase P	16417	958
CCR2	sem 177	3	125	RNase P	16961	1729
CCR3	diagnóstico	3	125	RNase P	14803	596
CCR3	0sem	3	125	RNase P	16746	1277
CCR3	14sem	3	125	RNase P	10361	829
CCR3	57sem	3	125	RNase P	16277	3750
PCR1	diagnóstico	3	125	RNase P	15973	3389
PCR1	sem 0	3	125	RNase P	14688	778
PCR1	sem 6	3	125	RNase P	13984	458
PCR1	sem 13	3	125	RNase P	15779	944
PCR1	sem 118	3	125	RNase P	14745	688
IR1	diagnóstico	2	125	RNase P	13768	906
IR1	sem 3	3	125	RNase P	14563	271
IR1	sem 9	3	125	RNase P	14350	1583
IR1	sem 13	1,2	50	RNase P	14587	1104
IR1	sem 40	3	125	RNase P	14367	31389
IR1	sem 46	3	125	RNase P	14217	1292
IR1	sem 84	3	125	RNase P	13912	458
IR2	diagnóstico	3	125	RNase P	11422	4063
IR2	week1	3	125	RNase P	16936	2083
IR2	week6	3	125	RNase P	9853	917
IR2	week8	3	125	RNase P	16170	1750
IR2	week13	3	125	RNase P	11169	6583
IR2	week52	3	125	RNase P	12428	1333
IR2	week124	3	125	RNase P	9308	729

* número de cópias por reação calculado pelo *software Quantasoft* e normalizado pelo volume de plasma e volume de eluição.

4.6 DETECÇÃO DO *DNA* TUMORAL CIRCULANTE NO PLASMA

Para detectar o *DNA* tumoral (*ctDNA*) dentre o *DNA* circulante, foram utilizados ensaios de *PCR* para duas variações estruturais exclusivas do tumor de cada paciente. Devido à baixa concentração do *ctDNA* no plasma, foi realizada a pré-amplificação do *DNA* circulante em uma reação multiplex, com iniciadores específicos para duas variações estruturais personalizadas e uma região controle (RNase P). A Figura 19 apresenta a comparação entre a eficiência de detecção do *ctDNA* por *ddPCR* em uma amostra sem realizar a etapa de pré-amplificação e a mesma amostra depois de pré-amplificada. É notável que essa estratégia aumenta a sensibilidade da detecção sem perder a especificidade, como pode ser verificado pelo aumento significativo do número de gotas positivas nas reações de *ddPCR* em que foi utilizado o *DNA* circulante, e a ausência de amplificação nas reações controle (reação sem *DNA* e reação com *DNA* de tecido normal) que também foram pré-amplificadas.

Assim, todas as amostras de *DNA* circulante foram pré-amplificadas e seus produtos foram utilizados para a detecção do *ctDNA*, em ensaios individuais para cada VE e RNase P, por *ddPCR*. Devido ao pequeno volume de *DNA* circulante (4-7 μ L) utilizado em cada reação de pré-amplificação, foram necessárias, em média, 16 reações para analisar ~75% do *DNA* circulante extraído de cada amostra de plasma.

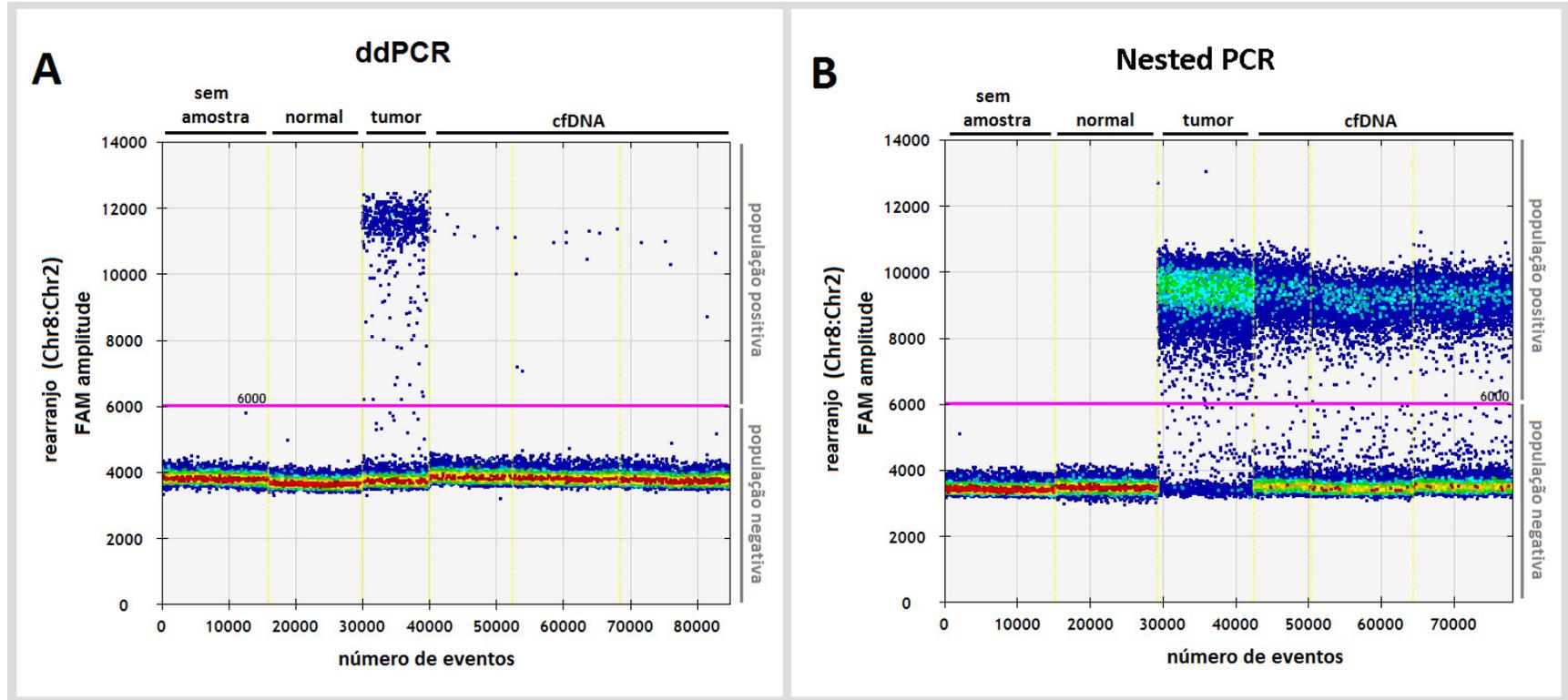


Figura 19 – Comparação de detecção do ctDNA usando ddPCR sem e com pré-amplificação. Representação gráfica do resultado da *PCR Digital*. Cada gota lida é representada por um ponto no eixo horizontal (número de eventos) e sua amplitude de fluorescência no eixo vertical. A região com maior densidade de gotas é apresentada por cores quentes, sendo a máxima densidade representada na cor vermelha e as menores na cor azul. O traço na cor rosa representa o limite entre a população de gotas positivas e negativas. **(A)** Resultado das reações de ddPCR com ensaio para detectar uma determinada VE, sem amplificação prévia das amostras. **(B)** Representa o mesmo ensaio utilizado em (A), porém as amostras foram previamente amplificadas em um *PCR multiplex* utilizando-se iniciadores externos aos utilizados na segunda amplificação.

Ao final da análise das reações de *ddPCR*, o número de cópias obtido foi normalizado pelo volume de plasma, volume de eluição e volume de *DNA* circulante utilizado na reação de pré-amplificação. Quando o número de cópias do *DNA* tumoral alvo por amostra foi igual a 1, para que essa detecção fosse considerada verdadeira, foi exigido um mínimo de três gotas positivas por amostra. Esse parâmetro foi calculado com base no limite de detecção de cada ensaio. Os resultados da análise de detecção de *ctDNA* em todas as amostras de plasma estão apresentados na Tabela 5 do APÊNDICE C.

Fragmentos de *ctDNA* reportando rearranjos cromossômicos, biomarcadores personalizados, foram detectados em amostras de plasma de quatro (67%) dos seis pacientes desse estudo. Os dois pacientes em que não foi possível detectar nenhum dos biomarcadores tinham tumores menos invasivos (T2), enquanto que os outros quatro pacientes apresentavam tumores T3. Para esses pacientes, a detecção dos biomarcadores foi compatível com a evolução clínica, com a exceção de uma coleta (semana 13) do paciente IR1. O Quadro 2 resume as principais características dos biomarcadores testados para cada paciente. A eficiência dos biomarcadores personalizados para monitorar a dinâmica tumoral nos pacientes com câncer de reto após QRTn estão apresentadas nas figuras 20-23.

O monitoramento do *ctDNA* do paciente IR2, diagnosticado com câncer de reto cT3N1, foi realizado através da detecção de dois biomarcadores personalizados, deleções D01 e D02, por *PCR* digital (Figura 20). Esses biomarcadores foram detectados nas

amostras de plasma no momento do diagnóstico. Uma semana após o término da QRTn, os níveis de *ctDNA* caíram consideravelmente e não foram detectados nas coletas durante o período de descanso (semana 6 e semana 8). No momento da reavaliação clínica foi o biomarcador D02 foi detectado, ainda que em baixa quantidade, indicando a presença de doença residual após o tratamento. Concordante com os resultados dos biomarcadores, os exames clínicos e radiológicos revelaram que esse paciente havia respondido incompletamente ao tratamento QRTn, e o paciente foi submetido a cirurgia radical (semana 13).

Quadro 2 – Relação dos biomarcadores testados no plasma

Paciente	VE	Regiões cromossômicas				Características da região	genes	Detecção no plasma	
		Cro A	Posição A	Cro B	Posição B				
CCR1	T01	19	30473830	17	50154848	-	intragênica	<i>URI1/CA10</i>	não
CCR1	I01	8	123281119	8	-125821045	>1 Mb	intergênica	-	não
CCR2	I01	12	1819278	12	-1870423	~50 Kb	intragênica	<i>ADIPOR2</i>	sim
CCR2	D02	17	39823076	17	39837384	~14 Kb	intragênica	<i>JUP</i>	sim
CCR3	D01	3	188150728	3	188409009	~250 Kb	intragênica	<i>LPP</i>	não
CCR3	D02	20	18031602	20	17582992	~450 Kb	intragênica	<i>OSTN, OVOL2</i>	não
PCR1	T01	8	93421404	6	13191227	-	intragênica	<i>PHACTR1</i>	sim
PCR1	T02	12	3607669	6	145470434	-	intragênica	<i>PRMT8</i>	sim
IR1	T01	17	35126102	1	223988874	-	intragênica	<i>TP53BP2</i>	sim
IR1	T02	8	40477264	2	152374971	-	intragênica	<i>ZMAT4, NEB</i>	sim
IR2	D01	3	60342822	3	60388671	~45 Kb	intragênica	<i>FHIT</i>	sim
IR2	D02	4	185436839	4	185913887	~500 Kb	intergênica	-	sim

(VE) Variação Estrutural, os candidatos estão identificados com as iniciais D (Deleção), I (Inversão), T (Translocação); as regiões cromossômicas são baseadas na sequência referência do genoma humano (hg19), o sinal negativo na coordenada sinaliza que é a fita menos (-) da região cromossômica de referência; (CroA) Cromossomo A, região *upstream* da VE; (CroB) cromossomo B, região *downstream* da VE. Na coluna (características da região) está informado o tamanho (bp) aproximado da região envolvida para as VE intracromossômicas. Nessa mesma coluna está informado se a região do ponto de quebra da VE é intergênica ou intragênica e quais são os genes truncados por essa VE.

Posteriormente, o exame histopatológico da peça cirúrgica revelou um grau de regressão tumoral 2 (GRT=2), no qual 50-75% das células tumorais presentes na lesão estão viáveis. Todas as amostras de sangue analisadas durante o período de seguimento (até semana 124) foram negativas para presença de *ctDNA*. O paciente em questão ainda está sob acompanhamento e, até a conclusão deste trabalho, está sem evidência de doença.

Essa correlação entre os dados clínicos e a detecção dos biomarcadores também foi observada no paciente PCR1, diagnosticado com câncer de reto cT3N1, para o qual foram avaliados os biomarcadores, translocações T01 e T02 (Figura 21). No diagnóstico, foram encontrados os dois biomarcadores analisados e seus níveis estavam reduzidos nas amostras após o término do tratamento, no qual T01 não foi detectado. Na semana 6 e na reavaliação clínica (semana 13) o *ctDNA* não foi detectado por nenhum dos dois ensaios.

No entanto, no momento da reavaliação clínica, a doença não pôde ser descartada pelos exames clínicos e radiológicos, devido à presença de cicatrizes e/ou fibrose presentes no local do tumor primário, e o paciente foi submetido à cirurgia radical. A análise histopatológica da peça removida revelou ausência de células tumorais viáveis, demonstrando que o paciente havia respondido completamente ao tratamento, corroborando os resultados negativos dos biomarcadores nos pontos seguintes ao término do tratamento. Esse paciente ainda encontra-se sob observação e não apresenta evidência de doença. Consistentemente, a amostra de plasma analisada durante o seguimento (semana 118) foi negativa quanto a presença de *ctDNA*.

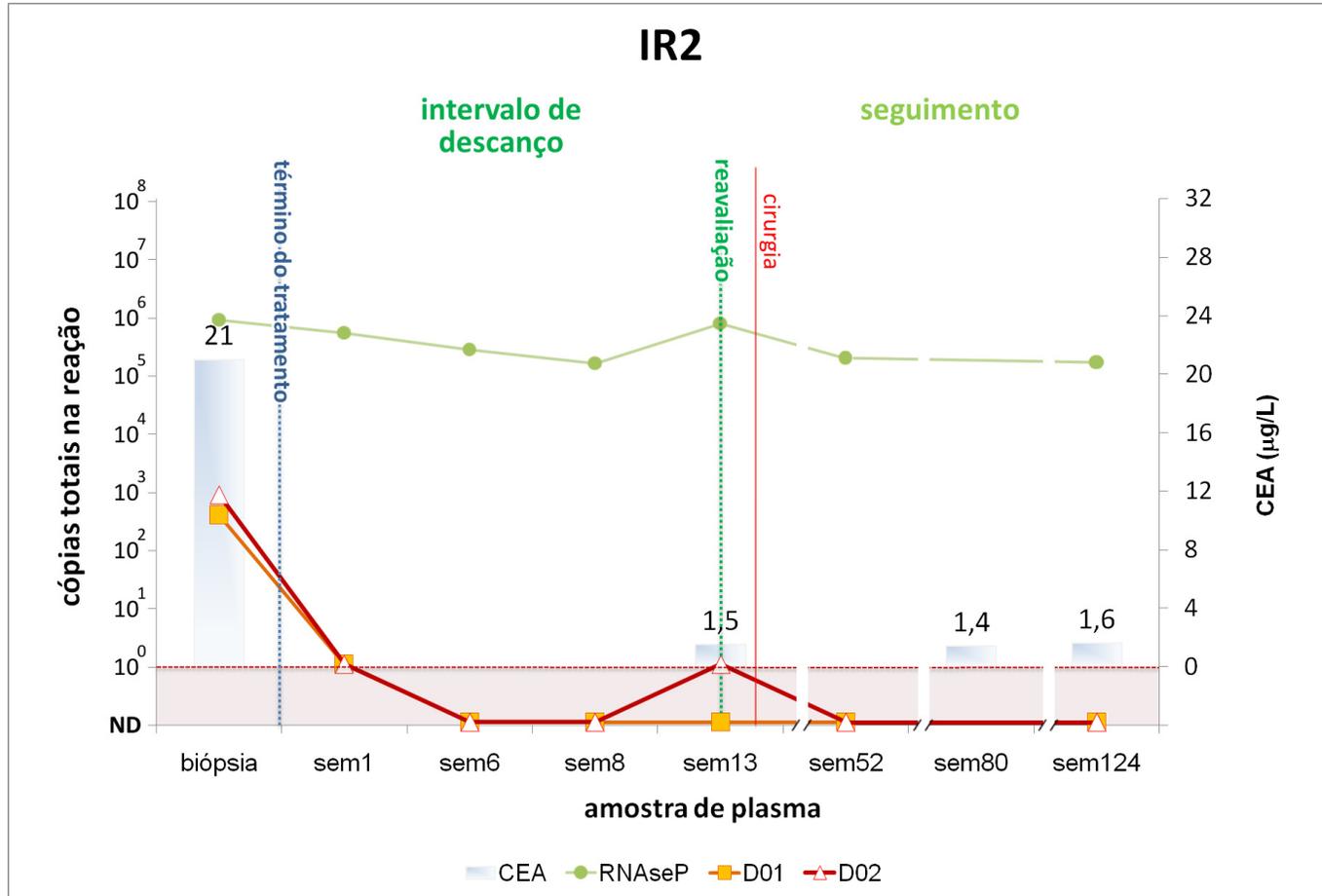


Figura 20 – Detecção dos biomarcadores personalizados por PCR digital: paciente com resposta incompleta. Para o monitoramento do *ctDNA* deste paciente, foram utilizados dois biomarcadores (D01 e D02). Na abscissa primária está apresentado o número total de cópias de *DNA* alvo amplificado nas reações de pré-amplificação para cada amostra de plasma analisada (ordenada). A abscissa secundária refere-se ao monitoramento sérico do marcador CEA (microgramas/L), que está apresentado no gráfico sob forma de colunas. As amostras foram nomeadas de acordo com o tempo (em semanas) em que foram coletadas, contado a partir do término do tratamento. O ensaio para a detecção da região controle (RNase P) reflete o *DNA* circulante total (linha verde). A linha vermelha tracejada (horizontal) demarca o limite de detecção do ensaio, abaixo dessa faixa os níveis são considerados não-detectáveis (ND). As amostras foram coletadas em diferentes momentos da doença e as linhas tracejadas na vertical separam três momentos específicos: diagnóstico (biópsia), intervalo de descanso após a QRTn e seguimento. A linha vermelha na vertical representa o momento da cirurgia em que o tumor primário foi ressecado.

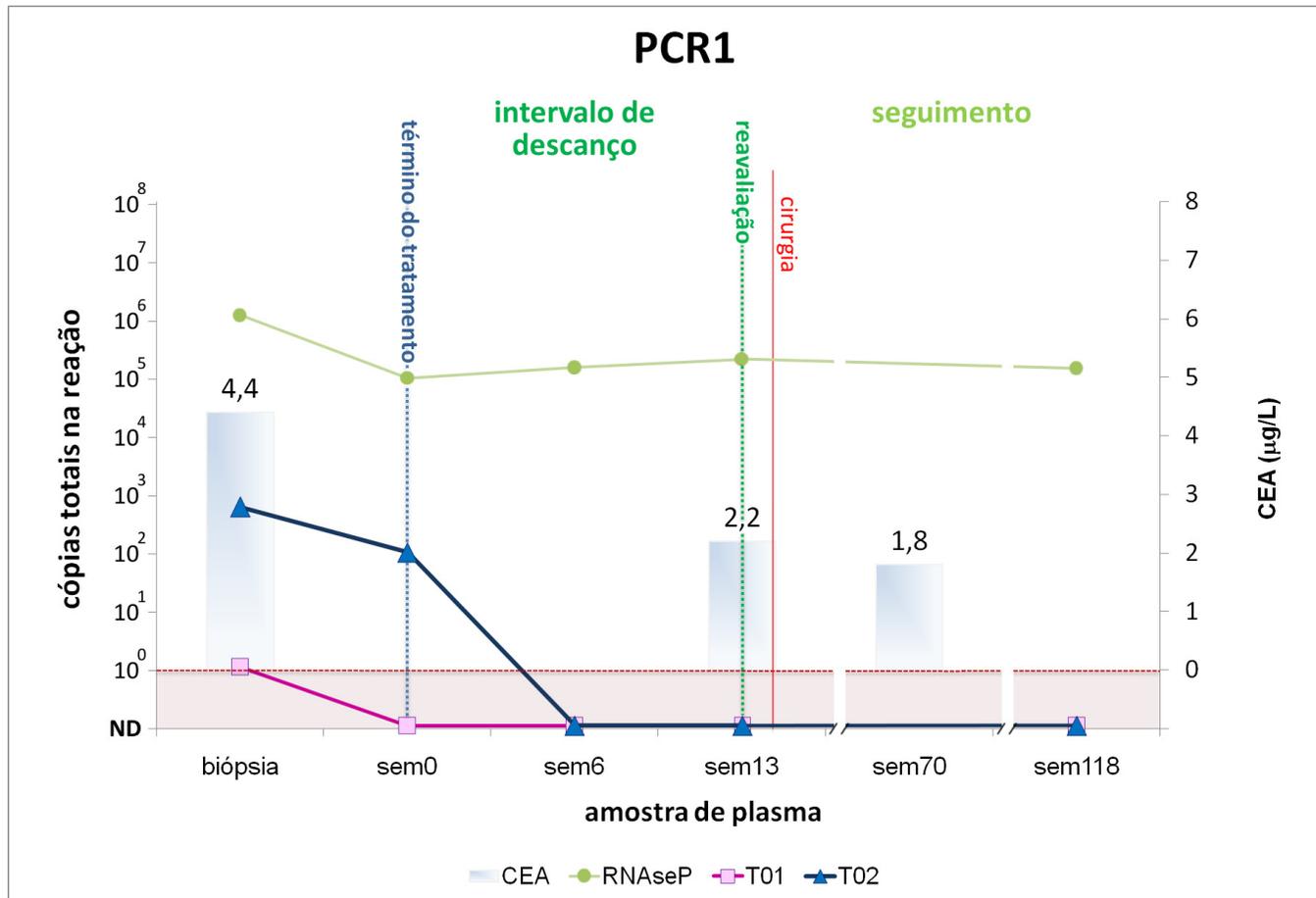


Figura 21 – Detecção dos biomarcadores personalizados por PCR digital: paciente com resposta patológica completa. Para o monitoramento do *ctDNA* deste paciente, foram utilizados dois biomarcadores (T01 e T02). Na abscissa primária está apresentado o número total de cópias de *DNA* alvo amplificado nas reações de pré-amplificação para cada amostra de plasma analisada (ordenada). A abscissa secundária refere-se ao monitoramento sérico do marcador CEA (microgramas/L), que está apresentado no gráfico sob forma de colunas. As amostras foram nomeadas de acordo com o tempo (em semanas) em que foram coletadas, contado a partir do término do tratamento. O ensaio para a detecção da região controle (RNase P) reflete o *DNA* circulante total (linha verde). A linha vermelha tracejada (horizontal) demarca o limite de detecção do ensaio, abaixo dessa faixa os níveis são considerados não-detectáveis (ND). As amostras foram coletadas em diferentes momentos e as linhas tracejadas na vertical separam três momentos específicos: diagnóstico (biópsia), intervalo de descanso após a QRTn e seguimento. A linha vermelha na vertical representa o momento que o paciente foi submetido à cirurgia e a peça retirada foi encaminhada para o exame histopatológico.

No monitoramento do paciente CCR2, diagnosticado com câncer de reto T3N0, foram utilizados dois biomarcadores na análise do *DNA* circulante, sendo duas deleções (D01 e D02) (*Figura 22*). No momento do diagnóstico, o *ctDNA* foi detectado com os dois biomarcadores. Logo após a QRTn houve uma redução nos níveis desses biomarcadores, mas ainda estava presente em níveis consideráveis no momento da reavaliação clínica. Após 12 semanas do término do tratamento, a avaliação clínica e radiológica indicou uma resposta completa clínica e o paciente foi recomendado para o protocolo *Wait and Watch*.

Curiosamente, 46 semanas após o término da QRTn, houve um aumento substancial nos níveis do *ctDNA*, evidenciando presença de doença neoplásica. Neste momento, os exames clínicos não revelaram recorrência local ou à distância da doença. Somente na avaliação seguinte do paciente, 62 semanas após a conclusão do tratamento, foi feito o diagnóstico da metástase hepática por meio de exames radiológicos convencionais. Assim, o paciente foi imediatamente submetido à quimioterapia sistêmica de primeira linha, iniciada na semana 73. Concordante com a boa resposta clínica apresentada inicialmente pelo paciente, os níveis de *ctDNA* reduziram significativamente na coleta da semana 84. No entanto, os exames clínicos detectaram a progressão da doença, e iniciou-se tratamento quimioterápico de segunda linha na semana 138. Condizente com a progressão da doença, na semana 158, foram observados grande aumento nos níveis de *ctDNA* em relação a coleta anterior. Essa tendência também foi observada na coleta seguinte (semana 170), indicando uma má resposta ao tratamento e a progressão da

doença. Com base nesses resultados, foi indicado o re-estadiamento precoce, que confirmou a progressão da doença e resultou no início do tratamento sistêmico de 3ª linha (semana 173). Os níveis dos biomarcadores na última coleta analisada (semana 175) sugeriu que, mesmo após o início do tratamento de 3ª linha, a doença ainda permanecia em evolução, o que foi confirmado pelos exames clínicos. Infelizmente, o paciente faleceu após algumas semanas (semana 186), devido ao agravamento do seu estado clínico.

Desta forma, a detecção dos biomarcadores no paciente CCR2, além de estar completamente condizente com o quadro clínico apresentado, foi capaz de detectar a doença metastática com 49 semanas de antecedência em comparação aos exames clínicos convencionais. O acompanhamento desses biomarcadores personalizados também puderam auxiliar na indicação do re-estadiamento precoce da doença, o que resultou no início de uma nova linha de tratamento sistêmico para tentar cessar a progressão da doença.

O monitoramento molecular do paciente IR1, diagnosticado com câncer de reto cT3N1, também foi capaz de detectar a doença metastática previamente à detecção pela avaliação clínica, com 36 semanas de antecedência. A análise do *ctDNA* foi realizada com dois biomarcadores (T01 e T02). O *DNA* tumoral circulante foi detectado no diagnóstico e após a QRTn, com pelo menos um biomarcador. Não foi possível detectar a presença de *ctDNA* na amostra coletada no momento da reavaliação clínica. No entanto, exames

clínicos e radiológicos realizados durante a reavaliação clínica, revelaram que o paciente havia respondido incompletamente, fato confirmado após a cirurgia radical, quando o exame histopatológico revelou a presença de células tumorais viáveis (GRT=3, no qual mais que 50% da massa tumoral é fibrose) na peça cirúrgica ressecada. A ausência de *ctDNA* na amostra no momento que o paciente ainda apresentava tumor residual, pode ser reflexo da falta de sensibilidade desse teste para a detecção de doença residual, e será discutida mais adiante.

Durante o seguimento, esse paciente foi diagnosticado com metástase no fígado 36 semanas após o término do tratamento e a hepatectomia foi realizada na semana 40. Os biomarcadores T01 e T02 foram detectados na coleta antes da cirurgia, concordante com a doença metastática. Nas semanas seguintes à remoção dos nódulos hepáticos, o nível do biomarcador T01 se manteve alto nas amostras de plasma, indicando que a doença ainda estava presente. Esse fato foi verificado clinicamente somente na semana 82, quando os exames de imagem diagnosticaram metástase no arco costal. Infelizmente, esse paciente não respondeu ao tratamento sistêmico e veio a óbito.

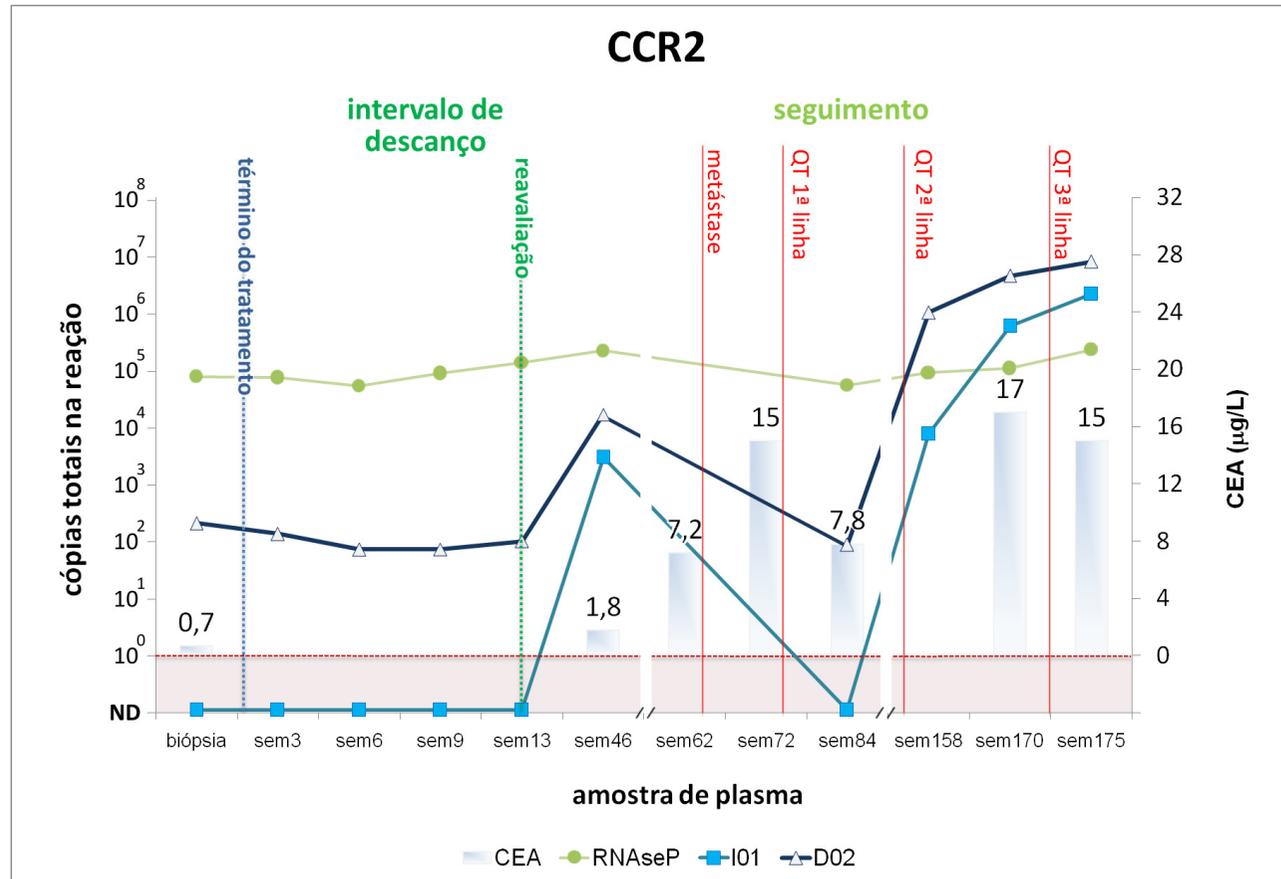


Figura 22 – Detecção dos biomarcadores personalizados por PCR digital: paciente com resposta clínica completa seguida de doença metastática. Para o monitoramento do *ctDNA* deste paciente, foram utilizados dois biomarcadores (I01 e D02). Na abscissa primária está apresentado o número total de cópias de *DNA* alvo amplificado nas reações de pré-amplificação para cada amostra de plasma analisada (ordenada). A abscissa secundária refere-se ao monitoramento sérico do marcador CEA (microgramas/L), que está apresentado no gráfico sob forma de colunas. As amostras foram nomeadas de acordo com o tempo (em semanas) em que foram coletadas, contado a partir do término do tratamento. O ensaio para a detecção da região controle (RNase P) reflete o *DNA* circulante total (linha verde). A linha vermelha tracejada (horizontal) demarca o limite de detecção do ensaio, abaixo dessa faixa os níveis são considerados não-detectáveis (ND). As amostras foram coletadas em diferentes momentos e as linhas tracejadas na vertical separam três momentos específicos: diagnóstico (biópsia), intervalo de descanso após a QRTn e seguimento. As linhas vermelhas verticais representam o momento que, através dos exames clínicos e de imagem, foi possível a detecção da metástase hepática (62 semanas após o término do tratamento) e os momentos que marcam os inícios das quimioterapias (QT) de 1ª linha (sem 73), 2ª linha (sem 138) e 3ª linha (sem 173) para tratamento sistêmico.

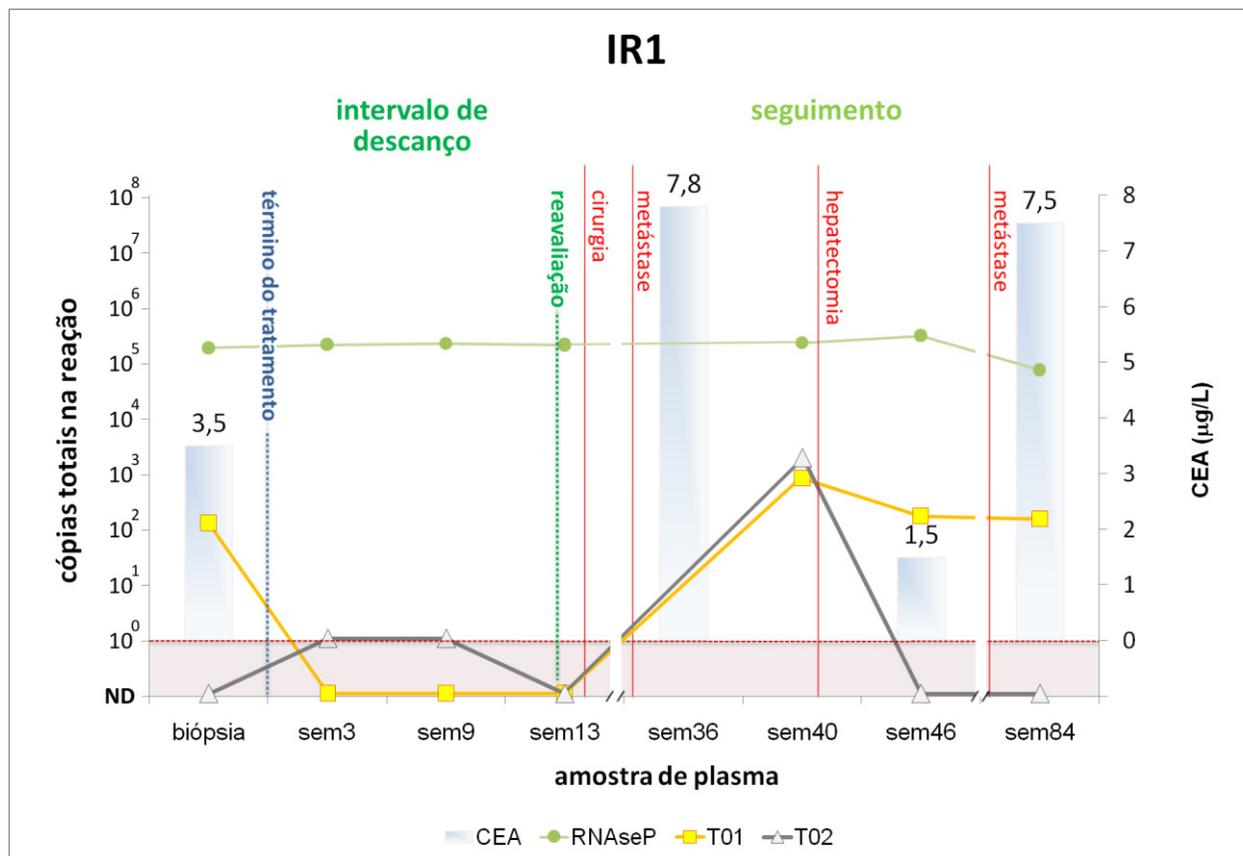


Figura 23 – Detecção dos biomarcadores personalizados por PCR digital: paciente com resposta incompleta seguida de metástase. Para o monitoramento do *ctDNA* deste paciente, foram utilizados dois biomarcadores (T01 e T02). Na abscissa primária, esta apresentada o número total de cópias de *DNA* alvo amplificado nas reações de pré-amplificação para cada amostra de plasma analisadas (ordenada). A abscissa secundária refere-se ao monitoramento sérico do marcador CEA (microgramas/L), que está apresentado no gráfico sob forma de colunas. As amostras foram nomeadas de acordo com o tempo (em semanas) em que foram coletadas, contado a partir do término do tratamento. O ensaio para a detecção da região controle (RNase P) reflete o *DNA* circulante total (linha verde). A linha vermelha tracejada (horizontal) demarca o limite de detecção do ensaio, abaixo dessa faixa os níveis são considerados não-detectáveis (ND). As amostras foram coletadas em diferentes momentos e as linhas tracejadas na vertical separam três momentos específicos: diagnóstico (biópsia), intervalo de descanso após a QRTn e seguimento. As linhas vermelhas na vertical representam o momento que o tumor primário foi ressecado (semana 13); em que a metástase hepática foi detectada (semana 36) nos exames clínicos e de imagem; quando realizada a remoção dos nódulos hepáticos (hepatectomia) e momento do diagnóstico da metástase no arco costal (semana 82), respectivamente.

5 DISCUSSÃO



5 DISCUSSÃO

O primeiro trabalho demonstrando a abordagem *Watch and Wait* para pacientes tratados com QRTn, foi publicado em 1998 (Habr-Gama, De Souza *et al.*, 1998). Desde então, a comunidade médica tem aumentado o interesse por essa conduta, mas, devido a subjetividade dos exames clínicos e limitações dos estudos por imagem, a prática dessa abordagem ainda é restrita e tem recebido algumas críticas (Glynne-Jones, Wallace *et al.*, 2008; Glynne-Jones e Hughes, 2012). No panorama atual, devido aos significativos avanços nas tecnologias de imagem investigativa, e também às experiências positivas de grupos independentes utilizando essa estratégia (Maas, Beets-Tan *et al.*, 2011; Smith, Ruby *et al.*, 2012), essa abordagem tem novamente chamado a atenção, pois poupa os pacientes com câncer de reto, que responderam completamente ao tratamento neoadjuvante, da cirurgia imediata, evitando assim os riscos e morbidades associadas à cirurgia.

Porém, ainda hoje não está disponível um marcador suficientemente específico e sensível que possa auxiliar na avaliação de resposta a QRTn, na detecção de doença residual mínima e no seguimento desses pacientes. A possibilidade de monitorar a carga tumoral dos pacientes submetidos ao tratamento neoadjuvante seria benéfica tanto para o paciente que responde completamente ao tratamento, uma vez que comprovada a inexistência de doença residual (situação na qual ele seria encaminhado com maior segurança para o protocolo *Watch and Wait*), como para o paciente que não responde ao tratamento e é submetido aos efeitos colaterais da quimiorradioterapia sem obter um de seus maiores benefícios (a resposta completa).

Durante o período de seguimento, a detecção dos biomarcadores personalizados mostra-se importante tanto para pacientes submetidos ao protocolo *Watch and Wait*, como para aqueles que receberam tratamento padrão, uma vez que, seria um grande auxílio na detecção precoce tanto de possíveis recidivas locais, nas quais seria realizada a cirurgia de resgate; como recidivas à distância, possibilitando o início do tratamento sistêmico. Também seria útil para monitorar a eficácia do tratamento sistêmico, podendo apontar quando é necessário antecipar o início de tratamentos de segunda ou terceira linhas e quando o paciente não estiver respondendo ao tratamento corrente.

O antígeno carcinoembrionário (*CEA*) é o biomarcador utilizado rotineiramente no seguimento dos pacientes com câncer colorretal (Newton, Newman *et al.*, 2012). Mas, essa proteína também é encontrada no soro de indivíduos sem doença neoplásica, ainda que em concentrações mais baixas e alterações nos níveis de *CEA* não são encontradas em uma parcela significativa (44-65%) dos pacientes com câncer avançado (Fakih e Padmanabhan, 2006).

Na ausência de marcadores sensíveis e específicos para detectar doença residual, a avaliação da resposta ao tratamento representa um desafio. Para a estratégia *Watch and Wait*, essa avaliação é baseada na rigorosa análise de um conjunto de exames, endoscópicos, radiológicos e clínicos, e quando indicam qualquer evidência de doença, é aconselhado o tratamento cirúrgico. Isto porque, os exames de imagem são bastante sensíveis para a detecção de até mesmo pequenos aglomerados de células, todavia, não são específicos o suficiente para distinguir células inflamadas e fibrose, das células

tumorais viáveis. Desta forma, na impossibilidade de descartar a presença de doença residual após o tratamento, parte dos pacientes que apresentam resposta completa patológica são submetidos à cirurgia desnecessariamente (Habr-Gama, Perez *et al.*, 2010; Maas, Nelemans *et al.*, 2010; Smith e Winter, 2014).

Nos últimos anos, importantes estudos comprovaram a sensibilidade dos biomarcadores baseados na detecção do *DNA* tumoral circulante no sangue de pacientes com câncer para diagnosticar, avaliar a resposta ao tratamento e detectar doença residual ou recidiva do tumor em diversos tipos de câncer. Especificamente, para tumores sólidos, nos quais alterações genéticas recorrentes não abrangem uma significativa parcela dos doentes (TCGA, 2012), surgiu uma nova categoria de biomarcadores que baseia-se nas características genômicas individuais do tumor (Mcbride, Orpana *et al.*, 2010; Dawson, Tsui *et al.*, 2013; Bettegowda, Sausen *et al.*, 2014). Esses trabalhos avaliaram a utilidade dos biomarcadores personalizados baseados no *ctDNA* em pacientes com câncer de mama, colorretal, osso, pâncreas, entre outros, mas nenhum deles avaliou especificamente pacientes com câncer de reto, sendo esse estudo o pioneiro.

Neste trabalho, foi avaliada a utilidade de biomarcadores personalizados baseados na detecção de variações estruturais, para manejo de pacientes com câncer de reto tratados com QRTn. Para a detecção do *ctDNA*, foram escolhidos rearranjos cromossômicos em vez de mutações pontuais e modificações epigenéticas, pois envolvem uma grande região do cromossomo, o que facilita a sua identificação e detecção (Mcbride, Orpana *et al.*, 2010). Cabe ainda ressaltar que a chance de

resultados falso-positivos é extremamente baixa, pois a produção de artefatos de *PCR* envolvendo regiões distantes do genoma é quase nula. Os resultados aqui apresentados corroboram essa hipótese, visto que não foram observadas amplificações nos ensaios para a detecção de *ctDNA* utilizando-se amostras do tecido normal, mesmo tendo sido realizada previamente uma etapa de pré-amplificação dessas amostras.

Adicionalmente, rearranjos cromossômicos muitas vezes estão associados a loci gênicos amplificados, o que aumenta as chances de detecção (Dawson, Tsui *et al.*, 2013; Bettegowda, Sausen *et al.*, 2014). Esse fato foi observado no presente estudo, quando demonstra a maior sensibilidade dos biomarcadores I01 e D02 para detectar o *ctDNA* nas amostras do paciente CCR2.

O fato dos rearranjos cromossômicos em câncer colorretal, na maioria das vezes, serem reportados em regiões que não envolvem genes codificadores de proteínas e com baixa taxa transcricional, sugere que tais mutações sejam passageiras, ou seja, que não estão diretamente envolvidos com o processo de iniciação e progressão do tumor (Drier, Lawrence *et al.*, 2013). Contudo, quando elas ocorrem nos estágios iniciais da tumorigênese possivelmente devem estar suficientemente representadas no tumor (Vogelstein, Papadopoulos *et al.*, 2013). Coerente com essa argumentação, a análise das variações estruturais, inicialmente identificadas no tumor primário do paciente CCR1 e IR1, puderam ser validadas nas amostras do tumor residual e tecido metastático, com a maioria das variações estruturais avaliadas.

Com a avaliação da estabilidade das variações estruturais nas amostras de tumor residual e metástase também foi possível inferir sobre a heterogeneidade tumoral, uma vez que algumas variações estruturais foram selecionadas negativamente durante a evolução do tratamento, como a VE T01 do paciente CCR1 e T07 e T11 do paciente IR1. Indicativos de seleção positiva também foram observados, pois durante a progressão da doença do paciente IR1 os níveis do biomarcador T01 foram detectados mesmo após a remoção cirúrgica dos nódulos metastático. Curiosamente, o ponto de quebra desta translocação ocorreu em uma região intrônica, truncando o gene que codifica para a proteína *TP53 Binding Protein II* e a perda funcional deste gene já foi relacionada com maior resistência à apoptose (Samuels-Lev, O'connor *et al.*, 2001; Bergamaschi, Samuels *et al.*, 2004). A expressão deste gene também foi encontrada regulada negativamente em câncer de mama metastático (Sgroi, Teng *et al.*, 1999). Juntos, estes fatos reforçam a importância da utilização de mais de um biomarcador personalizado para a detecção de *ctDNA* em cada paciente.

Para a identificação dos rearranjos personalizados foi utilizada uma metodologia com grande potencial de aplicação clínica, visando reduzir os custos e tempo de sequenciamento. Por esta razão, optamos pela *PARE (Personalized Analysis of Rearranged Ends)* e análise com a ferramenta ICRmax (Leary, Kinde *et al.*, 2010). Essa ferramenta de bioinformática foi desenvolvida pelo nosso grupo de pesquisa, com o objetivo de identificar alterações cromossômicas específicas do tumor através de um sequenciamento com baixa cobertura (~4 vezes) e sem a necessidade de sequenciamento de tecido normal pareado. Os resultados obtidos durante as

validações das variações estruturais foram importantes para a implementação de filtros essenciais utilizados nesse *pipeline*, resultando na co-autoria de um artigo científico aceito na revista *Genomics* (ANEXO B).

Quando se trata de ensaios customizados na prática clínica, sempre existe a preocupação referente ao custo e tempo proibitivos. A estimativa de custo associado ao material de consumo, nas dependências do Centro de Oncologia Molecular, para gerar dados de sequenciamento, utilizando a plataforma SOLiD 5500XL (Life Technologies), suficientes para uma cobertura real média de 4 vezes o genoma humano, somada a etapa de validação e produção dos ensaios para a detecção dos biomarcadores, foi de aproximadamente U\$ 3 000 (três mil dólares) por paciente e cerca de U\$ 400 dólares por amostra analisada com o *ddPCR*. Cabe ressaltar que esses custos foram calculados tendo como base esse estudo piloto e tais custos tendem a serem reduzidos significativamente quando aplicado em uma maior escala. Quanto ao tempo necessário, desde a obtenção da amostra (biópsia) até a produção dos ensaios para a detecção do *ctDNA* nas amostras de plasma, chegariam a perfazer 9 semanas. Esse período demonstra ser perfeitamente compatível com o intervalo de tratamento e avaliação clínica de resposta ao tratamento.

O *DNA* tumoral circulante é encontrado em baixas concentrações no sangue dos indivíduos, e varia entre 0,01% a 93% do *DNA* circulante total. No entanto, em pacientes com doença localizada, esses valores podem ser ainda mais baixos (Crowley, Di Nicolantonio *et al.*, 2013). A fim de aumentar a sensibilidade de detecção do *ctDNA*, que se baseou na reação em cadeia da polimerase (*PCR*), nesse trabalho foi realizada

uma estratégia que utiliza duas etapas de amplificação. Essa estratégia já havia sido descrita por McBride e colaboradores, para a detecção de variações estruturais no plasma de pacientes com câncer, e foi especialmente importante para esse projeto, uma vez que os pacientes incluídos no estudo tinham tumores localmente avançados. Na primeira etapa, o *ctDNA* alvo foi pré-amplificado para aumentar o número de moléculas alvo detectadas e quantificadas no segundo passo, em que foi utilizado o *PCR* digital. A especificidade foi garantida usando um par de iniciadores específicos para cada passo e também uma sonda específica para a região do ponto de quebra.

Foi possível monitorar com sucesso a dinâmica tumoral de todos os quatro pacientes com tumores de grau de invasão T3, usando dois biomarcadores personalizados. Com exceção de um ponto, semana 13 do paciente IR1 (Figura 23), todas as análises utilizando os biomarcadores personalizados foram concordantes com o quadro clínico dos pacientes. A ausência de biomarcadores na amostra no paciente IR1, que apresentou resposta incompleta ao tratamento e, portanto, no momento da reavaliação clínica (semana 13) apresentava tumor residual, pode ser explicada pelo menor volume de plasma utilizado nas análises. Para essa amostra, estava disponível apenas 1,2 mL de plasma e o padronizado para essa estratégia foi 3 mL. Outra possível explicação é que o *ctDNA* estava em baixas quantidades no plasma, devido ao grau de regressão tumoral atingido (GRT=3) no qual a maior parte da massa tumoral é fibrose; ou que o limite de sensibilidade do teste tenha sido atingido. Porém, para essa e outras amostras que o *ctDNA* não foi detectado mesmo na presença de doença, não é possível descartar que a qualidade da amostra tenha influenciado no resultado.

Para os pacientes CCR2 e IR1, os biomarcadores demonstraram um importante impacto clínico. Nesses pacientes, a doença sistêmica pode ser detectada de maneira sensível e específica com 49 e 36 semanas de antecedência dos exames clínicos, respectivamente. Para o CCR2 foi possível ainda indicar a progressão da doença durante a terapia sistêmica de segunda linha, o que resultou no re-estadiamento precoce e início da terceira linha de tratamento sistêmico.

A sensibilidade dos biomarcadores baseados em *ctDNA* para monitorar a dinâmica tumoral se mostrou superior a do marcador proteico *CEA*, atualmente utilizado na prática clínica. Essa diferença foi mais evidente na semana 46, após o término do tratamento, tanto para o paciente CCR2, como o IR1. Na semana 46, o paciente CCR2, apresentava níveis normais do *CEA*, enquanto que os biomarcadores personalizados já indicavam a presença de doença neoplásica, confirmada posteriormente. O paciente IR1 foi operado, na semana 40, para a remoção de nódulos metastáticos no fígado. O monitoramento do *CEA* na semana 46, apresentou níveis normais, entanto, nessa mesma semana já era possível detectar níveis de *ctDNA* no sangue deste paciente, provavelmente provenientes da metástase no arco costal, diagnóstica 36 semanas mais tarde. Esses resultados corroboram mais fortemente os estudos que mostram como os biomarcadores personalizados baseados em *ctDNA* são mais eficientes para monitorar a dinâmica tumoral (Dawson, Tsui *et al.*, 2013; Bettegowda, Sausen *et al.*, 2014).

Infelizmente, para os dois pacientes com tumor de reto avançado de grau de invasão T2 (CCR1 e CCR3), a detecção dos biomarcadores não foi útil e *ctDNA* não foi

detectado em nenhuma das amostras analisadas. Esses resultados estão de acordo com a literatura a qual relata que as concentrações de *ctDNA* estão diretamente correlacionadas com o grau de invasão do tumor e não apresentam correlação significativa com o volume tumoral (Diehl, Li *et al.*, 2005; Bettegowda, Sausen *et al.*, 2014). Esses pacientes no entanto, podem ser submetidos a cirurgia conservadora após o tratamento, o que implica em um menor impacto na qualidade de vida.

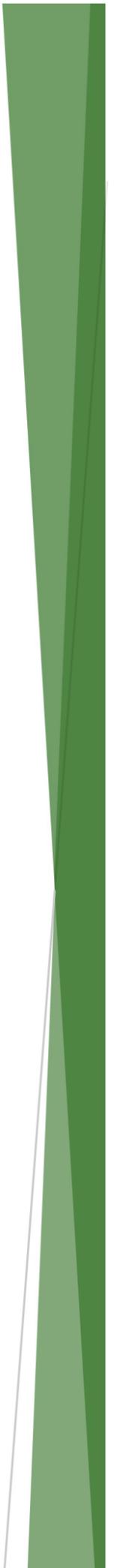
Na análise do presente estudo constatamos algumas limitações de estudo, dentre elas, primeiramente o número de pacientes incluídos. Isso deveu-se aos custos envolvidos, e várias dificuldades intrínsecas de um estudo prospectivo, como um frequente contato, e colaboração dos pacientes, para coletas de amostras durante intervalos definidos e por um longo período. Com relação às limitações técnicas, pode-se destacar o volume de plasma analisado. Foram analisados 3 mL de plasma, no entanto, o estudo publicado esse ano que avaliou *ctDNA* em amostras de 640 pacientes com câncer demonstrou que, em muitas vezes, a detecção do *DNA* tumoral em casos iniciais da doença só foi possível quando 5 mL ou mais de plasma foram analisados (Bettegowda, Sausen *et al.*, 2014).

O fato do *ctDNA* se encontrar altamente fragmentado no plasma, apresentando fragmentos que variam entre 130 e 180 pares de bases, o tamanho do fragmento que será amplificado (*amplicon*) através das reações de *PCR*, impacta diretamente na sensibilidade dos ensaios para a detecção dos biomarcadores personalizados. Nesse estudo, foram realizados ensaios em que o tamanho médio do *amplicon* foi de 178pb, para garantir que as condições dos ensaios de *PCR* estivessem próximas do ideal. Sendo

assim, a utilização de *amplicons* menores poderia aumentar a sensibilidade dos ensaios e quem sabe até ser útil para os pacientes com tumores mais iniciais ou menos invasivos.

Por último, mas não menos importante, a etapa de pré-amplificação convencional das moléculas reportando as alterações estruturais pode ter representado uma etapa limitante, pois possivelmente alguma amostra pode conter um número reduzido de moléculas de *ctDNA*, e nessas concentrações a eficiência de amplificação é reduzida drasticamente e com isso se teria chances de resultado de falso-negativo. No entanto, uma pequena alteração na metodologia pode ser eficiente para minimizar esse problema. A utilização de PCR em emulsão na etapa de pré-amplificação, a mesma tecnologia utilizada no *ddPCR*, que aproxima e confina as moléculas alvo dentro de gotas, o que torna a amplificação de moléculas bastante diluídas muito mais eficiente.

6 CONCLUSÃO



6 CONCLUSÃO

Nesse trabalho foi possível verificar que a identificação de biomarcadores personalizados, baseados em variações estruturais específicas do tumor de cada paciente, para detectar o *DNA* tumoral circulante (*ctDNA*) em amostras de plasma de pacientes com câncer de reto T3 submetidos a QRTn, tem grande potencial para aplicação clínica. O tempo necessário para o sequenciamento e identificação dos biomarcadores é compatível com a necessidade clínica, podendo ser realizado concomitantemente ao tratamento e intervalo de descanso, para ser utilizado no momento da avaliação de resposta e durante o seguimento dos pacientes. O custo atual dessa abordagem é praticável e sua eficiência pode justificar o investimento, já que os exames de imagem atualmente utilizados na clínica para o monitoramento desses pacientes envolvem um custo significativo.

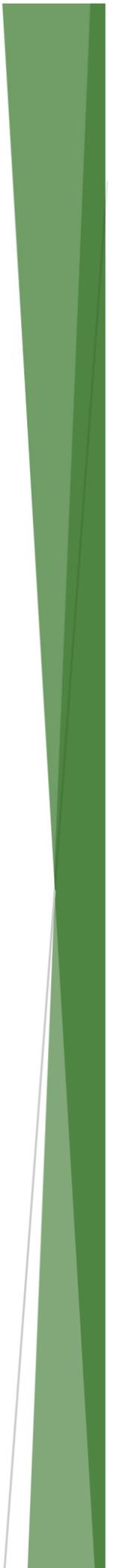
A utilidade dos biomarcadores personalizados no manejo de pacientes com câncer de reto localmente avançados, com nível de invasão T3, foi comprovada e, futuramente, pode auxiliar a implementação da abordagem *Watch and Wait* na rotina clínica. Essa abordagem molecular fornece uma alternativa mais sensível para o seguimento dos pacientes, através de um método não invasivo, que pode prover mais segurança à equipe médica durante o período de acompanhamento frequente desses pacientes.

Em suma, esse estudo demonstrou que biomarcadores personalizados baseados em *ctDNA* têm especificidade adequada para a detecção da doença em pacientes com câncer de reto e apresentou sensibilidade satisfatória para auxiliar na identificação da

resposta ao tratamento de QRTn. No entanto, os biomarcadores não foram eficientes para detectar a doença residual mínima. Além disso, esses resultados mostraram que os biomarcadores personalizados podem ser uma ferramenta útil para detecção precoce de recorrência sistêmica da doença, se mostrando mais sensível que os exames clínicos atualmente utilizados na rotina.

Este estudo, embora de carácter exploratório, mostrou dados relevantes e suficientes para justificar a realização de estudos adicionais para avaliar a utilidade do uso de biomarcadores personalizados na definição da conduta cirúrgica e no acompanhamento de pacientes com câncer de reto tratados com QRTn. Os estudos adicionais devem priorizar a inclusão de um maior número de pacientes, além de contemplar ajustes na metodologia para otimizar a sensibilidade do teste e, conseqüentemente, aumentar as taxas de detecção de doença residual mínima.

REFERÊNCIAS



REFERÊNCIAS

ALBERTSON, D. G.; COLLINS, C.; MCCORMICK, F.; GRAY, J. W. **Chromosome aberrations in solid tumors.** *Nat Genet*, v. 34, n. 4, p. 369-76, Aug 2003.

ALIX-PANABIERES, C.; SCHWARZENBACH, H.; PANTEL, K. **Circulating tumor cells and circulating tumor DNA.** *Annu Rev Med*, v. 63, p. 199-215, 2012.

BERGAMASCHI, D.; SAMUELS, Y.; JIN, B.; DURAISINGHAM, S. *et al.* **ASPP1 and ASPP2: common activators of p53 family members.** *Mol Cell Biol*, v. 24, n. 3, p. 1341-50, Feb 2004.

BERNE, R. M. K. B. M. S. B. A. *Berne & Levy physiology*. Philadelphia, PA: Mosby/Elsevier, 2010.

BETTEGOWDA, C.; SAUSEN, M.; LEARY, R. J.; KINDE, I. *et al.* **Detection of circulating tumor DNA in early- and late-stage human malignancies.** *Sci Transl Med*, v. 6, n. 224, p. 224ra24, Feb 19 2014.

CHANG, C. P. Y.; CHIA, R.-H.; WU, T.-L.; TSAO, K.-C. *et al.* **Elevated cell-free serum DNA detected in patients with myocardial infarction.** *Clinica Chimica Acta*, v. 327, n. 1-2, p. 95-101, 2003.

CHENG, X.; CHEN, V. W.; STEELE, B.; RUIZ, B. *et al.* **Subsite-specific incidence rate and stage of disease in colorectal cancer by race, gender, and age group in the United States, 1992-1997.** *Cancer*, v. 92, n. 10, p. 2547-54, Nov 15 2001.

COLLETTE, L.; BOSSET, J. F.; DEN DULK, M.; NGUYEN, F. *et al.* **Patients with curative resection of cT3-4 rectal cancer after preoperative radiotherapy or radiochemotherapy: does anybody benefit from adjuvant fluorouracil-based chemotherapy? A trial of the European Organisation for Research and Treatment of Cancer Radiation Oncology Group.** *J Clin Oncol*, v. 25, n. 28, p. 4379-86, Oct 1 2007.

CROWLEY, E.; DI NICOLANTONIO, F.; LOUPAKIS, F.; BARDELLI, A. **Liquid biopsy: monitoring cancer-genetics in the blood.** *Nat Rev Clin Oncol*, v. 10, n. 8, p. 472-84, Aug 2013.

DAWSON, S. J.; TSUI, D. W.; MURTAZA, M.; BIGGS, H. *et al.* **Analysis of circulating tumor DNA to monitor metastatic breast cancer.** *N Engl J Med*, v. 368, n. 13, p. 1199-209, Mar 28 2013.

DEVITA, V. T.; LAWRENCE, T. S.; ROSENBERG, S. A. *DeVita, Hellman, and Rosenberg's Cancer: Principles & Practice of Oncology*. Wolters Kluwer/Lippincott Williams & Wilkins, 2008. (v. 2).

DEVOS, T.; TETZNER, R.; MODEL, F.; WEISS, G. *et al.* **Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer.** *Clin Chem*, v. 55, n. 7, p. 1337-46, Jul 2009.

DIEHL, F.; LI, M.; DRESSMAN, D.; HE, Y. *et al.* **Detection and quantification of mutations in the plasma of patients with colorectal tumors.** *Proc Natl Acad Sci U S A*, v. 102, n. 45, p. 16368-73, Nov 8 2005.

DIEHL, F.; SCHMIDT, K.; CHOTI, M. A.; ROMANS, K. *et al.* **Circulating mutant DNA to assess tumor dynamics.** *Nat Med*, v. 14, n. 9, p. 985-90, Sep 2008.

DIEHL, F.; SCHMIDT, K.; DURKEE, K. H.; MOORE, K. J. *et al.* **Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients.** *Gastroenterology*, v. 135, n. 2, p. 489-98, Aug 2008.

DONNARD, E. R.; CARPINETTI, P. A.; NAVARRO, F. C.; PEREZ, R. O. *et al.* **ICRmax: An optimized approach to detect tumor-specific interchromosomal rearrangements for clinical application.** *Genomics*, Feb 7 2015.

DRIER, Y.; LAWRENCE, M. S.; CARTER, S. L.; STEWART, C. *et al.* **Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability.** *Genome Res*, v. 23, n. 2, p. 228-35, Feb 2013.

DWORAK, O.; KEILHOLZ, L.; HOFFMANN, A. **Pathological features of rectal cancer after preoperative radiochemotherapy.** *Int J Colorectal Dis*, v. 12, n. 1, p. 19-23, 1997.

EMMERTSEN, K. J.; LAURBERG, S. **Impact of bowel dysfunction on quality of life after sphincter-preserving resection for rectal cancer.** *Br J Surg*, v. 100, n. 10, p. 1377-87, Sep 2013.

ESPOSITO, A.; BARDELLI, A.; CRISCITIELLO, C.; COLOMBO, N. *et al.* **Monitoring tumor-derived cell-free DNA in patients with solid tumors: clinical perspectives and research opportunities.** *Cancer Treat Rev*, v. 40, n. 5, p. 648-55, Jun 2014.

FAKIH, M. G.; PADMANABHAN, A. **CEA monitoring in colorectal cancer. What you should know.** *Oncology (Williston Park)*, v. 20, n. 6, p. 579-87; discussion 588, 594, 596 passim, May 2006.

FERLAY, J.; SOERJOMATARAM, I.; DIKSHIT, R.; ESER, S. *et al.* **Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012.** *Int J Cancer*, Sep 13 2014.

GAHAN, P. B. **Circulating nucleic acids in plasma and serum: roles in diagnosis and prognosis in diabetes and cancer.** *Infect Disord Drug Targets*, v. 8, n. 2, p. 100-8, Jun 2008.

GALEAZZI, M.; MOROZZI, G.; PICCINI, M.; CHEN, J. *et al.* **Dosage and characterization of circulating DNA: present usage and possible applications in systemic autoimmune disorders.** *Autoimmunity Reviews*, v. 2, n. 1, p. 50-55, 2003.

GARCIA-AGUILAR, J.; CHEN, Z.; SMITH, D. D.; LI, W. *et al.* **Identification of a biomarker profile associated with resistance to neoadjuvant chemoradiation therapy in rectal cancer.** *Ann Surg*, v. 254, n. 3, p. 486-92; discussion 492-3, Sep 2011.

GLYNNE-JONES, R.; HUGHES, R. **Critical appraisal of the 'wait and see' approach in rectal cancer for clinical complete responders after chemoradiation.** *Br J Surg*, v. 99, n. 7, p. 897-909, Jul 2012.

GLYNNE-JONES, R.; WALLACE, M.; LIVINGSTONE, J. I.; MEYRICK-THOMAS, J. **Complete clinical response after preoperative chemoradiation in rectal cancer: is a "wait and see" policy justified?** *Dis Colon Rectum*, v. 51, n. 1, p. 10-9; discussion 19-20, Jan 2008.

HABR-GAMA, A.; DE SOUZA, P. M.; RIBEIRO, U., JR.; NADALIN, W. *et al.* **Low rectal cancer: impact of radiation and chemotherapy on surgical treatment.** *Dis Colon Rectum*, v. 41, n. 9, p. 1087-96, Sep 1998.

HABR-GAMA, A.; PEREZ, R. O.; NADALIN, W.; SABBAGA, J. *et al.* **Operative versus nonoperative treatment for stage 0 distal rectal cancer following chemoradiation therapy: long-term results.** *Ann Surg*, v. 240, n. 4, p. 711-7; discussion 717-8, Oct 2004.

HABR-GAMA, A.; PEREZ, R. O.; SABBAGA, J.; NADALIN, W. *et al.* **Increasing the rates of complete response to neoadjuvant chemoradiotherapy for distal rectal cancer: results of a prospective study using additional chemotherapy during the resting period.** *Dis Colon Rectum*, v. 52, n. 12, p. 1927-34, Dec 2009.

HABR-GAMA, A.; PEREZ, R. O.; WYNN, G.; MARKS, J. *et al.* **Complete clinical response after neoadjuvant chemoradiation therapy for distal rectal cancer: characterization of clinical and endoscopic findings for standardization.** *Dis Colon Rectum*, v. 53, n. 12, p. 1692-8, Dec 2010.

HABR-GAMA, A.; SABBAGA, J.; GAMA-RODRIGUES, J.; SAO JULIAO, G. P. *et al.* **Watch and wait approach following extended neoadjuvant chemoradiation for distal rectal cancer: are we getting closer to anal cancer management?** *Dis Colon Rectum*, v. 56, n. 10, p. 1109-17, Oct 2013.

INCA. Estimativas de Incidência e Mortalidade de Câncer no Brasil. *Base de dados na internet*: Rio de Janeiro: Instituto Nacional do Câncer, 2014.

KOSINSKI, L.; HABR-GAMA, A.; LUDWIG, K.; PEREZ, R. **Shifting concepts in rectal cancer management: a review of contemporary primary rectal cancer treatment strategies.** *CA Cancer J Clin*, v. 62, n. 3, p. 173-202, May-Jun 2012.

KROOK, J. E.; MOERTEL, C. G.; GUNDERSON, L. L.; WIEAND, H. S. *et al.* **Effective surgical adjuvant therapy for high-risk rectal carcinoma.** *N Engl J Med*, v. 324, n. 11, p. 709-15, Mar 14 1991.

LEARY, R. J.; KINDE, I.; DIEHL, F.; SCHMIDT, K. *et al.* **Development of personalized tumor biomarkers using massively parallel sequencing.** *Sci Transl Med*, v. 2, n. 20, p. 20ra14, Feb 24 2010.

LENGAUER, C.; KINZLER, K. W.; VOGELSTEIN, B. **Genetic instabilities in human cancers.** *Nature*, v. 396, n. 6712, p. 643-9, Dec 17 1998.

LEON, S. A.; SHAPIRO, B.; SKLAROFF, D. M.; YAROS, M. J. **Free DNA in the serum of cancer patients and the effect of therapy.** *Cancer Res*, v. 37, n. 3, p. 646-50, Mar 1977.

LIANG, P. S.; CHEN, T.-Y.; GIOVANNUCCI, E. **Cigarette smoking and colorectal cancer incidence and mortality: Systematic review and meta-analysis.** *International Journal of Cancer*, v. 124, n. 10, p. 2406-2415, 2009.

LIM, S. H.; BECKER, T. M.; CHUA, W.; CAIXEIRO, N. J. *et al.* **Circulating tumour cells and circulating free nucleic acid as prognostic and predictive biomarkers in colorectal cancer.** *Cancer Lett*, v. 346, n. 1, p. 24-33, Apr 28 2014.

LO, Y. M.; CORBETTA, N.; CHAMBERLAIN, P. F.; RAI, V. *et al.* **Presence of fetal DNA in maternal plasma and serum.** *Lancet*, v. 350, n. 9076, p. 485-7, Aug 16 1997.

LO, Y. M. D.; RAINER, T. H.; CHAN, L. Y. S.; HJELM, N. M. *et al.* **Plasma DNA as a Prognostic Marker in Trauma Patients.** *Clinical Chemistry*, v. 46, n. 3, p. 319-323, March 1, 2000 2000.

LOTFI, A. M.; SPENCER, R. J.; ILSTRUP, D. M.; MELTON, L. J., 3RD. **Colorectal polyps and the risk of subsequent carcinoma.** *Mayo Clin Proc*, v. 61, n. 5, p. 337-43, May 1986.

LYNCH, H. T.; DE LA CHAPELLE, A. **Hereditary colorectal cancer.** *N Engl J Med*, v. 348, n. 10, p. 919-32, Mar 6 2003.

MAAS, M.; BEETS-TAN, R. G.; LAMBREGTS, D. M.; LAMMERING, G. *et al.* **Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer.** *J Clin Oncol*, v. 29, n. 35, p. 4633-40, Dec 10 2011.

MAAS, M.; NELEMANS, P. J.; VALENTINI, V.; DAS, P. *et al.* **Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data.** *Lancet Oncol*, v. 11, n. 9, p. 835-44, Sep 2010.

MANDEL, P.; METAIS, P. **Les acides nucleiques du plasma sanguine chez l'homme.** *C. R. Acad. Sci. Paris*, v. 142, n. 3-4, p. 241-3, Feb 1948.

MCBRIDE, D. J.; ORPANA, A. K.; SOTIRIOU, C.; JOENSUU, H. *et al.* **Use of cancer-specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors.** *Genes Chromosomes Cancer*, v. 49, n. 11, p. 1062-9, Nov 2010.

MEACHAM, C. E.; MORRISON, S. J. **Tumour heterogeneity and cancer cell plasticity.** *Nature*, v. 501, n. 7467, p. 328-37, Sep 19 2013.

MEYERSON, M.; GABRIEL, S.; GETZ, G. **Advances in understanding cancer genomes through second-generation sequencing.** *Nat Rev Genet*, v. 11, n. 10, p. 685-96, Oct 2010.

MOREIRA, E. R. D. *Estudo de variações genômicas para a identificação de biomarcadores personalizados e novos alvos terapêuticos em tumores colorretais.* (2014). 109 f. (1) - Tese – Programa de Pós-graduação em Ciência Biológicas (Bioquímica), Instituto de Química da Universidade de São Paulo, São Paulo., 2014.

MOSER, L.; RITZ, J. P.; HINKELBEIN, W.; HOCHT, S. **Adjuvant and neoadjuvant chemoradiation or radiotherapy in rectal cancer--a review focusing on open questions.** *Int J Colorectal Dis*, v. 23, n. 3, p. 227-36, Mar 2008.

MOULIERE, F.; ROBERT, B.; ARNAU PEYROTTE, E.; DEL RIO, M. *et al.* **High fragmentation characterizes tumour-derived circulating DNA.** *PLoS One*, v. 6, n. 9, p. e23418, 2011.

NAWROZ, H.; KOCH, W.; ANKER, P.; STROUN, M. *et al.* **Microsatellite alterations in serum DNA of head and neck cancer patients.** *Nat Med*, v. 2, n. 9, p. 1035-7, Sep 1996.

NEGRINI, S.; GORGOULIS, V. G.; HALAZONETIS, T. D. **Genomic instability--an evolving hallmark of cancer.** *Nat Rev Mol Cell Biol*, v. 11, n. 3, p. 220-8, Mar 2010.

NEWTON, K. F.; NEWMAN, W.; HILL, J. **Review of biomarkers in colorectal cancer.** *Colorectal Dis*, v. 14, n. 1, p. 3-17, Jan 2012.

PAUN, B. C.; CASSIE, S.; MACLEAN, A. R.; DIXON, E. *et al.* **Postoperative complications following surgery for rectal cancer.** *Ann Surg*, v. 251, n. 5, p. 807-18, May 2010.

RAGO, C.; HUSO, D. L.; DIEHL, F.; KARIM, B. *et al.* **Serial assessment of human tumor burdens in mice by the analysis of circulating DNA.** *Cancer Res*, v. 67, n. 19, p. 9364-70, Oct 1 2007.

ROH, M. S.; COLANGELO, L. H.; O'CONNELL, M. J.; YOTHERS, G. *et al.* **Preoperative multimodality therapy improves disease-free survival in patients with carcinoma of the rectum: NSABP R-03.** *J Clin Oncol*, v. 27, n. 31, p. 5124-30, Nov 1 2009.

ROUKOS, D. H. **Next-generation, genome sequencing-based biomarkers: concerns and challenges for medical practice.** *Biomark Med*, v. 4, n. 4, p. 583-6, Aug 2010.

SAMUELS-LEV, Y.; O'CONNOR, D. J.; BERGAMASCHI, D.; TRIGIANTE, G. *et al.* **ASPP proteins specifically stimulate the apoptotic function of p53.** *Mol Cell*, v. 8, n. 4, p. 781-94, Oct 2001.

SAUER, R.; BECKER, H.; HOHENBERGER, W.; RODEL, C. *et al.* **Preoperative versus postoperative chemoradiotherapy for rectal cancer.** *N Engl J Med*, v. 351, n. 17, p. 1731-40, Oct 21 2004.

SCHWARZENBACH, H.; HOON, D. S.; PANTEL, K. **Cell-free nucleic acids as biomarkers in cancer patients.** *Nat Rev Cancer*, v. 11, n. 6, p. 426-37, Jun 2011.

SGROI, D. C.; TENG, S.; ROBINSON, G.; LEVANGIE, R. *et al.* **In vivo gene expression profile analysis of human breast cancer progression.** *Cancer Res*, v. 59, n. 22, p. 5656-61, Nov 15 1999.

SMITH, F. M.; WINTER, D. **Watch and wait for rectal cancer: where are we really at?** *Colorectal Dis*, v. 16, n. 5, p. 332-4, May 2014.

SMITH, J. D.; RUBY, J. A.; GOODMAN, K. A.; SALTZ, L. B. *et al.* **Nonoperative management of rectal cancer with complete clinical response after neoadjuvant therapy.** *Ann Surg*, v. 256, n. 6, p. 965-72, Dec 2012.

SOBIN, L. H.; GOSPODAROWICZ, M. K.; WITTEKIND, C. *TNM Classification of Malignant Tumours*. Wiley, 2009.

SOCIETY, A. C. Colorectal Cancer Prevention and Early Detection n. acessado em 27 de setembro de 20142014.

SORENSEN, G. D.; PRIBISH, D. M.; VALONE, F. H.; MEMOLI, V. A. *et al.* **Soluble normal and mutated DNA sequences from single-copy genes in human blood**. *Cancer Epidemiol Biomarkers Prev*, v. 3, n. 1, p. 67-71, Jan-Feb 1994.

TCGA. **Comprehensive molecular characterization of human colon and rectal cancer**. *Nature*, v. 487, n. 7407, p. 330-7, Jul 19 2012.

TEN BOSCH, J. R.; GRODY, W. W. **Keeping up with the next generation: massively parallel sequencing in clinical diagnostics**. *J Mol Diagn*, v. 10, n. 6, p. 484-92, Nov 2008.

THERASSE, P.; ARBUCK, S. G.; EISENHAUER, E. A.; WANDERS, J. *et al.* **New guidelines to evaluate the response to treatment in solid tumors**. **European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada**. *J Natl Cancer Inst*, v. 92, n. 3, p. 205-16, Feb 2 2000.

THIERRY, A. R.; MOULIERE, F.; GONGORA, C.; OLLIER, J. *et al.* **Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts**. *Nucleic Acids Res*, v. 38, n. 18, p. 6159-75, Oct 2010.

THOMAS, P. R.; LINDBLAD, A. S. **Adjuvant postoperative radiotherapy and chemotherapy in rectal carcinoma: a review of the Gastrointestinal Tumor Study Group experience**. *Radiother Oncol*, v. 13, n. 4, p. 245-52, Dec 1988.

TULCHINSKY, H.; SHMUELI, E.; FIGER, A.; KLAUSNER, J. M. *et al.* **An interval >7 weeks between neoadjuvant therapy and surgery improves pathologic complete response and disease-free survival in patients with locally advanced rectal cancer**. *Ann Surg Oncol*, v. 15, n. 10, p. 2661-7, Oct 2008.

VIRONEN, J. H.; KAIRALUOMA, M.; AALTO, A. M.; KELLOKUMPU, I. H. **Impact of functional results on quality of life after rectal cancer surgery**. *Dis Colon Rectum*, v. 49, n. 5, p. 568-78, May 2006.

VOELKERDING, K. V.; DAMES, S. A.; DURTSCHI, J. D. **Next-generation sequencing: from basic research to diagnostics.** *Clin Chem*, v. 55, n. 4, p. 641-58, Apr 2009.

VOGELSTEIN, B.; PAPADOPOULOS, N.; VELCULESCU, V. E.; ZHOU, S. *et al.* **Cancer genome landscapes.** *Science*, v. 339, n. 6127, p. 1546-58, Mar 29 2013.

WALSH, J. E.; TERDIMAN, J. P. **Colorectal cancer screening: Scientific review.** *JAMA*, v. 289, n. 10, p. 1288-1296, 2003.

WILLETT, W. C.; STAMPFER, M. J.; COLDITZ, G. A.; ROSNER, B. A. *et al.* **Relation of meat, fat, and fiber intake to the risk of colon cancer in a prospective study among women.** *N Engl J Med*, v. 323, n. 24, p. 1664-72, Dec 13 1990.

WU, A. H.; PAGANINI-HILL, A.; ROSS, R. K.; HENDERSON, B. E. **Alcohol, physical activity and other risk factors for colorectal cancer: a prospective study.** *Br J Cancer*, v. 55, n. 6, p. 687-94, Jun 1987.

APÊNDICES



APÊNDICE A – SEQUÊNCIAS DOS OLIGONUCLEOTÍDEOS (INICIADORES E SONDAS)

Quadro 3 – Sequências dos oligonucleotídeos utilizados para a validação dos candidatos a VE.

Paciente	VE	Sequências	
		Iniciador 1 (5'-->3')	Iniciador 2 (5'-->3')
CCR1	T01	CAAAGCCCCTGAATGGTTAC	TGGCCTACAGATCAAATCCAG
CCR1	T02	AAGCATTGGTGAAGGTAATCAG	TGGGCCTCACATAATCCTTC
CCR1	T03	GTAGGCATTGCTCCTTCAACC	TGTAGCTGTGTGAGGTGGTTG
CCR1	T04	GTCCGGTCATGGCTAAATTG	TTGCGTGTGGATCTGTAACC
CCR1	T05	CTCAGCCACCATAAAATTGTGC	CAGGCATGTGTGTCAAGGAC
CCR1	I01	AGGGCAATTATACCAATGTGC	GGAGACTGAAACCCATCCTG
CCR1	I02	AAGTACTGGCCTGCCATTTG	GCAGATTATCAGGCCAAAC
CCR1	D03	TTCTGCAGTTGAAAGCCAATTC	TCTATTGCTACTACCAACATTGC
CCR1	D04	GCCTCATGCAGTCTTGGTG	CAATCTGTCCGACAATGAGC
CCR1	D05	GGAGCCCTCACCAATGTAG	GTTGGGGCATAAAAAGATTCTG
CCR1	D06	TCATCCAACCTCCAGTCTACGC	TTCACCTCCCTTGAAACAC
CCR1	D07	GATGATCTTGACCCTATGTAGCC	AATAAACACCAAATCCCAATGTC
CCR1	D08	AGAACGCTGGAAGATTGGAG	AGGAACCCCAAGACAGTTC
CCR1	D09	CAGGGGTGAAAATAGCAACG	GCTGCATACACCCATTG
CCR1	D10	CCACAGCATCATATGTCAGAC	CCACAAACCCCAAGCTAGAC
CCR2	I01	CCATTTCTAGCCAAACCTG	GCATCCACTAGGGGTCTTAG
CCR2	D02	CCTTTCCCGTTACACACCAC	GGCCAGTCTGAATGAGGATG
CCR2	I03	CTGGGAAATGCAGGTGTAGG	ATGTTGGAGGCCTTTCTTCC
CCR2	T01	CAGCATGCTGATGAGGCTTG	CGAGAATTACTGCCATGAAGC
CCR2	T02	ACGATGTACCCTCAAGTCC	TCTGCCAGATAAGCTACAGG
CCR2	T03	TTCGAAAACGGTTGTTTGTGG	GGAGCCATGGCAGAAGAAC
CCR3	T01	GATCTATTGTGAGCATCTCATA	ATGGGAAAGGCAGATCACTG
CCR3	T06	CCAACACTTTTGCTAGGTCAAC	TAGGGAAATGCCAAGACAGG
CCR3	T10	TCATAGGCCTCGGTATTTCC	GAAATTCAGCTGGGATCTG
CCR3	T11	CGTGGTTCCTTACCTGCAC	GAGCCTGCTTGTGAGTTCC
CCR3	T19	GAGTGAGCACCTGGGTATGG	GCTTTCCAATTCCTCACACC
CCR3	T21	ACACTCCCTGCCTACCCTTC	ATATCGCCTTCCACTGAAGC
CCR3	T22	CAGGTGAGAGTGGAAGCTC	ATCCTCTGGGAAGGTAGG
CCR3	T23	CCTGAGGAGCTTTGGACAC	GTGGTGGTTACATGGGTGAC
CCR3	T24	ATGGCCTAATCGTCATCCAG	TCAGGACAATGCTCCACTTG
CCR3	T27	TTCTCACTTTCACGCAGCAG	TTGGAATCATTATCTAGAAGGG
CCR3	T31	GGACACGTTCTGCAGTCTC	CTCAGCCAGAGAACTCAC
CCR3	T34	GACTCCTTGTGGCAGTTTCC	CCTTAGGCACGCTTCAAATC
CCR3	T35	GGCTGAAGCAAGGACTGTTC	CAACAGGTGAGCACAGCAAG
CCR3	T37	CAATCACGGACTCACAATGC	CCTGGGATCTAGCAGATGCT
CCR3	T42	AAACATTGGTCTCCTGCAC	CATCTCAAGGTGTGCAAGTAGC
CCR3	T60	CGGTGAAAGAACCAGAATGC	CAAGAAAGCTAGCCAGGATTG
CCR3	T62	TCCCATGTCCTTACGAGTCC	TGGAGAAAAGCGGTAGGATG

Continua

Continuação

Paciente	VE	Sequências	
		Iniciador 1 (5'-->3')	Iniciador 2 (5'-->3')
CCR3	T63	CAACATCCTCAGGCCATCTC	TGGGTGCTAGACGATGTGTG
CCR3	D01	TGGGAAGAAGAGCTCAGTGC	GAGCACTCAAGGCACCTGAAG
CCR3	D02	AGTGTTCCTGGCTCTGATG	ATGGGCGGAGCTTGTATAG
CCR3	D03	GTGACCCTAAAGCAAACTGC	CGCGAATTTGGCCAGTCAC
CCR3	D04	AGAGCTCACATAGCAGATTG	AGTAGCCAACCATTCACC
CCR3	D05	CTCCTGCTCCTTCACTGG	TCATTCCGACTATTGCCTTGG
CCR3	D07	GGAGAACAAGAGAAACCCTTGG	TTCTCAGAGGAGGGGTGATG
CCR3	D10	CAAGCAACCCAACCTGAGTC	GGAGGTGTGGTTGGAGAATGG
CCR3	D11	ATCTTCCTAGCACTGCGTCTG	AATTATCATTAGCTGCTCCTTCC
CCR3	D13	GAGTAGAAGCCTCCAAAACG	TCCTTGACTTCTTGTGTCTGG
CCR3	D15	CACAGAGGTACGAAGACACAC	CTGCCAAGAGGTACAGAGAC
PCR1	T01	CAATTGCGTATCTGCAAAGAG	ATTACGCAGCAGCATGTGAC
PCR1	T02	GTTTAGATGCGCAGGGAGAG	CTGGAAGCCCTAAGTGTGG
PCR1	T03	TGATTTTGGTGACAGTATCATGG	TGGCTAGTAGATCTCTCTCC
IR1	T01	GGCTCCTTGCCGTATCAC	AAGAGTGCAGTTGAAGTCCTG
IR1	T02	CACTAATGGGAAACCGTTG	TGCCCTCCAATTAACCATTG
IR1	T03	TGAATCCACCGCAAACGAG	GGTACAGTCCATTGTTGAG
IR1	T05	GAGGTAATGGCCAGGTAAG	CTTCTTCCTTTAAGCCGATGC
IR1	T06	CTGTAGTTCCCAATCCTG	CTCTCCTGTGGAAGTGAATG
IR1	T07	AAGGAGCAAACAAGCACCTC	CTCCTGACTCCTGAAAGCTG
IR1	T08	TGCATTGGAGGCTAATCACATG	GGCAAGAACAAGAGGCAAAC
IR1	T11	ATTTCTGGGCATTCTGAGC	CTGCAATTCACGTCATCTC
IR1	T12	ATATCGCCCATTGCAATCTC	AGTGCAGTGCTTGTACAG
IR1	T13	TGTCAAGTGGAAAGGTAAGC	GAGGTCAGCAGGTCATCC
IR2	T01	CTTTCACGTGGAACCAAGTTG	ACCTCTGGGCACTGAAAAG
IR2	T03	TTCCAAAATATACCCAATTAAGT	TGTGGTGTGTTGTCTTCAAGTGG
IR2	T04	TTCACAGAGCACAAGGAGATCC	GCCAGTAATTGGGTATATTTTGG
IR2	D01	AACCACATTGTTGAGGAGAGCC	ACGCTTGGACTATTGGCTTCC
IR2	D02	CTCAGCTCTGGACATCTATGC	ACGTCCTGTCCCAAGAAC
IR2	D04	AATCCTCCACCAATTCATAAGC	AGTTGTTTATAGGGTTGCAATGG
IR2	D06	CAGCCAGCAGTTTTCTGTTCC	AGAATGTTCTGGTCTGGTCTCG
IR2	D07	GAAAACCTGCATTGCATGTGGTC	TAACATGCCCTGTTGTCTGAG
IR2	D09	GTGGACTGCTGCGGGTAAC	TTCCTCTGTACCTGGCCATTG
IR2	D11	CTCATCTGGAGCCTTGTTC	AGACAGGAAAGGACACCAAGC

Relação dos iniciadores utilizados na validação, por PCR e sequenciamento pelo método de Sanger, dos candidatos a variação estrutural para todos os pacientes.

Quadro 4 – Sequências dos oligonucleotídeos utilizados para a detecção do *DNA* tumoral circulante (*ctDNA*) no plasma dos pacientes.

Paciente	VE	Sequências dos iniciadores externos		<i>amplicon</i> AB	Sequências dos iniciadores internos		<i>amplicon</i> CD	Sonda MGB
		iniciador A (5'-->3')	iniciador B (5'-->3')		iniciador C (5'-->3')	iniciador D (5'-->3')		
CCR1	T01	GCTTGGGATTCACCTCTCTGG	ACAGCTACACCCCTGCATATTG	168pb	TTCTGGTACATAGCCTTGTGCATG	GCTACAATGGCAAAGTGGAGCA	87pb	TATCCAGGTGATTCTGT
CCR1	I01	GGGGTCTTTTCTAAAATAGGG	AGCAGCTGTCAAGGGTACATAAG	235pb	GGAAGGGCAATTATACCAAATGTG	TAAGGGCTGTTTCTTCATTTGACC	143pb	AAAGTGCTTTGTGCTCAT
CCR1	I02	TCCAGAAGTAATCAAGGCAGAAG	TCTATTGGCCTAAGGATCAAGG	171pb	AAGTACTGGCCTGCATTGTC	AGGGTGTGCCACTTTGTAGG	126pb	AGATATCAGCAGGACCTT
CCR2	T01	TGTAAGGCCAGGTGTTACAGAAGG	TTCTCCTCTTCTTCATTGCCTGAG	202pb	AGAAATGCCAAAAGCACCACAG	TGCCTGAGTAGCATTATCTGCAG	132pb	TCCTGACTAAAGATAGTG
CCR2	I01	CTCACACAGTTTTCTGTTGTTGG	AAACACAGCATATATAGGGTCCAG	229pb	ATGCGCCAATCAGCACTCTG	GGTCCAGCAGTATCCATGATTTCA	154pb	CTAGAGTTGGATAAACT
CCR2	D02	CATTTGCCACGCTTTGTGAC	GTCTCAAACCTCTCTGCTCAAG	143pb	GACAGGCCCTTTCCCGTTAC	CCTCAGCCTCCCAAATTTGTTG	93pb	ACACCTGCCCTGTGAA
CCR3	R01	CAGAGCGACAACCAGGAGAATG	GCAGCAGGTACTGAAAACAGCAG	145pb	CAGAGCGACAACCAGGAGAATG	GGAAGACCTGAGGGCAACTTC	106pb	CTTTGCTTGTGATATTTG
CCR3	D01	TGGGAGGATTGGACAAGAGC	TAATCCATGGCCTTCACCAGTC	160pb	AAGAAGAGCTCAGTGCACACAC	CATGGCCTTCACCAGTCAGC	155pb	AGGACCTTTGCACTAGA
CCR3	D02	AGTGTCCCTGGCTCTGATG	ACCTGCAAATTAAGGAAATACCG	220pb	GCAGCAGGTTACAGTGGAGCA	CCTAAACCATTAGCTGGCCTGTC	110pb	TCTGTTAGTTAAGTCCT
PCR1	T01	GACTCAGAATGTATTAGTCG	ATTACGCAGCAGCATGTGAC	159pb	TGGATGTGAACCTCTGGTGTAGG	ACGCAGCAGCATGTGACTGTAG	130pb	TACCAGGCAAAGTGG
PCR1	T02	AGTTTAGATGCGCAGGGAGAG	ACTACAGTACTCCTGAGAATGC	153pb	GCTGTCCCTGCCACAATTATGTCA	CTCCTGAGAATGCACTGAAGAGAAG	113pb	ACGGCACAGTATTGG
PCR1	T03	AACTAGCCATTGGGTGTTGC	TGGCTAGTAGATCTCTCTCC	262pb	CTTCTTAGCAGCCCCAACATTTTC	CCCTACACCAATCAACACCACAT	178pb	AAACATCATgtactttcc
IR1	T01	GCCAGAAGGTCCCAGACTG	AAGAGTGCAGTTGAAGTCTCTG	115pb	GAGTATACAGAAATGGGTGAGTGC	TGCAGTTGAAGTCTGGATTAAGG	80pb	CAGCAATGTGAAGTAAA
IR1	T02	AGGGAAGAGCTCATGCCAGC	GTTTGATCTTGACAATGGTGTCTG	143pb	ATCATTCTTAGCATCAGCTATGC	TGACAATGGTGTGTGTAACGTC	99pb	CCATGAgagaGAGAAG
IR1	T03	AACAACAACAATCCTCTTGCCC	AGATATTATTGGTGGGCTTCTGCT	180pb	CAACAATCCTCTTGCCCTAATTC	AGCCAGACATCATCTCAGCTCAG	106pb	AGCAGTCAATTGTGC
IR2	T01	CCTCCTTTGTTATGACGCTCATC	TTGAACAGTTGGTTTTAGTGGGTC	200pb	GCAAGTTTAGACAACGTAGGCCAT	ATTGAACAGTTGGTTTTAGTGGGTC	96pb	AAGGTACAAAATAAACAG
IR2	D01	AACCACATTGTTGAGGAGAGCC	GAATTTTGTCTGGGCATGAGC	133pb	AGAGCCATACAATACAGTCAGG	CTGGGCATGAGCATGACTGTAC	107pb	TCTGGTCCATCTTTC
IR2	D02	TCAATCAAATGGGCCTTTCAC	GACGTCTGTCCCAAGAAC	179pb	CTCAGCTCTGGACATCTCTATGC	ACAGTCACTGCAGATGTGATTGC	151pb	TCAGCATGTACCTTGTG

Tabela apresenta as sequências dos iniciadores externos (pré-amplificação), internos e sondas (ddPCR) utilizados para a detecção do *DNA* tumoral circulante no plasma. (**VE**) Variação Estrutural; (**amplicon**) tamanho do fragmento amplificado na reação de PCR; (**Sonda MGB**) seqüência complementar a região do ponto de quebra da variação estrutural. As letras minúsculas são inserções exclusivas da VE, sendo assim não são encontradas na seqüência referência das regiões cromossômicas envolvidas no rearranjo cromossômico.

APÊNDICE B – DETALHAMENTO DAS VALIDAÇÕES PARA TODOS OS PACIENTES

Quadro 5 – Relação de todos os candidatos a VE e detalhamento das validações

Paciente	VE	Suporte	Regiões cromossômicas				amplicon	validação
			Cro A	Posição A	Cro B	Posição B		
CCR1	T01	3	19	30473830	17	50154848	527pb	VERDADEIRO
CCR1	T02	3	12	-113887027	5	138654666	-	FALSO
CCR1	T03	10	20	21784409	2	21662554	-	FALSO
CCR1	T04	12	2	21662806	20	21784766	-	FALSO
CCR1	T05	2	4	-62950570	20	40588404	-	FALSO
CCR1	I01	11	8	123281119	8	-125821045	296pb	VERDADEIRO
CCR1	I02	16	8	-123281186	8	125821198	278pb	VERDADEIRO
CCR1	D03	8	5	19956445	5	19964837	-	FALSO
CCR1	D04	5	6	129737619	6	129747360	-	FALSO
CCR1	D05	15	8	135059421	8	135068172	-	FALSO
CCR1	D06	8	9	590553	9	608752	-	FALSO
CCR1	D07	4	10	58512976	10	58526973	-	FALSO
CCR1	D08	8	11	20844407	11	20849973	-	FALSO
CCR1	D09	4	13	106149087	13	106154458	-	FALSO
CCR1	D10	8	17	49756611	17	49746630	-	FALSO
CCR2	I01	32	12	1819278	12	-1870423	293pb	VERDADEIRO
CCR2	D02	4	17	39823076	17	39837384	742pb	VERDADEIRO
CCR2	I03	4	17	39113895	17	-39119434	-	FALSO
CCR2	T01	5	3	7252285	5	110411594	657pb	VERDADEIRO
CCR2	T02	4	13	21742463	11	108585788	-	FALSO
CCR2	T03	2	6	-93096822	3	187729086	-	FALSO
CCR3	T01	12	3	-183212528	1	120325297	517pb	VERDADEIRO
CCR3	T02	4	3	33137581	2	-147388834	-	não avaliado
CCR3	T03	3	1	160215419	13	54893995	-	não avaliado
CCR3	T04	4	7	-47240164	9	704159	-	não avaliado
CCR3	T05	3	5	120271940	11	-98538303	-	não avaliado
CCR3	T06	4	11	14081966	9	-135492204	-	FALSO
CCR3	T07	3	3	16816823	9	17237965	-	não avaliado
CCR3	T08	3	5	5070943	2	155538351	-	não avaliado
CCR3	T09	3	12	123518553	5	153569480	-	não avaliado
CCR3	T10	4	7	-15135800	6	50356521	-	FALSO
CCR3	T11	3	17	75554940	7	17287639	-	FALSO
CCR3	T12	4	7	17287961	17	75553984	-	não avaliado

Continua

Continuação

Paciente	VE	Suporte	Regiões cromossômicas				amplicon	validação
			Cro A	Posição A	Cro B	Posição B		
CCR3	T13	3	6	121159297	5	-72294542	-	não avaliado
CCR3	T14	3	6	-30347911	13	30930169	-	não avaliado
CCR3	T15	3	19	-1162677	11	8350478	-	não avaliado
CCR3	T16	4	14	75531533	20	-32057936	-	não avaliado
CCR3	T17	3	20	-32057636	14	75530586	-	não avaliado
CCR3	T18	3	7	96743820	6	-162109692	-	não avaliado
CCR3	T19	6	9	108600483	19	36202129	-	FALSO
CCR3	T20	5	19	36202141	9	108600066	-	não avaliado
CCR3	T21	3	17	16205625	13	-52380953	-	FALSO
CCR3	T22	3	13	-52379845	17	16205435	-	FALSO
CCR3	T23	4	11	76290875	1	26423121	-	FALSO
CCR3	T24	5	9	113035288	3	43284613	-	FALSO
CCR3	T25	3	11	131094921	3	14838189	-	não avaliado
CCR3	T26	5	4	128173375	6	115037477	-	não avaliado
CCR3	T27	5	6	-115036368	4	128172953	-	FALSO
CCR3	T28	3	14	64137302	10	-81765838	-	não avaliado
CCR3	T29	3	14	-64136866	10	81765519	-	não avaliado
CCR3	T30	8	20	-45625325	17	41496179	-	não avaliado
CCR3	T31	9	17	41496477	20	-45625939	-	FALSO
CCR3	T32	6	X	31348807	13	-99683154	-	não avaliado
CCR3	T33	3	8	-136960919	4	60407767	-	não avaliado
CCR3	T34	6	15	-63656548	3	67948043	-	FALSO
CCR3	T35	6	1	95783019	3	-29210420	-	FALSO
CCR3	T36	3	8	110661829	2	30748366	-	não avaliado
CCR3	T37	3	10	36624352	17	46809461	-	FALSO
CCR3	T38	3	2	23697776	1	-200067669	-	não avaliado
CCR3	T39	3	17	5370428	4	136304467	-	não avaliado
CCR3	T40	3	4	-174700609	20	18401256	-	não avaliado
CCR3	T41	4	2	45668154	3	49607881	-	não avaliado
CCR3	T42	6	5	-31049386	15	38063453	-	FALSO
CCR3	T43	3	13	85722111	4	-126424473	-	não avaliado
CCR3	T44	3	11	271377	12	5419031	-	não avaliado
CCR3	T45	3	10	74077352	2	135339239	-	não avaliado
CCR3	T46	4	19	7469846	6	44008490	-	não avaliado
CCR3	T47	4	19	46095775	13	-36533444	-	não avaliado
CCR3	T48	3	12	41104661	14	-59221041	-	não avaliado
CCR3	T49	5	13	90585218	17	41466137	-	não avaliado

Continua

Continuação

Paciente	VE	Suporte	Regiões cromossômicas				amplicon	validação
			Cro A	Posição A	Cro B	Posição B		
CCR3	T50	4	13	38984130	16	-48714023	-	não avaliado
CCR3	T51	3	16	-48712869	13	38983511	-	não avaliado
CCR3	T52	4	11	41683558	8	112940305	-	não avaliado
CCR3	T53	3	9	9702406	6	103030173	-	não avaliado
CCR3	T54	3	7	13347942	20	-22773590	-	não avaliado
CCR3	T55	3	4	34184275	2	12442802	-	não avaliado
CCR3	T56	4	3	-23739395	1	186161828	-	não avaliado
CCR3	T57	5	20	25474373	2	55021694	-	não avaliado
CCR3	T58	3	16	18152355	3	-177505172	-	não avaliado
CCR3	T59	3	6	-56409119	19	36437059	-	não avaliado
CCR3	T60	4	17	1283334	5	30418811	-	FALSO
CCR3	T61	4	4	170077559	10	81914501	-	não avaliado
CCR3	T62	10	10	81915020	4	170076973	-	FALSO
CCR3	T63	4	17	-5457357	13	98336419	-	FALSO
CCR3	T64	6	4	5475426	20	-8414374	-	não avaliado
CCR3	T65	4	8	29634497	11	-78474622	-	não avaliado
CCR3	T66	3	X	-121552359	5	153880179	-	não avaliado
CCR3	D01	6	3	188150728	3	188409009	201pb	VERDADEIRO
CCR3	D02	6	20	18031602	20	17582992	407pb	VERDADEIRO
CCR3	D03	12	2	79330891	2	79343275	-	FALSO
CCR3	D04	10	2	209942788	2	209949141	-	FALSO
CCR3	D05	14	3	30922129	3	49946114	-	FALSO
CCR3	D06	5	9	29092095	9	29098015	-	não avaliado
CCR3	D07	5	9	35303532	9	35310157	-	FALSO
CCR3	D08	9	9	97680013	9	97686290	-	não avaliado
CCR3	D09	12	13	57752445	13	57788865	-	não avaliado
CCR3	D10	9	16	6389412	16	6500824	337pb	VERDADEIRO
CCR3	D11	8	16	6810984	16	6824389	266pb	VERDADEIRO
CCR3	D12	9	16	6903990	16	7208828	-	não avaliado
CCR3	D13	9	16	62544263	16	62550717	-	FALSO
CCR3	D14	9	18	64958987	18	64967332	-	não avaliado
CCR3	D15	9	19	39411317	19	39932371	133pb	VERDADEIRO
PCR1	T01	5	8	93421404	6	13191227	778pb	VERDADEIRO
PCR1	T02	3	12	3607669	6	145470434	648pb	VERDADEIRO
PCR1	T03	4	6	65543456	1	-62053749	1011pb	VERDADEIRO
PCR1	T04	2	19	40086883	14	-65447676	-	não avaliado
PCR1	D01	7	3	121356366	3	121382720	-	não avaliado

Continua

Continuação

Paciente	VE	Suporte	Regiões cromossômicas				amplicon	validação
			Cro A	Posição A	Cro B	Posição B		
PCR1	D02	9	5	8591553	5	8598032	-	não avaliado
PCR1	D03	6	5	54798967	5	125145693	-	não avaliado
PCR1	D04	9	10	79265480	10	88087970	-	não avaliado
PCR1	D05	10	10	79265512	10	88092435	-	não avaliado
PCR1	D06	7	16	6109988	16	6116508	-	não avaliado
PCR1	D07	5	16	6298413	16	6312549	-	não avaliado
PCR1	D08	7	16	6329791	16	6909675	-	não avaliado
PCR1	D09	6	16	78371178	16	78384996	-	não avaliado
PCR1	D10	7	20	14897971	20	14938767	-	não avaliado
PCR1	D11	5	20	14926929	20	15109439	-	não avaliado
IR1	T01	9	17	35126102	1	223988874	215pb	VERDADEIRO
IR1	T02	6	8	40477264	2	152374971	536pb	VERDADEIRO
IR1	T03	5	2	122159723	8	127649237	577pb	VERDADEIRO
IR1	T04	2	8	-73783752	11	40398761	-	não avaliado
IR1	T05	5	2	30169357	8	-127636126	458pb	VERDADEIRO
IR1	T06	7	7	127318226	14	53067743	524pb	VERDADEIRO
IR1	T07	5	11	-10847237	1	16502493	707pb	VERDADEIRO
IR1	T08	3	14	53067736	7	127318227	644pb	VERDADEIRO
IR1	T09	3	8	74841613	2	-195681531	-	não avaliado
IR1	T10	4	1	53615346	5	167865176	-	não avaliado
IR1	T11	5	1	-119766787	17	35387373	954pb	VERDADEIRO
IR1	T12	3	4	-156768839	11	87895463	262pb	VERDADEIRO
IR1	T13	4	8	74814865	2	128394833	-	FALSO
IR1	D01	6	1	58338901	1	104111466	-	não avaliado
IR1	D02	11	1	61444642	1	155822926	-	não avaliado
IR1	D03	5	1	155870136	1	222019855	-	não avaliado
IR1	D04	5	2	221000576	2	122992576	-	não avaliado
IR1	D05	19	6	53928657	6	53934870	-	não avaliado
IR2	T01	4	14	85145043	8	74493953	1016pb	VERDADEIRO
IR2	T02	3	13	102813872	2	-135875336	-	não avaliado
IR2	T03	8	22	29065971	2	238177710	~650pb	VERDADEIRO
IR2	T04	5	2	238177423	22	29065747	~760pb	VERDADEIRO
IR2	D01	9	3	60342822	3	60388671	251pb	VERDADEIRO
IR2	D02	11	4	185436839	4	185913887	150pb	VERDADEIRO
IR2	D03	5	X	6831906	X	6908289	-	não avaliado
IR2	D04	7	1	210077275	1	210086003	-	FALSO
IR2	D05	5	2	106879453	2	106885973	-	não avaliado

Continua

Continuação

Paciente	VE	Suporte	Regiões cromossômicas				<i>amplicon</i>	<i>validação</i>
			Cro A	Posição A	Cro B	Posição B		
IR2	D06	6	5	9207770	5	9435181	213pb	VERDADEIRO
IR2	D07	5	5	58772843	5	58858549	492pb	VERDADEIRO
IR2	D08	5	6	32630672	6	32723496	-	não avaliado
IR2	D09	14	6	77016657	6	77029247	-	FALSO
IR2	D10	5	9	23362336	9	23377749	-	não avaliado
IR2	D11	9	18	51205947	18	51210677	-	FALSO

Tabela contém informações de todos os candidatos a variação estrutural. **(VE)** Variação Estrutural, os candidatos estão identificados com as iniciais D (Deleção), I (Inversão), T (Translocação); as regiões cromossômicas são baseadas na sequência referência do genoma humano (GRCh37/hg19), o sinal negativo na coordenada sinaliza que é a fita menos (-) da região cromossômica de referência; **(Suporte)** número de sequências pareadas que suportam a presença da VE, após a aplicação de todos os filtros do *pipeline*; **(Cro A)** Cromossomo A, região a montante (*upstream*) da VE; **(Cro B)** cromossomo B, região a jusante (*downstream*) da VE; **(amplicon)** tamanho do fragmento (pares de bases) amplificado por PCR e confirmado por sequenciamento de Sanger; **(Verdadeiro)** Candidato a VE validado, confirmando a as regiões cromossômicas envolvidas e a especificidade ao tecido tumoral ; **(Falso)** Candidato a VE que não produziu um produto de PCR ou que o fragmento amplificado foi resultado de uma região germinativa. As coordenadas cromossômicas dos candidatos validados são exatamente as coordenadas do ponto de quebra.

APÊNDICE C – RESULTADOS DOS ENSAIOS DE DETECÇÃO DO *ctDNA* PARA TODOS OS ENSAIOS E AMOSTRAS

Tabela 5– Resultados dos ensaios de detecção do *ctDNA* para todos os ensaios e amostras

Paciente	Coleta	Volume de DNA circulante analisado (µL)	Ensaio	nº gotas analisadas	nº gotas positivas	Ensaio	nº gotas analisadas	nº gotas positivas	Deteccção <i>ctDNA</i> *
CCR1	diagnóstico	112	T01	259049	0	I01	242197	2	não
CCR1	sem 0	105	T01	217474	2	I01	251326	2	não
CCR1	sem 6	56	T01	113286	0	I01	100711	0	não
CCR1	sem 9	56	T01	125543	1	I01	113739	2	não
CCR1	sem 12	84	T01	197769	1	I01	188527	1	não
CCR1	sem 22	56	T01	117316	0	I01	120527	0	não
CCR1	sem 53	112	T01	239402	2	I01	185614	3	não
CCR2	diagnóstico	44	I01	147158	0	D02	155873	14499	sim
CCR2	sem 3	55	I01	166120	0	D02	167290	12776	sim
CCR2	sem 6	48	I01	75989	0	D02	89458	4415	sim
CCR2	sem 9	56	I01	104309	0	D02	109229	6320	sim
CCR2	sem 13	90	I01	240070	0	D02	228506	12997	sim
CCR2	sem 46	54	I01	124984	109531	D02	132785	132784	sim
CCR2	sem 84	105	I01	162833	0	D02	185962	12877	sim
CCR2	sem 158	112	I01	223646	223252	D02	241682	241682	sim
CCR2	sem 170	80	I01	139513	106932	D02	140040	140039	sim
CCR2	sem 175	40	I01	118251	117734	D02	119466	119460	sim
CCR3	diagnóstico	112	D01	250651	0	D02	226091	0	não
CCR3	0sem	112	D01	232379	0	D02	232204	0	não
CCR3	14sem	112	D01	234417	2	D02	247247	0	não
CCR3	57sem	112	D01	236922	2	D02	226953	0	não

Continua

Continuação

Paciente	Coleta	Volume de DNA circulante analisado (μ L)	Ensaio	nº gotas analisadas	nº gotas positivas	Ensaio	nº gotas analisadas	nº gotas positivas	Deteção ctDNA*
PCR1	diagnóstico	45	T01	124679	7	T02	137620	41496	sim
PCR1	sem 0	60	T01	170754	0	T02	178569	10434	sim
PCR1	sem 6	90	T01	197250	0	T02	256007	1	não
PCR1	sem 13	100	T01	295680	2	T02	293765	1	não
PCR1	sem 118	96	T01	226729	0	T02	231798	0	não
IR1	diagnóstico	56	T01	105351	10680	T02	72944	2	sim
IR1	sem 3	112	T01	211725	1	T02	196538	3	sim
IR1	sem 9	105	T01	309660	0	T02	323169	3	sim
IR1	sem 13	28	T01	93337	0	T02	83943	0	não
IR1	sem 40	12	T01	53415	13763	T02	48624	24628	sim
IR1	sem 46	48	T01	77174	8930	T02	98049	0	sim
IR1	sem 84	48	T01	86158	8856	T02	82655	0	sim
IR2	diagnóstico	56	D01	98010	27114	D02	118797	60645	sim
IR2	sem 1	112	D01	226334	3	D02	213476	6	sim
IR2	sem 6	112	D01	237734	2	D02	229014	0	não
IR2	sem 8	112	D01	237892	0	D02	242787	0	não
IR2	sem 13	112	D01	216029	0	D02	254090	11	sim
IR2	sem 52	56	D01	98074	1	D02	124013	1	não
IR2	sem 124	112	D01	217152	0	D02	222525	0	não

*Quando pelo menos um dos biomarcadores foi detectado na amostra.

ANEXOS



LISTA DE ANEXOS

Anexo A – *Abstract* do manuscrito em fase final de preparação: The use of personalized biomarkers and liquid biopsies to monitor treatment response and disease recurrence in locally advanced rectal cancer after neoadjuvant chemoradiation

Anexo B – Manuscrito aprovado para publicação no periódico *Genomics*: ICRmax: an optimized approach to detect tumor-specific InterChromosomal Rearrangements for Clinical Application

Anexo C – Termo de Consentimento Livre e Esclarecido aplicado ao paciente

Anexo D – Aprovação do projeto no Comitê de Ética do Hospital Alemão Oswaldo Cruz

Anexo E – Súmula Curricular

ANEXO A – *ABSTRACT* DO MANUSCRITO EM FASE FINAL DE PREPARAÇÃO

The use of personalized biomarkers and liquid biopsies to monitor treatment response and disease recurrence in locally advanced rectal cancer after neoadjuvant chemoradiation.

Paola A. Carpinetti^{1,2}, Elisa R. Donnard^{1,2}, Fabiana Bettoni¹, Paula F. Asprino¹, Jorge Sabbaga¹, Angelita Haber-Gama⁴, Rodrigo O. Perez^{3,4}, Raphael B. Parmigiani¹, Pedro A. F. Galante¹, Anamaria A. Camargo^{1,3}

1- Centro de Oncologia Molecular - Hospital Sírio-Libanês – São Paulo-SP – Brazil.

2- Departamento de Bioquímica, Instituto de Química - Universidade de São Paulo – São Paulo-SP – Brazil.

3- Instituto Ludwig de Pesquisa sobre o Câncer – São Paulo-SP – Brazil

4- Hospital Alemão Oswaldo Cruz – São Paulo-SP – Brazil

Abstract

Neoadjuvant chemoradiotherapy (nCRT) followed by surgery is the standard treatment for patients with locally advanced rectal cancer. Variable degrees of tumor regression are observed after nCRT and a significant proportion of patients develop complete tumor regression. Alternative treatment strategies, including close surveillance without any immediate surgery, have been investigated to spare patients with complete tumor regression from potentially unnecessary adverse outcomes of radical surgery. However, clinical and radiological assessment of tumor response to nCRT do not allow the accurate identification of patients with complete response and surgical resection after nCRT is the cornerstone of curative therapy, despite the considerable risk of postoperative functional disorders. In the present study, we report the use of liquid biopsies and personalized biomarkers to monitor the treatment response, detect residual disease and early recurrence in locally advanced rectal cancer after nCRT. We sequenced the whole-genome of four rectal tumors and identified patient-specific chromosomal rearrangements that were used to monitor the circulating tumor DNA (ctDNA) in liquid biopsies serially collected at the time of diagnosis, during nCRT and follow-up. ctDNA was quantified in a total of 28 serial plasma samples using a nested-amplification strategy and Droplet Digital PCR. We compared ctDNA levels with the radiographic imaging of tumors and CEA levels. Altogether, our results indicate that the use of personalized biomarkers and liquid biopsies are not sensitive enough for the detection of residual disease after nCRT, but can be efficiently used to monitor treatment response and detect disease recurrence, preceding radiological diagnosis even among patients with excellent primary response to nCRT.

Similar good results were obtained when assessing tumor response to adjuvant therapy and disease progression. Although this proof-of principle study points towards the use of personalized biomarkers and liquid biopsies as an important tool to tailor the management of rectal cancer patients, it has limitations that need to be addressed prior to the replication in a larger prospective cohort and introduction of this strategy into clinical practice.

Keywords: Personalized biomarker, circulating tumor DNA, rectal cancer, rearrangement chromosomal, chemoradiotherapy management.

ANEXO B – MANUSCRITO ACEITO PARA PUBLICAÇÃO NO PERIÓDICO
GENOMICS



Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Methods

ICRmax: An optimized approach to detect tumor-specific interchromosomal rearrangements for clinical application

Elisa R. Donnard^{a,b}, Paola A. Carpinetti^{a,b}, Fábio C.P. Navarro^{a,b}, Rodrigo O. Perez^{c,d}, Angelita Habr-Gama^c, Raphael B. Parmigiani^a, Anamaria A. Camargo^{a,d}, Pedro A.F. Galante^{a,*}

^a Centro de Oncologia Molecular, Hospital Sírio-Libanês, São Paulo, Brazil

^b Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

^c Instituto Angelita & Joaquim Gama, São Paulo, Brazil

^d Ludwig Institute for Cancer Research, São Paulo, Brazil

ARTICLE INFO

Article history:

Received 26 August 2014

Accepted 26 January 2015

Available online xxx

Keywords:

Cancer biomarkers

Chromosomal rearrangements

Clinical application

Bioinformatics tool

ABSTRACT

Somatically acquired chromosomal rearrangements occur at early stages during tumorigenesis and can be used to indirectly detect tumor cells, serving as highly sensitive and tumor-specific biomarkers. Advances in high-throughput sequencing have allowed the genome-wide identification of patient-specific chromosomal rearrangements to be used as personalized biomarkers to efficiently assess response to treatment, detect residual disease and monitor disease recurrence. However, sequencing and data processing costs still represent major obstacles for the widespread application of personalized biomarkers in oncology. We developed a computational pipeline (ICRmax) for the cost-effective identification of a minimal set of tumor-specific interchromosomal rearrangements (ICRs). We examined ICRmax performance on sequencing data from rectal tumors and simulated data achieving an average accuracy of 68% for ICR identification. ICRmax identifies ICRs from low-coverage sequenced tumors, eliminates the need to sequence a matched normal tissue and significantly reduces the costs that limit the utilization of personalized biomarkers in the clinical setting.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Somatically acquired chromosomal rearrangements, including duplications, inversions, deletions, insertions and translocations, constitute a key feature of tumor genomes [1–3]. They occur at early stages during tumor formation and persist throughout tumor progression [4]. These rearrangements are not present in normal cells from cancer patients and can thus be used to unequivocally detect tumor cells, serving as highly sensitive and specific tumor biomarkers to approach clinically relevant endpoints, such as assessment of response to therapy and detection of disease recurrence [5–7].

Technically, PCR detection of translocated DNA sequences that originate from different chromosomes or thousands of base pairs apart is a straightforward process when compared to PCR discrimination of single-base alterations and, therefore, tumor-specific chromosomal rearrangements represent ideal markers for monitoring tumor burden [5,8]. Highly sensitive and specific assays developed to detect recurrent chromosomal translocations in hematological tumors have become standard practice to monitor residual disease and predict relapse to targeted therapy, allowing individualized therapeutic choices [9,10].

Unfortunately, a similar use of known chromosomal rearrangements in solid tumors has been hampered by the absence of recurrent rearrangements in these tumors [3]. Recently, however, whole-genome sequencing has become efficient and affordable enough to allow genome-wide identification of patient-specific somatic chromosomal rearrangements [2,11–13] that could be used as personalized biomarkers to detect tumor cells [5].

Currently, analysis of short read paired-end or mate-pair sequences aligned against the human reference genome is the most efficient strategy to detect somatic chromosomal rearrangements present in tumor genomes. However, due to the repetitive nature of the human genome and the presence of structural variations, assigning the correct mapping positions to short reads and calling somatic chromosomal rearrangements are not straightforward processes [1,14], and a large number of false positive candidates are usually identified [3], generating unacceptable levels of noise. In order to reduce the high number of false positives candidates and analysis complexity, sequencing of a matched normal tissue DNA sample is usually required [15,16], increasing both the sequencing and computational costs, and further limiting the use of this strategy in actual clinical practice. Some of the most used software currently available for structural variation (SV) detection show good precision and high sensitivity for detecting inter-chromosomal rearrangements, such as BreakDancer [15], CREST [17], HYDRA-SV [18]

* Corresponding author.

E-mail address: pgalante@mochsl.org.br (P.A.F. Galante).

and others. But these algorithms either rely on matched normal DNA sequencing, or dispense with the matched normal sequencing but require high sequence coverage to identify correct breakpoints through split mapped read analysis, hampering the clinical application of these tools.

Here we present a pipeline (ICRmax) that allows the efficient identification of tumor-specific interchromosomal rearrangements. ICRmax has sets of filters that efficiently eliminate the need to sequence a matched normal genome and to allow the efficient identification of interchromosomal rearrangements from low-coverage ($\sim 3\times$) sequenced tumor genomes. The use of ICRmax significantly reduces the sequencing and computational costs associated with the identification of tumor-specific chromosomal rearrangements and will certainly contribute to the widespread use of personalized biomarkers in the routine clinical practice of patients with solid tumors.

2. Results and discussion

2.1. ICRmax implementation and filters

ICRmax was developed to efficiently identify a minimal set of reliable interchromosomal rearrangements from low coverage sequenced tumor genomes without the need to sequence a matched normal genome, reducing the sequencing cost and creating an opportunity to implement the use of personalized biomarkers in the routine clinical management of solid tumors. The ICRmax pipeline, described with

details in Methods, is summarized in Fig. 1 and a step-by-step command line are provided in the supplementary material.

2.2. Identification of interchromosomal rearrangements in rectal tumors with ICRmax

In order to show how our pipeline works, we used ICRmax to identify interchromosomal rearrangements in six rectal tumor genomes (RT1–RT6). For each sample, mate-pair libraries with average insert size of 600 bp were generated using tumor genomic DNA and sequenced in a SOLiD sequencing platform to varying depths (see Table 1). Sequenced reads were aligned against the human genome reference sequence using Bioscope (Applied Biosystems) and only high quality ($Q > 20$) alignments were selected for further analysis. Sequence coverage varied from 4 to $9\times$ and the calculated physical coverage based on the average insert size of sequenced fragments varied between 15 to $62\times$. The insert size of ~ 600 bp for the libraries yields a higher physical coverage with lower sequencing depth and facilitates the identification of structural variation [1] also contributing to lowering the cost associated to this protocol since smaller insert size libraries (~ 200 bp) will probably require higher sequencing depth. Sequencing and coverage information for each tumor sample is presented in Table 1. We also sequenced three matched normal genomes (N1, N5 and N6) and submitted the sequenced reads through the same pipeline to evaluate their contribution to exclude false positive events that were not filtered out by ICRmax.

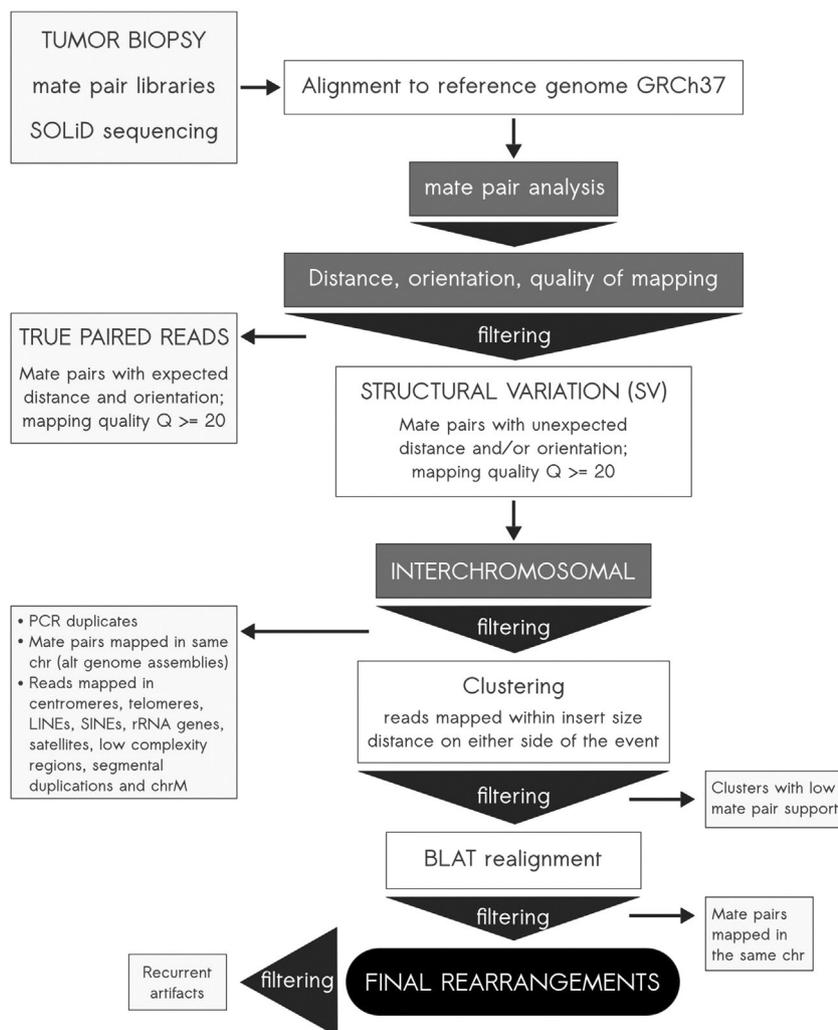


Fig. 1. Bioinformatics pipeline showing the main steps for whole tumor genome analysis and selection of sequences indicating the presence of structural variation such as interchromosomal rearrangements.

Table 1

Mapping results for rectal tumor samples and paired normal tissue samples submitted through whole genome sequencing.

Sample	Total reads	Mapped nucleotides	Physical coverage	Sequence coverage
RT1	1,035,604,016	25,967,977,794 (50%)	180,593,981,498 (62×)	9.0×
RT2	393,756,912	12,232,074,899 (62%)	57,714,657,702 (20×)	4.2×
RT3	385,789,584	11,616,923,871 (60%)	53,821,366,172 (19×)	4.0×
RT4	398,436,826	12,231,994,959 (61%)	59,355,360,167 (20×)	4.2×
RT5	425,460,416	15,150,461,585 (51%)	42,211,056,664 (15×)	5.2×
RT6	991,625,036	20,921,867,961 (42%)	77,491,127,440 (27×)	7.2×
N1	728,574,754	21,532,212,957 (49%)	78,716,757,426 (27×)	7.4×
N5	682,517,714	26,824,738,972 (65%)	94,020,024,357 (32×)	9.3×
N6	305,101,394	9,129,663,072 (60%)	43,929,159,747 (15×)	3.2×

Sequence and physical coverage for the matched normal genomes varied from 3.2 to 9.3× and from 15 to 32×, respectively (Table 1).

ICRmax pipeline was followed as described in the Methods section and supplementary material. Some of the filtering steps involve regions that cause misalignment and are also included in other pipelines for detecting structural variations [15,18]. However, they are usually not as comprehensive or strict since there are repetitive genomic regions that can yield uniquely mapped reads. In our experience, detection of somatic structural variation benefits from this strict filtering approach.

As expected, a variable number of putative interchromosomal rearrangements were identified in each tumor genome (ranging from 9 to 105 events per sample, average of 32). Interestingly, when we compared the set of ICR candidates in each tumor genome we found a significant number of recurrent events (Table S1). Since solid tumors are well known for their lack of recurrent tumor-specific rearrangements [3] we assumed that these recurrent events (Table S1) might correspond to artifacts that could be removed by comparison with the matched normal samples [19–21]. The presence of recurrent structural variation breakpoints was recently observed in a study of complex genomic rearrangements [22] and as a validation strategy only breakpoints present in a single tumor were considered somatic. Accordingly, we selected eight tumor-recurrent events for validation by PCR amplification with primers flanking the putative breakpoint region determined by the alignment of mate pair reads. As expected for such false positive somatic events, amplification with specific primers was obtained when using both the tumor and matched normal DNA (Figure S2) and these cases are therefore not candidates for tumor-specific rearrangements. We therefore implemented this inter-tumor comparison as a last filtering step in our ICRmax pipeline, removing all the identified recurrent artifacts in our tumor genomes from the final set of rearrangement candidates. The final sets of rearrangements found are represented in Fig. 2 and Fig. S3 using Circos [23].

Table 2

Validation results.

Sample	Number of clusters	Rearrangements after recurrent filter	Tested rearrangements	Validated as tumor specific
RT1	27	10	6	3 (50.0%)
RT2	18	15	10	9 (90.0%)
RT3	9	4	4	3 (75.0%)
RT4	105	96	11	1 (9.0%)
RT5	14	2	1	1 (100.0%)
RT6	19	14	4	1 (25.0%)

We generated a list of recurrent artifacts that is available as supplementary data (Table S2), or directly at <http://www.bioinfo.mochsl.org.br/icrmax> (file recurrent_artifacts.bed).

Thus, ICRmax does not rely on the need to sequence matched normal/tumor genomes due to the implementation of additional filters to remove recurrent events and to eliminate false-positive rearrangements, reducing by at least half the sequencing cost, but still allowing the identification of a reliable minimal set of tumor specific ICRs for clinical application. To the best of our knowledge, this is the first pipeline for detection of somatic structural variations that dispenses with the sequencing of matched normal genome.

After removing recurrent artifacts and false-positive candidates, the number of identified events for each sample varied from 2 to 96, with an average of 23 per sample (detailed in Table 2). Rearrangements present in each sample are graphically represented in Fig. 2 and Supplementary Fig. 3. A subset of these putative rearrangements was then selected for PCR validation followed by Sanger sequencing (Table 2). Of the 36 putative interchromosomal rearrangements selected for validation, 18 (50%) were confirmed as somatic tumor-specific events and their exact breakpoints were determined for most cases after sequencing (PCR amplification results are shown in Supplementary Fig. 2). Validation efficiency between different samples varied from 9 to 90%, and some patients, like RT4, provided more difficulty in the validation stage, possibly due to the presence of different levels of genetic instability and tumor heterogeneity. Fig. 3 illustrates the genomic region, as well as the sequence coverage, of a confirmed rearrangement between chromosomes 1 and 17 detected in sample RT2. The exact breakpoint was determined by PCR amplification followed by Sanger sequencing of the amplified fragment (see Supplementary Fig. 4). Interestingly, this rearrangement is located in the *TP53BP2* gene region, and loss of this functional gene product has been associated with gastric and other cancer susceptibility [24].

These validated rearrangements can be used in further investigation as biomarkers for their corresponding patients. The validation stage is essential in distinguishing tumor specific events from germline polymorphisms or false positives candidates. Furthermore, one could

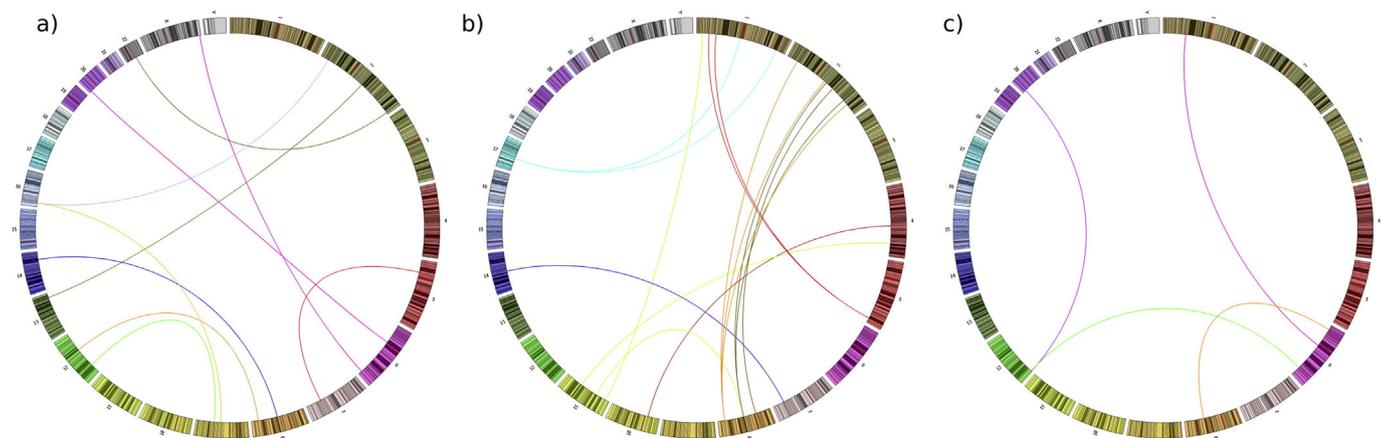


Fig. 2. Circos representation of the interchromosomal rearrangements found in 3 tumor genomes. a) RT1; 10 rearrangements b) RT2; 15 rearrangements and c) RT3; 4 rearrangements.

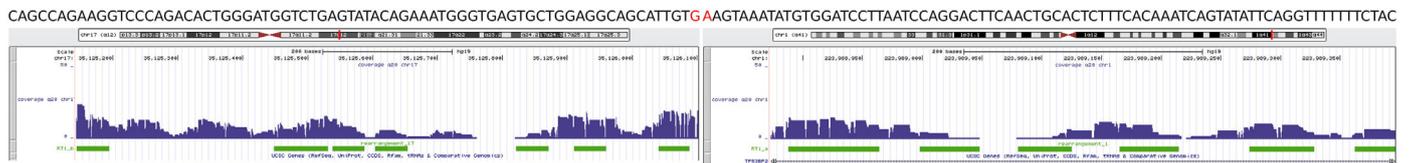


Fig. 3. Genome browser view of reads (green) indicating one of RT2 rearrangements between chromosome 17 (left) and chromosome 1 (right). Region shown for chromosome 17 ends in the breakpoint found after Sanger sequencing. Region shown for chromosome 1 also starts in the exact breakpoint nucleotide. Read orientation indicates that the rearrangement is structured as shown and confirmed through Sanger sequencing. The nucleotide sequence of the breakpoint region is shown above. Last nucleotide of chromosome 17 and first nucleotide of chromosome 1 are marked in red. Sequence coverage is shown in dark blue, and only reads with mapping quality ≥ 20 are used. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

argue that studies can rely mainly on this step to select for tumor-specific events. However, the costs of primer design and the time spent on breakpoint amplification can be a limiting factor when identifying personalized chromosomal rearrangements in the clinical setting. The optimization of the bioinformatics analysis is therefore critical for defining a concise set of reliable rearrangements to substantially increase the efficiency of the entire process.

2.3. Comparison to a tumor-normal paired sequencing

As previously mentioned, to assess the contribution of sequencing matched normal genomes when using ICRmax, we sequenced matched normal DNA for three of our tumor samples. Sequencing and coverage information for all normal samples is presented in Table 1. We found that most (62%) of the ICR candidates detected in both normal and tumor DNA were also detected in other tumor samples and therefore present in our list of recurrent artifacts. For instance, 13 events were detected in both N1 and RT1 samples derived from the same patient, but 10 of these rearrangements were also present in our list of tumor-recurrent artifacts. Interestingly the number of events filtered by the recurrent artifact list was for some of the samples greater than the number of events removed exclusively by comparison to the matched normal genomes. For example, for patient #1 a total of 3 rearrangements that were not identified by sequencing the matched normal genome were filtered out using the list of tumor-recurrent rearrangements and one extra event was removed by comparison with one of the other normal tissues sequenced. Similar results were observed for the other two patients for whom we have sequenced the matched normal DNA (Table S1). Noteworthy, for patient #6, sequencing the matched normal DNA (N6) removed only four rearrangements found in the tumor genome (and only one of these could not be removed by the tumor-recurrent list), likely due to the low coverage obtained for the normal sample (Table 1), suggesting that sequencing paired normal tissue with lower coverage is not an effective option to remove false-positive candidates. The matched normal tissue contribution therefore does not justify the increase in the sequencing cost, since a search for recurrent events in different tumor genomes allows the efficient identification of artifactual events.

2.4. Expansion of the recurrent artifacts list

One of the main conclusions from our analysis of structural variations in rectal tumor genomes was that removing recurrent events reduces substantially the cost and time of analysis for the validations in the personalized approach, dispensing with the sequencing of matched normal DNA. Currently, a large amount of data is available from large-scale genome sequencing projects (normal and tumor). We sought to expand our recurrent artifact list and improve our capability to exclude these events from future analysis by applying our pipeline to the 2362 individuals from the 1000 Genomes project [25]. The results from our analysis reveal a total of 2800 events indicating interchromosomal rearrangements that are recurrent in three or more individuals. From the 31 cases of recurrent artifacts identified in our six rectal tumor genomes, 16 are present in this set found for individuals of the 1000

Genomes project. There are also three events in the 1000 Genome recurrent set that are present in our final candidate set for the patients (not eliminated by any filter). Two of these events had in fact been tested by PCR to validate them as tumor-specific and resulted in normal DNA amplification. Therefore, this analysis of 2362 normal genomes substantially increases our list of recurrent artifacts to exclude from future analysis of tumor-specific structural variations. The list of 2800 recurrent interchromosomal events can also be downloaded from <http://www.bioinfo.mochsl.org.br/icrmax>. We expect that even more recurrent artifacts can be identified and excluded through the analysis of data available from tumor genomes already sequenced and we intend to continue to increase this list.

2.5. Simulated rearrangement datasets

Overall, our pipeline was effective when applied to the tumor genomes sequenced and allowed the identification of a minimal subset of personalized interchromosomal rearrangements. However, tumor heterogeneity and genetic instability prevents both the complete identification of rearrangements present in the tumor genome as well as a true estimate of our method's accuracy. In order to better evaluate our pipeline we simulated three human genomes (based on hg19) containing different numbers of ICRs. We randomly created 20, 30 and 40 ICR events for each genome (RG1, RG2 and RG3, respectively; for details, see Methods and Supplementary data). From each genome (RG1, RG2 and RG3) we also generated three sets of randomly sampled reads (mate pairs of 50 nt and insert size of 700 nt, on average), which represent a simulated physical coverage of $44\times$, $25\times$ and $13\times$, respectively (see Table 3). Using ICRmax and increasing the minimal mate pair support to five reads, we were able to correctly identify 42 out of the 90 simulated rearrangements (47% specificity) and eight additional false positive candidates (84% accuracy; Table 4). As expected, for the low sequence coverage genomes ($1.9\times$ and $3.8\times$; Table 3), ICRmax performed with a lower sensitivity (42.5%), but a remarkable accuracy (100%). When we decrease the mate pair support requirement to three reads in the RG2 and RG3 simulated genomes, we detected a higher number of true simulated rearrangements (18/30 for RG2 and 20/40 for RG3) but a number of false positives also appear in the sets (14 for RG2 and 3 for RG3). The results presented in Table 4 suggest that our pipeline achieved an experimentally acceptable accuracy and sensitivity for ICR detection and performed well even under low sequencing ($1.9\times$) and physical ($13\times$) coverage (for RG3, for instance). Further details on the simulated reads can be found in the Methods section and results regarding our simulated data are presented in Tables 3 and 4 and Supplementary Fig. 3.

Table 3
Mapping results for genomes containing randomly simulated ICRs.

Simulated genomes	Reads generated	Mapped reads	Sequence coverage	Physical coverage
RG1	402,771,620	336,548,722 (83.5%)	$5.8\times$	$44\times$
RG2	263,391,883	221,636,923 (84.1%)	$3.8\times$	$25\times$
RG3	131,695,960	110,878,325 (84.2%)	$1.9\times$	$13\times$

Table 4
Identification of simulated events by ICRmax.

Simulated genomes	Simulated ICRs	Found ICRs (final set)	True positives (accuracy)	Sensitivity
RG1	20	21	13 (62%)	65% (13/20)
RG2	30	15	15 (100%)	50% (15/30)
RG3	40	14	14 (100%)	35% (14/40)
Total	90	50	42 (84%)	47% (42/90)

We used the same simulated datasets as input to the HYDRA-SV algorithm [18]. HYDRA successfully detected 58 out of the 90 simulated rearrangements, yielding higher sensitivity (64%) but poorer accuracy in all three sets (Table S4). There is an increase in the number of breakpoints detected that do not correspond to the simulated rearrangements, including in RG2 and RG3, for which the ICRmax pipeline only detected the true simulated breakpoints. Similar to ICRmax, HYDRA performed with higher accuracy in the simulated datasets that had lower coverage.

2.6. Pipeline test in a different tumor type and different sample source

To better establish the robustness of our pipeline and its performance with data generated using different sequencing platforms, tissue sources and tumor type, we used data from a study on breast cancer genome evolution [26]. For four of the patients in the original study (P1, P4, P5 and P6), we downloaded whole genome sequencing data from FFPE samples corresponding to invasive ductal carcinoma (IDC) or matched normal tissue from lymph nodes (Lymph). After alignment and duplicate read removal, the number of reads per sample was on average ~7 million, and the insert size for these libraries was ~300–400 bp according to the original study. We submitted these reads through the ICRmax pipeline, and detected interchromosomal rearrangements for each patient (4, 6, 7, and 5 respectively). We then performed either our final step in the pipeline, the inter-tumor comparison, or the usual matched normal comparison. Overall, we detected 4 recurrent artifacts in two or more of the patients. After removal of these artifacts, the final sets contained 2, 2, 5 and 2 rearrangements respectively. If instead we perform a usual tumor vs. normal comparison in the initial set, we identify 2 artifacts and remove them from each paired sample. The resulting sets contain 3, 3, 5 and 2 rearrangements. Consequently, with the inter-tumor comparison we were able to detect and remove two additional artifacts. The artifacts detected are not the same. Of the 4 artifacts identified by the inter-tumor comparison, only, one was identified by the matched normal comparison. Therefore there is a single artifact identified by matched normal comparison that cannot be removed otherwise, and three artifacts removed by the inter-tumor comparison that are not detected in the matched normal.

If we compare the final set of results from the IDC samples to the recurrent artifact list we generated from 1000 Genomes, we find an additional 2 artifacts that can be removed from the set thus reinforcing the value of building a database with recurrent artifacts found in both tumor and normal samples.

2.7. Cost analysis

Based on the data presented here, we estimated the final cost associated with the immediate application of this protocol in the clinical setting. At our facilities, the sequencing cost in the SOLiD 5500XL platform is approximately US\$1 per 20 million bases. To obtain a 4× sequencing coverage for each patient sample, one needs to generate ~20 billion bases (400 million reads with 50 nt), totalizing ~US\$1000 per patient. For each patient the final cost for the development of personalized chromosomal rearrangements should include additional costs for sample processing and mate-pair library construction (~US\$470 per patient) as well as for PCR validation. Testing 10 putative rearrangements will

cost ~US\$200 in primer synthesis and ~US\$10 for all PCR reactions. Assuming a 50% validation rate, subsequent Sanger sequencing of the five PCR-validated rearrangements for breakpoint determination would add ~US\$50 reaching the final cost of ~US\$1730 per patient.

3. Conclusion

Recent advances in high throughput sequencing technology have allowed the genome-wide identification of patient-specific chromosomal rearrangements [2,3,5,6,16]. These personalized biomarkers are especially useful to assess response to treatment, detect residual disease and monitor disease recurrence and are expected to have a widespread use in clinical oncology [5,6]. Unfortunately, current methods for chromosomal rearrangement detection require the sequencing of both tumor and matched normal genomes increasing the sequencing and computational costs and precluding the implementation of personalized biomarkers in clinical practice [15,16]. ICRmax aims to eliminate this need, relying on a set of strict mapping filters and inter-tumoral comparison to greatly reduce false positive candidates. However it is important to emphasize that the main goal is not to obtain the full set of ICR events present in the tumor genome, but to efficiently identify a minimal set of rearrangements which in any case can only represent part of a much larger set of tumor alterations due to the percentage of a solid tumor collected in a biopsy, for example. The minimal set of identified rearrangements can be validated and included into practice, reducing the cost and time required for the identification of clinically useful personalized biomarkers.

4. Methods

4.1. Sample preparation and whole genome sequencing

Biopsies from six locally advanced rectal tumors were obtained from patients treated at the Instituto Angelita & Joaquim Gama/Hospital Alemão Oswaldo Cruz (<http://www.hospitalalemao.org.br>). Tumor samples were collected after informed consent and approval of the institution's ethics committee. Blood samples from the same patients were also collected and circulating leucocytes were used as a source of normal matched germline DNA. Genomic DNA was extracted using a Trizol based protocol for simultaneous DNA/RNA extraction [27]. Mate-pair libraries were generated for the SOLiD platform according to the manufacturer's instructions. In brief, 5 ug of whole genome amplified (WGA) DNA samples were randomly sheared into 0.6–1.0 kb fragments using the Covaris S2 System. Fragmented DNA was size selected on 0.8% agarose gels and used as template in emulsion PCR. DNA fragments were coupled to bead via an adapter sequence and clonally amplified. Amplified DNA fragments were then covalently attached to a glass slide. Sequencing primers hybridized to the adapter sequence and fluorescently labeled di-base probes were used in ligation-based sequencing generating 50 nt mate-pair reads.

4.2. Read alignment and selection of discordant mate pairs

ICRmax starts from sequence alignment data against the human reference genome in .bam format allowing users to apply their algorithm of choice for the alignment of mate-pair or pair-end reads. For the rectal tumor samples, all color space reads were aligned using standard BioScope mapreads (Applied Biosystems) algorithm against hg19/GRCh37 reference sequence downloaded from UCSC Genome Browser (<http://genome.ucsc.edu/>) [28]. Only sequences with unique mapping and mapping quality greater than or equal to 20 ($Q \geq 20$) are used in subsequent steps. Once the reads are paired, their orientation and mapping position are analyzed. Read pairs presenting an alignment against the human reference genome in the expected orientation and within expected distance (mean insert size \pm 2 s.d.) are discarded while aberrant pairs with reads mapping on different chromosomes are retained

in the analysis to identify interchromosomal rearrangements. Other mate pairs with anomalous mapping patterns can be selected to search for intrachromosomal rearrangements such as insertions, deletions and inversions [29].

In order to minimize reference genome assembly errors and structural polymorphisms, selected read pairs are submitted through a second mapping step using alternative human genome assemblies and read pairs mapping on the same chromosome are excluded. Three alternative genome assemblies were used as reference: HuRef (J. Craig Venter Institute) [30], GRCh37_alt (partial reference genome with alternative representations – Genome Reference Consortium) and CRA (human chr7 complete sequence – The Center for Applied Genomics) [31]. After mapping, all cases where both reads in a mate pair map in the same chromosome were removed from further analysis.

All reads with identical mapping coordinates (identical start or end positions) are removed from further analysis, since they likely result from PCR duplicates generated during the library construction and amplification steps and can provide sequencing support to a false positive candidate [32–34].

4.3. Rearrangement identification

Sequences mapped inside centromeric (+1 Mb) and telomeric regions (+1 Mb), downloaded from UCSC Genome Browser – <http://genome.ucsc.edu>, are removed from the bed files with paired reads using Bedtools [35] as detailed in the supplementary material. Read pairs with sequences mapping to some of the regions defined by RepeatMasker [36] (LINEs, SINEs, low complexity regions, satellites and rRNA genes), which may result in ambiguous mapping, are also removed. Another filtering step removes read pairs mapped in regions corresponding to segmental duplications [37], in order to avoid false-positive rearrangements caused by the misalignment of read pairs in the parental and duplicated regions. Lastly we removed mate pairs with reads mapping to the mitochondrial genome since we observed that the presence of nuclear copies of mitochondrial DNA (*numts*) in the nuclear genome [38] leads to misalignments and the identification of false-positives.

Finally, read pairs are grouped (clustered) based on their genomic coordinates. Reads from different pairs mapped within a minimal distance corresponding to the average library insert size + 2 standard deviation (s.d.) on either side of the putative rearrangement are grouped. For interchromosomal rearrangements, mate pairs with reads mapping in the same two chromosomes are compared and reads from each pair should map within a predetermined window to indicate the same event (windows are calculated by adding 2 s.d. to mean insert size). Once clusters are formed, the number of mate pairs supporting the event is evaluated. Clusters composed of a minimum of three and a maximum of 80 pairs are selected, reducing the impact of incorrect mapping and other artifacts. The rationale for setting the minimum and maximum read support is related to the sequence coverage. Our suggestion is to keep the “minimum” equal to or half of the sequence coverage and the “maximum” (to avoid cluster with repetitive elements, which are hard to validate) of $\sim 20\times$ the sequence coverage. Both the clustering and this cutoff reduce substantially the number of reads indicating rearrangement events and simplify the last steps of the pipeline. For the evaluation of the recurrent events, mate pair support was lowered to two reads for the comparison between patients, keeping only events in the final set that could not be found in other patients even with the lower support.

For SOLiD platform sequencing data, remaining reads are submitted through a third and final mapping step using BLAT [39] to remove further unreliable read mapping that may result from the initial alignment step for color space reads. Reads from a single mate pair showing other mapping possibilities in the same chromosome and within expected distance from their mate pairs are removed from the final set. After removing these mate pairs, the minimum read pair support should once

again be evaluated and only clusters with at least two reads respecting the same orientation pattern should be maintained (see Supplementary Fig. 1). BLAT parameters used were the same indicated for reproducing the webtool results and can be found in Genome Browser [28,40].

4.4. 1000 Genomes data analysis

Whole genome sequencing data from 2362 individuals included in the 1000 Genomes project were accessed through the Open Science Data Cloud (www.opensciencedatacloud.org) and submitted through our ICRmax pipeline. We increased the minimum mate-pair support for an interchromosomal event to 5 and the average size of insert between mate-pairs considered was 300 nt as described [25].

4.5. Simulated datasets

Based on the human genome sequence (hg19), three sets (RG1, RG2 and RG3) of ICR were randomly generated by using Perl scripts. Briefly, we first randomly selected a chromosome and a genomic coordinate, followed by a second selection of another breakpoint in a different chromosome. The FASTA sequences of both chromosomes involved in the rearrangements were joined, creating a new chromosome. This process was repeated $20\times$ for RG1, $30\times$ for RG2 and $40\times$ for RG3 (Table S3). Chromosomes that were not involved in any simulated rearrangement were also kept in the final file. Finally, these genomes were then used to simulate color space sequenced reads with 50 bp and 700 bp insert size between mate pair reads. In order to simulate the real sequencing, the generated reads contain 1% sequencing errors. The number of reads generated was calculated based on the sequence and physical coverage we wished to obtain for each simulation relative to the size of the reference human genome (hg19).

The HYDRA-SV algorithm was downloaded from code.google.com/p/hydra-sv/. Following the example workflow in the documentation, the parameters calculated were – mno 1350 ($10\times$ m.a.d. of the DNA fragment libraries as suggested), – mld 1000 (the same overlap distance used in ICRmax clustering) and – m 5 (at least five mate-pairs supporting the breakpoint).

4.6. Analysis of FFPE breast cancer samples

Whole genome sequencing data was downloaded from SRA (accession: PRJNA193652). We used a subset of the available data, 5 files from each sample (IDC and Lymph) for each of 4 patients. We excluded patient 2 from the analysis since the authors found mouse genome contamination in three libraries. We did not use patient 3 data for a more homogeneous dataset, since the normal tissue available for this patient was not derived from lymph node like the other 4 used. Downloaded files were converted to FASTq files using the SRA Toolkit and reads were aligned to hg19 using Bowtie2 [41] with the very-sensitive parameter. Reads with mapping quality greater than or equal to 20 were submitted through the ICRmax pipeline.

4.7. Orientation patterns and primer design

Read orientation is also considered when defining the candidates for interchromosomal rearrangements. For the true positive candidates, mate pairs spanning a single breakpoint will show equal patterns of orientation. In case of SOLiD sequencing, mate pairs indicating the same breakpoint can be classified into four different patterns detailed in Figure S1. After dividing remaining mate pairs into different orientation patterns clusters still containing two or more mate pairs are chosen for validation. Clusters can end up divided into two sets of orientations, which most likely represent two separate breakpoints for an insertion or translocation. After observing read orientation, support for the final clusters is lowered to at least two mate pairs indicating the rearrangement.

4.8. Breakpoint validation by PCR and sequencing

Primers were designed for a randomly chosen subset of the identified rearrangements for each patient. Mate-pair sequences flanking predicted breakpoints were used as target sequences for primer design using Primer3 when possible, with further manual evaluations necessary to adjust the sequences to the target regions. Mate-pair sequence order and orientation were used to guide primer design. When primers could not be designed from mate-pair sequences, upstream genomic sequence up to 200 bp was used for primer design. Primers were used for PCR on tumor and matched normal samples to confirm the somatic origin and tumor-specificity of the rearranged fragments. Sanger sequencing of PCR products was used to confirm the specificity of the PCR amplification and to map breakpoint sequences.

5. Data access

The sequences used in this analysis will be submitted to The European Nucleotide Archive (ENA; www.ebi.ac.uk/ena) under accession number PRJEB4781. Additional files can be obtained at <http://www.bioinfo.mochsl.org.br/icrmax>.

Disclosure declaration

The authors declare they have no competing interests.

Acknowledgments

We thank Fabiana Bettoni, Paula Asprino, Fernanda Koyama, Bruna S de Quevedo, and Natalia Felicio for their technical supports on the sample collecting, sample preparation or on the DNA sequencing. We also thank Daniel T. Ohara for technical support on the bioinformatics pipelines. This work made use of the Open Science Data Cloud (OSDC) which is an Open Cloud Consortium (OCC)-sponsored project. AAC and ED are supported by FAPESP (11/50684-8 and 10/12658-2, respectively). This work was also supported by the Ludwig Institute for Cancer Research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ygeno.2015.01.009>.

References

- [1] M. Meyerson, S. Gabriel, G. Getz, Advances in understanding cancer genomes through second-generation sequencing, *Nat. Rev. Genet.* 11 (2010) 685–696.
- [2] P.J. Campbell, P.J. Stephens, E.D. Pleasance, S. O'Meara, H. Li, T. Santarius, L.A. Stebbings, C. Leroy, S. Edkins, C. Hardy, J.W. Teague, A. Menzies, D.J. Turner, C.M. Cleve, M.A. Quail, A. Cox, C. Brown, R. Durbin, M.E. Hurler, P.A. Edwards, G.R. Bignell, M.R. Stratton, P.A. Futreal, Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing, *Nat. Genet.* 40 (2008) 722–729.
- [3] A.J. Bass, M.S. Lawrence, L.E. Brace, A.H. Ramos, Y. Drier, K. Cibulskis, C. Sougnez, D. Voet, G. Saksena, A. Sivachenko, R. Jing, M. Parkin, T. Pugh, R.G. Verhaak, N. Stransky, A.T. Boutin, J. Barretina, D.B. Solit, E. Vakiani, W. Shao, Y. Mishina, M. Warmuth, J. Jimenez, D.Y. Chiang, S. Signoretti, W.G. Kaelin, N. Spardy, W.C. Hahn, Y. Hoshida, S. Ogino, R.A. Depinho, L. Chin, L.A. Garraway, C.S. Fuchs, J. Baselga, J. Taberner, S. Gabriel, E.S. Lander, G. Getz, M. Meyerson, Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion, *Nat. Genet.* 43 (2011) 964–968.
- [4] D. Hanahan, R.A. Weinberg, Hallmarks of cancer: the next generation, *Cell* 144 (2011) 646–674.
- [5] R.J. Leary, I. Kinde, F. Diehl, K. Schmidt, C. Clouser, C. Duncan, A. Antipova, C. Lee, K. McKernan, F.M. De La Vega, K.W. Kinzler, B. Vogelstein, L.A. Diaz Jr., V.E. Velculescu, Development of personalized tumor biomarkers using massively parallel sequencing, *Sci Transl Med* 2 (2010) 20ra14.
- [6] S.J. Dawson, D.W. Tsui, M. Murtaza, H. Biggs, O.M. Rueda, S.F. Chin, M.J. Dunning, D. Gale, T. Forshew, B. Mahler-Araujo, S. Rajan, S. Humphray, J. Becq, D. Halsall, M. Wallis, D. Bentley, C. Caldas, N. Rosenfeld, Analysis of circulating tumor DNA to monitor metastatic breast cancer, *N. Engl. J. Med.* 368 (2013) 1199–1209.
- [7] F. Diehl, K. Schmidt, M.A. Choti, K. Romans, S. Goodman, M. Li, K. Thornton, N. Agrawal, L. Sokoll, S.A. Szabo, K.W. Kinzler, B. Vogelstein, L.A. Diaz Jr., Circulating mutant DNA to assess tumor dynamics, *Nat. Med.* 14 (2008) 985–990.
- [8] M.E. Talkowski, C. Ernst, A. Heilbut, C. Chiang, C. Hanscom, A. Lindgren, A. Kirby, S. Liu, B. Muddukrishna, T.K. Ohsumi, Y. Shen, M. Borowsky, M.J. Daly, C.C. Morton, J.F. Gusella, Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research, *Am. J. Hum. Genet.* 88 (2011) 469–481.
- [9] S. Branford, Chronic myeloid leukemia: molecular monitoring in clinical practice, *Hemat. Am. Soc. Hematol. Educ. Program* (2007) 376–383.
- [10] T. Flohr, A. Schrauder, G. Cazzaniga, R. Panzer-Grumayer, V. van der Velden, S. Fischer, M. Stanulla, G. Basso, F.K. Niggli, B.W. Schafer, R. Sutton, R. Koehler, M. Zimmermann, M.G. Valsecchi, H. Gadner, G. Masera, M. Schrappe, J.J. van Dongen, A. Biondi, C.R. Bartram, Minimal residual disease-directed risk stratification using real-time quantitative PCR analysis of immunoglobulin and T-cell receptor gene rearrangements in the international multicenter trial AIEOP-BFM ALL 2000 for childhood acute lymphoblastic leukemia, *Leukemia* 22 (2008) 771–782.
- [11] E.D. Pleasance, R.K. Cheetham, P.J. Stephens, D.J. McBride, S.J. Humphray, C.D. Greenman, I. Varela, M.L. Lin, G.R. Ordóñez, G.R. Bignell, K. Ye, J. Alipaz, M.J. Bauer, D. Beare, A. Butler, R.J. Carter, L. Chen, A.J. Cox, S. Edkins, P.I. Kokko-Gonzales, N.A. Gormley, R.J. Grocock, C.D. Haudenschild, M.M. Hims, T. James, M. Jia, Z. Kingsbury, C. Leroy, J. Marshall, A. Menzies, L.J. Mudie, Z. Ning, T. Royce, O.B. Schulz-Trieglaff, A. Spiridou, L.A. Stebbings, L. Szajkowski, J. Teague, D. Williamson, L. Chin, M.T. Ross, P.J. Campbell, D.R. Bentley, P.A. Futreal, M.R. Stratton, A comprehensive catalogue of somatic mutations from a human cancer genome, *Nature* 463 (2010) 191–196.
- [12] E.D. Pleasance, P.J. Stephens, S. O'Meara, D.J. McBride, A. Meynert, D. Jones, M.L. Lin, D. Beare, K.W. Lau, C. Greenman, I. Varela, S. Nik-Zainal, H.R. Davies, G.R. Ordóñez, L.J. Mudie, C. Latimer, S. Edkins, L. Stebbings, L. Chen, M. Jia, C. Leroy, J. Marshall, A. Menzies, A. Butler, J.W. Teague, J. Mangion, Y.A. Sun, S.F. McLaughlin, H.E. Peckham, E.F. Tsung, G.L. Costa, C.C. Lee, J.D. Minna, A. Gazdar, E. Birney, M.D. Rhodes, K.J. McKernan, M.R. Stratton, P.A. Futreal, P.J. Campbell, A small-cell lung cancer genome with complex signatures of tobacco exposure, *Nature* 463 (2010) 184–190.
- [13] E.R. Mardis, New strategies and emerging technologies for massively parallel sequencing: applications in medical research, *Genome Med.* 1 (2009) 40.
- [14] L.A. Garraway, E.S. Lander, Lessons from the cancer genome, *Cell* 153 (2013) 17–37.
- [15] K. Chen, J.W. Wallis, M.D. McLellan, D.E. Larson, J.M. Kalicki, C.S. Pohl, S.D. McGrath, M.C. Wendt, Q. Zhang, D.P. Locke, X. Shi, R.S. Fulton, T.J. Ley, R.K. Wilson, L. Ding, E.R. Mardis, BreakDancer: an algorithm for high-resolution mapping of genomic structural variation, *Nat. Methods* 6 (2009) 677–681.
- [16] Y. Drier, M.S. Lawrence, S.L. Carter, C. Stewart, S.B. Gabriel, E.S. Lander, M. Meyerson, R. Beroukhim, G. Getz, Somatic rearrangements across cancer reveal classes of samples with distinct patterns of DNA breakage and rearrangement-induced hypermutability, *Genome Res.* 23 (2013) 228–235.
- [17] J. Wang, C.G. Mullighan, J. Easton, S. Roberts, S.L. Heatley, J. Ma, M.C. Rusch, K. Chen, C.C. Harris, L. Ding, L. Holmfeldt, D. Payne-Turner, X. Fan, L. Wei, D. Zhao, J.C. Obenauer, C. Naevae, E.R. Mardis, R.K. Wilson, J.R. Downing, J. Zhang, CREST maps somatic structural variation in cancer genomes with base-pair resolution, *Nat. Methods* 8 (2011) 652–654.
- [18] A.R. Quinlan, R.A. Clark, S. Sokolova, M.L. Leibowitz, Y. Zhang, M.E. Hurler, J.C. Mell, I.M. Hall, Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome, *Genome Res.* 20 (2010) 623–635.
- [19] P.A. Galante, R.B. Parmigiani, Q. Zhao, O.L. Caballero, J.E. de Souza, F.C. Navarro, A.L. Gerber, M.F. Nicolas, A.C. Salim, A.P. Silva, L. Edsall, S. Devalle, L.G. Almeida, Z. Ye, S. Kuan, D.G. Pinheiro, I. Tojal, R.G. Pedigoni, R.G. de Sousa, T.Y. Oliveira, M.G. de Paula, L. Ohno-Machado, E.F. Kirkness, S. Levy, W.A. da Silva Jr., A.T. Vasconcelos, B. Ren, M.A. Zago, R.L. Strausberg, A.J. Simpson, S.J. de Souza, A.A. Camargo, Distinct patterns of somatic alterations in a lymphoblastoid and a tumor genome derived from the same individual, *Nucleic Acids Res.* 39 (2011) 6056–6068.
- [20] M. Ray, S. Goldstein, S. Zhou, K. Potamouis, D. Sarkar, M.A. Newton, E. Esterberg, C. Kendziorski, O. Bogler, D.C. Schwartz, Discovery of structural alterations in solid tumor oligodendroglioma by single molecule analysis, *BMC Genomics* 14 (2013) 505.
- [21] L. Yang, L.J. Luquette, N. Gehlenborg, R. Xi, P.S. Haseley, C.H. Hsieh, C. Zhang, X. Ren, A. Protopopov, L. Chin, R. Kucherlapati, C. Lee, P.J. Park, Diverse mechanisms of somatic structural variations in human cancer genomes, *Cell* 153 (2013) 919–929.
- [22] A. Malhotra, M. Lindberg, G.G. Faust, M.L. Leibowitz, R.A. Clark, R.M. Layer, A.R. Quinlan, I.M. Hall, Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms, *Genome Res.* 23 (2013) 762–776.
- [23] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M.A. Marra, Circos: an information aesthetic for comparative genomics, *Genome Res.* 19 (2009) 1639–1645.
- [24] H. Ju, K.A. Lee, M. Yang, H.J. Kim, C.P. Kang, T.S. Sohn, J.C. Rhee, C. Kang, J.W. Kim, TP53BP2 locus is associated with gastric cancer susceptibility, *Int. J. Cancer* 117 (2005) 957–960.
- [25] C. Genomes Project, G.R. Abecasis, A. Auton, L.D. Brooks, M.A. DePristo, R.M. Durbin, R.E. Handsaker, H.M. Kang, G.T. Marth, G.A. McVean, An integrated map of genetic variation from 1,092 human genomes, *Nature* 491 (2012) 56–65.
- [26] D.E. Newburger, D. Kashef-Haghighi, Z. Weng, R. Salari, R.T. Sweeney, A.L. Brunner, S.X. Zhu, X. Guo, S. Varma, M.L. Troxell, R.B. West, S. Batzoglou, A. Sidow, Genome evolution during progression to breast cancer, *Genome Res.* 23 (2013) 1097–1108.
- [27] S. Chevillard, A method for sequential extraction of RNA and DNA from the same sample, specially designed for a limited supply of biological material, *Biotechniques* 15 (1993) 22–24.

- [28] W.J. Kent, C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, D. Haussler, The human genome browser at UCSC, *Genome Res.* 12 (2002) 996–1006.
- [29] P. Medvedev, M. Stanciu, M. Brudno, Computational methods for discovering structural variation with next-generation sequencing, *Nat. Methods* 6 (2009) S13–S20.
- [30] S. Levy, G. Sutton, P.C. Ng, L. Feuk, A.L. Halpern, B.P. Walenz, N. Axelrod, J. Huang, E.F. Kirkness, G. Denisov, Y. Lin, J.R. MacDonald, A.W. Pang, M. Shago, T.B. Stockwell, A. Tsiamouri, V. Bafna, V. Bansal, S.A. Kravitz, D.A. Busam, K.Y. Beeson, T.C. McIntosh, K.A. Remington, J.F. Abril, J. Gill, J. Borman, Y.H. Rogers, M.E. Frazier, S.W. Scherer, R.L. Strausberg, J.C. Venter, The diploid genome sequence of an individual human, *PLoS Biol.* 5 (2007) e254.
- [31] S.W. Scherer, J. Cheung, J.R. MacDonald, L.R. Osborne, K. Nakabayashi, J.A. Herbrick, A.R. Carson, L. Parker-Katirae, J. Skaug, R. Khaja, J. Zhang, A.K. Hudek, M. Li, M. Haddad, G.E. Duggan, B.A. Fernandez, E. Kanematsu, S. Gentles, C.C. Christopoulos, S. Choufani, D. Kwasnicka, X.H. Zheng, Z. Lai, D. Nusskern, Q. Zhang, Z. Gu, F. Lu, S. Zeesman, M.J. Nowaczyk, I. Teshima, D. Chitayat, C. Shuman, R. Weksberg, E.H. Zackai, T.A. Grebe, S.R. Cox, S.J. Kirkpatrick, N. Rahman, J.M. Friedman, H.H. Heng, P.G. Pelicci, F. Lo-Coco, E. Belloni, L.G. Shaffer, B. Pober, C.C. Morton, J.F. Gusella, G.A. Bruns, B.R. Korf, B.J. Quade, A.H. Ligon, H. Ferguson, A.W. Higgins, N.T. Leach, S.R. Herrick, E. Lemyre, C.G. Farra, H.G. Kim, A.M. Summers, K.W. Gripp, W. Roberts, P. Szatmari, E.J. Winsor, K.H. Grzeschik, A. Teebi, B.A. Minassian, J. Kere, L. Armengol, M.A. Pujana, X. Estivill, M.D. Wilson, B.F. Koop, S. Tosi, G.E. Moore, A.P. Boright, E. Zlotorynski, B. Kerem, P.M. Kroisel, E. Petek, D.G. Oscier, S.J. Mould, H. Dohner, K. Dohner, J.M. Rommens, J.B. Vincent, J.C. Venter, P.W. Li, R.J. Mural, M.D. Adams, L.C. Tsui, Human chromosome 7: DNA sequence and biology, *Science* 300 (2003) 767–772.
- [32] I. Kozarewa, Z. Ning, M.A. Quail, M.J. Sanders, M. Berriman, D.J. Turner, Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G + C)-biased genomes, *Nat. Methods* 6 (2009) 291–295.
- [33] R. Nielsen, J.S. Paul, A. Albrechtsen, Y.S. Song, Genotype and SNP calling from next-generation sequencing data, *Nat. Rev. Genet.* 12 (2011) 443–451.
- [34] R.E. Handsaker, J.M. Korn, J. Nemes, S.A. McCarroll, Discovery and genotyping of genome structural polymorphism by sequencing on a population scale, *Nat. Genet.* 43 (2011) 269–276.
- [35] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics* 26 (2010) 841–842.
- [36] A. Smit, Hubley, R & Green, P, RepeatMasker Open-3.0, in, 1996–2010.
- [37] J.A. Bailey, A.M. Yavor, H.F. Massa, B.J. Trask, E.E. Eichler, Segmental duplications: organization and impact within the current human genome project assembly, *Genome Res.* 11 (2001) 1005–1017.
- [38] E. Hazkani-Covo, R.M. Zeller, W. Martin, Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes, *PLoS Genet.* 6 (2010) e1000834.
- [39] W.J. Kent, BLAT—the BLAST-like alignment tool, *Genome Res.* 12 (2002) 656–664.
- [40] B. Rhead, D. Karolchik, R.M. Kuhn, A.S. Hinrichs, A.S. Zweig, P.A. Fujita, M. Diekhans, K.E. Smith, K.R. Rosenbloom, B.J. Raney, A. Pohl, M. Pheasant, L.R. Meyer, K. Learned, F. Hsu, J. Hillman-Jackson, R.A. Harte, B. Giardine, T.R. Dreszer, H. Clawson, G.P. Barber, D. Haussler, W.J. Kent, The UCSC Genome Browser database: update 2010, *Nucleic Acids Res.* 38 (2010) D613–D619.
- [41] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, *Nat. Methods* 9 (2012) 357–359.

ANEXO C – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO



**TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO
BTS - HAOC**

I - DADOS DE IDENTIFICAÇÃO DO SUJEITO DA PESQUISA OU RESPONSÁVEL LEGAL

1. NOME DO PACIENTE

DOCUMENTO DE IDENTIDADE Nº : SEXO : .M F

DATA NASCIMENTO:/...../.....

ENDEREÇO Nº..... APTO:

BAIRRO.....CIDADE.....

CEP:..... TELEFONE: DDD (.....)

2.RESPONSÁVEL LEGAL.....

NATUREZA (grau de parentesco, tutor, curador etc.)
.....

DOCUMENTO DE IDENTIDADE :SEXO: M F

DATA NASCIMENTO:/...../.....

ENDEREÇO: NºAPTO:

BAIRRO: CIDADE:.....

CEP: TELEFONE: DDD (.....)

II - DADOS SOBRE A PESQUISA CIENTÍFICA

1. TÍTULO DO PROTOCOLO DE PESQUISA: Banco de Tecido e Seroteca do Hospital Alemão Oswaldo Cruz (BTS-HAOC).

2. INVESTIGADOR PRINCIPAL: Dra. Angelita Habr-Gama, Dr Rodrigo Oliva Perez (contato)
CARGO/FUNÇÃO: Médicos Cirurgiões e Pesquisadores do Departamento de Gastroenterologia FMUSP e do Hospital Alemão Oswaldo Cruz (HAOC)

INSCRIÇÃO CONSELHO REGIONAL Nº 93950

PESQUISADORES ASSOCIADOS: Dr. Joaquim Gama Rodrigues, Dra. Anamaria Camargo, Dr. Sandro de Souza, Jeanne Bronzatti, Lucia Pedras, Igor Proscurshim, Carlos Eduardo Jacob

GERENTE DO PROJETO: Dra. Juçara Parra (contato)

COORDENADOR DO PROJETO: Dr. Joaquim Gama Rodrigues

3. AVALIAÇÃO DO RISCO DA PESQUISA:

SEM RISCO RISCO MÍNIMO RISCO MÉDIO RISCO BAIXO RISCO MAIOR

(probabilidade de que o indivíduo sofra algum dano como consequência imediata ou tardia do estudo)

III - INFORMAÇÕES AO PACIENTE

Este é um documento informativo e possui 12 páginas. Por favor, certifique-se disso. Se o português não é sua língua, avise, pois será fornecido o documento em outra linguagem.

Antes que seja assinado este documento, você poderá levá-lo para casa e discutir com sua família a decisão de participar ou não da pesquisa, doando seu tecido (retirado durante exames ou cirurgias já pré-estabelecidas por seu tratamento) para pesquisa do câncer.

A amostra de tecido a ser armazenada é a sobra do que foi retirado pelo cirurgião e que seria desprezada. Antes do procedimento, você deverá assinar um Formulário de Consentimento para o Banco de Tecido e Seroteca do Hospital Alemão Oswaldo Cruz (BTS-HAOC). Ao assiná-lo você concorda com as informações aqui colocadas e dá o consentimento para doar tecido para o Banco de Tecidos BTS-HAOC.

Todo este material será identificado através de números, e o nome dos pacientes não será revelado, para que se possa assegurar sua privacidade e de seus familiares.

Você tem toda a liberdade para decidir sobre a inclusão de seu tecido no BTS-HAOC assim como em retirar o seu consentimento a qualquer momento e de deixar de participar do estudo sem que isto traga algum tipo de prejuízo ao seu acompanhamento médico.

Uma cópia do consentimento lhe será fornecida, e o original será mantido nos arquivos médicos.

Em caso de dúvidas, solicite uma consulta adicional com o Dr. Rodrigo Oliva Perez. Não tenha pressa para entender este informativo escrito e as explicações e informações orais dadas pelo médico que assina este formulário. Você também pode falar com seu médico particular. Você pode entrar em contato com o Dr. Rodrigo Oliva Perez ou com a profa. Angelita Habr-Gama na Rua Manuel da Nóbrega 1564, Ibirapuera; Tel: 3887-1757; Fax: 3884-8845. Este estudo foi aprovado pelo Comitê de Ética em pesquisa do HAOC. Você pode contatar o Comitê de Ética através do telefone (11) 3549-0000 para mais informações.

CONSENTIMENTO PÓS-ESCLARECIDO

CONSENTIMENTO DO PACIENTE:

- Eu li ou foi lido para mim em uma língua que eu entendi a folha de INFORMAÇÕES AO PACIENTE este Termo de Consentimento
- Eu tive oportunidade para questionar e estou satisfeito com as respostas que eu recebi.
- Eu livremente concordei em doar meu tecido ao Banco de Tecido e Seroteca do Hospital Alemão Oswaldo Cruz (BTS-HAOC) de acordo com as condições na folha de informação ao paciente, após procedimento cirúrgico terapêutico ou para diagnóstico executado por indicação médica.
- Eu autorizo testar minhas amostras quanto a genes que são transmitidos hereditariamente e aumentam os riscos de doenças? () SIM () NÃO
- Eu recebi uma cópia da folha de Informações ao Paciente e receberei uma cópia assinada do Termo de Consentimento para guardar.
- Eu entendi que esta pesquisa pode levar muitos anos para ser concluída e qualquer informação obtida não beneficiará a mim ou minha família pessoalmente ou financeiramente.
- Declaro que, após convenientemente esclarecido pelo pesquisador e ter entendido o que me foi explicado, consinto em participar do presente Protocolo de Pesquisa.

Nome do Participante: _____ Data: / /

Assinatura: _____

Nº Registro: _____

Nome da testemunha: _____ Data: / /

DECLARAÇÃO DA PESSOA PROVEDORA DA INFORMAÇÃO:

Eu dei uma explicação verbal dos objetivos do Banco de Tecido, seus possíveis procedimentos e riscos e acredito que o participante entendeu minhas explicações.

Nome e Título: _____

Assinatura: _____ Data / /

(Nota: Todas as partes que assinam o Termo de Consentimento devem datar sua própria assinatura)

1) O que é um Banco de Tecidos?

R: O Banco de Tecidos é um serviço sem fins lucrativos idealizado com a intenção de facilitar ética e cientificamente a pesquisa usando tecidos humanos. O Banco de Tecidos tem a aprovação da Comissão de Ética para coletar e armazenar sangue e outros fluidos, amostras de tecidos e informações sobre indivíduos saudáveis e pacientes que possam ter câncer ou que possuam risco de desenvolver câncer (por exemplo devido a um histórico familiar de câncer). A coleta da amostra só é feita em pacientes submetidos a terapêutica cirúrgica ou para diagnóstico por indicação médica.

Quando um pesquisador decide estudar certos tipos de câncer ele poderá recorrer ao Banco de Tecidos para obter as amostras necessárias à sua pesquisa, desde que esta esteja dentro dos parâmetros éticos e científicos exigidos sendo aprovada pelo Comitê Diretor Científico do BTS-HAOC e pelo Comitê de Ética em Pesquisa do Hospital Alemão Oswaldo Cruz (CEP HAOC). Portanto, mesmo que você não tenha câncer, o tecido doado ou a biópsia podem beneficiar a pesquisa médica.

O BTS-HAOC localiza-se no Instituto Ludwig de Pesquisa Sobre o Câncer: Hospital Alemão Oswaldo Cruz - Rua João Julião, 245 - Prédio de Apoio - 1º andar - Paraíso CEP: 01323-903 - São Paulo - SP – Brasil / Telefone: 11 - 3388 3200

2) O que é requisitado de mim?

R: Cientistas e médicos que estudam os aspectos bioquímicos e genéticos de certos tipos de câncer buscando novas formas de diagnóstico, tratamento e prevenção precisam realizar pesquisas com o sangue e tecidos de um grande número de pacientes com doença semelhante. Este tipo de pesquisa depende de pessoas como você, que doam, não remuneradamente, amostras de sangue, fluidos e tecido para a pesquisa. O Banco de Tecidos apenas coleta, armazena e distribui amostras com a permissão da pessoa que as doou. Portanto, estamos pedindo sua permissão para realizar cinco procedimentos: 1. Coletar 25-50ml de sangue de uma veia em seu braço. Essa amostra deve ser coletada o mais próximo possível de sua cirurgia. 2. Coletar amostras adicionais durante seu seguimento após o procedimento. 3. Após a cirurgia, o tecido que foi removido é enviado ao laboratório para testes diagnósticos, conforme previsto no procedimento Nós precisaríamos de uma pequena amostra do

tecido que não foi enviado para testes diagnósticos e que seria descartada. 4. Gostaríamos de seguir sua evolução após o procedimento, coletando informações de agências de saúde, prontuários médicos, etc. Essas informações são importantes para caracterizar melhor a amostra de tecido que você doou. 5. Você será requisitado a preencher uma pesquisa ou questionário que fornecem aos pesquisadores informações importantes para suas pesquisas.

3) Mas para que seja feito o diagnóstico correto não é necessário todo o tecido retirado?

R: Normalmente, não. Se existe um tumor, o cirurgião vai removê-lo, juntamente com tecido normal ao seu redor para garantir que todo o tecido doente foi retirado. Este espécime é enviado ao laboratório de patologia, onde um patologista o examina. O patologista seleciona áreas do tecido, que serão processadas em blocos de parafina e depois cortadas em fatias muito finas, para permitir a visualização no microscópio; fazendo então o diagnóstico e enviando um laudo ao médico. Após o diagnóstico ter sido feito, normalmente ainda resta tecido nesses blocos, que é então armazenado no laboratório.

4) Se eu doar tecido, quanto o Banco de Tecidos retirará?

R: O Banco de Tecidos retirará apenas tecido que não é necessário ao patologista para estabelecer seu diagnóstico final. Este tecido é considerado “excesso” e normalmente é destruído quando não foi doado a algum Banco de Tecidos. Não há nenhuma garantia de que seu tecido será retirado pelo Banco de Tecidos, isso é o patologista que decide. Se há tecido que pode ser retirado pelo Banco de Tecidos, ele será armazenado a fresco ou em blocos de parafina, para pesquisas futuras. Normalmente, recebemos pedaços de tecido a fresco, de tecido normal e anormal ou tumoral, no tamanho de uma moeda de 5 centavos. Se você concordar em ser um doador(a), não será realizado nenhum procedimento adicional a menos que você concorde em ter amostras de sangue retiradas durante o seguimento.

Medula óssea: Se seu médico requisitou uma biópsia de medula óssea como parte do planejamento de seu tratamento clínico, nós pedimos que você doe uma quantidade

extra de medula óssea para a finalidade de pesquisas científicas. Somente um procedimento será realizado. A coleta da quantidade extra de medula óssea (2-10ml) será realizada ao mesmo tempo que a coleta para o exame diagnóstico requisitado.

5) Por que vocês necessitam de sangue ou medula óssea?

R: Se você nos permitir coletar o sangue, 25-50ml de sangue serão retirados de uma veia de seu braço. É o mesmo procedimento de um exame de sangue normal, com a diferença que você não receberá nenhum resultado. Tentaremos realizar essa coleta ao mesmo tempo em que outras amostras serão retiradas para exames que seu médico requisitou. O mesmo acontecerá com a biópsia de medula óssea. Seu médico apenas retirará uma quantidade extra (2-10ml) de medula óssea durante a coleta. Todo esse material será utilizado por pesquisadores para avaliar proteínas e o material genético (DNA e RNA) presente em suas células. Muitos tipos de câncer e doenças sanguíneas não podem ser detectados precocemente, mas os cientistas estão trabalhando arduamente para desenvolver testes sanguíneos que possam alertar as pessoas sobre tais cânceres enquanto eles ainda estão em seus estágios iniciais, em que há uma boa chance de serem tratados e curados. Para desenvolver novos testes, os pesquisadores precisam testar centenas de amostras de sangue de doadores como você. Uma forma de fazer isto é coletar e comparar amostras sanguíneas múltiplas do mesmo paciente ao longo de um período. Pesquisadores podem então estudar e comparar os níveis de determinada substância (ex. proteínas e hormônios) antes e após a cirurgia. Esses resultados podem ser utilizados para descobrir as bases genéticas da doença ou determinar importantes biomarcadores que podem nos dar informações sobre a doença, como prognóstico. Se você permitir que sejam coletadas amostras sanguíneas em seu seguimento, coletaremos 25-50ml em intervalos. Esse procedimento normalmente ocorre durante sua consulta, e é totalmente voluntário, de forma que você pode parar de doar essas amostras quando desejar.

6) Quem usará meu tecido?

R: Pesquisadores nacionais ou estrangeiros podem requisitar tecidos do BTS-HAOC, uma vez que seus projetos tenham aprovação da Comissão de Ética responsável.

7) Que tipos de pesquisas serão realizadas com meu sangue e tecido?

R: O Banco de Tecidos e Seroteca do HAOC fornece amostras para pesquisadores envolvidos em vários aspectos da pesquisa sobre câncer, de tal forma que não podemos dizer exatamente em quais projetos suas amostras serão utilizadas agora ou no futuro. Alguns cientistas podem estudar o material genético presente em suas células, e outros as proteínas por ela produzidas. De qualquer forma, suas amostras serão utilizadas para estudar causas de câncer e/ou desenvolver métodos para a detecção, diagnóstico, monitoramento e/ou tratamento do câncer. Outros cientistas podem ainda estudar como as doenças se expressão nas famílias, comparando material genético (proteínas ou genes) presentes em células normais e tumorais em grupos semelhantes de pacientes, o que é chamado de pesquisa genética.

Algumas vezes o pesquisador precisa ter acesso aos blocos de parafina arquivados nos laboratórios de patologia e que foram utilizados para fazer o diagnóstico de câncer. Isso ocorre normalmente quando o Banco de Tecidos não tem material suficiente ou quando o material disponível já foi oferecido a outro pesquisador. Nós somente poderemos ter acesso a esses arquivos através de sua permissão e da permissão do Departamento de Patologia.

Células obtidas de seu sangue ou tecido podem ser utilizadas para estabelecer linhagens celulares. Uma linhagem celular consiste em células que crescem em meio de cultura no laboratório por um longo período, permitindo a criação de uma grande fonte de material para pesquisa, e permitindo que o material seja compartilhado com outros grupos, e permita a comparação de resultados entre eles. Somente pesquisadores com um certificado específico podem estabelecer culturas de células ou realizar pesquisas genéticas.

Avanços rápidos na tecnologia tornam impossível prever quais testes ou estudos serão possíveis no futuro. Dessa forma, quando um pesquisador tiver a intenção de utilizar seu tecido para pesquisa no futuro, uma comissão de ética avaliará a pesquisa e decidirá se ela pode continuar ou não.

8) Meu tecido será vendido para outra companhia com fins lucrativos?

R: O Banco de Tecidos é um serviço sem fins lucrativos dedicado a fornecer fontes de pesquisa a cientistas envolvidos na pesquisa médica. Seus tecidos serão fornecidos a pesquisadores de locais como hospitais, universidades, institutos de pesquisa médica, instituições governamentais e organizações comerciais como companhias farmacêuticas. Seu tecido não será vendido. O Banco de Tecidos pode eventualmente cobrar dos pesquisadores uma tarifa para cobrir gastos com armazenamento e administração de suas coleções de tecidos, mas o tecido nunca é vendido. Mesmo que o conhecimento adquirido com a pesquisa médica leve a descobertas de valor comercial para o pesquisador e sua instituição, não haverá nenhum benefício financeiro para você, suas famílias ou para o Banco de Tecidos.

9) Eu poderei saber dos resultados das pesquisas que utilizaram meus tecidos?

R: Você tem o direito legal de acessar seus registros médicos quando desejar. No decorrer do tratamento você irá receber de seu médico resultados sobre o seu procedimento. Esta informação é parte de seu registro médico permanente, no entanto você não receberá resultados específicos das pesquisas. Isto ocorre porque as pesquisas são de caráter experimental, podem levar vários anos e usar amostras e informações de um grande número de pacientes. Dessa forma, o que o pesquisador descobre provavelmente não tem valor direto para você, sua família ou sua saúde futura.

10) O que acontecerá se o pesquisador descobrir algo de importância médica para mim ou minha família?

R: A grande maioria dos projetos de pesquisa para os quais seu tecido será destinado não revelarão algo de importância médica para você ou sua família. Entretanto, se isso ocorrer e for realizada uma descoberta que possa interferir em sua condição médica, o pesquisador deve informar a Comissão de Ética responsável e o Banco de Tecidos. A comissão analisará a pesquisa e decidirá se é importante que você seja contatado. Se o novo conhecimento for importante para você e/ou sua família o Banco de Tecidos entrará em contato com o seu médico e você poderá saber mais detalhes assim que

desejar. Se não conseguirmos entrar em contato com seu médico, tentaremos nos comunicar diretamente com você, nos endereços e telefones fornecidos.

11) Quais são os riscos e benefícios da doação de tecidos e pesquisa genética?

R: A pesquisa genética envolve o estudo do material genético (DNA), que é compartilhado com seus parentes consanguíneos. Esse tipo de pesquisa pode levantar importantes questões. É improvável que tais questões apareçam, mas você deve pensar sobre elas cuidadosamente. A pesquisa genética é realizada com o objetivo de descobrir métodos mais precisos de verificar o risco de doenças em um grupo de pessoas ou em pessoas com uma forte história familiar ou predisposição à doença. Para realizar tais estudos o pesquisador deve demonstrar para a Comissão de Ética que os procedimentos são para fornecer aconselhamento genético.

Benefícios:

É improvável que doando seu tecido haja algum benefício direto para você ou sua família. Planejamos armazenar seu tecido por um longo período, e pode ser que ele seja utilizado apenas daqui vários anos. Portanto, o benefício maior será para as futuras gerações de pacientes com câncer.

A pesquisa genética pode fornecer meios de se identificar pessoas em uma família com um risco aumentado de desenvolver doenças para as quais existe um tratamento conhecido, ou seja, possível a cura.

Essa pesquisa pode também permitir que um indivíduo ou sua família tomem atitudes para prevenir a doença antes que ela aconteça, fazendo, por exemplo, testes de “screening” se disponíveis ou realizando mudanças de hábitos de vida, se necessário.

Esclarecimentos:

O tecido que coletamos é obtido após ter sido removido de seu corpo e seria normalmente destruído, de tal forma que não existe risco físico para você. Quando o sangue é coletado você pode ter um certo desconforto no braço e desenvolver um pequeno hematoma. Se a amostra adicional de medula óssea é retirada, pode causar desconforto ou dor no local da aspiração, o que dura poucos segundos. Após o procedimento diagnóstico terminar pode haver alguma dor por dois ou três dias. Se

Se você resolver participar da pesquisa, pode se sentir aborrecido se, por exemplo, a pesquisa é realizada em um momento em que você não se sente bem ou se sente ansioso (a). É claro que você pode parar de responder a pesquisa, ou não responder alguma questão em particular se assim desejar.

Conhecer os resultados de uma pesquisa genética pode trazer incertezas e ser inquietante se, por exemplo, é identificado o risco de desenvolver alguma doença sem tratamento, prevenção ou cura conhecidos.

Algumas pessoas podem descobrir informações sobre doenças hereditárias envolvendo filhos, irmãos e irmãs, o que pode interferir nas relações familiares. Você pode se deparar com questões como “Devo contar à família?”, já que outros membros da família podem desejar não descobrir tais informações. É importante entender que os resultados de uma pesquisa genética normalmente não indicam que você tem alguma doença, ou que você vá desenvolver alguma; e indica somente que você tem um risco de desenvolver determinada doença, de tal forma que não há certeza de que você desenvolverá o quadro.

Qualquer resultado que seja de importância para você ou sua família deve ser confirmado, o que pode incluir ter outra amostra de sangue coletada para testes. Essa é uma prática padrão para pacientes que recebem resultados de testes genéticos e será realizada gratuitamente para você. Aconselhamento genético também será realizado gratuitamente se necessário. Antes do teste ser repetido para verificar os resultados da pesquisa você será informado dos riscos envolvidos. Isto é especialmente importante para as pessoas que descobrem que tem uma mutação que aumenta o risco de desenvolver alguma doença, como o câncer. Saber de tais informações pode levar a discriminações no trabalho ou dificuldades em conseguir uma cobertura de planos de saúde, por exemplo. Porém, todos os dados e informações individuais, ou que caracterizem um indivíduo desta pesquisa serão mantidas em segredo, não haverá violação da privacidade de nenhum paciente. No formulário de Consentimento Informado você poderá optar por ter ou não suas amostras testadas quanto a genes que são transmitidos hereditariamente e aumentam os riscos de doenças. Se você não deseja ter suas amostras testadas, selecione “Não” quando for perguntado no Consentimento Informado. Qualquer que seja sua escolha, ela será respeitada.

12) Que outras informações serão necessárias?

R: O Banco de Tecidos coletará informações de seus registros médicos como peso, altura e pressão arterial antes do procedimento e outros detalhes sobre seu estado, como diagnóstico e resultado de exames. Também gostaríamos de acompanhar sua evolução após o procedimento cirúrgico através das informações coletadas por seu médico nas consultas de seguimento. Seu médico anotará informações tais como: quais medicamentos você está tomando e sua dieta. Você não precisará comparecer a consultas extras por isso. Também gostaríamos de sua permissão para enviar pesquisas ou questionários aprovados pela Comissão de Ética, o que nos permitirá coletar informações padronizadas sobre todos os doadores do Banco de Tecidos. São questionários sobre seu estilo de vida, perguntando por exemplo sobre o hábito de fumar ou de realizar exercícios físicos. Gostaríamos também de obter e/ou verificar suas informações armazenadas em agências de saúde ou outros registros médicos. Nenhuma informação pessoal será coletada sem sua permissão.

13) Como é garantida minha privacidade?

R: Suas informações pessoais serão mantidas em sigilo, exceto aquelas requeridas por lei. Apesar das informações coletadas pelo Banco de Tecidos sobre você serem identificáveis (elas ficam relacionadas ao seu nome ou registro médico) o Banco de Tecidos as manterá em sigilo para manter sua privacidade. Informações sobre você não serão fornecidas para ninguém sem seu consentimento por escrito. As informações fornecidas aos pesquisadores não incluem informações pessoais como seu nome ou data de nascimento, no lugar é colocado um código, que pode ser reconhecido somente pelos funcionários autorizados do Banco de Tecidos. Os tecidos e informações são fornecidos aos pesquisadores codificados para que não possa se identificar você ou sua família. Para garantir todo esse processo o Banco de Tecidos passa por auditorias regulares.

14) O que acontecerá com meus tecidos armazenados se eu vier a falecer?

R: No caso de morte, continuaremos armazenando seus tecidos e disponibilizando-os para os pesquisadores. Sua doação continuará a ser utilizada nas pesquisas sob as mesmas condições éticas e legais aqui discutidas. No Consentimento Informado você

pode nomear um representante que será contatado quando precisarmos de alguma decisão. Essa pessoa pode ser um parente consanguíneo ou esposo(a). Discuta as possíveis implicações da doação com a pessoa que você nomeará como seu representante.

15) E se eu não doar tecido?

R: A sua decisão de doar ou não tecido não afetará em nada seu tratamento agora ou no futuro. Você não tem nenhuma obrigação em doar tecido ou sangue. Se não doar, o tecido retirado durante o procedimento e que não é necessário para o diagnóstico será destruído, da maneira usual.

16) E se eu mudar minha decisão?

R: Sua participação é completamente voluntária. Você pode mudar de ideia e retirar seu consentimento a qualquer momento, e isso não afetará seu tratamento. Para retirar seu consentimento simplesmente contate um representante do Banco de Tecidos. As amostras armazenadas serão destruídas.

17) Quem eu posso contatar se tiver mais dúvidas?

R: O Banco de Tecidos fica aberto no horário comercial. Se você tem qualquer dúvida técnica sobre como suas amostras estão sendo armazenadas ou dúvidas quanto às informações que estão sendo coletadas ou quanto aos projetos de pesquisa em andamento, você pode solicitar uma consulta adicional com o Dr. Rodrigo Oliva Perez. Não tenha pressa para entender este informativo escrito e as explicações e informações orais dadas pelo médico que assina este formulário. Você também pode falar com seu médico particular.

A qualquer momento você pode entrar em contato com o Dr. Rodrigo Oliva Perez ou com o prof. Angelita Habr-Gama através do endereço e telefone abaixo:

Endereço: Manuel da Nóbrega 1564, Ibirapuera; Tel: 3887-1757; Fax: 3884-8845

O estudo foi analisado e aprovado pelo Comitê de Ética do Hospital Alemão Oswaldo Cruz. Se você tiver quaisquer dúvidas sobre a ética da pesquisa, seus direitos como um paciente de pesquisa ou caso tenha uma lesão relacionada ao estudo, você pode contatar o Comitê de Ética através do telefone (11) 3549-0000 para mais informações.

18) Quem eu posso contatar que é independente do Banco de Tecidos?

R: Prof. Dr. José Fernando Perez (Diretor presidente da Recepta Biopharma)

Tel: 11 37092600

É de extrema importância que haja um representante independente do Banco de Tecidos para defender os interesses do sujeito de pesquisa, caso haja necessidade. Por esta razão solicitou-se a anuência do Prof. José Fernando Perez – ex-diretor científico da FAPESP por três mandatos. Professor Titular de FÍSICA da Universidade de São Paulo e atualmente trabalhando em Pesquisa Biológica e Translacional.

ANEXO D – APROVAÇÃO DO PROJETO NO COMITÊ DE ÉTICA

Parecer Consubstanciado

Título da Pesquisa: "Criação do banco de tecidos e seroteca do Hospital Alemão Oswaldo Cruz"

Pesquisador Principal: Dr. Joaquim José Gama Rodrigues

Instituição: Hospital Alemão Oswaldo Cruz – SP

Sumário Geral do Protocolo: Recebemos para avaliação o Projeto: Criação do Banco de Tecidos e Seroteca do Hospital Alemão Oswaldo Cruz (BTS-HAOC). Entendemos que o BTS-HAOC tem como objetivo coletar e armazenar amostras de tecidos (amostras de tumores, tecidos controle, sangue, plasma, medula óssea) de pacientes submetidos a tratamento cirúrgico por neoplasia do sistema digestivo.

O BTS-HAOC disponibilizará material biológico e dados epidemiológicos, psico-sociais e clínico-patológicos, para estudos relacionados a características genéticas e moleculares de tumores e dos pacientes com possíveis desdobramentos na prática clínica de prevenção, diagnóstico, estadiamento e tratamento destas doenças;

Julgamos, entretanto, a criação do Banco de Tecidos e Seroteca medida de grande importância na ampliação do horizonte da pesquisa clínica em oncologia ao nível molecular do Hospital Alemão Oswaldo Cruz, haja vista a demanda crescente de projetos nesta área.

Dois projetos de pesquisa, relacionados à criação do BTS-HAOC, "Estudo das micrometástases na medula óssea de pacientes com câncer do cólon e do reto" e "Pesquisa da expressão gênica relacionada à resposta do adenocarcinoma do reto ao tratamento neoadjuvante com radioterapia associada à quimioterapia", estão no momento sendo submetidos à apreciação do CEP-HAOC.

Por não se tratar de projeto de pesquisa *per se*, mas um projeto de uma estrutura que beneficiará diversos projetos de pesquisa, a norma ou regulamento deverá ser submetido à aprovação do CEP, conforme a Resolução CNS no 347. No entanto, toda nova pesquisa a ser feita com o material biológico do BTS-HAOC deverá ser submetida à aprovação do CEP.

Considerações/Comentários:

As pendências relacionadas no parecer de 11/08/2008 foram atendidas conforme solicitado. Mediante tal resposta, o projeto de pesquisa foi analisado e considerado **APROVADO**.

- APROVADO
 APROVADO COM RECOMENDAÇÕES
 NÃO APROVADO
 COM PENDÊNCIAS

(A ausência de resposta em 30 dias acarretará em arquivamento do processo por falta de interesse do pesquisador)

TEMÁTICA ESPECIAL SIM NÃO
CONEP SIM NÃO Completo Folha de rosto

- Lembramos que os pesquisadores têm como dever enviar relatórios anuais e o relatório final ao CEP-HAOC.
- O CEP também deve ser informado sobre qualquer "alteração" ou "emenda" ao referido projeto, e sobre quaisquer efeitos adversos.

Data do 1º Relatório Parcial: 22/03/09

Data do 2º Relatório Parcial: 22/09/09


Dr. José Marcio Neves Jorge
Coordenador do Comitê de Ética em Pesquisa
Hospital Alemão Oswaldo Cruz

ANEXO E – SÚMULA CURRICULAR

SÚMULA CURRICULAR

DADOS PESSOAIS

Nome: Paola de Avelar Carpinetti-Oliveira

Local e data de nascimento: Caraguatatuba-SP, 25/03/1984

FORMAÇÃO

Agosto/2010 - Janeiro/2015

Doutorado no Departamento de Bioquímica do Instituto de Química – Programa de Ciências Biológicas (Bioquímica)
Universidade de São Paulo

Maio/2006 - Julho/2010

Bacharelado em Bioquímica
Universidade Federal de Viçosa

ATIVIDADES COMPLEMENTARES

Dezembro/2013 – Treinamento QX 200 – PCR Digital (Bio-Rad)

Fevereiro - Julho/2013 – Monitoria na disciplina de Bioquímica (QBQ 0313)
Departamento de Bioquímica do Instituto de Química da Universidade de São Paulo

Julho - Dezembro/12 – Monitoria na disciplina de Bioquímica Experimental (QBQ 0316)
Departamento de Bioquímica do Instituto de Química da Universidade de São Paulo

Fevereiro - Julho/12 – Monitoria na disciplina de Bioquímica e Bio. Mol. (QBQ 0102)
Departamento de Bioquímica do Instituto de Química da Universidade de São Paulo

Março - Julho/2010 – Tutoria em Bioquímica Fundamental (BQI 096)
Departamento de Bioquímica da Universidade Federal de Viçosa

Outubro/2009 - Julho/2010 – Monitoria na disciplina Bioquímica Fisiológica (BQI 241)
Departamento de Bioquímica da Universidade Federal de Viçosa

Agosto/2008 - Julho/2010 – Iniciação Científica sob orientação da professora Elizabeth Pacheco Batista Fontes, bolsista CNPq. Laboratório de Biologia Molecular de Plantas (INCT/IPP)

OCUPAÇÃO

Bolsista de Doutorado, CAPES, 01/08/2010 - 31/07/2014

PUBLICAÇÕES

1. ICRmax: an optimized approach to detect tumor-specific InterChromosomal Rearrangements for Clinical Application
Donnard ER, **Carpinetti PA**, Navarro FCP, Perez RO, Habr-Gama A, Parmigiani RB, Camargo AA and Galante PAF. *Genomics*, Feb 7 2015. [doi:10.1016/j.ygeno.2015.01.009]