

**VALIDAÇÃO DE CLASSIFICADORES MOLECULARES
PREDITORES DE RESPOSTA À QUIMIOTERAPIA
COMBINADA COM RADIOTERAPIA EM CARCINOMA
EPIDERMÓIDE DE LARINGE E HIPOFARINGE
LOCALMENTE AVANÇADO**

BIANCA DIANA BARRETO

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

Área de Concentração: Oncologia

Orientador: Dr. Luiz Fernando Lima Reis

Co-Orientador: Dr. Alex Fiorini Carvalho

São Paulo

2009

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Barreto, Bianca Diana

Validação de classificadores moleculares preditores de resposta à quimioterapia combinada com radioterapia em carcinoma epidermóide de laringe e hipofaringe localmente avançado / Bianca Diana Barreto – São Paulo, 2009.

154p.

Tese (Doutorado)-Fundação Antônio Prudente.

Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.

Orientador: Luiz Fernando Lima Reis

Descritores: 1. ESTUDOS DE VALIDAÇÃO COMO ASSUNTO. 2. PERFIL DA EXPRESSÃO GÊNICA. 3. MARCADORES BIOLÓGICOS. 4. CÂNCER DE CABEÇA E PESCOÇO. 5. CARCINOMA EPIDERMÓIDE. 6. TGF-ALFA. 7. QUIMIOTERAPIA. 8. RADIOTERAPIA. 9. VALIDAÇÃO INDEPENDENTE. 10. CLASSIFICADOR MOLECULAR. 11. PREDITOR DE RESPOSTA.

“The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality.”

~Albert Einstein, EUA (nascido na Alemanha), físico (1879 - 1955)

“É imoral pretender que uma coisa desejada se realize magicamente, simplesmente porque a desejamos. Só é moral o desejo acompanhado da severa vontade de prover os meios da sua execução.”

~José Ortega y Gasset, Espanha, filósofo (1883-1955)

DEDICATÓRIA

*Aos meus pais,
pela formação, valores e princípios que me ensinam,
pelo carinho e apoio constantes que me dão.*

*Ao Sergio, pelo seu grande amor e amizade,
por ter te encontrado e fazer parte da minha vida.*

AGRADECIMENTOS

Ao meu orientador Dr. Luiz Fernando Lima Reis, desde o mestrado, por todas as lições e ensinamentos, e por tornar o ambiente de trabalho agradável. Por transmitir calma e alívio nos momentos de maior desespero. Pela confiança em mim e pela oportunidade.

Ao meu co-orientador Dr. Alex, por ser sempre acessível e mostrar o lado positivo das coisas, pelos ensinamentos experimentais e ajudas com problemas técnicos.

Ao médico Dr. Luiz Paulo Kowalski, pelos ensinamentos clínicos e idéias que contribuíram para a realização deste projeto.

Ao Departamento de Cabeça e Pescoço do Hospital AC Camargo, em especial ao Dr. Ulisses Nicolau e ao ex-membro Dr. Olavo Feher, que, em 1999, iniciou o ensaio clínico fase II, quando eu estava iniciando a faculdade, que originou todo o resto....

Aos membros do Instituto de Matemática e Estatística da USP, Jordão e Ana, e aos ex-membros Hirata, Gustavo, Elier e Lucas, por todas as análises, por todo o aprendizado e principalmente pela paciência de ensinar a “matemática além do alcance”. Especialmente à Ana, por corrigir a tese.

Ao Departamento de Oncologia Clínica do Hospital AC Camargo, em especial a Ludmila, pela ajuda no encaminhamento dos pacientes e ao residente Thiago, pelo acompanhamento dos mesmos.

Ao Departamento de Radioterapia e Quimioterapia do Hospital AC Camargo.

Ao Departamento de Anatomia Patológica do Hospital AC Camargo, especialmente Dr. Fernando Soares e Dra. Isabela, pelas ajudas com as microdissecções e análises imunoistoquímicas. À Miyuki, Carlos e Severino, pelas lâminas histológicas.

Ao Laboratório de Bioinformática do Hospital AC Camargo, em especial à Dra. Helena, Renato, Luiz Paulo e Daniel, pela ajuda com as análises de controle de qualidade dos dados.

Aos membros do Comitê de Ética e Pesquisa do Hospital AC Camargo, pelo cuidado para com os projetos.

Aos funcionários do SAME, pela retirada para pesquisa e organização dos prontuários médicos.

Às bibliotecárias do Hospital AC Camargo, especialmente a Suely, pela ajuda com os artigos e formatações.

Ao Departamento de Pós-Graduação, especialmente Ana Maria, Luciana e a ex-funcionária Márcia.

Aos integrantes da Banca de Qualificação, pelas críticas construtivas.

Aos funcionários e estudantes do Hospital AC. Camargo, da Fundação Antônio Prudente e do ex-instituto Ludwig.

Ao Dr. Ricardo Brentani, pela brilhante direção do Hospital AC Camargo e por crer na Pesquisa.

À Fundação Antônio Prudente do Hospital AC Camargo, instituição de excelência, pela possibilidade de realização deste trabalho.

Ao Hospital Araújo Jorge, em Goiânia, pela colaboração neste projeto, especialmente ao Dr. José Carlos.

Ao Hospital de Câncer de Barretos - Fundação Pio XII, pela colaboração neste projeto, especialmente aos Dr. Carlos Roberto, Dr. André, Cristiane e Tamara.

Ao Hospital do Câncer do Ceará, em Fortaleza, pela colaboração neste projeto, especialmente aos Dr. Ricardo e Dra. Rosane.

À FAPESP, pelo suporte financeiro e credibilidade.

Aos pacientes, que aceitaram doar fragmento de suas biópsias para o estudo, mesmo em um momento tão difícil de suas vidas. É por acreditar que uma vida pode ser melhorada que “isso tudo” valeu e vale a pena.

A todos os integrantes do laboratório, pelos momentos de muito aprendizado e ajudas, e ainda de muita diversão, alegria e amizade, Ana Helena, Nair, Vladimir, Juliana, Camila, Graziela, Anderson e Alex, e aos ex-membros, Aline, Adriana, Ana Coló, Patrícia, Tati, Beatriz, Bárbara, Luciana, Lara, Sibebe, Regina, Suzana, Eduardo, Letícia, Sarah, Mari S, Mari M, Cham e Waleska. Cada um contribui da sua forma e de modo muito importante e especial para todo este trabalho se realizar. Sem dúvida a união faz a força e os amigos que nos resgatam nos momentos de fraqueza fazem toda a diferença.

À super secretária Aline, pela presteza e pelas ajudas de todos os tipos e horas.

A todos os meus amigos, especialmente a Ana Paula, Giovana, Renata, Erika e Nancy. Apesar de não me ajudarem diretamente no laboratório, ajudam indiretamente. Me ajudam na vida com o ombro e coração amigos de todas as horas. Não importa a distância nem o tempo que se passe, são eternos.

Aos meus amigos de Houston, especialmente à Mirabela, Paromita, Rachid, Ulass, Abdullah, Jay, Thea e Marlon, por amenizar a dor dos momentos de saudade. Graças a vocês o que parecia impossível se tornou possível.

Aos meus familiares, em especial a minha avó Yolanda e ao meu avô João *in memoriam*, por serem maravilhosos e muito especiais.

À minha mãe Júlia, ao meu pai Percival e ao meu irmão Fábio, por tantas coisas..., por tudo. Por estarem sempre comigo e torcendo por mim. Por serem os responsáveis por parcela grande e importante do que sou. Acreditam tanto em mim que sinto até medo...

Ao Sergio, meu namorado e meu melhor amigo que se tornou meu marido, pelo amor, incentivo e ajudas constantes. Por ser a pessoa que é. Por estar sempre lá e ser meu porto seguro.

RESUMO

Barreto BD. **Validação de classificadores moleculares preditores de resposta à quimioterapia combinada com radioterapia em carcinoma epidermóide de laringe e hipofaringe localmente avançado.** São Paulo; 2009. [Tese de Doutorado-Fundação Antônio Prudente].

Quimiorradioterapia combinada se tornou o padrão ouro de tratamento para carcinoma epidermóide (CE) de laringe e hipofaringe localmente avançado. Mas, pacientes não respondedores ao tratamento sofrem com o avanço da doença, ressecção posterior mais invasiva, toxicidade evitável e maior custo do tratamento. Até o momento, não existem preditores de resposta suficientemente validados para possibilitar a utilização na rotina clínica. Assim, se torna importante a identificação de classificador que discrimine, a partir de biópsia prévia, pacientes respondedores (R) de não respondedores (NR). Em estudo prévio, analisamos 35 pacientes com CE de laringe e hipofaringe, localmente avançado, tratados com quimiorradioterapia. Através de *microarray*, do método do Discriminante Linear de Fisher e posterior validação cruzada, identificamos 4 trios de genes e 27 quadras, capazes de classificar pacientes R e NR com 100% de acerto. Para avaliar se os padrões discriminatórios identificados eram devido ao acaso e sua possibilidade de utilização na rotina clínica, é necessário realizar validação independente dos classificadores. Assim, novo grupo de 28 pacientes, com as mesmas características do estudo anterior, teve biópsia coletada e sua expressão gênica analisada. As regras de classificação identificadas anteriormente foram aplicadas a este novo conjunto, e os trios e quadras de genes classificadores foram validados; descartando assim a identificação de padrão discriminatório devido ao acaso; o que corrobora para sua utilização na rotina clínica. Foi ainda identificada árvore de decisão, capaz de discriminar os grupos R e NR com maiores taxas de sensibilidade (90,4%) e especificidade (71,4%), comparadas aos trios e quadras isoladamente. Finalmente, análise de *array* apresentou TGFA superexpresso em ambos os grupos de NR, seguido de validação imunoistoquímica. Este gene pode estar relacionado com a resistência dos pacientes ao tratamento.

SUMMARY

Barreto BD. [**Validation of molecular classifiers for chemoradiotherapy response prediction of locally advanced larynx and hypopharynx tumors**]. São Paulo; 2009. [Tese de Doutorado-Fundação Antônio Prudente].

Concurrent chemoradiotherapy has become the gold standard for the treatment for the locally advanced SCC of the larynx and the hypopharynx. However, non-responders suffer from disease progression, a more aggressive surgery and preventable toxicity, besides an increased treatment cost. Up to the present date, there are no accurately validated predictors of response to treatment to allow their use in the routine clinical practice. Thus, it is important the identification of a classifier that it able to discriminate, from a previous biopsy, responder (R) patients from non-responder (NR) ones. In a preceding work, we analyzed samples from 35 patients with locally advanced SCC of the larynx and the hypopharynx, treated with chemoradiotherapy. Using microarray, Fisher's Linear Discriminant analysis and subsequent cross validation analysis (leave-one out), we identified 4 trios of genes and 27 quartets, being them able to discriminate R and NR patients with 100% accuracy. An independent validation of these classifiers is required in order to access overfitting as well as their potential use in routine clinical practice. Thus, a new set of 28 patients, all with the same characteristics of the ones in the previous study, had a biopsy sample obtained and their gene expression profiling analyzed. The previous identified classification rules were then applied to the independent group, and the trios and quartets of genes were validated; excluding overfitting and corroborating their use in routine clinical practice. Furthermore, classification trees were identified, being the best one able to discriminate R and NR groups with superior rates of sensitivity (90,4%) and specificity (71,4%), in comparison with the trios and quartets in isolation. At last, microarray analysis showed TGFA overexpressed in NR patients of both groups analyzed, followed by immunohistochemical validation. This gene could be related to the patient's resistance to the treatment.

LISTA DE FIGURAS

Figura 1	Avaliação da integridade do RNA total extraído de amostras tumorais.....	55
Figura 2	Avaliação da integridade do RNA referência extraído.....	58
Figura 3	Avaliação da integridade do RNA referência amplificado em um ciclo.....	58
Figura 4	Avaliação da integridade do RNA referência amplificado em dois ciclos.....	62
Figura 5	Avaliação da integridade de RNAs amplificados das amostras tumorais.....	68
Figura 6	Avaliação da integridade de RNAs amplificados das amostras tumorais.....	68
Figura 7	Lâminas de <i>microarray</i> hibridizadas.....	75
Figura 8	Quantificação de lâmina.....	78
Figura 9	<i>Cluster</i> hierárquico.....	92
Figura 10	Gráficos <i>MM plot</i>	94
Figura 11	Árvore A de decisão.....	100
Figura 12	Árvore B de decisão.....	102
Figura 13	Gráficos <i>boxplot</i> dos genes das vias.....	105

Figura 14	Fotomicrografias de imunistoquímica para TGFA.....	108
Figura 15	Gráficos <i>boxplot</i> dos genes EGFR e TGFA.....	110
Figura 16	Comparação dos <i>clusters</i> hierárquicos.....	117
Figura 17	Comparação da classificação de amostras por trio classificador molecular em <i>plots</i> 3D.....	120
Figura 18	Esquema do objetivo geral.....	124
Figura 19	Esquema do objetivo secundário.....	127

LISTA DE TABELAS

Tabela 1	Características clínicopatológicas dos pacientes do projeto de Mestrado.....	31
Tabela 2	Trios de genes do projeto de Mestrado.....	32
Tabela 3	Quadras de genes do projeto de Mestrado.....	33
Tabela 4	Modelo do cálculo do intervalo de confiança dos classificadores identificados em projeto de Mestrado.....	50
Tabela 5	Cálculo do intervalo de confiança.....	50
Tabela 6	Características clínicopatológicas das 28 amostras.....	51
Tabela 7	Avaliação do rendimento e pureza do RNA total extraído.....	53
Tabela 8	Linhagens celulares do <i>pool</i> de RNA referência.....	57
Tabela 9	Avaliação do rendimento e pureza do RNA amplificado em um ciclo.....	59
Tabela 10	Avaliação do rendimento e pureza do RNA referência amplificado em segundo ciclo.....	60
Tabela 11	Avaliação do rendimento e pureza do RNA das amostras tumorais, amplificado em um ciclo.....	67
Tabela 12	Listagem das hibridizações de 29 amostras em duplicata.....	76
Tabela 13	Lista de anticorpos.....	83

Tabela 14	Características clínicopatológicas dos pacientes.....	88
Tabela 15	Nove genes mais diferencialmente expressos.....	95
Tabela 16	Trios de genes validados.....	96
Tabela 17	Quadras de genes validados.....	98
Tabela 18	Aplicação das regras de classificação das quadras de genes.....	101
Tabela 19	Cálculo de sensibilidade e especificidade dos melhores trio, quadras e árvores de decisão.....	103
Tabela 20	Seis genes mais diferencialmente expressos nas 3 vias.....	104
Tabela 21	Avaliação da marcação do anticorpo em imunoistoquímica.....	107
Tabela 22	Acertos e erros das árvores de decisão.....	122

LISTA DE ABREVIATURAS

ASA	<i>American Society of Anesthesiology</i>
BSA	albumina sérica bovina
cDNA	DNA complementar
cDNA Array	arranjos de cDNA
CE	carcinoma epidermóide
CECP	carcinoma epidermóide de cabeça e pescoço
cm	centímetro
DAB	diaminobenzidina
dCTP	desoxicitosina trifosfato
DEPC	dietilpirocarbonato
dL	decilitro
DMSO	dimetilsulfóxido
DNA	ácido desoxirribonucléico
dNTPs	mistura de desoxinucleotídeos trifosfato
EC	estádio clínico
EDTA	ácido etilenodiaminotetracético
e.g.	por exemplo (<i>exempli gratia</i>)
g	força g centrífuga
g	grama
cGy	<i>centigray</i>
HE	hematoxilina-eosina
HRP	peroxidase de raiz forte (<i>Horseradish</i>)
IC	intervalo de confiança
i.v	intravenoso
M	molar
MeV	mili elétron volt
mg	miligrama
ml	mililitro
mm²	milímetro quadrado

MMP	metaloproteinase
ng	nanograma
NR	paciente não respondedor
ORESTES	<i>open reading frame ESTs</i>
pb	par de base
pH	potencial hidrogeniônico
PBS	tampão fosfato em solução salina
PCR	reação em cadeia da polimerase
PoliA	cauda poli(A) do RNAm
R	paciente respondedor
RNA	ácido ribonucléico
RNAa	RNA amplificado
RNAm	RNA mensageiro
rNTPs	mistura de ribonucleotídeos trifosfato
rpm	rotações por minuto
RT	transcrição reversa
S	coeficiente de sedimentação, unidade de Svedberg
SCC	<i>squamous cell carcinoma</i>
SDS	lauril sulfato de sódio
SSC	solução salina com citrato de sódio
SVD	<i>singular value decomposition</i>
TA	temperatura ambiente
TGO	transaminase glutâmico oxalacética
TGP	transaminase glutâmico pirúvica
TNM	extensão do tumor primário; metástase em linfonodo regional; metástase à distância
Tris	tris (hidroximetil) aminometano
µg	micrograma
µL	microlitro
µm	micrometro
UV	ultravioleta
VC	validação cruzada

ÍNDICE

1	INTRODUÇÃO	1
1.1	O Câncer.....	1
1.2	Estruturas Laríngea e Hipofaríngea.....	6
1.3	Câncer de Laringe e Hipofaringe.....	8
1.4	Tratamento do Câncer de Laringe e Hipofaringe.....	10
1.5	Agentes quimioterápicos Cisplatina e Paclitaxel.....	13
1.6	Fatores de Risco.....	15
1.7	Classificadores Moleculares ou Biomarcadores.....	18
1.8	Análise da Expressão Gênica em Larga Escala.....	22
1.9	cDNA <i>Microarray</i>	23
1.10	Utilização de cDNA <i>Microarray</i> no Estudo do Câncer.....	24
1.11	Ensaio Clínico fase II Previamente Realizado.....	28
1.12	Identificação de Classificadores Moleculares de Resposta em Estudo de Mestrado.....	30
1.13	Validação dos Classificadores Moleculares.....	36
2	JUSTIFICATIVA	40
3	OBJETIVOS	41
3.1	Objetivo Geral.....	41
3.2	Objetivos Secundários.....	41
4	MATERIAL E MÉTODOS	42
4.1	Objetivo geral: Validação de Classificadores Moleculares.....	42
4.1.1	Critérios de Inclusão da População de Estudo.....	42
4.1.2	Critérios de Exclusão da População de Estudo.....	43
4.1.3	Quimioterapia da População de Estudo com Paclitaxel.....	43
4.1.4	Quimioterapia da População de Estudo com Cisplatina.....	44
4.1.5	Radioterapia da População de Estudo.....	44
4.1.6	Coleta de Amostras, de Dados Clínicos e Seguimento dos Pacientes.....	45
4.1.7	Análise e Microdissecção das Biópsias.....	47
4.1.8	Avaliação do Número Total de Amostras.....	49

4.1.9	Extração do RNA total das Amostras.....	52
4.1.10	Desenho Experimental e RNA Referência.....	56
4.1.11	Amplificação do RNA Mensageiro.....	62
4.1.12	Amplificação: Síntese de cDNA.....	64
4.1.13	Amplificação: Transcrição <i>in vitro</i>	65
4.1.14	Marcação indireta do cDNA.....	69
4.1.15	Incorporação dos corantes ao cDNA.....	70
4.1.16	Confecção, Pré-Hibridização, Hibridização e Lavagem das Lâminas de <i>Array</i>	72
4.1.17	Escaneamento das Lâminas de <i>Array</i> e Quantificação dos Dados.....	74
4.1.18	Análises dos Dados Gerados por <i>Microarray</i> e Análises Matemáticas.....	78
4.2	Objetivo Secundário: Estudo de Vias Metabólicas e Genes.....	81
4.2.1	Populações de Estudo e Vias Analisadas.....	81
4.2.2	Genes Analisados e Reações de Imunoistoquímica.....	82
5	RESULTADOS	87
5.1	Objetivo geral: Validação de Classificadores Moleculares.....	87
5.1.1	Seguimento Clínico dos Pacientes.....	87
5.1.2	Análises dos Dados Gerados por <i>Microarray</i> e Análises Matemáticas.....	90
5.2	Objetivo Secundário: Estudo de Vias Metabólicas e Genes.....	103
6	DISCUSSÃO	114
7	CONCLUSÃO	129
8	REFERÊNCIAS BIBLIOGRÁFICAS	131

APÊNDICES

Apêndice 1 Composição do *Second Strand Buffer*

Apêndice 2 Cálculo dos p-valores para a diferença identificada entre pacientes respondedores e não respondedores por imunoistoquímica

Apêndice 3 Cálculo dos p-valores para a correlação entre variáveis clinicopatológicas e resposta clínica dos pacientes

1 INTRODUÇÃO

1.1 O CÂNCER

O câncer ocasiona grandes perdas para a sociedade, além de um efeito devastador nos pacientes e seus familiares; além disso, os custos econômicos causados por esta doença são enormes, tanto em termos de cuidados médicos para o tratamento quanto em perdas humanas devido à mortalidade precoce (ETZIONI et al. 2003).

Em 2005, de um total de 58 milhões de mortes ocorridas no mundo, o câncer foi responsável por 7,6 milhões, o que representa 13% de todas as mortes (Ministério da Saúde 2007). Do total de óbitos ocorridos por câncer em 2005, mais de 70% ocorreram em países de média ou baixa renda (World Health Organization -WHO- 2006).

O câncer é a segunda maior causa de morte nos Estados Unidos da América, sendo responsável por 23,1% de todas as mortes ocorridas no ano de 2006. Um total de 1.479.350 novos casos e 562.340 mortes por câncer estão projetadas para ocorrer nos EUA em 2009. Apenas doenças cardíacas matam mais indivíduos do que câncer, as quais são responsáveis por 26% de todas as mortes ocorridas em mesmo ano e país (JEMAL et al. 2009).

Recentemente, nos EUA, as taxas de incidência por câncer diminuíram, tanto para o sexo masculino (1,8% por ano de 2001 a 2005) como para o sexo feminino (0,6% por ano de 1998 a 2005), principalmente por causa do decréscimo na

incidência dos 3 principais sítios acometidos no gênero masculino (pulmão, próstata e colorretal) e nos 2 principais no feminino (mama e colorretal). As taxas de morte causada por câncer também diminuíram, tanto em homens (19,2% entre 1990 e 2005) como em mulheres (11,4% entre 1991 e 2005). Mas, apesar do progresso ocorrido, o câncer ainda foi o maior agente causador de morte, nos EUA, entre mulheres de 1 a 39 anos de idade e acima de 80 anos, no ano de 2006. O mesmo ocorreu para homens de 40 a 59 anos e acima de 80 anos (JEMAL et al. 2009).

No Brasil, as estimativas para o ano de 2008, válidas também para o ano de 2009, apontam que ocorrerão 466.730 novos casos de câncer, 231.860 para o gênero masculino e 234.870 para o feminino. A distribuição dos novos casos de câncer segundo a localização primária é bem heterogênea entre estados e capitais do País. As regiões Sul e Sudeste, de maneira geral, apresentam as maiores taxas, enquanto as regiões Norte e Nordeste mostram as menores taxas. As taxas da região Centro-Oeste apresentam um padrão intermediário (Ministério da Saúde 2007).

Cânceres são classificados de acordo com tecido e tipo de célula que lhe originaram, desta forma, os que são originados de células epiteliais recebem a nomenclatura de carcinomas e aqueles que são originados de tecidos conjuntivos ou musculares são denominados sarcomas. Existem ainda tipos de cânceres que não se enquadram nestas duas categorias e incluem os vários tipos de leucemias, derivadas de células hematopoiéticas e cânceres derivados do sistema nervoso (ALBERTS et al. 1994).

As células dos tumores malignos, ou câncer, podem apresentar diversos graus de diferenciação: bem, moderadamente, pouco ou indiferenciados (ou anaplásicos). Quanto maior o grau de diferenciação de um tumor, maior é a semelhança

histológica ao tecido normal maduro de origem da neoplasia. A taxa de crescimento dos tumores geralmente está inversamente correlacionada com seu nível de diferenciação e diretamente correlacionada com sua agressividade. Assim sendo, tumores mais agressivos costumam ser mais indiferenciados e costumam crescer mais rapidamente que os benignos (COTRAN et al. 1999).

O desenvolvimento tumoral ocorre via processo análogo à evolução Darwiniana, no qual sucessões de mudanças genéticas, cada uma conferindo um tipo ou outro de vantagem de crescimento, leva à conversão progressiva de células normais em células cancerosas (FOULDS 1954; NOWELL 1976).

Acredita-se que células cancerosas resultem de uma manifestação de seis características típicas das células tumorais, comuns a mais de cem tipos de cânceres. Essas seis características correspondem a sucessivas adaptações ambientais sofridas pela célula normal, até sua transformação em célula tumoral, com capacidade de migração e alojamento em outros tecidos. As seis características são: auto-suficiência de estímulos de crescimento; insensibilidade a estímulos inibitórios de crescimento; potencial de replicação indefinido; evasão da apoptose; angiogênese e invasão de tecidos e metástase (HANAHAN e WEINBERG 2000).

A primeira característica é a habilidade de produzir, com independência, moléculas que estimulam o crescimento celular. Ou seja, a auto-suficiência de sinais de crescimento permite às células tumorais reduzirem a dependência de estímulos do seu microambiente de células normais. Por isso, acredita-se que as vias de sinalização de crescimento estejam desreguladas em todos os tumores humanos.

A segunda habilidade é a evasão de moléculas que inibem a proliferação celular. As células humanas apresentam limites intrínsecos em sua multiplicação,

mas, a via que controla potencial de replicação nas células cancerígenas parece agir de modo independente das outras vias metabólicas.

A terceira característica é o potencial de se replicar indefinidamente, bloqueando os mecanismos que normalmente limitariam a divisão celular. A quarta habilidade é a evasão da morte celular programada, mecanismo disparado pelas células sempre que algo anormal é detectado, como divisão acelerada. Adquirir resistência à apoptose é uma habilidade de provavelmente todos os tipos de câncer, e existem diversas estratégias para fazê-lo.

A quinta habilidade é a indução de formação de novos vasos sanguíneos, responsáveis por trazer sangue com nutrientes e oxigênio, indispensáveis para o tumor em crescimento. A neovascularização é um pré-requisito para a rápida expansão clonal (monoclonal) necessária para a formação do tumor macroscópico. Deste modo, a indução da angiogênese é um evento que provavelmente ocorre entre o início e a metade do surgimento do câncer.

A sexta característica é a habilidade de invadir outros tecidos. A metástase é o principal evento responsável por transformar um tumor em maligno, e envolve mudanças na adesão entre as células no seu microambiente e a ativação de proteases extracelulares.

Essas seis características novas adquiridas devem ocorrer na maioria e talvez em todos os tipos de tumores humanos, deste modo, essa multiplicidade de eventos explica porque o câncer é um evento relativamente raro durante a vida humana (HANAHAN e WEINBERG 2000).

Enquanto mudanças genéticas e epigenéticas em células neoplásicas são provavelmente iniciadoras de carcinogênese, existem várias evidências que indicam

que as respostas de células estromais do tecido neoplásico promovem e/ou regulam a progressão à malignidade e/ou subsequente metástase. As células estromais que contribuem com esse evento são (mio)fibroblastos, células vasculares, leucócitos infiltrantes e células especializadas no suporte mesenquimal, próprias de cada órgão (DENARDO et al. 2008).

Atualmente, acredita-se que o microambiente inflamatório pode desempenhar papel crítico na iniciação e progressão do câncer, sendo então considerado a sétima característica de células tumorais. Células imunes cronicamente ativadas, ou seja, inflamação, surge como regulador funcionalmente significativo do desenvolvimento do câncer primário, assim como de sobrevivência e de proliferação em tecidos ectópicos. Isto ocorre através da imune-regulação da angiogênese, da imune-regulação do remodelamento tecidual e intravasamento e do preparo imune de sítios pré-metastáticos (DENARDO et al. 2008).

OBERMUELLER et al. (2004) observaram que células inflamatórias, principalmente neutrófilos, contribuem com a angiogênese, através do remodelamento da matriz extracelular pela secreção de MMP9 e MMP13. Outros estudos observaram que a expressão abundante de TGF-beta e MMPs no microambiente tumoral pode promover invasão e metástase (BIERIE et al. 2008 e 2009; YANG et al. 2008).

Alguns tipos de leucócitos, principalmente células imune inatas, como mastócitos, células mielóides inatas, granulócitos e macrófagos potenciam a progressão tumoral (COUSSENS e WERB 2002; THEOHARIDES e CONTI 2004; BALKWILL et al. 2005; DE VISSER et al. 2006; CONDEELIS e POLLARD 2006) e aumentam a sobrevivência neoplásica. O microambiente inflamatório pró-tumor

promove não apenas transformação maligna e desenvolvimento de tumores sólidos, mas também promove a disseminação de células neoplásicas na corrente sanguínea, através da capacidade de invasão de células malignas, da expansão de angiogênese e da entrada nos vasos sanguíneos e linfáticos (BERGERS e BENJAMIN 2003; WYCKOFF et al. 2004; GOSWAMI et al. 2005; WYCKOFF et al. 2007).

1.2 ESTRUTURAS LARÍNGEA E HIPOFARÍNGEA

Nos animais primitivos, a laringe era um pouco mais que um esfíncter para proteger o trato respiratório inferior da inalação de substâncias nocivas e corpos estranhos, mas, durante o processo evolutivo, a mesma assumiu a função da fonação (BENJAMIN et al. 1996). Nos humanos, a laringe é um órgão altamente especializado.

Este órgão apresenta um arcabouço protetor osteocartilaginoso e membranoso, que inclui o osso hióide e as cartilagens epiglótica, tireóide e cricóide. Este arcabouço, em conjunto com o par de cartilagens aritenóideas e as pequenas cartilagens corniculadas e cuneiformes, está inteiramente conectado por uma série de membranas e ligamentos. O par de cartilagens aritenóideas em forma de pirâmide se articula com a superfície articular superior das lâminas posteriores das cartilagens cricóide. Esta articulação sinovial é vital para vários movimentos complexos das pregas vocais, incluindo a abdução e a adução durante a respiração e as variações de tensão e massa das pregas vocais durante a fonação. Os músculos intrínsecos da laringe são inervados pelo nervo laríngeo recorrente, que comanda as funções da respiração, fonação, deglutição e fechamento da laringe (BENJAMIN et al. 1996).

A laringe pode ser dividida em 3 regiões; laringe **supraglótica**, que compreende a epiglote, falsas cordas, ventrículos, pregas ariepiglóticas e aritenóides; **glote**, que são as cordas vocais verdadeiras e comissuras anterior e posterior; e por fim **subglote**, que se inicia um cm abaixo das cordas vocais verdadeiras e se estende até o primeiro anel traqueal (BENJAMIN et al. 1996). A maioria dos tumores de laringe, aproximadamente 66%, acomete a glote (PARKIN et al. 1990).

A drenagem linfática da região supraglótica é rica, conseqüentemente, tumores nessa localização apresentam altos índices de metástases linfáticas regionais, entre 50-60%, e freqüentemente apresentam incidência bilateral. De forma contrária, os tumores glóticos apresentam baixa incidência de metástases linfáticas cervicais e, quando ocorrem, geralmente são ipsilaterais. As metástases linfáticas de tumores subglóticos são mais freqüentes e envolvem preferencialmente as cadeias paratraqueais, cervicais baixas e mediastinais altas (LINDBERG 1972; BYERS et al. 1988; MATOS DE SÁ 1997).

A hipofaringe é uma das subdivisões da faringe e também é denominada laringo-faringe, a região mais inferior da faringe. A hipofaringe pode ser delimitada pela prega faringoepiglótica e pela prega cricofaríngea, sendo um canal importante na passagem do alimento (BARBOSA 1979).

A hipofaringe pode ser dividida em **seios piriformes**, localizados um de cada lado do estojo laríngeo e em **porção inter-crico-vertebral** ou **retrocricóide**, disposta entre a face posterior da cricóide, para diante, e o plano vertebral, para trás. Os seios piriformes são sedes preferenciais dos tumores de hipofaringe, sendo o sítio acometido em 80-90% dos casos (BARBOSA 1979).

Esta região apresenta ampla e rica distribuição linfática, o que ocasiona elevada incidência de metástases; além disso, esses tumores costumam exibir menor grau de diferenciação que os da laringe (BARBOSA 1979).

Os sintomas dos pacientes são indicativos da localização da lesão, sendo que odinofagia sugere tumor supraglótico e rouquidão indica tumor glótico e/ou subglótico. O câncer supraglótico geralmente é acompanhado de outros sinais e sintomas, como alteração na qualidade de voz, disfagia leve e sensação de “caroço” na garganta. Nas lesões avançadas das cordas vocais, além de rouquidão, pode ocorrer odinofagia, disfagia e ainda dispnéia.

1.3 CÂNCER DE LARINGE E HIPOFARINGE

Tumores malignos de Cabeça e Pescoço são o sexto tipo de câncer mais comum no mundo, com incidência anual acima de 640.000 casos em todo o mundo (SEIWERT et al. 2009), e com 48.010 novos casos estimados para 2009, apenas nos EUA (JEMAL et al. 2009).

Nos EUA são estimados 12.290 novos casos de câncer de laringe para o ano de 2009, sendo que destes, 80,71% acometem o gênero masculino. Para o mesmo ano, a estimativa é de 3.660 mortes, sendo que 79,23% para o sexo masculino. Já para câncer de faringe, estão estimados 12.610 novos casos, sendo que o gênero masculino é acometido por 79,46% dos casos. A estimativa é de 2.230 mortes, sendo que 73,54% são para o sexo masculino (JEMAL et al. 2009).

Carcinoma epidermóide de cabeça e pescoço compõe um grupo heterogêneo de tumores que se originam no trato aerodigestivo superior (ROEPMAN et al. 2005).

O câncer de laringe é um dos mais comuns a atingir a região da cabeça e pescoço, representando aproximadamente 25% dos tumores malignos que acometem esta área e 2% de todas as doenças malignas (Ministério da Saúde 2009).

Aproximadamente 90% dos cânceres de laringe e hipofaringe apresentam origem epitelial. Essas células são encontradas no tecido que forma a superfície da pele, o revestimento dos órgãos ocos e as passagens dos tratos respiratório e digestivo. Desta forma, o tipo predominante é carcinoma epidermóide (ou carcinoma de células escamosas), incluindo os ceratinizantes ou não-ceratinizantes, de graus bem diferenciado, moderadamente ou pouco. Os adenocarcinomas de origem salivar menor, os tumores neuroendócrinos, os mesenquimais e os melanomas também podem ser encontrados, porém em padrão infreqüente (OLIVEIRA et al. 1997).

Contrastando com câncer de mama, câncer de cabeça e pescoço se dissemina quase que exclusivamente por via linfática, na qual os linfonodos cervicais constituem a primeira localização (LEEMANS et al.1993). Uma explicação possível é que na cascata de metástase linfática apresentada em CECP, ocorrem alterações adicionais nos linfonodos, as quais são necessárias para as células tumorais se disseminarem, depois, para sítios distantes via hematogénea (PANTEL e BRAKENHOFF 2004).

Estudo demonstrou que, através da análise do padrão de expressão gênica de metástases linfonodais de CECP, por cDNA *microarray*, a metástase reflete tanto o perfil de expressão de genes globais como o de genes metastáticos referentes aos tumores primários de que se originaram. Esses achados corroboram o conceito de que as células tumorais disseminadas alteram o tecido ao seu redor para um

microambiente que se parece com o microambiente do tumor primário de origem (ROEPMAN et al. 2006a).

1.4 TRATAMENTO DO CÂNCER DE LARINGE E HIPOFARINGE

O tratamento dos cânceres de cabeça e pescoço pode causar problemas nos dentes, fala e deglutição; além disso, os danos físicos e funcionais podem levar à queda da auto-estima, depressão, perda de emprego e retiro do convívio social. A laringectomia total implica na perda da voz fisiológica e em traqueostomia definitiva. Todos os pacientes perdem a capacidade de comunicação oral imediatamente após a cirurgia e somente 30% dos pacientes passam a ter fala esofágica compreensível (Ministério da Saúde 2009). Assim, quanto mais precoce for o diagnóstico, menor é a possibilidade do tratamento causar deformidades e problemas psico-sociais.

A adoção de um regime quimioterápico associado à radioterapia vinha mostrando resultados promissores desde a década de 1980 (BATAINI et al. 1982; PIGNON et al. 2000), tanto na forma de tratamento adjuvante para pacientes de alto risco (VERMUND 1970; VANDENBROUCK et al. 1987), quanto na forma de tratamento neoadjuvante, em uma estratégia para preservação de órgãos (MENDENHALL et al. 1987).

A laringectomia total associada ou não à radioterapia pós-operatória era o tratamento padrão para carcinomas de laringe e hipofaringe localmente avançados (KOWALSKI et al. 1996; SANTOS et al. 1998), até o início da década de 1990.

Desde o início da década de 1990, este tratamento vem sendo substituído pela combinação de quimioterapia com radioterapia, após demonstração em estudos

prospectivos, por *Veterans Affairs Laryngeal Cancer Study Group* (Anonymus 1991) e LEFEBVRE et al. (1996) de que este tratamento oferecia as mesmas chances de sobrevida, mas com possibilidade de preservação da laringe em 30 a 60% dos casos.

Para a maioria dos pacientes, existem diversas opções de tratamento, como radioterapia, quimiorradioterapia, radioterapia com cetuximab, cirurgia ou ainda uma combinação dessas modalidades (PRAMANA et al. 2007).

Quimioterapia combinada à radioterapia se tornou padrão pra a maioria dos pacientes portadores de CECP em estágios avançados, apesar de aproximadamente 30% desenvolverem recorrências loco-regionais (BALM et al. 2004). Atualmente, a administração combinada de ambas as modalidades é o enfoque recomendado pela maioria dos especialistas (National Comprehensive Cancer Network NCCN 2007).

Esta estratégia alcança taxas mais altas de controle loco-regional, quando comparada à quimioterapia de indução antes do tratamento e também comparada à quimioterapia adjuvante após a radioterapia (WOLF e FISHER 1992; LEFEBVRE et al.1996; PIGNON et al. 2000; FORASTIERE et al. 2003; PFISTER et al. 2006).

No entanto, pacientes não respondedores ao tratamento clínico e recorrência loco-regionais são limitações consideráveis do método (SINGH e PFISTER 2008). O tratamento cirúrgico seguinte à falha ou à recorrência, frequentemente requer laringectomia ou laringofaringectomia totais, associadas a um amplo esvaziamento cervical para o controle local do tumor primário ou da recidiva. A conseqüente perda da fala e o uso permanente de traqueostoma é considerado um prejuízo funcional grave que afeta dramaticamente a qualidade de vida dos pacientes (WEYMULLER et al. 2000; HANNA et al. 2004; VARTANIAN et al. 2004).

Inúmeros tipos de tratamentos têm sido realizados nos últimos anos, como quimioterapia com cisplatina associada à radioterapia (WEBER et al. 2004). Outra alternativa é o uso de quimioterapia exclusiva com os fármacos cisplatina, paclitaxel e ifosfamida combinadas (SHIN et al. 1998) ou de radioterapia exclusiva (WEBER et al. 2004), quando existe, por exemplo, contra-indicação clínica para a quimioterapia.

Em ensaio clínico conduzido por FORASTIERE et al. (2003) utilizando cisplatina ministrada concomitantemente à radioterapia, demonstrou-se melhora significativa na preservação de órgãos quando comparada à quimioterapia neoadjuvante seguida de radiação ou radioterapia exclusiva. Em carcinomas epidermóides avançados de cabeça e pescoço quimiorradioterapia com paclitaxel apenas ou combinado com cisplatina demonstrou melhora significativa nas taxas de resposta, quando comparadas à radioterapia exclusiva (FORASTIERE et al. 2001; GARDEN et al. 2004).

A maioria dos pacientes se apresenta com doença em estágio avançado, o que requer terapia agressiva. Devido à natureza debilitante da terapia agressiva e à habilidade limitada de identificar pacientes de alto risco para a falha do tratamento, melhores biomarcadores de prognóstico, assim como tratamentos mais efetivos e menos tóxicos são urgentemente necessários (CHUNG et al. 2006).

Deste modo, no estudo realizado por FEHER et al. (2002), que originou trabalho de Mestrado de BARRETO (2005, Proc FAPESP 02/10891-5) e no presente estudo de doutorado, os pacientes portadores de carcinoma epidermóide de laringe e hipofaringe localmente avançado foram tratados com quimioterapia, realizada com cisplatina e paclitaxel combinados, concomitante à radioterapia.

1.5 AGENTES QUIMIOTERÁPICOS CISPLATINA E PACLITAXEL

Os agentes quimioterápicos agem basicamente de 3 formas nas células: causando danos ao DNA, o que impede a posterior replicação celular; inibindo a síntese do DNA com o bloqueio de nucleotídeos necessários (fase S do ciclo celular), o que também impede a posterior replicação e, por fim, existem fármacos que agem parando o processo mitótico (fase M do ciclo celular), impedindo a posterior divisão celular.

Embora o fármaco de escolha para a quimioterapia associada à radioterapia seja a Cisplatina, estudos demonstraram resultados promissores com a utilização do Paclitaxel, para o tratamento de formas avançadas de carcinoma epidermóide (ROWINSKY et al. 1992; ROSE 1992; KEARNS et al. 1995).

O composto *cis*-PtCl₂(NH₃)₂ foi descrito em 1845 e sua estrutura deduzida em 1893. Cisplatina (*cis*-diaminodicloroplatina II) é um complexo inorgânico divalente hidrossolúvel sintetizado, que contém platina. Existem mais de 1.000 compostos deste tipo, mas a cisplatina foi a mais ativa dessas substâncias em sistemas tumorais experimentais e se mostrou de grande valor clínico (ROSENBERG 1973).

A cisplatina é um agente antineoplásico alquilante de bases nitrogenadas, que entra na célula por difusão, reage com o DNA e forma principalmente ligações cruzadas intrafilamentares entre guaninas adjacentes, levando à inibição da replicação e transcrição do DNA (GILMAN et al. 1996). Este fármaco age em

qualquer fase do ciclo celular, mas é mais eficiente durante a síntese de DNA (fase S).

A cisplatina apresenta ampla atividade como agente antineoplásico e é especialmente útil no tratamento de malignidades epiteliais. Este fármaco se tornou a base de esquemas curativos para câncer testicular avançado e tem notável atividade contra câncer ovariano e cânceres da cabeça e pescoço, bexiga, esôfago e pulmão.

Paclitaxel (taxol) foi descoberto em 1967, ao ser isolado da casca do Teixo do Pacífico (*Taxus brevifolia*) e ter sua atividade antitumoral notada, em ampla variedade de tumores em roedores. A estrutura deste alcalóide vegetal, (C₄₇H₅₁NO₁₄), extremamente complexa, foi determinada em 1970.

Paclitaxel apresenta atividade antineoplásica e exhibe ações farmacológicas únicas como inibidor da mitose (fase celular M). Seu mecanismo de ação se baseia na sua interação com microtúbulos, especificamente à subunidade β da tubulina, impedindo a desmontagem dos microtúbulos, evento essencial para a continuidade do processo mitótico, levando à interrupção da mitose e posterior morte celular (GILMAN et al. 1996). Outras células também são afetadas adversamente pelo taxol, mas, como as células cancerígenas se dividem muito mais rapidamente do que as normais, estas são mais susceptíveis à ação do quimioterápico.

O fármaco foi aprovado para tratamento de câncer ovariano refratário à cisplatina em 1992 (ROWINSKY et al. 1993; ROWINSKY e DONEHOWER 1995) e apresenta atividade promissora contra cânceres de cabeça e pescoço, mama, pulmão, esôfago e sarcoma de Kaposi, apresentando taxas significativas de resposta.

1.6 FATORES DE RISCO

A etiologia do câncer de cabeça e pescoço é multifatorial (MAIER et al. 1992), o que torna bastante relevante o estudo de expressão gênica diferencial, além de fatores de risco ambiental.

O CECP ocorre mundialmente em frequência crescente, devido ao aumento da expectativa de vida global e ao aumento do consumo de álcool e tabaco (REID et al. 2000; BRAY e MOLLER 2006).

O Brasil é um dos países com maior incidência desta neoplasia maligna e este fato está, provavelmente, relacionado à alta frequência de tabagismo e etilismo na comunidade, que são os principais fatores de risco (MAIER et al. 1992; GOLDENBERG et al. 2003).

Como demonstrado em inúmeros estudos, existe forte associação entre ingestão excessiva de álcool e tabagismo com o desenvolvimento de câncer no trato aerodigestivo superior, sendo o tabagismo o maior fator de risco para o desenvolvimento de câncer na laringe e hipofaringe (Ministério da Saúde 2009).

MAIER et al. (1992) concluíram que a utilização de álcool e tabaco aumenta o risco de desenvolvimento de CE de laringe, de modo dose-dependente; e ainda que o uso combinado resulta em efeito sinérgico, aumentando o risco de modo multiplicativo, e não apenas aditivo.

O risco de câncer associado ao tabaco tem sido explicado pelo fato de sua fumaça conter mais de 30 compostos diferentes, comprovadamente carcinogênicos, como hidrocarbonetos aromáticos policíclicos, nitrosaminas específicas do tabaco,

aldeídos, ácidos nitrosaminos e metais (HOFFMANN et al. 1985; BOFFETTA et al. 2008).

Estudos demonstraram que o etanol é o principal agente carcinogênico em bebidas alcoólicas, tanto por causar a queima direta do epitélio (HSU 1992), como por produzir acetaldeído através da reação do etanol com a saliva (TILLONEN et al. 1999). O álcool é metabolizado em acetaldeído, substância tóxica e carcinogênica, pelas enzimas álcool desidrogenase, catalase e enzimas do sistema citocromo P450. O acetaldeído é o principal produto do metabolismo do etanol, sendo uma molécula altamente reativa e eletrolítica. Grupos eletrolíticos altamente reativos podem formar aductos de DNA, que quando escapam do sistema de reparo celular podem causar mutações em pontos críticos, iniciando-se a carcinogênese.

O consumo crônico de etanol aumenta a atividade dessas enzimas procarcinogênicas (GARRO et al. 1981; FARINATI et al. 1985), além de afetar a carcinogênese, através da inibição do reparo de DNA (MUFTI et al. 1988). A administração aguda e crônica de etanol ainda diminui a produção e secreção salivar (KISSIN e KALEY 1974; MAIER et al. 1986 e 1988), o que resulta em maior concentração local de procarcinógenos e carcinógenos, prejudicando a superfície mucosa da cavidade oral (MAIER et al. 1992).

O mate, conhecido no Brasil como chimarrão, é uma erva da árvore *Ilex paraguariensis*, espécie da família Aquifoliaceae, e seu consumo foi associado com o aumento da taxa de cânceres orais, orofaríngeos, laríngeos e no esôfago. Estudo demonstrou que o mate é carcinogênico e que seu consumo aumenta a taxa de desenvolvimento de câncer na laringe e faringe em 2,1 vezes (GOLDENBERG et al. 2003). Este fato se torna importante uma vez que a incidência de câncer na laringe

em indivíduos do gênero masculino no Brasil está entre as maiores do mundo (MUIR e WEILAND 1995).

O mate pode agir como solvente dos carcinógenos químicos encontrados no tabaco, além disso, os compostos fenólicos encontrados no mate podem agir como promotores do câncer. Soma-se a isto o fato do mate conter taninos e compostos N-nitrosos, que também são suspeitos de serem carcinógenos (GOLDENBERG et al. 2003).

Foram também encontradas substâncias na fração de óleo do mate, que se mostraram carcinogênicas em animais de laboratório (VASSALLO et al. 1985). Outros estudos ainda concluíram que o mate é genotóxico, causando mutagênese em bactérias e aumentando a frequência de aberrações cromossômicas em linfócitos periféricos humanos tratados com mate (FONSECA et al. 2000).

LÖNING et al. (1985) foram os primeiros autores a sugerirem relação entre papilomavírus humano (HPV) e CECP, em 1985, e desde então evidências desta possível associação têm aumentado.

Os HPVs dos tipos 16, 18 e 33, também conhecidos como HPVs de alto risco, produzem proteínas oncogênicas, E6 e E7, que se ligam e inativam p53 e pRb, respectivamente, que são importantes proteínas supressoras de tumor; promovendo a imortalização celular, desregulação do ciclo celular e apoptose (ZUR HAUSEN 2000; VAN HOUTEN et al. 2001).

De forma similar ao câncer cervical, o HPV também parece ser responsável por causar o desenvolvimento de CECP (GILLISON et al. 2000). Meta-análise de ensaios publicados incluiu 5.046 amostras de CECP e demonstrou 26% de

prevalência de HPV, sendo que a grande maioria foi do tipo 16 (KREIMER et al. 2005).

A localização predominante de tumores HPV-associados foi orofaringe, com predileção pelos pacientes não fumantes. De forma similar ao câncer cervical, a detecção de HPV em CECP está associada ao histórico sexual, sendo a exposição direta a causa da infecção (D'SOUZA et al. 2007). Outro estudo ainda demonstrou que a imunossupressão aumenta o risco para a infecção e para o desenvolvimento de CECP relacionado ao HPV (KUTLER et al. 2003). VAN HOUTEN et al. (2001) também demonstraram que HPV 16 são, de fato, responsáveis por uma parcela dos carcinomas epidermóides na hipofaringe.

1.7 CLASSIFICADORES MOLECULARES OU BIOMARCADORES

As duas últimas décadas testemunharam um crescimento explosivo na quantidade de dados genômicos e proteômicos gerados, o que tem conduzido a uma nova era da medicina molecular, na qual detecção de doença, diagnóstico e tratamento serão cada vez mais individualizados, para cada perfil molecular individual (GINSBURG e MCCARTHY 2001; JAIN 2002; LITTLE et al. 2008; ALLISON 2008).

Esta revolução é baseada na disponibilidade e aplicação de novos biomarcadores de predição de comportamento da doença, de tecnologias avançadas para rápida detecção e diagnóstico, de novas terapias com alvos moleculares e celulares e de tecnologias computacionais para análise e gerenciamento de dados (PHAN et al. 2009).

Marcadores biomoleculares, ou assinaturas moleculares, incluem grupos de genes, genes alterados ou mutados, transcritos de RNAm, proteínas, lipídios, carboidratos, metabólitos e estados alterados de expressão de tais marcadores que possam ser correlacionados com um comportamento biológico ou uma resposta clínica (LIOTTA e PETRICOIN 2000; PETRICOIN et al. 2002; NEGM et al. 2002; LUDWIG e WEISTEIN 2005; NILSSON et al. 2009).

MicroRNAs, ou miRNA, são pequenas moléculas de RNA, de aproximadamente 22 bases, que são processadas a partir de transcritos primários longos (pri-miRNA) e que regulam expressão gênica em nível pós-transcricional. Essas moléculas também estão sendo utilizados como biomarcadores, no desenvolvimento de ensaios de diagnóstico molecular com enfoque primário no câncer, por empresas como Asuragen e Exiqon.

Biomarcadores normalmente são divididos em marcadores de prognóstico, marcador preditivo ou marcador de resposta terapêutica. Biomarcadores prognósticos permitem a predição do curso natural de um câncer individual, tornando possível, por exemplo, distinguir tumores indolentes de tumores agressivos. Biomarcadores preditivos são utilizados para avaliar a probabilidade de um paciente se beneficiar de um tratamento específico. Por fim, biomarcadores farmacodinâmicos medem os efeitos do tratamento, a curto prazo, que uma droga apresenta em determinado tumor, e estes são utilizados para guiar a seleção da dose nos estágios iniciais do desenvolvimento clínico de novas drogas (PHAN et al. 2009).

Inúmeros estudos vêm buscando identificar marcadores que diferenciem pacientes respondedores de não respondedores, tanto em relação à radioterapia

quanto à quimioterapia. Esta descoberta é muito importante, uma vez que a predição da resposta do paciente irá facilitar a escolha do tratamento adequado.

Assim, alguns trabalhos buscam a identificação de preditores de resposta terapêutica (WOLF et al. 1990; WOLF e CAREY 1992), de marcadores de metástase à distância, de metástase em linfonodo, de recorrência, de prognóstico e de sobrevida livre de doença, entre outros.

Foi demonstrado que, pacientes portadores de CECP que expressam p53 mutada e com baixa taxa de proliferação celular apresentam grande probabilidade de não responderem à radioterapia; enquanto que a expressão de p53 do tipo selvagem e alta taxa de proliferação são excelentes indicativos de resposta ao tratamento radioterápico, até mesmo com doença em estado avançado. O mesmo estudo avaliou outros possíveis marcadores em tumores primários, como HSP-27, c-erb B-2 e GST π , mas nenhuma relação foi encontrada quanto à resposta radioterápica (RAYBAUD-DIOGÈNE et al. 1997).

A superexpressão de p53 mutada foi, novamente, correlacionada com pobre resposta à radioterapia em CECP, e detectada em aproximadamente 60% dos casos (FIELD et al. 1993a e b).

Carcinomas epidermóides de cabeça e pescoço freqüentemente geram metástases para linfonodos regionais, assim, MAULA et al. (2003) estudaram a expressão de moléculas de adesão das células de CECP, e concluíram que a expressão de CD44 estava associada com prognóstico ruim, com risco aumentado de recorrências locais e com maior resistência à radioterapia.

HANNA et al. (2001) analisaram, através de cDNA *array*, o perfil de expressão de 1.187 genes, incluindo oncogenes, genes supressores de tumor, genes

do ciclo celular e genes relacionados com apoptose, em amostras de CECP resistentes e sensíveis à radioterapia. Como resultado, foram identificados 60 genes preditores de resposta à radioterapia.

Foi realizado estudo na tentativa de se identificar marcadores de resposta à quimioterapia com Cisplatina/Paclitaxel ou Cisplatina/5-fluorouracil em pacientes com CECP. Como resultado, nenhum dos possíveis marcadores, p53, GST π , MRP, c-erb B2, TS e GSH-S, foram significativamente correlacionados com resposta clínica à quimioterapia. Apesar disto, a expressão de p53 mutada e a superexpressão de GST π e TS se mostraram bons preditores de menor sobrevida (SHIGA et al. 1999).

A maioria dos estudos tem se focado em processos biológicos associados ao prognóstico, e não tem se focado especificamente em resposta ao tratamento (PRAMANA et al. 2007). Alguns estudos identificaram assinaturas associadas ao prognóstico, em uma vasta gama de situações, como por exemplo assinatura de “cicatrização de ferida” (CHANG et al. 2005), de hipóxia (CHI et al. 2006), de células tronco (GLINSKY et al. 2005) e de instabilidade cromossomal (CARTER et al. 2006).

No entanto, essas assinaturas parecem ser monitoras gerais de comportamento maligno ou agressivo, pois foram aplicadas a diferentes tipos de câncer e a pacientes que receberam uma variedade de tratamentos diferentes. Assim sendo, elas não são nem doença-específico nem tratamento-específico (PRAMANA et al. 2007).

1.8 ANÁLISE DA EXPRESSÃO GÊNICA EM LARGA ESCALA

A análise diferencial da expressão gênica tem sido realizada por diversas metodologias que comparam o conteúdo de RNAm de diferentes células ou tecidos. Dentre estas técnicas pode-se citar hibridização subtrativa (LEE et al. 1991); DDRT-PCR (RT-PCR por amostragem diferencial) (LIANG e PARDEE 1992); RAP-PCR *fingerprinting* (Impressões Digitais por PCR arbitrário de RNA) (WELSH et al. 1992); RDA (Análise por Representação Diferencial) (HUBANK e SCHATZ 1994); técnica de SAGE (Análise em Série da Expressão Gênica) (VELCULESCU et al. 1995) e, bastante utilizada atualmente, análise global de genes por microarranjos de cDNAs (SCHENA et al. 1995; DERISI et al. 1996). Mais recentemente, a expressão de miRNA também tem sido utilizada (DUMONT e TLSTY 2009; HURST et al. 2009; MATHÉ et al. 2009).

A possibilidade de utilização de técnicas robotizadas permitiu a análise da expressão diferencial, em grande escala, de genes de duas ou mais populações distintas. Através do DNA *microarray*, um grande número de fragmentos de cDNAs pode ter os níveis de expressão de milhares de genes analisados em um único experimento. Esta metodologia permite a análise da expressão funcional de um grande número de genes, até mesmo aqueles cujas funções ainda não foram elucidadas, ou que não tiveram sua seqüência completa de nucleotídeos determinada.

1.9 cDNA *Microarray*

Embora o primeiro estudo envolvendo *microarray* denomine o material marcado ou flutuante de sonda, ou *probe*, (SCHENA et al. 1995), a maior parte dos trabalhos subseqüentes consideram sonda os fragmentos fixados e alvo o material marcado (SOUTHERN et al. 1999). Esta nomenclatura é baseada fundamentalmente no fato de o material fixado apresentar seqüência conhecida, como as sondas utilizadas em experimentos de *Southern* e *Northern blot*.

Os DNA *microarrays*, também conhecidos como *chips* de DNA, são superfícies de vidro ou nylon que contêm milhares de fragmentos de DNA distribuídos em espaços mínimos. Os arranjos de cDNA são feitos através da deposição robótica do DNA em lâmina de vidro, em pontos com diâmetro entre 50-150 μm , e geralmente são utilizados produtos amplificados de clones de cDNA. Esta tecnologia é importante na determinação da expressão gênica de dezenas de milhares de genes em um único experimento.

Os chips podem ser diferentes quanto ao tamanho dos fragmentos de DNA fixados, métodos de arranjo, ligações existentes entre o DNA e o *chip*, métodos de hibridização ou de detecção. Os formatos mais conhecidos para os *chips* são arranjo de cDNA (SCHENA et al. 1995) e arranjo de oligonucleotídeos sintetizados *in situ* (PEASE et al. 1994).

O formato de *biochip* em vidro possui vantagens sobre as grandes membranas flexíveis (SOUTHERN 1975; LENNON e LEHRACH 1991), uma vez que reduz a área de difusão necessária para que o cDNA alvo encontre a sonda, reduzindo assim o volume de hibridização. Deste modo, é necessária uma menor quantidade de RNA

das amostras de análise, tornando possível o estudo de material proveniente de biópsia ou que apresente quantidade limitada.

O *microarray* pode ser utilizado tanto para monitorar a expressão gênica quanto para detectar polimorfismos ou mutações no DNA genômico. Neste último caso, são utilizados *biochips* onde oligonucleotídeos são fixados no suporte sólido. Assim sendo é possível, por exemplo, analisar regulação gênica envolvida na progressão de uma doença, estudar polimorfismos que predisõem a uma determinada doença e desenvolver tratamentos mais efetivos contra doenças.

1.10 UTILIZAÇÃO DE cDNA *MICROARRAY* NO ESTUDO DO CÂNCER

A utilização de cDNA *microarray* em análises comparativas de expressão gênica tem crescido muito nos últimos anos. A capacidade de estudar milhares de genes simultaneamente tem justificado o abandono de técnicas de menor escala, especialmente em situações nas quais células ou tecidos comparados são bastante semelhantes.

Esta técnica, utilizada na determinação de padrões de expressão gênica, está gerando quantidade abundante de novas informações que ajudarão no diagnóstico do câncer e na escolha da melhor forma de terapia a ser utilizada. Através da utilização de *microarray* centenas de genes podem servir como sondas para detectar quais genes estão ativos em diferentes tipos de células, o que gera uma quantidade de informação sem precedentes na história; o que seria muito difícil, ou praticamente impossível utilizando o enfoque gene por gene. Em alguns casos é possível

determinar quais cânceres apresentarão resposta frente à terapia e quais não, o que é extremamente vantajoso e apresenta aplicação clínica direta (MARX 2000).

A utilização de *microarrays* tem gerado enorme quantidade de informação em estudos relacionados ao câncer, com o objetivo da identificação tanto de genes potencialmente envolvidos no processo de transformação maligna quanto para aplicação clínica, quer seja no diagnóstico ou no prognóstico (LAKHANI et al. 2001), além da identificação de marcadores preditivos em diversos tipos tumorais (VAN DE VIJVER et al. 2002; SINGH et al. 2002; VALK et al. 2004).

A técnica de DNA *arrays* (SCHENA et al. 1995) permite a análise da expressão de centenas a milhares de genes em um único ensaio, e o resultado combinado de experimentos distintos permite o estabelecimento de perfis gênicos para cada condição patológica, identificando genes sub ou superexpressos (QUACKENBUSH 2001). Além disso, o agrupamento de genes em classes de expressão fornece pistas sobre a função biológica e a relevância de genes ainda desconhecidos (QUACKENBUSH 2001).

Os estudos realizados até o momento podem ser divididos principalmente em três grandes grupos: os que buscam identificar perfis de tumores, a maioria visando melhoras no diagnóstico de subtipos de tumores morfológicamente homogêneos (LAKHANI e ASWORTH 2001); os que identificam subtipos correlacionados com prognósticos diferentes e ainda os que buscam alvos terapêuticos ou novos marcadores moleculares (MACOSKA 2002).

A resistência à radioterapia é determinada por múltiplos eventos celulares que são controlados por um grande grupo de genes e suas interações, deste modo, os

arrays de cDNA surgem como ótimas ferramentas para tal comparação (HANNA et al. 2001).

Estudo anterior realizado por nosso grupo identificou genes diferencialmente expressos entre tecido gástrico normal e tumoral utilizando *array* de 4.500 clones ORESTES. Avaliando as razões de expressão entre os dois tipos de tecidos e a variabilidade dessa expressão entre as diversas amostras de um mesmo tecido foram selecionados 80 genes capazes de separar perfeitamente, por *Self-Organizing Map* (SOM), tecidos normais e malignos. Diversos trios de genes selecionados dentre os 80 foram também capazes de realizar essa discriminação (MEIRELES et al. 2003).

Em outro estudo, nosso grupo identificou série de classificadores moleculares que distinguem amostras cancerosas gástricas das não cancerosas, através da análise de Discriminante Linear de Fisher. Também foram identificadas metaplasias intestinais com perfis de expressão gênica semelhantes aos de adenocarcinoma. Este dado pode ser útil no seguimento dos pacientes, a fim de determinar seu potencial como teste prognóstico para transformação maligna (MEIRELES et al. 2004).

CHOI e CHEN (2005) revisaram alguns estudos sobre câncer de cabeça e pescoço, e a maioria estava focado nas diferenças genéticas entre CECP e epitélio normal. CHUNG et al. (2004) definiram 4 subtipos de CECP baseado no padrão de expressão gênica associado à resposta clínica, e também foram capazes de prever o estatus do linfonodo cervical metastático. Em seguida, os autores identificaram classificador composto por 75 genes, capaz de separar grupo de alto e baixo risco para recorrência, em pacientes tratados cirurgicamente, com ou sem radioterapia pós operatória ou quimioterapia (CHUNG et al. 2006).

WARNER et al. (2004), O'DONNELL et al. (2005) e ROEPMAN et al. (2005) também identificaram assinaturas capazes de prever a formação de metástases regionais em CECP. Utilizando perfil de expressão gênica, BRAAKHUIS et al. (2006) tentaram prever metástase à distância em CECP, mas sem sucesso.

Apesar do desenvolvimento da tecnologia em genômica e proteômica ter tornado possível a análise simultânea de grande número de variáveis, possibilitando pesquisa baseada na descoberta, ainda existem grandes desafios em relação à validação dos dados identificados pela metodologia de *microarray*. Neste tipo de pesquisa, grande quantidade de dados podem ser analisados sem uma hipótese, a fim de buscar padrões discriminatórios úteis (STEARNS et al. 2003; RANSOHOFF 2003).

Mas, a principal dificuldade é que os resultados podem não ser reprodutíveis, como no caso de *overfitting*, que é a identificação de padrão discriminatório devido ao acaso, que pode ocorrer quando grande número de variáveis são avaliadas para um pequeno número de resultados. A solução para este problema é a avaliação correta da reprodutibilidade (RANSOHOFF 2004).

Deste modo, atualmente, o maior desafio da metodologia de *array* é a validação dos dados identificados. Para isto, o modelo de reconhecimento do padrão, que é derivado do grupo de identificação ou treinamento, deve ser aplicado a um grupo de validação independente, composto por indivíduos que não foram utilizados no grupo anterior. Se a discriminação não puder ser reproduzida novamente neste segundo grupo, *overfitting* é a explicação mais plausível (RANSOHOFF 2004), invalidando os dados identificados no primeiro experimento.

Assim, a validação é o maior desafio deste método e a validação independente ocorre em apenas 10% das pesquisas com *microarray* (NTZANI e IOANNIDIS 2003).

1.11 ENSAIO CLÍNICO FASE II PREVIAMENTE REALIZADO

É de extrema importância o fato de que, em todo tratamento oncológico conservador que não apresente 100% de eficácia, existe o risco de nos pacientes não respondedores ao tratamento ocorrer avanço da neoplasia. Isso irá, inevitavelmente, diminuir as chances de sobrevivência do paciente.

Pacientes não respondedores ainda sofrem frequentemente de toxicidade, atraso na cirurgia e riscos de complicações. Além disso, existe um aumento significativo no custo do tratamento. Neste sentido, vários estudos vêm sendo realizados na tentativa de se identificar marcadores moleculares que possam diferenciar, a priori, os pacientes respondedores dos não respondedores a determinado tratamento clínico.

A partir desta problemática, e com o objetivo de preservação de órgãos, em 1999, o Hospital AC Camargo realizou um ensaio clínico fase II para preservação de laringe e hipofaringe em carcinoma epidermóide avançado. A terapia ministrada foi baseada na quimioterapia combinada com Paclitaxel e Cisplatina concomitante à radioterapia.

Este ensaio clínico (FEHER et al. 2002a e b) admitiu, originalmente, 43 pacientes, e, como critérios principais de inclusão, os mesmos deveriam ser portadores de carcinoma epidermóide estádios clínicos III ou IV, sem metástase à

distância, virgem de quimio ou radioterapia prévias e sem presença de outro tumor primário.

Os pacientes foram tratados com doses semanais de Paclitaxel, 30mg/m², e Cisplatina, 20mg/m², concomitantemente à radioterapia, realizada na dose de 7040 cGy em frações diárias de 180 cGy. Os pacientes que não apresentaram resposta completa ou parcial com a dose de radiação de 5040 cGy, com tumor residual no final da radioterapia ou no momento da recorrência local foram submetidos à laringectomia. A retirada de linfonodos ocorreu em todos os pacientes linfonodos positivos, independente da resposta à quimiorradioterapia.

Após seguimento clínico, que variou de 3 a 5 anos, os pacientes foram classificados em dois grupos, respondedores (R) ou não respondedores (NR). Os R apresentaram resposta completa ao tratamento, ou seja, remissão da doença local por completo após tratamento quimiorradioterápico, confirmado por tomografia computadorizada, endoscopia e, sempre que necessário, biópsia. Este grupo obteve a preservação do órgão.

O grupo dos NR engloba três subtipos diferentes de resposta dos pacientes. O primeiro subtipo não apresentou nenhuma diminuição no tamanho da doença local, e o segundo apresentou resposta parcial, o que significa que houve diminuição do tumor, mas não o seu desaparecimento completo. Já o último subtipo apresentou recorrência local, casos em que inicialmente houve resposta completa com posterior recidiva local, na laringe ou hipofaringe. Este grupo não obteve preservação do órgão, que era o objetivo do ensaio clínico; por isso esses três subtipos foram classificados como NR.

1.12 IDENTIFICAÇÃO DE CLASSIFICADORES MOLECULARES DE RESPOSTA EM ESTUDO DE MESTRADO

A falha do tratamento em pacientes que foram submetidos à quimiorradioterapia com o objetivo de preservar a laringe pode ser atribuída a múltiplos fatores, como extensão extralaringea e envolvimento subglótico, mas, a dificuldade de prever a ausência de resposta permanece (WEBER et al. 2004).

Assim sendo, torna-se extremamente relevante a busca de marcadores que possam prever determinada resposta de pacientes a determinado tipo de tratamento, ou seja, prever o prognóstico. Assim, nós idealizamos estudo de Mestrado (BARRETO 2005, Proc FAPESP 02/10891-5) com a finalidade de identificar classificadores moleculares que distinguíssem pacientes R dos NR.

No início do ensaio clínico fase II foi realizada biópsia incisional nos pacientes, através de microlaringoscopia, para posterior confirmação histopatológica da doença. Após a biópsia, um fragmento de tecido foi imediatamente congelado em nitrogênio líquido; assim, amostras de 35 pacientes foram armazenadas para serem posteriormente utilizadas no projeto de Mestrado com objetivo de identificação de classificadores moleculares.

Ao final do seguimento clínico dos 35 pacientes (Tabela 1), que variou de 3 a 5 anos, 21 (60%) foram classificados como R e 14 (40%) classificados como NR.

Assim sendo, após ensaio clínico fase II, foi realizado Projeto de Mestrado, que identificou, através da metodologia de cDNA *microarray*, 4 trios (Tabela 2) e 27 quadras de genes (Tabela 3) que predizem, com 100% de sensibilidade e 100% de especificidade, os 35 pacientes classificados como R ou NR. Os classificadores se

baseiam no perfil de expressão gênica combinada de cada trio ou quadra de genes para classificar cada amostra.

Tabela 1 - Características clinicopatológicas dos 35 pacientes incluídos no projeto de Mestrado sobre identificação dos classificadores, estádios III ou IV, sem metástase à distância e sem tratamento prévio.

Características	%
Idade (anos)	
Média	57,5
Intervalo	39-74
Gênero	
Masculino	83
Feminino	17
Sítio Tumoral Acometido	
Laringe Supraglótica	31
Laringe Glótica	31
Laringe Subglótica	0
Laringe Transglótica ^(a)	6
Hipopfaringe - Seio Piriforme	29
Hipopfaringe - Retro-Cricóide	3
Tabagismo	
Positivo	94
Etilismo	
Positivo	83
Diagnóstico Histológico	
CE bem diferenciado	26
CE moderadamente diferenciado	51
CE pouco diferenciado	11
CE basalóide	9
CE basalóide com diferenciação neuroendócrina	3
T	
2	3
3	66
4	31
N	
0	51
1	20
2a-c	20
3	9

^(a)Laringe translótica é tumor que acomete glote e supraglote.

Tabela 2 - Trios de genes que separam pacientes respondedores de não respondedores com 100% de sensibilidade e 100% de especificidade, identificados em projeto de Mestrado. Os trios estão ordenados por ordem decrescentes de SVD.

Trio	Gene 1	Gene 2	Gene 3	SVD R/NR^(b)	VC^(c)
3	ATP9B	KIAA1691	Hs.218008	10,926	35
4	FAD104	PPP1R9B	IFIT1	10,882	35
5	KDEL3	KIAA1691	TRIM9	10,684	35
18	LAMA3	KIAA0819	TRIM2	9,984	35

^(b)*singular value decomposition* ^(c)validação cruzada por *leave-one-out*

Tabela 3 - Quadras de genes que separam pacientes respondedores de não respondedores com 100% de sensibilidade e 100% de especificidade, identificados em projeto de Mestrado. As 27 quadras estão ordenadas por ordem decrescentes de SVD.

Quadra	Gene 1	Gene 2	Gene 3	Gene 4	SVD_R/N ^(b)	VC ^(c)
1	KIAA1691	TRIM9	ATP9B	Hs.218008	16,176	35
4	Hs.218008	KIAA1691	FLJ10648	CGBP	14,825	35
5	ZNF151	KIAA1691	TRIM9	ATP9B	14,762	35
6	ATP9B	KIAA1691	Hs.218008	GTF2H4	14,514	35
8	DFFA	KIAA1691	TRIM9	CDCA4	14,327	35
9	KIAA0819	KIAA1691	TRIM37	HLA-DOB	14,300	35
12	FAD104	PPP1R9B	IFIT1	FLJ20321	14,133	35
13	NOS3	KIAA1691	TRIM9	ATP9B	14,121	35
15	SMCY	TPT	C3AR1	C14orf130	13,830	35
16	DKK4	KIAA1691	HLA-DOB	CLASP2	13,769	35
18	KDEL3	KIAA1691	TRIM9	EIF2B3	13,661	35
19	Hs.289123	TRIM9	KIAA1691	ATP9B	13,536	35
27	DKFZP566C1	RALA	ITGB5	KIAA0095	13,191	35
29	Jade-1	KIAA1691	TRIM9	SCAMP4	13,176	35
30	CGBP	KIAA1691	KIAA0819	TRIM37	13,149	35
32	SDCCAG10	ATP9B	TRIM9	KIAA1691	13,075	35
38	APP	DKFZP566C1	RBM12	PIGL	12,795	35
40	ATP5L	SMCY	AKAP10	DBT	12,769	35
42	DMTF1	KIAA1691	TRIM9	ATP9B	12,712	35
47	BBS2	KIAA1691	KIAA0335	ZDHHC6	12,520	35
53	Hs.379757	KIAA1691	TRIM9	KDEL3	12,257	35
54	IAN4L1	KIAA1691	DKK4	ATP9B	12,233	35
55	AHCYL1	KIAA1691	TRIM9	DUSP18	12,210	35
56	ITGB5	SMCY	ATP9B	DKFZP434C171	12,197	35
62	DKFZP564O043	FAD104	KIAA1170	HMGCL	11,927	35
71	E2IG4	KIAA0819	KIAA1691	PP2135	11,690	35
100	FLJ12565	DKFZP434A1315	KLF8	WWP2	9,980	35

^(b)singular value decomposition ^(c)validação cruzada por *leave-one-out*

Após alguns estudos, percebe-se que a análise do tipo Discriminante Linear de Fisher (HASTIE et al. 2001) consegue manter uma boa relação entre facilidade e desempenho, tornando-se uma ótima escolha para estudo de identificação. Deste modo, a identificação dos classificadores foi realizada pelo método do Discriminante Linear de Fisher, através da estratégia de busca seqüencial, com o algoritmo Busca e

Escolha (CRISTO 2003), utilizando os parâmetros k , l , com $k=100$ e $l=4600$, e teste-t na seleção dos genes para compor os trios e quadras. Esta busca seguiu a mesma estratégia descrita por nosso grupo, em MEIRELES et al. (2004).

Inicialmente, foram identificados os 100 genes com menor p-valor obtido do teste t, para a comparação de respondedores e não respondedores, dentre todos os presentes na lâmina de *array* estudada. Esses 100 genes foram então combinados, um a um, com os (4600-1) genes restantes, a fim de se identificar os 100 pares que melhor separassem os pacientes R dos NR. Em seguida, esses 100 melhores pares foram novamente combinados, um a um, com os (4600-2) genes restantes e foram identificados agora os 100 trios que melhor separaram ambos os grupos.

Finalmente, os 100 melhores trios foram combinados, um a um, com os (4600-3) genes restantes e desta forma foram identificadas as 100 quadras que melhor separaram os grupos R dos NR. Não foi identificado nenhum par de gene capaz de classificar corretamente as amostras.

O método de Fisher calcula uma estatística conhecida como SVD, *singular value decomposition*, número que representa a razão dos desvios padrões inter sobre intragrupos, dos valores de expressão dos genes projetados no sistema de coordenadas encontrados pelo Discriminante de Fisher. O SVD indica quão bem ficam separados os dois grupos estudados (intergrupos) com o conjunto de genes utilizados; medindo o quão compacta estão as duas nuvens de pontos, ou seja, a dispersão dos pontos, e também medindo a distância entre as duas nuvens.

Após, foi aplicada a técnica estatística de validação cruzada (VC), *leave-one-out*, que consiste em processo iterativo, onde em cada iteração é retirada uma amostra do conjunto, o classificador é recalculado, a amostra retirada é classificada e

então o número de acertos é contado. O processo de classificação dessas amostras foi repetido por 35 vezes sendo que, em cada um dos testes, uma das 35 amostras foi deixada de fora do processo de determinação da regra do classificador e a 35ª amostra foi usada como validação.

Quanto maior for o valor de VC, mais eficiente é o classificador; neste caso, eram 35 amostras, então $VC = 35$ indica um classificador excelente, por classifica corretamente 100% dos pacientes respondedores e não respondedores.

Desta forma, os 100 trios e 100 quadras de classificadores moleculares identificados através do método do Discriminante Linear de Fisher foram validados por análise do tipo validação cruzada, que resultou em 4 trios e 27 quadras de genes classificadoras com 100% de acerto ($VC = 35$). Esses classificadores ainda tiveram sua significância estatística determinada por validação do tipo *bootstrap*.

A identificação de classificadores é a identificação de uma regra de classificação, o que torna possível sua aplicação em outro banco de dados, quando se desejar analisar o comportamento do classificador prévio em conjunto diferente de dados.

Apesar da validação matemática realizada, deve-se ainda validar esses classificadores em grupo independente de pacientes, a fim de que os mesmos possam ser utilizados na rotina clínica. Isso ocorre principalmente devido ao fato dos dados terem sido identificados através da metodologia de *array*. Os classificadores identificados também devem ser validados em estudo prospectivo com coorte maior.

1.13 VALIDAÇÃO DOS CLASSIFICADORES MOLECULARES

Se a acurácia do preditor for avaliada no mesmo conjunto de dados que foram utilizados no desenvolvimento da assinatura, o resultado conterà viés. Por esta razão, cada assinatura necessita validação. Existem 3 métodos de validação: validação amostra-independente (*independent-sample validation*); validação com partição do grupo amostral (*split-sample validation*) e validação interna (*internal validation*) (GÖNEN 2009).

Validação independente é considerada o padrão ouro das técnicas de validações de dados gerados por *microarray*, e requer uma obtenção de dados independentes e que não foram utilizados no desenvolvimento da assinatura (GÖNEN 2009).

Validação com partição é a técnica mais semelhante à validação interna. Nesta técnica as amostras são randomicamente divididas em dois grupos, nos quais um é utilizado para o desenvolvimento da assinatura e o outro para a validação. A principal crítica a este método é a ausência de captura da variabilidade entre os estudos, sendo que a validação independente é capaz de identificar essa variabilidade. Por esta razão, a estimativa obtida por este método superestima a real performance preditiva. Além disso, este método requer um número inicial de amostras suficientemente grande para que, quando seja dividido em dois grupos haja amostras suficientes para efetivamente realizar os modelos de desenvolvimento e de validação (GÖNEN 2009).

Mas, a diminuição efetiva do grupo amostral é certamente um incômodo, desta forma, foi desenvolvida validação interna. Esta técnica pode ser realizada via

validação cruzada (*cross-validation*) ou *bootstrap*. Validação cruzada tenta mimetizar a técnica de validação com partição, através da criação de múltiplas divisões do mesmo grupo de dados e repetindo o processo de desenvolvimento do modelo e validação. Validação *bootstrap* é uma variação do mesmo tema, baseada na reclassificação do grupo de dados, no qual as amostras são classificadas do dado original com substituição (GÖNEN 2009).

Validação interna é aceitável durante o desenvolvimento inicial de uma assinatura (SIMON 2005), no entanto, antes de sua introdução no uso clínico, deve-se insistir em validação independente ou validação com partição. Apesar de parecer óbvio, é bastante grave o fato de que validação deficiente foi um dos três erros estatísticos mais comuns em estudos de perfil de expressão por *microarray* (DUPUY e SIMON 2007).

Nas últimas décadas, o desenvolvimento da tecnologia em genômica e proteômica vem tornando possível a análise simultânea de grande número de variáveis (*high-throughput analysis*), possibilitando pesquisa baseada na descoberta. Neste tipo de pesquisa, grande quantidade de dados podem ser analisados sem uma hipótese, a fim de buscar padrões discriminatórios úteis (STEARNS et al. 2003; RANSOHOFF 2003), como por exemplo entre grupos de pessoas com diferentes diagnósticos, prognósticos ou respostas à terapia. Mas, a principal dificuldade é que os resultados podem não ser reprodutíveis (RANSOHOFF 2004).

Overfitting é a identificação de padrão discriminatório devido ao acaso, que pode ocorrer quando grande número de variáveis são avaliadas para um pequeno número de amostras. A solução para este problema é a avaliação correta da reprodutibilidade. (RANSOHOFF 2004).

Para isto, o modelo de reconhecimento do padrão, que é derivado do grupo de treinamento, deve ser aplicado a um grupo de validação independente, composto por indivíduos que não foram utilizados no grupo anterior. Se a discriminação não puder ser reproduzida novamente neste segundo grupo, *overfitting* é a explicação mais plausível (RANSOHOFF 2004).

Apesar disso, validação independente ocorre em apenas 10% das pesquisas com *microarray* (NTZANI e IOANNIDIS 2003). Um dos principais problemas no momento de validar esses experimentos ocorre quando o grupo de validação não é independente. Estudo que avaliou perfil de expressão de 25.000 genes, para prever prognóstico em câncer de mama, utilizou 295 indivíduos no grupo de validação, mas, 61 destes também tinham sido anteriormente utilizados no grupo de treinamento (VAN DE VIJVER et al. 2002; VAN'T VEER et al. 2002).

Outro problema também freqüente nesses estudos é a apresentação de grupo de validação muito reduzido. Um segundo estudo avaliou o perfil de expressão de 12.000 genes, a fim de prever prognóstico em câncer de mama, em apenas 13 indivíduos (HUANG et al. 2003), número reduzido demais para ser significativo (BAKER et al. 2002; SIMON et al. 2003). Outro estudo, que analisou 24.000 genes, apresentou 24 indivíduos no grupo de treinamento e apenas 6 no de validação (CHANG et al. 2003).

PRAMANA et al. (2007), através da análise da expressão gênica, identificaram preditor de resposta à quimiorradioterapia (cisplatina) em CECP em estágio avançado. No entanto, quando as amostras foram divididas em grupos de treino e de validação, nenhum classificador robusto pode ser identificado.

Deste modo, é necessário que seja realizada validação independente, pra que os classificadores moleculares identificados previamente em estudo de Mestrado possam ser utilizados na rotina clínica. Após, ainda deve ser realizada validação em estudo prospectivo com coorte maior.

2 JUSTIFICATIVA

A disponibilidade de classificadores moleculares, que possam prever a resposta de tumores ao tratamento, apresentam grande importância na escolha das condutas clínicas a serem oferecidas aos pacientes. Além disso, em todo tratamento oncológico conservador que não apresente 100% de eficácia existe o risco de, nos pacientes não respondedores ao tratamento, ocorrer avanço da doença. Isso irá, inevitavelmente, diminuir as chances de sobrevivência do paciente. Em nosso projeto prévio de Mestrado, foi identificado e validado matematicamente conjuntos de genes cujos níveis de expressão combinados podem ser utilizados como classificadores moleculares. Esses classificadores são capazes de agrupar, com 100% de precisão, pacientes R e NR ao tratamento clínico, realizado com quimioterapia concomitante à radioterapia, em carcinoma epidermóide de laringe e hipofaringe localmente avançado. Mas, para que esses classificadores sejam introduzidos na rotina clínica, é necessário validá-los em grupo independente de amostras, uma vez que validação independente é o padrão ouro das validações de dados gerados por *microarray*. Após, também se faz necessária a realização de estudo prospectivo com coorte maior. Uma vez devidamente validados, os classificadores poderão guiar o paciente para o tratamento mais eficaz e de forma individualizada, já que os R serão encaminhados para quimioterapia concomitante à radioterapia e apresentarão preservação de órgão; enquanto que os NR serão encaminhados de imediato para cirurgia, evitando progressão da doença, ressecção mais invasiva e toxicidade desnecessária, além da diminuição dos custos do tratamento.

3 OBJETIVOS

3.1 OBJETIVO GERAL

Validar classificadores moleculares preditores de resposta (identificados em grupo de 35 amostras, em projeto de Mestrado) à quimiorradioterapia (cisplatina e paclitaxel), em carcinoma epidermóide de laringe e hipofaringe localmente avançado, em grupo de 28 amostras.

3.2 OBJETIVOS SECUNDÁRIOS

- Estudar vias metabólicas potencialmente envolvidas na ação de cisplatina/taxol concomitante à radioterapia, através da análise da expressão gênica diferencial em grupo de 35 pacientes (projeto de Mestrado), classificados em respondedores ou não respondedores.
- Validar os genes identificados, nas 28 amostras do grupo da validação, através da análise do perfil de expressão gênica diferencial entre respondedores e não respondedores.

4 MATERIAL E MÉTODOS

4.1 OBJETIVO GERAL: VALIDAÇÃO DOS CLASSIFICADORES MOLECULARES

4.1.1 Critérios de inclusão da população de estudo

Apenas os pacientes que preencheram todos os critérios abaixo foram incluídos no estudo da validação dos classificadores moleculares, anteriormente identificados em estudo de Mestrado.

Esses pacientes do presente estudo da validação receberam o mesmo tratamento clínico que os pacientes do estudo prévio da identificação. Da mesma forma, foram utilizados os mesmos critérios de inclusão e exclusão dos pacientes para ambos os estudos, o da identificação e o da validação dos classificadores moleculares.

- Diagnóstico histológico de carcinoma epidermóide provenientes das regiões laringe ou hipofaringe;
- Faixa etária entre 18 e 70 anos;
- Estádios T3 ou T4, N0 a N3, sem invasão maciça de partes moles;
- Candidatos à laringectomia *near total*, total ou faringolaringectomia como tratamento locorregional;
- Metástases linfonodais ressecáveis;
- Ausência de tratamentos prévios do tumor como cirurgia, radioterapia ou quimioterapia; biópsia prévia da lesão e traqueostomias são permitidas;
- Ausência de doença metastática à distância;
- Condição de desempenho (escala de Karnofsky) de 70 ou mais;

- Expectativa de vida estimada de pelo menos 6 meses;
- Reserva medular adequada: leucócitos $\geq 3.500/\text{mm}^3$, neutrófilos $\geq 1.500/\text{mm}^3$, plaquetas $\geq 100.000/\text{mm}^3$ e hemoglobina $\geq 9,0 \text{ g/dL}$;
- Função hepática adequada: bilirrubina total $\leq 1,5$ vezes o limite superior normal; TGO e TGP ≤ 3 vezes o limite superior normal;
- Função renal adequada: creatinina $\leq 1,5$ vezes o limite superior normal;
- Aderência do paciente e proximidade geográfica que permitam acompanhamento adequado;
- Assinatura de consentimento pós-informado;
- Clinicamente ASA I a III.

4.1.2 Critérios de exclusão da população de estudo

Pacientes que preencheram qualquer um dos critérios abaixo foram excluídos do estudo:

- Doença cardíaca ativa ou não compensada por tratamento, ou infarto agudo do miocárdio nos 6 meses prévios;
- Infecção ativa presente (ao critério do investigador);
- Doenças sistêmicas concomitantes consideradas sérias pelo investigador (e.g. diabetes, hipertensão, etc.);
- Outra neoplasia primária (exceto carcinoma *in situ* da cérvix uterina ou CE adequadamente tratado), exceto se após tratamento curativo, há mais de 5 anos;
- Presença de doença psiquiátrica severa;
- Participação em outro protocolo experimental.

4.1.3 Quimioterapia da população de estudo com paclitaxel

Os pacientes foram tratados com doses semanais de $30\text{mg}/\text{m}^2$ de Paclitaxel. O fármaco foi diluído em 500 mL de solução salina 0,9% ou de glicose 5%, em frascos

de poliolefina ou vidro, com equipo de nitroglicerina polietinilada. Filtros em linha de acetato de celulose com poros de 0,22 μm foram utilizados.

Premedicação para o Paclitaxel foi utilizada com o intuito profilático, para reações de hipersensibilidade ao cremofor (diluyente). Foi administrada precedendo o Paclitaxel por 1 hora: Dexametasona - 20 mg i.v.; Ranitidina - 50 mg i.v. e Difenidramina - 50 mg i.v.

Paclitaxel foi administrado em uma hora. Os ciclos de Paclitaxel se repetiram a cada semana por 5 ciclos, na fase inicial (neoadjuvante). A dose foi calculada com base na superfície corpórea, utilizando-se o peso real do paciente.

4.1.4 Quimioterapia da população de estudo com cisplatina

Os pacientes foram tratados com doses semanais de 20mg/m² de Cisplatina. O fármaco foi diluído em 500 mL de solução salina 0,9% contendo 22,5 g de manitol. A infusão foi precedida de hidratação com 500 mL de solução salina 0,9% e foi precedida de medicação antiemética profilática (e.g. ondansetron). O tempo de infusão foi de 1 hora.

4.1.5 Radioterapia da população de estudo

Foi realizada radioterapia 2D, com simulador convencional. A dose total foi de 7040 cGy, sendo administrada 180 cGy em fração única diária por 5 dias/semana, durante 7 1/2 semanas. Foi efetuada em acelerador linear de 4 MeV em campos englobando a região cervical até 5040 cGy, com proteção medular aos 4500 cGy, e, em seguida, 2 reduções sobre a lesão primária aos 5040 e 6400 cGy. O *boost* final, após 6400 cGy, englobou a lesão primária com margens de 1 cm.

Os pacientes que não apresentaram resposta completa ou parcial com a dose de radiação de 5040 cGy foram submetidos à laringectomia.

4.1.6 Coleta de amostras, de dados clínicos e seguimento dos pacientes

Um importante avanço clínico na escolha dos pacientes a serem submetidos ao tratamento de preservação de órgão, ou à cirurgia, seria a disponibilidade de classificadores que pudessem indicar, através de uma biópsia prévia, os pacientes que responderiam à quimiorradioterapia. Assim, seria evitada progressão tumoral em pacientes que não responderiam ao tratamento clínico, pois estes seriam encaminhados de imediato à cirurgia. Deste modo, ocorreriam aumento na sobrevida global, melhora na qualidade de vida, reduções na morbidade e custo do tratamento dos pacientes portadores de CE.

Apesar da validação matemática dos classificadores previamente identificados ter sido realizada em estudo anterior, deve-se validá-los em um grupo independente de amostras, uma vez que validação independente é considerada o padrão ouro das técnicas de validação em experimentos de *microarray* (GÖNEN 2009).

Esta etapa é crítica para afastar as possibilidades de achados ao acaso, uma preocupação sempre constante com dados gerados por experimentos de *cDNA array*, onde o número de genes testados e por conseqüência o número de possíveis classificadores, é desproporcionalmente maior que o número de amostras testadas (RANSOHOFF 2004). Assim sendo, se apropriadamente validados, existe grande possibilidade dos classificadores serem utilizados na rotina médica, o objetivo principal deste estudo.

É importante ressaltar que, a fim de que os classificadores sejam validados, o tratamento clínico recebido pelos pacientes deste estudo deve ser exatamente o mesmo recebido pelo grupo anterior. Da mesma forma, a biópsia do paciente deve ser coletada previamente ao início de qualquer quimiorradioterapia.

Assim, foi realizada colaboração com outras três instituições brasileiras, e o Projeto de Preservação de Órgãos foi re-escrito, atualizado, enviado e aprovado pelos devidos Comitês de Ética e Pesquisa das três instituições colaboradoras.

As instituições colaboradoras foram o Hospital do Câncer do Ceará, Fortaleza, com 9 amostras coletadas; o Hospital de Câncer de Barretos -Fundação Pio XII-, São Paulo, com 15 amostras; o Hospital Araújo Jorge, de Goiânia, com 3 amostras coletadas e 8 amostras em nossa instituição, Hospital AC Camargo.

Tubos contendo *RNAlater* -Ambion- foram enviados para as instituições e, após a confirmação dos critérios de inclusão, aceite do paciente e assinatura do termo de consentimento pós-informado, foi realizada biópsia incisional através de microlaringoscopia, para posterior confirmação histopatológica da doença. Um fragmento de tecido foi imediatamente armazenado em tubo contendo *RNAlater*.

RNAlater é um reagente que estabiliza e protege o RNA celular em amostras de tecido que não estão congeladas. A vantagem deste reagente e o principal motivo de sua utilização é a eliminação da necessidade de processamento ou congelamento imediatos da amostra. Além disso, não existe prejuízo na qualidade e nem na quantidade do RNA que é posteriormente extraído. A amostra deve ter no máximo 0,5 cm em qualquer dimensão e ser embebida em 5 volumes de *RNAlater*, desta forma, a substância penetrará e conservará toda a amostra (<http://www.ambion.com>). Após, a amostra deve permanecer um dia a 4°C ou à TA e, somente após, ser

armazenada a -80°C , para que o tecido seja completamente penetrado pela substância.

No momento da coleta das amostras, também foram coletados todos os dados clínico-patológicos desses pacientes, através da revisão dos prontuários em cada centro colaborador. Após o início do projeto, todos os pacientes tiveram sua evolução clínica acompanhada, através do seguimento dos prontuários.

4.1.7 Análise e microdissecção das biópsias

Após a coleta da biópsia, uma lâmina de cada amostra foi corada por HE e posteriormente analisada pelo Serviço de Anatomia Patológica do Hospital AC Camargo, para confirmação do diagnóstico histológico e realização de semimicrodissecção manual de amostras que apresentassem fragmentos indesejáveis, como excesso de tecido normal ou de infiltrado inflamatório, e ainda necrose.

Apesar de alguns trabalhos utilizarem microdissecção a laser (LCM) (ALEVIZOS et al. 2001; YAMABUKI et al. 2006), estudos do perfil de expressão de tumores sólidos geralmente empregam a secção tumoral inteira, contendo células tumorais e microambiente ao redor, o que inclui componentes da matriz extracelular e células estromais, como fibroblastos e células de resposta imune (BISSELL e RADISKY 2001).

O padrão de expressão gênica do tumor é derivado tanto de células tumorais como de estromais, e também de fatores inflamatórios (ROSS et al. 2000); deste modo, é importante considerar que a inclusão de células estromais influencia a resposta de estudos do perfil de expressão tumoral (ROEPMAN et al. 2006b). Isso ocorre porque células cancerosas são entremeadas por estroma benigno, vasos

sanguíneos e células inflamatórias (HEPPER 1984; LIU et al. 2004; STEEG 2008; WU et al. 2008).

Estudo demonstrou que o perfil de expressão gênica estromal contém tanto genes tumorais quanto estromais, o que indica que as células tumorais disseminadas alteram seu estroma adjacente para um microambiente “metastático”. Este é similar ao microambiente do tumor primário, no qual as células tumorais podem sobreviver e proliferar (MUELLER e FUSENING 2004).

A retirada de tecido normal em excesso foi realizada a fim de enriquecer o material com células tumorais, reduzindo a heterogeneidade das amostras que terão seu RNA utilizado nos experimentos de *microarray*. Apenas amostras com no mínimo 70% de tecido tumoral (CE) e com ausência de displasia ou carcinoma *in situ* foram utilizadas.

ROEPMAN et al. (2006b) demonstraram que a utilização de material para experimentos de *microarray* com proporções reduzidas de células tumorais, no caso menos de 50%, resultou em decréscimo da acurácia do preditor e também que a porcentagem tumoral ótima é de 60% a 70%. A distribuição assimétrica de assinaturas gênicas dos vários componentes explica a perda de acurácia em amostras com pequenas quantidades de células tumorais.

Deste modo, após a realização e visualização das lâminas de HE com o auxílio de microscópio, 2 de 9 amostras provenientes de Fortaleza foram descartadas. Uma delas apresentava tecido inflamatório infiltrando toda a amostra e tecido normal, sem a presença de tumor; enquanto que a segunda apresentava bastante tecido normal e apenas tumor *in situ*.

Dentre 15 amostras provenientes de Barretos 2 foram descartadas, uma por não apresentar CE invasor e outra por apresentar RNA degradado após a extração. Nenhuma das 3 amostras provenientes de Goiânia foi descartada.

Das 8 amostras provenientes de São Paulo, 2 tiveram necessidade de ser descartadas, uma por apresentar RNA degradado após extração e a outra devido à saída do paciente do Protocolo.

Desta forma, dentre as 35 amostras coletadas no total, 29 estavam sendo utilizadas no Projeto. Mas, após a finalização dos experimentos de *microarray*, a amostra número 3 teve que ser descartada, pois o paciente não finalizou o tratamento clínico. Assim, 28 amostras totalizaram o projeto de validação.

4.1.8 Avaliação do número total de amostras

O objetivo deste projeto consiste em validar os classificadores previamente identificados em Estudo de Mestrado (BARRETO 2005, Proc FAPESP 02/10891-5). Trinta e cinco pacientes fizeram parte deste projeto de Mestrado, ou seja, 35 amostras. É importante lembrar que os classificadores prévios classificaram 100% das amostras corretamente; ou seja, a sensibilidade foi de 100% (verdadeiro positivo) e sem nenhum falso negativo, levando a 100% de especificidade (Tabela 4).

Com o objetivo de avaliar a quantidade de amostras necessárias para tornar o presente projeto viável, foi realizado cálculo de intervalo de confiança (Tabela 5) com diversos números de amostras. O cálculo foi realizado com auxílio do *Website Vassar* (<http://faculty.vassar.edu/lowry/clin1.html>), onde é possível realizar diversas simulações.

Tabela 4 - Modelo do cálculo do Intervalo de Confiança, para classificadores moleculares identificados previamente em projeto de Mestrado. Sensibilidade = $VP/TP = 60/60 = 100\%$. Especificidade: $VN/TN = 40/40 = 100\%$.

		Condição:	
		Ausente	Presente
Ensaio clínico (classificador)	Teste +	Erro I: 0% Falso Positivo (FP)	Acerto: 60% Verdadeiro Positivo (VP)
	Teste -	Acerto: 40% Verdadeiro Negativo (VN)	Erro II: 0% Falso Negativo (FN)

Tabela 5 - Cálculo do intervalo de confiança para IC 95% de sensibilidade e para IC 95% de especificidade.

Número de Amostras	IC 95% Sensibilidade	IC 95% Especificidade
20	[69,87-100%]	[59,7-100%]
25	[74,65-100%]	[65,5-100%]
28	[77,07-100%]	[67,8-100%]
30	[78,12-100%]	[69,87-100%]

A sensibilidade é a capacidade do teste em identificar corretamente os R; enquanto que a especificidade é a capacidade de classificar corretamente os NR. Quanto maior forem os valores de ambos, melhor será o teste.

O número final de amostras utilizadas na validação dos classificadores moleculares foi de 28 (Tabela 6); e, com este número amostral, torna-se possível obter sensibilidade e especificidade acima de 60%.

Neste caso, valores acima de 60% podem ser considerados como de um bom classificador, uma vez que a distribuição dos pacientes do projeto de Mestrado foi de 60% R e 40% NR. Assim sendo, se todos os pacientes fossem encaminhados à

quimiorradioterapia, 60% responderiam ao tratamento e obteriam o benefício da preservação de órgão; enquanto que 40% dos pacientes teriam prejuízo.

Tabela 6 - Características clinicopatológicas das 28 amostras, ou pacientes, incluídos no presente projeto.

Características	%
Idade (anos)	
Média :	57,5
Mediana:	58,5
Intervalo:	41-68
Gênero	
Masculino	85,7
Feminino	14,3
Sítio Tumoral Acometido	
Laringe Glótica	7,15
Laringe Transglótica ^(a)	28,5
Laringe Supraglótica	7,15
Laringe Supraglótica e Hipofaringe	25,0
Laringe Transglótica e Hipofaringe	17,9
Hipofaringe	14,3
Tabagismo	
Positivo	100
Etilismo	
Positivo	85,7
Diagnóstico Histológico	
CE bem diferenciado	14,3
CE moderadamente diferenciado	60,7
CE pouco diferenciado	25,0
T	
2	3,6
3	75,0
4	21,4
N	
0	57,1
1	32,1
2a-c	10,8

^(a)Laringe Transglótica é tumor que acomete glote e supraglote.

4.1.9 Extração do RNA total das amostras

A etapa seguinte foi a extração do RNA total das 31 amostras provenientes dos quatro centros colaboradores. Para isso, os fragmentos de biópsia foram transferidos para tubos contendo de 2 a 3 mL de TRIZOL[®] (Life Technologies, USA), de acordo com o tamanho da biópsia, e triturados com o auxílio de homogenizador (Polytron[®], Kinematica AG, Switzerland).

A trituração foi realizada em tubo imerso em gelo e até não existir mais partículas visíveis. Um mL do triturado foi distribuído em novo tubo e deixado por 5 minutos à TA. Após, foi centrifugado a 12.000g / 4°C por 10 minutos e o sobrenadante foi coletado em tubo contendo 200 µL de clorofórmio. Foi homogeneizado por inversão e deixado por 5 minutos à TA. Em seguida, foi centrifugado a 12.000g / 4°C por 15 minutos e o sobrenadante foi coletado em tubo contendo 500 µL de isopropanol. Foi novamente deixado à TA por 10 minutos e, após, foi centrifugado a 12.000g / 4°C por 10 minutos e o sobrenadante foi descartado. Em seguida, foi adicionado 1 mL de etanol 70% (em água tratada com DEPC -Merck, Germany) e o *pellet* foi solto e lavado, após, foi realizada nova centrifugação a 7.600g / 4°C por 5 minutos e o sobrenadante foi novamente descartado. O *pellet* foi lavado da mesma forma por mais duas vezes.

Terminada a lavagem, o *pellet* foi seco em centrífuga a vácuo e ressuspendido em 15 a 140 µL de água tratada com DEPC, de acordo com seu tamanho. Os tubos contendo RNAs foram então aquecidos a 55°C por 10 minutos e ressuspendidos por pipetagem; em seguida, foram armazenados em freezer -80°C.

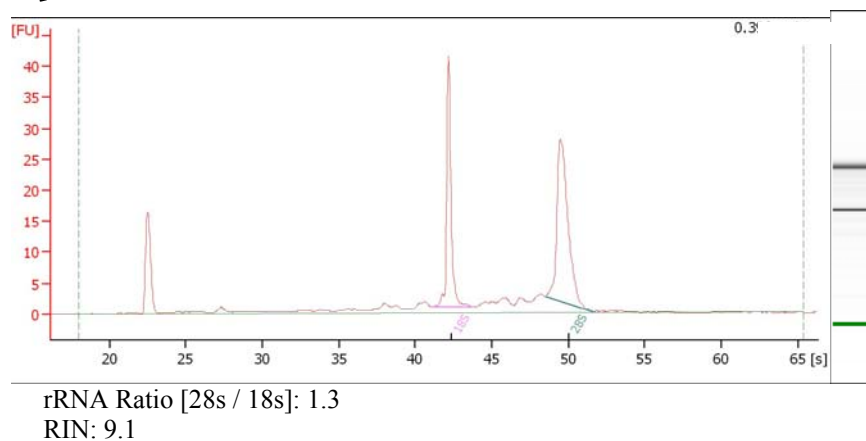
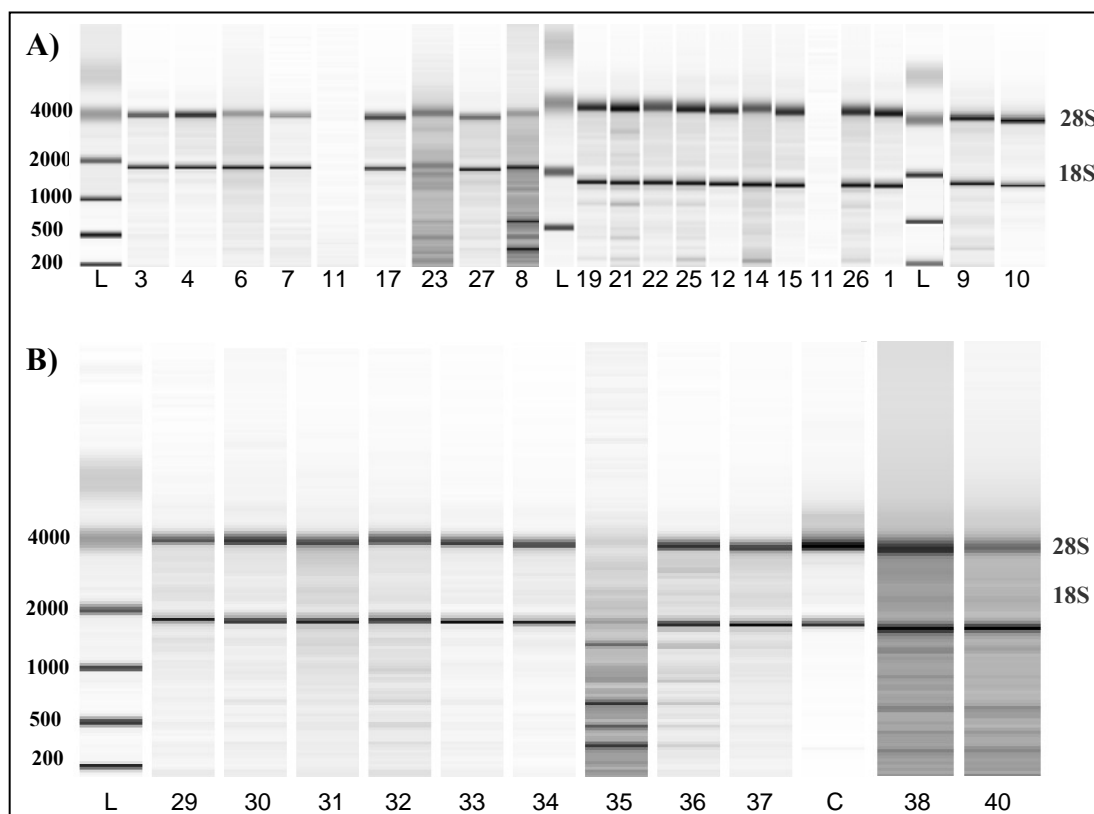
A concentração e pureza dos produtos foram aferidas através de espectrofotômetro NanoDrop (ND-1000 Spectrophotometer), utilizando-se 1 μ L de RNA total (Tabela 7).

Tabela 7 - Avaliação do rendimento e pureza do RNA total extraído das 31 amostras. Esta análise foi realizada com 1 μ L de RNA total em NanoDrop.

Amostra	[ng/ul]	260/280	260/230	Rendimento RNAt extraído: ug totais
3 FO	906,11	1,94	2,25	63,427
4 FO	909,97	1,85	2,26	72,7976
12 SP	852,32	1,92	2,05	119,28
19 BA	1137,8	1,98	1,7	125,67
21 BA	1133,2	1,94	2,13	95,12
27 FO	902,12	1,59	2,43	90,21
26 SP	576,1	1,91	1,57	8,64
1 FO	1284,27	1,88	2,17	51,37
6 FO	727,93	1,49	1,25	27,66
7 FO	768,48	1,84	2,02	24,59
9 SP	3152	1,93	1,89	15,13
14 SP	717,33	1,69	1,26	21,519
17 BA	1026,78	1,93	1,86	82,1424
22 BA	844,93	1,85	1,22	25,347
23 BA	993,86	1,98	1,88	59,6316
8 FO	367,88	1,07	1,03	11,03
10 SP	1008	1,94	1,95	15,02
15 BA	717,17	1,91	1,22	10,757
25 BA	825,85	1,9	1,46	12,387
11 SP	724	2,02	1,4	5,375
29 GO	1518,66	2,02	2,09	91,11
30 GO	784,5	2	1,86	70,6
31 GO	1064,34	2,08	1,81	53,21
32 BA	831,52	2,05	1,85	33,26
33 BA	1579,29	2,06	2,04	78,96
34 BA	1841,68	2,03	2,14	110,5
35 BA	211,15	2	1,33	35,338
36 BA	1242,91	2,07	2,03	49,71
37 BA	1414,86	2,04	2,05	70,74
38 SP	1291,09	1,96	1,9	77,46
40 BA	1064,28	1,94	1,92	63,85

Em seguida, foi realizada análise de um segundo controle de qualidade do RNA (Figura 1 A e B), através do perfil eletroforético realizado em equipamento Agilent 2100 Bioanalyzer e utilizando-se 500 ng de RNA total, seguindo as instruções dadas pelo fabricante.

As amostras que obtiveram a relação $28S/18S \geq 1$ e o RIN (SCHROEDER et al. 2006), número de integridade de RNA ≥ 5 , são classificadas como tendo boa qualidade e aptas para serem utilizadas em experimentos de *microarray*. Dentre o total de 31 amostras, 2 não apresentaram qualidade compatível com os experimentos de *array*, números 11 e 35, devido à degradação dos RNA. A análise em Bioanalyzer foi repetida, para confirmação da degradação da amostra 11, e novamente o mesmo dado foi constatado. Desta forma, após a extração do RNAt, foram totalizadas 29 amostras para prosseguir com os experimentos de *microarray*.



Legenda: Perfil eletroforético avaliado em Bioanalyzer, das bandas 28S e 18S ribossomais, do RNA total extraído de 20 amostras na figura A e de 11 amostras na figura B. As amostras números 11 e 35 apresentam-se degradadas. As canaletas L são referentes ao marcador de tamanho *ladder* e C ao controle HB4. Abaixo, exemplificação de curva de eletroesferograma, com picos das bandas 28S e 18S ribossomais do RNA total.

Figura 1 - Avaliação da integridade do RNA extraído das 31 amostras.

4.1.10 Desenho experimental e RNA referência

De acordo com o desenho experimental deste estudo e seu objetivo de determinar o perfil de expressão gênica por cDNA *microarray* de várias amostras biológicas, torna-se necessário normalizar os dados coibridizando-os com algum RNA referência. A utilização da referência não influencia na qualidade dos resultados gerados; assim sendo, todas as amostras deste estudo serão coibridizadas com uma referência comum, como proposto por POLLACK (2003), de modo a diminuir a quantidade total de hibridizações necessárias.

A referência é uma amostra de RNA que é marcada com um fluorocromo e hibridizada juntamente com cada um dos RNAs de cada amostra de interesse, por sua vez marcado com um fluorocromo distinto. Em seguida é realizado um segundo experimento com inversão dos corantes, que funciona como duplicata do experimento e biológica. Assim, a quantidade relativa de cada RNAm em uma amostra será calculada como a razão entre a intensidade da amostra sobre a intensidade da referência; e todas as amostras poderão ser comparadas entre si através da razão das razões (YANG e SPEED 2002).

Segundo POLLACK (2003), a referência ideal é aquela que produz um sinal mínimo em todos os elementos do *array*, além de ser facilmente produzida e reproduzível, não só no tempo, mas também em diferentes laboratórios. Desta modo, o que mais se aproxima da referência ideal é um conjunto de linhagens celulares, como o utilizado na Universidade de Stanford (PEROU et al. 1999, 2000; ROSS et al. 2000). Entretanto, as 11 linhagens citadas nesses trabalhos não estão disponíveis em nossa instituição, assim, foi optado por um novo conjunto de 15 linhagens celulares (Tabela 8), com o cuidado de selecionar tipos celulares equivalentes.

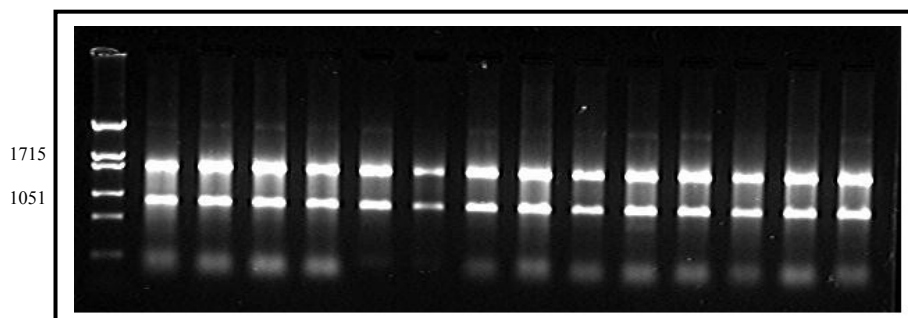
Tabela 8 - Quinze linhagens celulares que totalizaram o *pool* de RNA referência.

Linhagem	Descrição	Método de extração de RNA	Meio de Cultura
Daudi	Linfoma de Burkitt	TRIZOL	RPMI 1640 10% SFB
DLD-1	Adenocarcinoma de Cólon	TRIZOL	RPMI 1640 10% SFB
DU 145	Carcinoma de Próstata	TRIZOL	MEM 10% SFB
FaDU	Carcinoma Epidermóide de Faringe	TRIZOL	MEM 10% SFB
GM 637	Fibroblasto	RNeasy Midi	DMEM 10% SFB
H 1080	Fibrossarcoma	TRIZOL	DMEM 10% SFB
H 146	Carcinoma de Não Peq. Células de Pulmão	RNeasy Midi	RPMI 1640 10% SFB
HB4a	Célula Luminal de Mama	Cloreto de Césio	RPMI 1640 10% SFB
HEK 293	Rim Embrionário	RNeasy Midi	MEM 10% SFB
Jurkat	Leucemia Aguda de Células T	TRIZOL	RPMI 1640 10% SFB
Saos-2	Osteossarcoma	TRIZOL	McCoy's 5a medium (modified) + 15% SFB
SK-BR-3	Adenocarcinoma de Mama	TRIZOL	McCoy's 5a medium (modified) + 10% SFB
SK-MEL-28	Melanoma	TRIZOL	MEM 10% SFB
T24	Carcinoma de Bexiga	TRIZOL	MEM 10% SFB
T98G	Glioblastoma	TRIZOL	MEM 10% SFB

Essas linhagens foram cultivadas até atingir quantidade suficiente, e em seguida tiveram seu RNA total extraído (Figura 2), com o mesmo protocolo do TRIZOL que descrito anteriormente para as amostras de interesse; ou, no caso da utilização de *kit* de extração, foram seguidas as instruções dadas pelo fabricante.

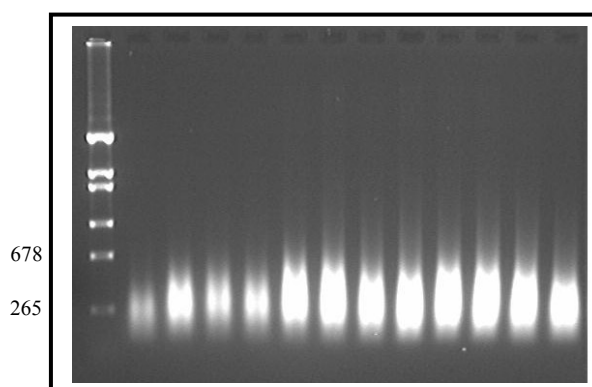
A fim de evitar inserção de viés, todo o material extraído foi reunido em um único *pool*, totalizando um estoque de 1,816 $\mu\text{g}/\mu\text{L}$. Em seguida, foram feitas alíquotas na concentração de 450 $\text{ng}/\mu\text{L}$ e armazenadas em freezer -80°C , em 3 volumes de etanol 100%.

Após, o RNA referência foi amplificado em um ciclo (Figura 3), com o mesmo protocolo descrito posteriormente para as amostras de interesse.



Legenda - Perfil eletroforético das bandas 28S e 18S ribossomais, do RNA total extraído de 14 de 15 linhagens celulares do RNA referência. Um μg de RNA total de amostra foi fracionado por eletroforese em gel de agarose 1% sob condições desnaturantes, e corado por brometo de etídeo. A primeira canaleta é referente ao marcador de tamanho em base, pel Hind III.

Figura 2 - Avaliação da integridade do RNA referência extraído.



Legenda: Perfil eletroforético de RNA referência amplificado em um ciclo, de 12 de 15 amostras. Um μg de RNAa foi fracionado por eletroforese em gel de agarose 1% sob condições desnaturantes, e corado por brometo de etídeo. A primeira canaleta é referente ao marcador de tamanho em base, pel Hind III.

Figura 3 - Avaliação da integridade do RNA referência amplificado em um ciclo.

Após finalizada extração, amplificação e marcação do RNA das amostras de interesse e do RNA referência, as hibridizações foram iniciadas utilizando-se o RNA referência descrito acima. Mas, a incorporação dos corantes a esta leva de RNA referência não foi satisfatória e conseqüentemente as hibridizações não atingiram a qualidade esperada. Assim, ao invés de utilizar 6 μg de RNAa referência por

amostra, foi optado por utilizar 10 μg de RNAa. Para isso, seria necessária amplificação de uma nova leva de RNA referência, já que não haveria quantidade suficiente de RNAa para este súbito aumento de 4 μg por amostra.

É importante lembrar que sempre o mesmo *pool* de RNAa referência deve ser utilizado em todo o projeto e em todas as hibridizações que serão comparadas entre si, a fim de evitar inserção de viés.

Assim, foram então realizados 2 ciclos de amplificação, a fim de se obter quantidade de RNA referência suficiente para ser utilizada em todo o projeto. A massa inicial de RNA total amplificado foi de 18 μg .

Inicialmente, o RNA referência que estava armazenado em etanol foi precipitado, através da adição de 1 μL de NaOAc (3 M, pH 5,4) e resfriado a -20°C por 12 horas. Após a centrifugação e lavagem do *pellet* de RNA com etanol 70%, o *pellet* foi ressuscitado em 12 μL de água tratada com DEPC e este volume foi então dividido igualmente em 6 tubos, para que cada um contivesse 3 μg de RNA total, quantidade ideal para realização do protocolo de amplificação.

Em seguida, o primeiro ciclo de amplificação do RNA referência (Tabela 9) ocorreu de maneira idêntica à descrita posteriormente para as amostras de RNA dos 29 pacientes de interesse.

Tabela 9 - Avaliação do rendimento e pureza do RNA amplificado em um ciclo, a partir do RNA total da referência. Esta análise foi realizada em aparelho NanoDrop e com 1 μL de RNAa.

Tubo	260/280	260/230	[RNA]	Total (18 μL)
1	2,21	2,79	400 ng/ μL	7,2 μg
2	2,27	2,66	1650 ng/ μL	29,7 μg
3	2,23	2,69	1373 ng/ μL	24,7 μg
4	2,27	2,72	1258 ng/ μL	22,6 μg
5	2,23	2,73	1234 ng/ μL	22,2 μg
6	2,22	2,68	1228 ng/ μL	22,1 μg

A etapa seguinte foi o segundo ciclo de amplificação. Apenas RNAs dos tubos 1 e 2 citados acima foram amplificados novamente, uma vez que esta massa de RNA seria suficiente para obtenção de massa de RNA desejada ao final do segundo ciclo.

Foram então realizadas 9 reações e, em cada uma delas, a massa inicial de RNA amplificado (1 ciclo) foi igual a 800 ng. Nos tubos 1 a 6 (Tabela 10) do segundo ciclo foram utilizados RNA do tubo 1 do primeiro ciclo e nos tubos 7 a 9, foram utilizados RNA do tubo 2 do primeiro ciclo.

Tabela 10 - Avaliação do rendimento e pureza do RNA referência amplificado em segundo ciclo. Esta análise foi realizada em aparelho NanoDrop e com 1 μ L de RNAs.

Tubo	260/280	260/230	[RNA]	Total (28 μ L)
1	2,18	2,67	2678 ng/ μ L	75 μ g
2	2,13	2,66	3082 ng/ μ L	86 μ g
3	2,19	2,71	3005 ng/ μ L	84 μ g
4	2,14	2,69	3100 ng/ μ L	87 μ g
5	2,17	2,67	3032 ng/ μ L	85 μ g
6	2,15	2,71	3015 ng/ μ L	84 μ g
7	2,14	2,68	3050 ng/ μ L	85 μ g
8	2,13	2,69	3005 ng/ μ L	84 μ g
9	2,14	2,67	3250 ng/ μ L	91 μ g

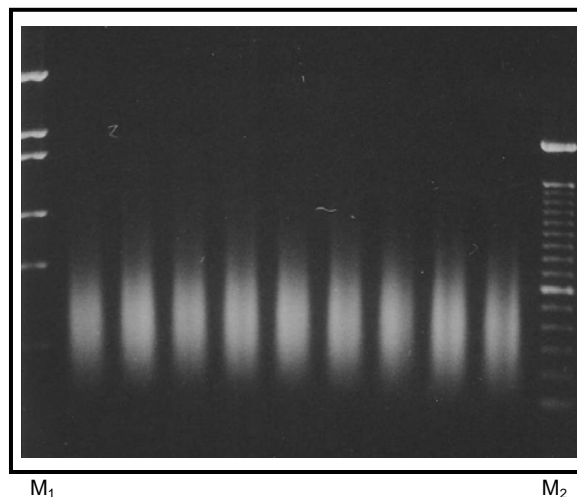
A síntese da segunda fita de cDNA foi realizada utilizando-se 0,2 μ L de RNase H [5 U/ μ L] para [Final = 1U], incubação a 37°C por 10 minutos, desnaturação a 95°C por 2 minutos seguidos de resfriamento em gelo. Em seguida, foram adicionados 0,15 μ g de Oligo dT24-T7 [0,5 μ g/ μ L] e a reação foi aquecida a 42°C por 10 minutos. Foram então adicionados 16,4 μ L de água tratada com DEPC; 10 μ L de 2nd *strand buffer* [5x] caseiro (Apêndice 1) para [Final = 1x]; 1 μ L de dNTP mix [10 mM] para [Final = 0,33 mM]; 0,5 μ L de *E. coli* DNA Ligase [10 U/ μ L] para [Final = 5U] e 1,3 μ L de DNA *Polymerase* I [10 U/ μ L] para [Final = 13U] seguida

de incubação a 16°C por 2 horas. Imediatamente, foi adicionado 1 µL de T4 DNA *Polymerase* [5 U/µL] para [Final = 5U] e as amostras foram novamente incubada a 16° C por 10 minutos.

As etapas de purificação do cDNA, transcrição *in vitro* e purificação do RNAa foram realizadas conforme descrito, posteriormente, no protocolo do primeiro ciclo de amplificação referente as 29 amostras de interesse. No entanto, o RNA referência foi ressuspensionado em 30 µL de água tratada com DEPC, sendo 1 µL utilizado para quantificação em espectrofotômetro NanoDrop (Tabela 10) e 1 µL utilizado para visualização de perfil do RNA, através de fracionamento em gel de agarose (Figura 4).

Para a utilização do RNA referência amplificado em 2 ciclos foi feito um *pool* com todo o material obtido nos 9 tubos, a fim de uniformizar ao máximo o RNA referência para evitar a inserção de vies.

Após finalizada a amplificação do RNA referência em 2 ciclos, foi realizada síntese do cDNA com 10 µg de RNAa referência por amostra. Em seguida, foi realizada marcação e hibridização; os protocolos utilizados novamente foram os mesmos que utilizados para as amostras de interesse.



Legenda: Perfil eletroforético de RNA referência amplificado em 2 ciclos, de 9 amostras. Um μg de RNAa foi fracionado por eletroforese em gel de agarose 1% sob condições desnaturantes, e corado por brometo de etídeo. As canaletas M são referentes aos marcadores de tamanho em base, M₁ é pel Hind III e M₂ é 100 pb.

Figura 4 - Avaliação da integridade do RNA referência amplificado em dois ciclos.

4.1.11 Amplificação do RNAm

A etapa seguinte foi a amplificação do RNAm, em 1 ciclo, a partir de 3 μg de RNA total de cada amostra de interesse.

FELDMAN et al. (2002) demonstrou que a correlação entre duas hibridizações de RNA amplificado foi bastante superior a de duas hibridizações de RNA total, e que um número maior de genes diferenciais foi identificado no primeiro caso, sugerindo que a amplificação do RNA melhora a qualidade da hibridização e aumenta a capacidade de identificação de marcadores.

GOMES et al. (2003) validaram a metodologia de amplificação de RNA e sua utilização em experimentos de cDNA *microarray*, demonstrando a reprodutibilidade dos dados, a ausência de viés em relação à amostra de RNA não amplificada e a obtenção de maior sinal dos *spots* e de menor sinal de *background*, quando comparado ao RNA não amplificado. Deste modo, este procedimento propicia não apenas aumentar a quantidade, mas também a pureza do RNA.

O protocolo de amplificação anteriormente utilizado em nosso laboratório era baseado na metodologia descrita por WANG et al. (2000) com algumas modificações descritas por GOMES et al. (2003). Este protocolo utilizava a técnica de *switch-template* e amplificação guiada pela T7 RNA polimerase. No entanto, nos últimos anos, o protocolo de GOMES et al. (2003) foi adaptado pelo protocolo descrito por VAN GELDER et al. (1990), com algumas modificações. O protocolo descrito por VAN GELDER et al. (1990) não se baseia na técnica de *switch-template* e realiza uma síntese convencional da segunda fita de cDNA.

WANG et al. (2003) fez uma comparação entre o protocolo baseado em *switch-template* e a síntese convencional da segunda fita de cDNA, concluindo que, embora ambos os protocolos produzam resultados confiáveis e reprodutíveis, o segundo resulta em maior massa de RNA, sendo então recomendado quando a amplificação parte de uma massa muito pequena de RNA. Todas as amostras deste projeto são provenientes de biópsia, o que de fato limita a quantidade de RNA existente para experimentos.

A amplificação em 1 ciclo foi iniciada através da síntese da primeira fita de cDNA, a partir de 3 µg de RNA total (6,7 µL), 500 ng de oligo dT(24)-T7, *primer* que contém a seqüência do promotor da T7-RNA polimerase na extremidade 5' (5'-GGC CAG TGA ATT GTA ATA CGA CTC ACT ATA GGG AGG CGG TTT TTT TTT TTT TTT TTT TTT-3') e quantidade de água tratada com DEPC suficiente para 7,7 µL; seguida de desnaturação a 70°C por 10 minutos.

Após, foi adicionado 4,8 µL de MgCl₂ [25 mM] para [Final = 6 mM]; 4 µL de 5X *buffer* Impron II para [Final = 1x]; 2 µL de dNTP mix [10 mM] para [Final =

1 mM]; 0,5 μL de RnaseIN [40 U/ μL] para [Final =20 U] e 1 μL de Impron II *Reverse Transcriptase*, em reação de 20 μL e incubado a 42°C por 2 horas.

Os rendimentos das reações em 1 ciclo de ampliações de RNAm variaram de 29 a 220 vezes, com média de 129 vezes; o que foi suficiente para prosseguir com os experimentos; desta forma, não foi necessário realizar dois ciclos de amplificação. Para o cálculo de rendimento, se considera que 5% do RNA total é composto por RNAm e que se parte de 3 μg de RNA total de amostra. A variação nos rendimentos obtidos é atribuída às diferenças na pureza do RNA total das amostras utilizadas.

4.1.12 Amplificação: síntese de cDNA

O passo seguinte se baseou na síntese da segunda fita de cDNA, através da adição, ao produto da reação anterior, de 17,08 μL de água tratada com DEPC; 10 μL de 2nd *strand buffer* [5x] caseiro (Apêndice 1) para [Final = 1x]; 1 μL de dNTP mix [10 mM] para [Final = 0,2 mM]; 0,12 μL RNase H [5 U/ μL] para [Final = 0.6 U]; 0,5 μL de *E. coli* DNA Ligase [10 U/ μL] para [Final = 5U] e 1,3 μL de DNA *Polymerase* I [10 U/ μL] para [Final = 13U]. Após a adição destes reagentes a reação foi incubada a 16°C por 2 horas. Em seguida, foi adicionado 1 μL de T4 DNA *Polymerase* [5 U/ μL] para [Final = 5U] e incubado a 16°C por 10 minutos.

Finalizada a reação anterior, foram realizadas purificação e precipitação do cDNA dupla-fita, através da adição, ao produto da reação anterior, de 102 μL de água tratada com DEPC e 152 μL de fenol pH 8,0/clorofórmio/álcool isoamílico [25:24:01]; em seguida, o produto foi homogeneizado por inversão e centrifugado a 20.000g / TA por 5 minutos. Após, o sobrenadante foi coletado e foi adicionado 152

μL de clorofórmio/álcool isoamílico [24:01]; em seguida, foi homogeneizado por inversão e centrifugado a 20.000g / TA por 5 minutos.

No passo seguinte, o sobrenadante foi coletado em novo tubo e foi então adicionado $1 \mu\text{L}$ de acrilamida [5 mg/mL] para [Final = $5 \mu\text{g}$]; $0,5\text{X}$ volume de NH_4Ac [$7,5 \text{ M}$] e $2,5\text{X}$ volumes de etanol 100% gelado. Este produto foi homogeneizado por inversão e, em seguida, incubado em freezer -20°C por pelo menos 1 hora.

Após, foi realizada centrifugação a 20.000g / 4°C por 30 minutos, o sobrenadante foi desprezado, o *pellet* foi lavado com $1,5 \text{ mL}$ de etanol 75% e centrifugado a 20.000g / 4°C , por 8 minutos. A etapa da lavagem foi repetida por mais 2 vezes. Terminada a lavagem, o *pellet* foi seco em centrífuga a vácuo e ressuspendido em $10 \mu\text{L}$ de água DEPC.

4.1.13 Amplificação: transcrição *in vitro*

Terminada a purificação e precipitação do cDNA dupla fita, foi realizada a transcrição *in vitro* (*Kit RiboMAX – Promega*), através da adição, ao tubo de reação, de $10 \mu\text{L}$ de cDNA dupla-fita resultante da etapa anterior; $7,5 \mu\text{L}$ de rNTP mix [25 mM] para [Final = $7,5 \text{ mM}$]; $5 \mu\text{L}$ de T7 *transcription buffer* [5x] pra [Final = 1x] e $2,5 \mu\text{L}$ de *enzyme mix T7 RNA polymerase* em reação de $25 \mu\text{L}$. Esta reação foi incubada por 6 horas a 37°C .

A última etapa foi a purificação do RNAa, na qual ocorre a adição de $0,5 \text{ mL}$ de Trizol (Invitrogen) em cada tubo de transcrição e posterior homogeneização, por fim o produto foi incubado por 5 minutos em TA. Em seguida foi adicionado $100 \mu\text{L}$ de clorofórmio, foi realizada homogeneização e o produto foi incubado por 3

minutos em TA. Após, foi realizada centrifugação a 20.000g / 4°C por 30 minutos e a fase aquosa foi transferida para novo tubo e foi adicionado 250 µL de isopropanol. O produto foi homogeneizado e incubado por 10 minutos a TA e depois por 10 minutos a 4°C. Após, foi realizada centrifugação a 20.000g / 4°C por 30 minutos, o *pellet* foi lavado com 1,5 mL de etanol 70% e centrifugado a 8000g / 4°C por 10 minutos. Esta lavagem foi repetida mais 2 vezes e o *pellet* foi seco em centrífuga a vácuo.

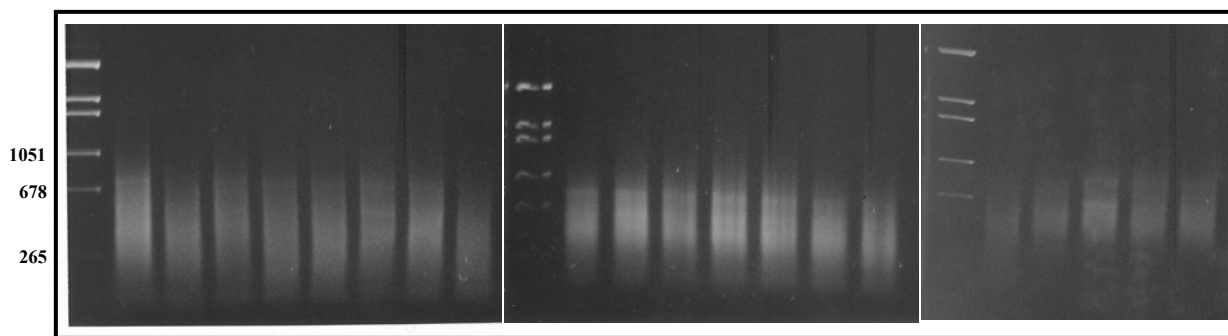
Para finalizar, o *pellet* foi ressuscitado em 11-20 µL de água DEPC, dependendo de seu tamanho, e, posteriormente, aquecido a 55°C por 5 minutos.

Uma vez que o RNA foi amplificado, 1 µL foi utilizado para quantificação em aparelho NanoDrop (Tabela 11) e 500 ng do RNA foram utilizados para verificação do perfil do RNAa de 20 amostras através de eletroforese em gel de agarose 1% (Figura 5). As 11 amostras seguintes amplificadas tiveram 500 ng do RNAa utilizados para verificação do perfil (Figura 6) em aparelho Bioanalyzer. Como esperado, a degradação das amostras número 11 e 35 foi novamente confirmada.

Assim sendo, as 31 amostras coletadas foram amplificadas, embora a reação não tenha funcionado para as amostras números 11 e 35. A amostra número 8 apresentou rendimento insuficiente para o prosseguimento das reações seguintes, deste modo, a reação de amplificação foi repetida e o resultado foi então satisfatório.

Tabela 11 - Avaliação do rendimento e pureza do RNA amplificado em um ciclo, a partir do RNA total de 31 amostras. Esta avaliação foi realizada em aparelho NanoDrop e com 1µL de RNAA.

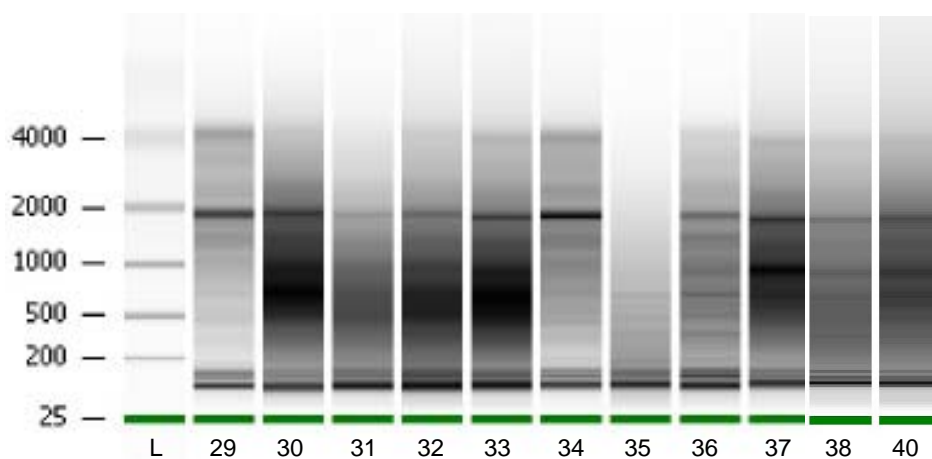
Amostra	RNAA ng/ul	260/280	260/230	RNAA -1 Round- ug totais
3	1111,72	2,1	1,89	16,675
4	1630,72	2,1	2,16	24,46
12	1304,39	2,11	1,47	19,565
19	1765,79	2,13	2,57	26,486
21	1730,67	2,1	1,35	25,96
27	1090,5	2,13	1,81	16,357
1	2204,99	2,19	1,9	33,075
6	1190,66	2,2	2,51	17,86
7	1485,78	2,19	2,46	22,287
9	715,52	2,23	2,42	10,732
14	1128,32	2,21	2,19	16,924
17	1497,86	2,18	2,49	22,468
22	1243,34	2,21	1,56	18,649
23	892,44	2,23	2,62	13,386
8	965,70	2,10	1,90	10,176
10	1270,14	2,13	2,24	19,05
15	1747,17	2,11	2,07	26,208
25	1083,07	2,13	1,81	16,246
26	1182,52	2,15	1,92	17,737
11	897,63	2,17	2,45	13,464
29	1945,19	2,17	2,19	23,34
30	1617,08	2,17	2,49	19,4
31	1974,25	2,23	1,73	23,69
32	1734,49	2,19	1,42	20,81
33	1994,61	2,17	1,41	23,93
34	1881,31	2,18	1,87	22,57
35	211,15	2,0	1,33	35,33
36	1440,65	2,2	0,81	17,28
37	1393,53	2,15	1,21	16,722
38	1219,21	1,99	2,69	19,5
40	1186,52	2,01	2,29	18,98



M 1 6 7 9 14 17 22 23 M 3 4 12 19 21 27 15 M 11 8 10 25 26

Legenda: Perfil eletroforético de RNA amplificado de 20 amostras. Quinhentos ng de RNAa foram fracionados por eletroforese em gel de agarose 1% sob condições desnaturantes, e corado por brometo de etídeo. As canaletas M são referentes ao pel Hind III, marcador de tamanho em base.

Figura 5 - Avaliação da integridade de RNAs amplificados.



Legenda: Perfil eletroforético de RNA amplificado de 11 amostras. Quinhentos ng de RNAa foram avaliados em aparelho Bioanalyzer. A canaleta L é referente ao *ladder*, marcador de tamanho em base.

Figura 6 - Avaliação da integridade de RNAs amplificados.

4.1.14 Marcação indireta do cDNA

O fluorocromo utilizado no projeto foi o Alexa Fluor, uma vez que este é superior quando comparado ao corante *Cyanine*, já que este é mais fotoestável e barato, além de emitir sinal de 2,5 a 3 vezes maior (WILD SMITH et al. 2001).

Nesta etapa, o RNA amplificado das amostras foi utilizado na síntese de cDNA marcado pela incorporação indireta de fluoróforos Alexa 555 e Alexa 647 (Molecular Probes, EUA), baseado no protocolo descrito por DERISI et al. (1996). Neste protocolo, o RNA amplificado é submetido à transcrição reversa na presença de *Random Primers Hexamers*, transcriptase e *aminoallyl*-dUTP. Após hidrólise alcalina do RNA, ocorre o acoplamento dos fluoróforos ao *aminoallyl*. O produto é então purificado e a eficiência da marcação é verificada em espectrofotômetro.

Assim, 6 µg de RNA amplificado de cada amostra, quantidade de água DEPC suficiente para 10,4 µL e 3 µL de *Random Hexamers* pd(N)₆ (GE Healthcare, RU) [5 µg/µL] para [Final = 15 µg] foram desnaturados a 70°C, por 10 minutos, e mantidos em gelo por 5 minutos. Em seguida foi realizada transcrição reversa, ou seja, síntese de cDNA, na presença de 9,6 µL de MgCl₂ [25 mM] para [Final = 6 mM]; 4 µL de dNTP mix *low T* (Eppendorf, Alemanha) com *Aminoallyl*-dUTP (Sigma-Aldrich, EUA) [dATP, dCTP e dGTP = 5 mM; dTTP = 2 mM e *Amininoallyl* dUTP = 3 mM] para [dATP, dCTP e dGTP Final = 0,52 mM; dTTP Final = 0,21 mM e *Amininoallyl* dUTP Final = 0,31 mM], 8 µL de 5X Buffer Improm II pra [Final = 1x]; 1 µL de RNasin [40 U/µL] para [Final = 40U] e 2 µL de Improm II RT (Promega, Madison, WI), em volume final de reação de 38 µL. Esta reação foi então mantida a 25°C por 10 minutos e aquecida a 42°C por 2 horas. A etapa seguinte foi a degradação do

RNA com 20 μL de solução NaOH [1,0 M] e EDTA [0,25 M], a 70 °C por 20 minutos. Após, a reação foi neutralizada com 20 μL de HCL [1M].

O passo seguinte foi a precipitação do cDNA com a adição de 300 μL de isopropanol; 2 μL de glicogênio [20 $\mu\text{g}/\mu\text{L}$]; 2 μL de acrilamida [100 $\mu\text{g}/\mu\text{L}$] e 40 μL de acetato de sódio [7,5 M]. Esta reação foi então incubada à TA por 20 minutos, centrifugada a 12.000g / TA por 30 minutos e o sobrenadante foi então descartado. Foi adicionado 1,5 mL de etanol 75% para lavar o *pellet* formado, e, em seguida, foi realizada centrifugação a 7.600g / TA, por 5 minutos. A lavagem do *pellet* foi realizada no total de 3 vezes.

4.1.15 Incorporação dos corantes ao cDNA

Em seguida, foi realizado o acoplamento dos fluoróforos Alexa Dye 555 ou 647 (Invitrogen, EUA) ao cDNA anteriormente sintetizado. Para esta reação, o *pellet* de cDNA foi ressuspensionado em 16 μL de tampão bicarbonato 0,0625 M pH 9,0 e em seguida foi aquecido a 55°C, por 5 minutos. Nesta etapa, o volume de cada amostra foi dividido em duas alíquotas de 8 μL ; uma referente ao *main* e outra ao *swap*.

Na reação *main* a amostra é marcada com Alexa 555 e conseqüentemente a referência com Alexa 647; enquanto que na reação *swap* ocorre a inversão dos fluorocromos. A inversão dos corantes é realizada a fim de evitar introdução de viés por incorporação diferencial dos corantes, além de funcionar como duplicata biológica do experimento.

Após, foi adicionado 5 μL de DMSO de alta qualidade ao tubo do fluorocromo e este foi vortexado por 15 segundos; em seguida, 5 μL do corante ressuspensionado foi então adicionado a cada tubo contendo 8 μL de amostra ou 8 μL de

referência. A reação foi incubada à TA por 1 hora e 30 minutos, ao abrigo da luz, uma vez que os fluorocromos são fotossensíveis.

O próximo passo foi a purificação da reação com o *kit* QIAquick PCR *Purification* (Qiagen, Alemanha), que também ocorre ao abrigo da luz. Foi adicionado à reação 40 µL de NaOAc [100mM] pH 5,2 e 500 µL de *Binding Buffer* PBI. Este produto foi então aplicado ao centro da coluna do *kit*, centrifugado a 13.000 rpm / TA por 1 minuto e o *flow-thru* descartado. Em seguida, foi adicionado 750 µL de etanol 80%, foi centrifugado a 13.000 rpm / TA por 1 minuto e o *flow-thru* foi novamente descartado. Esta lavagem com etanol foi realizada por 2 vezes. Novamente foi realizada centrifugação a 13.000 rpm / TA, por 1 minuto, e a coluna foi transferida para novo tubo. Foram adicionados 50 µL de tampão de eluição, *Elution Buffer*, previamente aquecido a 70°C, foram aguardados 2 minutos e centrifugado a 13.000 rpm / TA por 1 minuto. Este último passo foi realizado novamente.

A eficiência da marcação dos corantes ao cDNA foi verificada em espectrofotômetro (ND-1000 Spectrophotometer, NanoDrop, EUA), através da medição da fluorescência total. Dentre 31 amostras marcadas, as números 11 e 35 não apresentaram incorporação desejável de fluorocromo, o que de fato era esperado, devido à degradação do RNA.

Cada amostra e sua referência, marcadas com corantes distintos, foram então misturadas e transferidas para tubo contendo 2 µL de poli(A) DNA (Amersham Pharmacia, EUA) [2 µg/µL] e 2 µL de Cot1 DNA (Gibco, BRL) [1 µg/µL]. O volume foi diminuído até 12,5 µL, em centrífuga a vácuo.

Após, foram adicionados 2 μL de DNA de esperma de salmão (Amersham Biosciences, EUA) [5 $\mu\text{g}/\mu\text{L}$]; 40 μL de Solução de Hibridização 2X previamente aquecida a 70°C e filtrada em filtro de 0,22 μm (10 mL de SSC 20X; 400 μL de SDS 10%; 4 mL de solução Denhardt's 50X e 5,6 mL de água DEPC,) e 50 μL de formamida deionizada. A reação foi incubada a 45°C até o momento do uso, após, foi realizada desnaturação a 95°C por 5 minutos e a solução foi mantida aquecida a 70°C até o momento da aplicação em estação de hibridização.

4.1.16 Confeção, pré-hibridização, hibridização e lavagem das lâminas de array

As lâminas, ou *biochips*, que foram utilizadas neste projeto (4.8K) foram confeccionadas pelo nosso laboratório e tiveram produtos de PCR dos clones ORESTES fixados em sua superfície.

As lâminas utilizadas contêm 4.800 elementos, ou sondas de cDNA, escolhidos a partir de seqüências ORESTES geradas pelo Projeto Genoma do Câncer, a partir de bibliotecas de cabeça e pescoço (língua, laringe e hipofaringe), tireóide, pulmão, bexiga, rim, cólon, próstata, epidídimo, testículo, mama, tumores ginecológicos (vulva, vagina, cérvix e ovário), tumores abdominais (estômago, esôfago, fígado e pâncreas), melanoma, sarcoma, retinoblastoma, leiomiossarcoma, medula óssea e sistema nervoso.

Os fragmentos de ORESTES imobilizados foram selecionados de acordo com cinco critérios principais: seqüências maiores que 300 pb; seqüências que correspondam a genes com seqüência completa conhecida (*full lenght*) e depositadas no banco de dados RefSeq; seqüências que só apresentem um único mapeamento

(*hit*) no DNA genômico; seqüências que não apresentem similaridade maior que 85% em pelo menos 100 pb com qualquer outro *cluster* presente no banco de dados UNIGENE; e, finalmente, seqüências que correspondam à extremidade mais 3' do gene orientado no genoma, sendo que os últimos 100 pb são evitados a fim de que a seqüência se localize a montante do sítio de poliadenilação.

Seqüências ORESTES contidas em plasmídios pUC foram amplificadas em placas de 96 poços e o produto foi purificado em placa com resina G50. Alíquotas destes produtos foram depositadas em lâminas de vidro silanizadas (Corning), na qual os elementos foram fixados com o auxílio do robô Flexys (Genomic Solutions). As lâminas foram tratadas e submetidas a *crosslink* em luz UV.

Uma vez que as lâminas de *microarray* foram confeccionadas, foi realizada pré-hibridização em solução de SSC 5X; SDS 0,2%; BSA 1% e solução Denhardt's 5X, a 42°C *overnight*. Após, a mistura do RNA da amostra de interesse com o RNA referência foram então hibridizadas a 42°C, por 16-18 horas, na estação Gene TAC Hybridization Station (Genomic Solutions, RU).

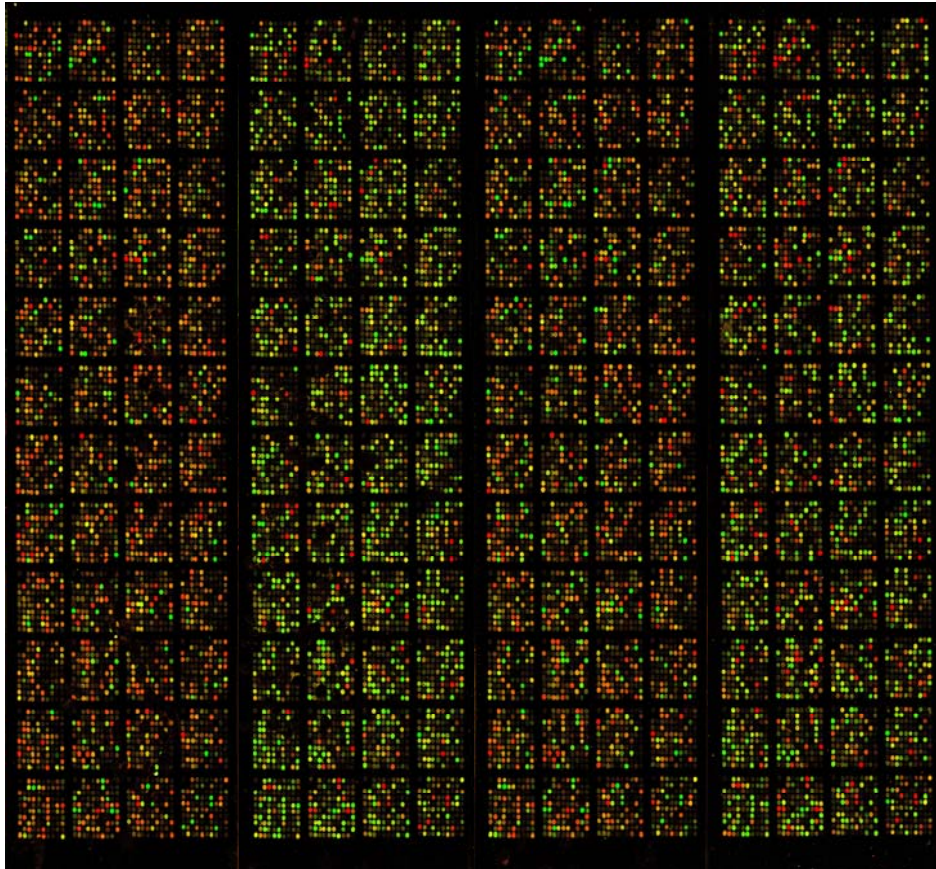
Após a hibridização, as lâminas foram lavadas sob agitação em 3 cubas diferentes, com solução de SSC 2X a 42°C por 10 minutos cada; após, as lâminas foram lavadas 2 vezes em solução de SSC 0,1X SDS 0,1%, a 42°C por 10 minutos cada; e, finalmente, a lavagem ocorreu em solução de SSC 0,1X, a 42°C por 10 minutos.

4.1.17 Escaneamento das lâminas de *array* e quantificação dos dados

Após a lavagem, as lâminas foram centrifugadas, escaneadas e as imagens foram então digitalizadas, utilizando *scanner* a laser confocal (*ScanArray Express Microarray Scanner* - *Packard Bioscience*, EUA), que captura os valores de intensidade dos *spots* e digitaliza os sinais gerados pelas hibridizações. Foram capturadas imagens digitais para cada um dos dois comprimentos de onda (fluorocromo) utilizados.

Para cada amostra referente a um único paciente foram hibridizadas duas lâminas, *main* e *swap* (Figura 7), assim, 66 lâminas foram hibridizadas no total (Tabela 12), sendo que 4 amostras foram refeitas devido a artefatos de hibridização.

Apesar da hibridização da amostra 3 ter sido finalizada, ela não foi incluída nas análises seguintes, pois o paciente abandonou o tratamento antes de seu término. Deste modo, o número final de pacientes participantes do presente projeto foi de 28.



Legenda: Visualização de quatro lâminas de vidro 4.8K, após hibridização e escaneamento. A primeira lâmina é referente à amostra número 6 *main*, e a segunda lâmina à amostra 6 *swap*. A terceira lâmina é referente à amostra 27 *main*, e a quarta à amostra 27 *swap*.

Figura 7 - Quatro lâminas de *microarray* hibridizadas.

Tabela 12 - Hibridização de 29 amostras em duplicata, *main* e *swap*, totalizando 66 lâminas hibridizadas e escaneadas. As 4 amostras com a anotação (2) foram refeitas.

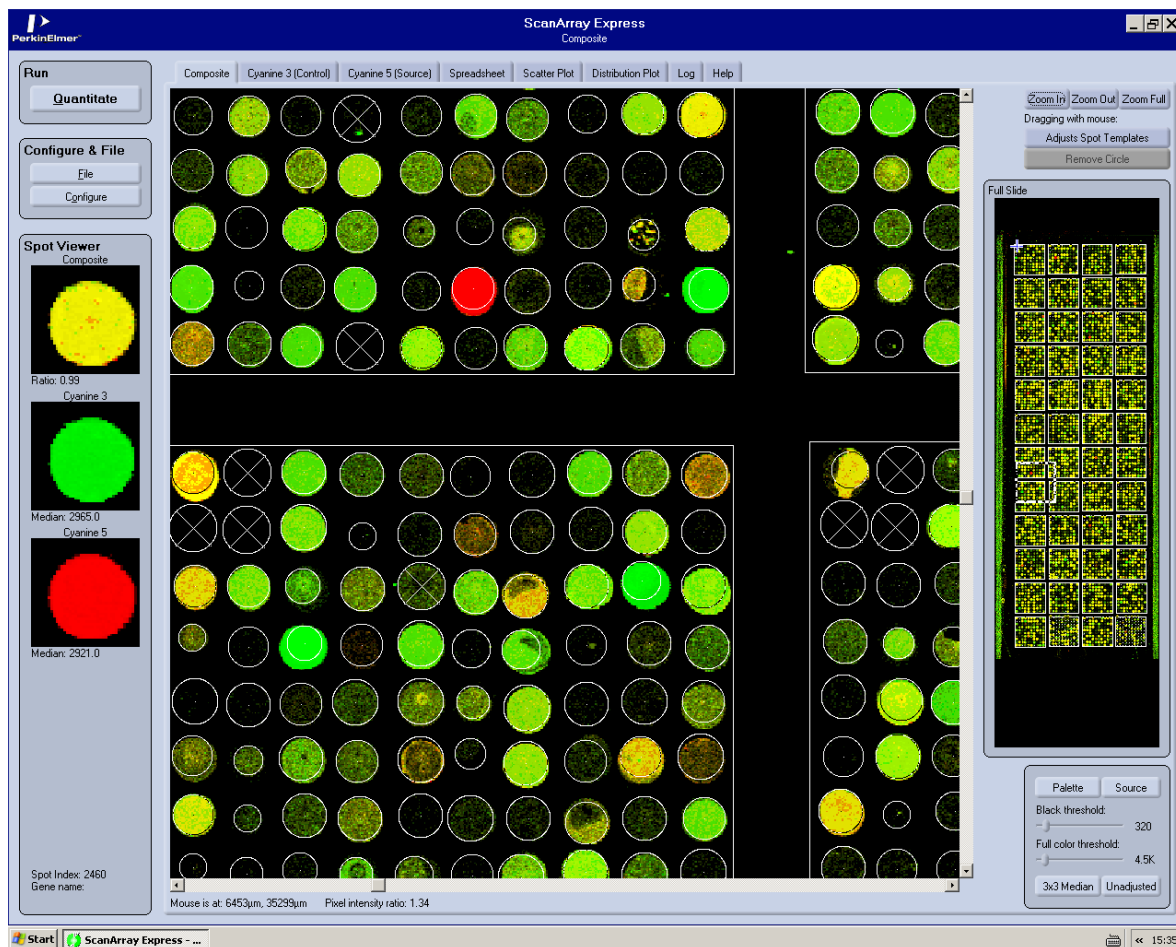
Número da Amostra	Cor e Nome do Corante da Amostra	Código de Barras da Lâmina de Array	Lote da Lâmina
3	rosa, alexa 555 -C3-	13631663	57
3'	azul, alexa 647 -C5-	13631664	57
4	rosa, alexa 555 -C3-	13631691	58
4'	azul, alexa 647 -C5-	13631692	58
12	rosa, alexa 555 -C3-	13631665	57
12'	azul, alexa 647 -C5-	13631666	57
19	rosa, alexa 555 -C3-	13631662	57
19'	azul, alexa 647 -C5-	13631678	57
21(2)	rosa, alexa 555 -C3-	13631952	59
21'(2)	azul, alexa 647 -C5-	13636451	46
27	rosa, alexa 555 -C3-	13631697	57
27'	azul, alexa 647 -C5-	13631698	57
1(2)	rosa, alexa 555 -C3-	13631813	42
1'(2)	azul, alexa 647 -C5-	13631814	42
6	rosa, alexa 555 -C3-	13631695	57
6'	azul, alexa 647 -C5-	13631696	57
7	rosa, alexa 555 -C3-	13632129	57
7'	azul, alexa 647 -C5-	13632130	57
9	rosa, alexa 555 -C3-	13631954	58
9'	azul, alexa 647 -C5-	13631955	58
14	rosa, alexa 555 -C3-	13631689	57
14'	azul, alexa 647 -C5-	13631690	57
17	rosa, alexa 555 -C3-	13631693	58
17'	azul, alexa 647 -C5-	13631699	58
22	rosa, alexa 555 -C3-	13631819	58
22'	azul, alexa 647 -C5-	13631820	58
23	rosa, alexa 555 -C3-	13632098	57
23'	azul, alexa 647 -C5-	13632103	57
8	rosa, alexa 555 -C3-	13632101	56
8'	azul, alexa 647 -C5-	13632115	56
10	rosa, alexa 555 -C3-	13636763	56
10'	azul, alexa 647 -C5-	13631957	58
15	rosa, alexa 555 -C3-	13631961	58
15'	azul, alexa 647 -C5-	13631963	58
25	rosa, alexa 555 -C3-	13631964	58
25'	azul, alexa 647 -C5-	13631966	58
26	rosa, alexa 555 -C3-	13631949	58
26'	azul, alexa 647 -C5-	13631967	58
29	rosa, alexa 555 -C3-	13636626	40
29'	azul, alexa 647 -C5-	13636628	40
30	rosa, alexa 555 -C3-	13636629	40
30'	azul, alexa 647 -C5-	13636630	40
31(2)	rosa, alexa 555 -C3-	13636456	46
31'(2)	azul, alexa 647 -C5-	13636457	46
32	rosa, alexa 555 -C3-	13636624	40
32'	azul, alexa 647 -C5-	13636625	40

Cont/ Tabela 12

33	rosa, alexa 555 -C3-	13636631	40
33'	azul, alexa 647 -C5-	13636632	40
34	rosa, alexa 555 -C3-	13636634	40
34'	azul, alexa 647 -C5-	13636644	40
36	rosa, alexa 555 -C3-	13636627	40
36'	azul, alexa 647 -C5-	13636645	40
37	rosa, alexa 555 -C3-	13636440	46
37'	azul, alexa 647 -C5-	13636441	46
38(2)	rosa, alexa 555 -C3-	13636465	46
38'(2)	azul, alexa 647 -C5-	13636466	46
40	rosa, alexa 555 -C3-	13636459	46
40'	azul, alexa 647 -C5-	13636468	46

Após finalização das hibridizações, foi realizada quantificação dos dados gerados por *microarray*. Os valores de intensidade dos *spots* de todos os *arrays* gerados nas hibridizações foram capturados, utilizando-se o *software* ScanArray Express Microarray Scanner (Packard Bioscience, EUA) (Figura 8).

As quantificações das marcações de *spot* e *background* foram obtidas pelo método de histograma, utilizando a mediana dos valores dos pixels e percentis de 80 a 95 para o sinal e 5 a 20 para o *background*, utilizando-se o mesmo programa, ScanArray Express. Este método constrói um histograma dos valores dos pixels de cada *spot* e permite fixar os percentis que serão utilizados para os cálculos de intensidade de sinal e de *background*. Este método é simples, estável e exclui pixels de características extremas do cálculo das intensidades do *spot* e do *background*.



Legenda: Imagem da quantificação de uma lâmina utilizando-se o *software* ScanArray Express. As imagens obtidas separadamente dos dois corantes são sobrepostas, formando então a imagem composta.

Figura 8 - Quantificação de lâmina.

4.1.18 Análises dos dados gerados por *microarray* e análises matemáticas

Após quantificação das 64 lâminas, todas as tabelas de intensidade de sinal geradas foram depositadas via *Web Page*, em banco de dados, *Array Manager*, para o Laboratório de Bioinformática do Hospital AC Camargo, coordenado pela Dra. Helena Brentani. Os mesmos dados também foram depositados em banco de dados, Maiges (Mathematical Analysis for Interacting Gene Expression Systems), do

Instituto de Matemática e Estatística da Universidade de São Paulo, IME/USP, com a coordenação do Dr. Eduardo Jordão Neves.

A abordagem matemática e estatística para análise dos dados gerados por *microarray* têm sido amplamente discutida na literatura e utilizada com sucesso por nosso grupo (GOMES et al. 2003; MEIRELES et al. 2003, 2004; STOLF et al. 2003; 2005 e 2006).

A correção dos valores dos *spots* foi realizada subtraindo-se do sinal de intensidade de cada *spot* o valor de *background* deste mesmo *spot*; isso foi realizado para cada *spot* individualmente. A seguir, todos os genes que foram utilizados como controles positivos e os *spots* em branco foram removidos da análise, restando um conjunto de 4.600 genes.

Após, foi calculada a razão entre canal 1 (Alexa 555) e canal 2 (Alexa 647), para todos os genes de todas as lâminas, e foi então realizada uma transformação nesses valores, aplicando logaritmo na base dois (\log_2), e fazendo em seguida, a normalização dos *swaps*, que neste caso é a média aritmética (para todos os genes) entre os valores de razão citados acima, ou seja, $(\text{controle} + \text{swap}) / 2$.

A análise dos dados foi realizada com o pacote MaigesPack (ESTEVEZ 2007), que está disponível para *download* via projeto Bioconductor (<http://www.bioconductor.org>). Ambos utilizam a plataforma estatística R (<http://www.r-project.org>) como ambiente computacional. A normalização dos dados foi realizada utilizando-se o método de loess, que é um método de regressão não linear localmente ponderado, com *span* 0,4 e grau 2, conforme implementado nos pacotes do R e Bioconductor.

A normalização visa igualar a intensidade relativa dos dois fluoróforos, devido à incorporação diferencial dos corantes. Deste modo, a normalização ajusta diferenças na marcação e na eficiência da detecção dos dois corantes, bem como diferenças nas quantidades iniciais das amostras (QUACKENBUSH 2001). Métodos de regressão local como lowess ou loess são ferramentas promissoras para gerar normalizações robustas e flexíveis, por corrigirem variações sistemáticas locais ao invés de adotarem normalizações globais (DUDOIT et al. 2002b). Por esta razão foi adotado o método de loess para a normalização dos dados deste estudo.

Após a normalização, foram realizadas análises de controle de qualidade das lâminas. Assim, foram gerados *cluster* hierárquico e gráficos do tipo MM *plot*, além da análise do número de *spots* utilizáveis de cada lâmina. Também foram analisados os genes mais diferencialmente expressos do conjunto de *array*.

Com o intuito de se avaliar o desempenho dos classificadores anteriormente identificados no Mestrado (100 melhores trios e 100 melhores quadras), realizou-se uma centralização tanto nos dados de expressão gênica gerados no Mestrado quanto nos dados atuais.

Essa centralização faz-se necessária devido à utilização de diferentes RNAs referência entre ambos os estudos. Após a centralização, dos dados, a regra de classificação que anteriormente identificou os 100 melhores trios e as 100 melhores quadras foi agora aplicada aos dados gerados pelas amostras do grupo da validação

Para finalizar, foi aplicada a estratégia de árvore de decisão (DUDOIT et al. 2002a), a fim de otimizar o resultado dos classificadores validados.

4.2 OBJETIVO SECUNDÁRIO: ESTUDO DE VIAS METABÓLICAS E GENES

4.2.1 Populações de estudo e vias analisadas

Esta etapa do estudo foi constituída pelas 35 amostras provenientes de pacientes portadores de CE de laringe e hipofaringe, localmente avançado e que foram tratados com quimioterapia, realizada com cisplatina e paclitaxel, concomitante à radioterapia. Estes pacientes foram classificados em R ou NR frente à quimiorradioterapia. Este grupo de amostras foi anteriormente utilizado em projeto de Mestrado, o qual foi responsável pela identificação dos classificadores moleculares de resposta.

Desta forma, com o objetivo secundário de estudar vias metabólicas potencialmente envolvidas com o mecanismo de ação de quimiorradioterapia, a expressão gênica diferencial dos 35 pacientes, classificados em R ou NR, foi analisada e correlacionada com a expressão das mesmas proteínas; com a finalidade de identificar alterações possivelmente relacionadas com resistência dos pacientes ao tratamento clínico.

Em seguida, a validação dos genes identificados no grupos das 35 amostras foi realizada no grupo das 28 amostras provenientes do estudo presente, que compõem o grupo da validação.

Assim, foram selecionadas três vias para estudo, Ras-Raf-Map-kinase (ALROY e YARDEN 1997; LEWIS et al. 1998); a via PI3K e AKT (CHAN et al. 1999; BLUME-JENSEN e HUNTER 2001; VIVANCO e SAWYERS 2002) e por último a via *kinase C* e Jak/Stat (YARDEN e SLIWKOWSKI 2001); vias estas

relacionadas ao EGFR, ao mecanismo de ação de drogas e de conhecida importância em câncer de cabeça e pescoço.

A principal rota de sinalização *downstream* da família ErbB é através da via Ras-Raf-MAP-kinase (ALROY e YARDEN 1997). Ativação de Ras inicia cascada de fosforilação de múltiplos passos, que levam à ativação de proteínas kinases (MAPKs) ativadas por mitose, ERK1 e ERK2 (LEWIS et al. 1998). ERK1/2 regula transcrição de moléculas que estão ligadas à proliferação celular, sobrevivência, e à transformação em estudos de laboratório.

Outro alvo importante na sinalização de EGFR é fosfatidilinositol 3-kinase (PI3K) e a proteína *downstream* serina/treonina kinase AKT (CHAN et al. 1999; BLUME-JENSEN e HUNTER 2001; VIVANCO e SAWYERS 2002). AKT transcreve sinal que desencadeia cascada de resposta, desde crescimento celular e proliferação até sobrevivência e motilidade (VIVANCO e SAWYERS 2002).

Uma terceira via de sinalização é através da via kinase de proteína ativada por estresse, envolvendo as proteínas kinase C e Jak/Stat. A ativação desta via se traduz no núcleo, na expressão de distintos programas transcricionais que medeiam uma variedade de respostas celulares, incluindo divisão celular, sobrevivência (ou morte), motilidade, invasão, adesão e reparo de DNA (YARDEN e SLIWKOWSKI 2001).

4.2.2 Genes analisados e reações de imunoistoquímica

Através da análise das três vias selecionadas e do Teste de Wilcoxon, foram identificados 6 genes, IL4R; NTRK2 ou TRKB; GNA12; EGFR; CMET e TGFA (Tabela 4) diferencialmente expressos entre as comparações R/NR no grupo das 35 amostras, a partir dos dados gerados por *microarray*.

Após, foi realizada extensa pesquisa e levantamento bibliográfico para conhecimento e escolha dos melhores anticorpos a serem empregados no estudo (Tabela 13) realizado com as 35 amostras. Até o presente momento, não existia anticorpo disponível no mercado para imunistoquímica para a proteína do gene humano GNA12. Já para o gene TRKB existem dois anticorpos disponíveis, um que reconhece a proteína fosforilada e outro que reconhece a proteína quando ela não está fosforilada, assim sendo, ambos os anticorpos foram adquiridos.

Tabela 13 - Anticorpos adquiridos para detecção das proteínas por imunistoquímica, no grupo das 35 amostras do projeto de Mestrado, de CE de laringe e hipofaringe, em pacientes R e NR

Anticorpo	Catálogo	Fabricante	Título
<i>Rabbit polyclonal to TrkB</i> (phospho Y515)	Ab51187	Abcam Inc.	1:50
<i>Rabbit polyclonal to TrkB</i>	Ab51190	Abcam Inc.	1:25
<i>Monoclonal anti-human to IL-4R</i>	MAB230	R&D Systems	1:25
<i>Rabbit polyclonal to c-MET</i>	sc-161	Santa Cruz Biotech.	1:400
<i>Mouse monoclonal to EGFR</i>	NCL-EGFR-384	Novo Castra	1:400
<i>Epitope specific rabbit to TGF-α</i>	RB-9241	NeoMarkers	1:100

Concomitantemente, foi realizado levantamento no Hospital AC Camargo, em arquivo de material emblocado em parafina dos 35 pacientes utilizados no projeto de Mestrado. Foram recuperados 25 blocos dos 35 pacientes do estudo prévio, para serem utilizados no estudo presente.

Após corte e coloração HE, uma amostra foi descartada por não conter área tumoral suficiente para possibilitar análise confiável. Assim, foram totalizadas 24 amostras, 13 referentes a pacientes R e 11 a NR. Em seguida, foram montadas lâminas para cada uma das 24 amostras.

Cortes de 3 µm de espessura do material incluído em parafina foram aplicados em lâminas de vidro cobertas com silano SuperFrost Plus (#EP-SFPLUS), e deixadas por 24 horas em estufa a 60°C para desparafinização. Para completar esta etapa o material foi mergulhado em xilol a 60°C por 20 minutos; xilol à TA por 20 minutos; etanol 100% por 30 segundos; etanol 95% por 30 segundos e, para finalizar, etanol 70% por 30 segundos. Após, as lâminas foram lavadas em água destilada corrente.

Em seguida foi iniciada recuperação dos antígenos; para isso, a solução tampão citrato [10 mM] pH 6,0 foi fervida e as lâminas foram mergulhadas, a fim de não agredir o tecido e colaborar na retirada dos grupos formaldeídos ainda presentes no mesmo. Após, as lâminas foram lavadas em água destilada corrente.

O próximo passo foi o bloqueio da peroxidase endógena com H₂O₂ 3% para evitar que o sistema de revelação por polímeros (HRP), realizada em etapas posteriores, apresentasse reação de fundo com a peroxidase endógena. Em seguida, foi realizada lavagem com água destilada corrente e posteriormente lavagem com solução salina tamponada com fosfatos (PBS) [10 mM] pH 7,4 por 5 minutos, para que o tecido voltasse à sua qualidade fisiológica.

As lâminas foram então incubadas com os títulos pré-estabelecidos de cada anticorpo, como demonstrado na Tabela 5, em tampão PBS contendo BSA 1% (Sigma, A9647, EUA) e azida sódica 0,1%, a 4°C e por 18 horas em câmara úmida. Após a incubação, as lâminas foram lavadas em tampão PBS com 3 trocas, de 3 minutos cada.

O sistema de detecção escolhido foi o de polímeros, que não contém biotina ou estreptavidina, sendo selecionados o *Kit* NovoLink (NovoLink Max Polymer,

cód. RE7260-k, Reino Unido). Por se tratar de um sistema onde não há biotina ou estreptavidina, espera-se minimizar eventual *background* resultante da atividade de biotina endógena.

Deste modo, foi iniciado o sistema de amplificação e detecção, sendo realizada primeira incubação a 37°C, por 30 minutos, com Post Primary Block (NovoLink). Em seguida, foi realizada lavagem com tampão PBS com 3 trocas de 3 minutos cada e nova incubação com NovoLink Polymer (NovoLink) a 37°C por 30 minutos. Foi finalizado com nova lavagem em tampão PBS, com 3 trocas de 3 minutos cada.

A última etapa foi a revelação com cromógeno e contracoloração. As lâminas foram incubadas em solução substrato de DAB 60 mg% (Sigma, d-5637, EUA), 1 mL de DMSO, 1 mL de H₂O₂ 6% e 100 mL de PBS, a 37°C por 5 minutos, ao abrigo da luz. Nesta etapa deve ser observado ao microscópio, nas lâminas controles, o desenvolvimento de precipitado castanho dourado, como produto final da reação. Em seguida, as lâminas foram lavadas em água destilada corrente por 3 minutos, e depois, contracoradas com Hematoxilina de Harris por 1 minuto. Após, as lâminas foram lavadas em água destilada corrente, imersas em água amoniacal (NH₄OH 0,5%) e por fim, lavadas novamente em água destilada corrente.

As lâminas foram então desidratadas em etanol 80% por 30 segundos; etanol 95% por 30 segundos; etanol 100% por 30 segundos cada, repetido 2 vezes; xilol 4 vezes por 30 segundos cada, e finalmente montadas em *Entellan neu lamínulas* (Merck, 1.07961, Alemanha), uma resina que visa proteger a lâmina contra eventuais atritos com a lamínula.

Após analisar os resultados das reações de imunistoquímicas, o teste de Wilcoxon foi novamente aplicado, mas agora nas 28 amostras pertencentes ao grupo da validação dos classificadores, sendo possível analisar os dados de expressão dos genes de escolha, gerados pela técnica de *microarray*.

Essa etapa específica do projeto, ou objetivo secundário, foi realizada com a finalidade de analisar vias metabólicas, identificar genes diferencialmente expressos entre R e NR e comparar o comportamento dos genes identificados em ambos os grupos de amostras (projeto de Mestrado e projeto da validação).

5 RESULTADOS

5.1 OBJETIVO GERAL: VALIDAÇÃO DE CLASSIFICADORES MOLECULARES

5.1.1 Seguimento clínico dos pacientes

Foram utilizadas amostras de pacientes provenientes de 4 instituições diferentes, o que certamente torna mais difícil o exercício de validação, devido ao maior número de variações existentes. Devido a isso, uma vez validados, os classificadores serão ainda mais robustos.

Fizeram parte deste estudo 28 amostras (Tabela 14), sendo que são 21R (75%) e 7NR (25%). O tempo médio de seguimento dos pacientes R, finalizado em dezembro de 2009, foi de 20,5 meses, com mediana de 18 meses, sendo que o intervalo variou de 6 a 93 meses.

O paciente R que apresentou o menor tempo de seguimento, 6 meses, teria maior probabilidade de recidivar, devido ao pequeno tempo de seguimento. Caso isso ocorresse, sua classificação mudaria de R para NR, e, conseqüentemente, o resultado identificado até o momento não seria válido.

Mas, este paciente apresentou linfonodo palpável e suspeita de lesão residual. Desta forma, foi realizado esvaziamento cervical e 13 linfonodos foram ressecados. Mas, após análise anatomopatológica, nenhum linfonodo apresentou atipia; conseqüentemente, o paciente permaneceu, corretamente, classificado como R.

Tabela 14 - Características clinicopatológicas dos 28 pacientes e respectivos tumores incluídos no projeto, sem metástase à distância e sem quimiorradioterapia prévias. P-valores calculados pelo teste exato de Fisher.

Características	%	p-valor
Idade (anos)		1,0
Média :	57,5	
Mediana:	58,5	
Intervalo:	41-68	
Gênero		1,0
Masculino	85,7	
Feminino	14,3	
Sítio Tumoral Acometido		0,39
Laringe Glótica	7,15	
Laringe Transglótica ^(a)	28,5	
Laringe Supraglótica	7,15	
Laringe Supraglótica e Hipofaringe	25,0	
Laringe Transglótica e Hipofaringe	17,9	
Hipofaringe	14,3	
Tabagismo		-
Positivo	100	
Etilismo		0,54
Positivo	85,7	
Diagnóstico Histológico		0,57
CE bem diferenciado	14,3	
CE moderadamente diferenciado	60,7	
CE pouco diferenciado	25,0	
T		1,0
2	3,6	
3	75,0	
4	21,4	
N		0,032
0	57,1	
1	32,1	
2a-c	10,8	

^(a)Laringe Transglótica é tumor que acomete glote e supraglote.

Já o tempo médio de recidiva dos pacientes NR foi de 7,8 meses, com mediana de 5 meses, sendo que o intervalo variou de 2 a 24 meses. Um único paciente recidivou após 24 meses, o que não é o mais comum. Dados da literatura demonstraram que 80% das recorrências ocorrem em até dois anos (JONES et al. 1992; TAKES et al. 2001).

Analisando os mesmos dados das amostras anteriores, utilizadas na identificação dos classificadores, tem-se que o tempo médio de recidiva apresentado foi de 12,7 meses, com mediana de 11 meses.

Esta comparação entre tempo médio de recidiva anterior com o tempo médio de seguimento atual demonstra que o mesmo é suficiente para a análise acurada dos classificadores, uma vez que o tempo médio de seguimento atual é de 20,5 meses, superior ao tempo médio de recidiva apresentado anteriormente, de 12,7 meses.

Os fatores preditivos de resposta mais importante são volume tumoral (DUBBEN et al. 1998; MANCUSO et al. 1999; VAN DEN BROEK et al. 2004) e classificação TNM. Apesar disso, nem o comportamento biológico e nem resposta à terapia podem ser totalmente explicados por estes fatores (PRAMANA et al. 2007).

Dessa forma, foram calculados os p-valores para todas as variáveis (Tabela 14), tanto através da utilização do teste exato de Fisher quanto do teste qui-quadrado. Ambos podem ser utilizados quando as variáveis são categóricas. O teste estatístico exato de Fisher é útil quando o número amostral é pequeno, independente das características das amostras. Além disso, quando o número amostral é pequeno, é significativamente mais acurado do que o teste qui-quadrado. Já o teste estatístico qui-quadrado não deve ser utilizado com número pequeno de amostras ou dados desbalanceados, pois pode levar a conclusões opostas. Dessa forma, foram considerados os valores do teste de Fisher.

Foram considerados dois grupos, R x NR, com hipótese nula H_0 de que a variável clinicopatológica não tem correlação com resposta do paciente ao tratamento clínico, com coeficiente de confiança de 95% (Apêndice 3).

A única variável que apresentou significância estatística (p-valor: 0,032) foi metástase em linfonodo regional, N, (Tabela 14 e Apêndice 3), e 100% dos pacientes NR são N0. Este dado não parece apresentar um padrão, pois no grupo da identificação dos classificadores apenas 50% das amostras NR eram N0 (p-valor: 0,32).

Todas as outras variáveis não apresentam significância estatística, inclusive cidade de origem dos pacientes (p-valor: 1,0). Neste caso, a hipótese nula não é rejeitada.

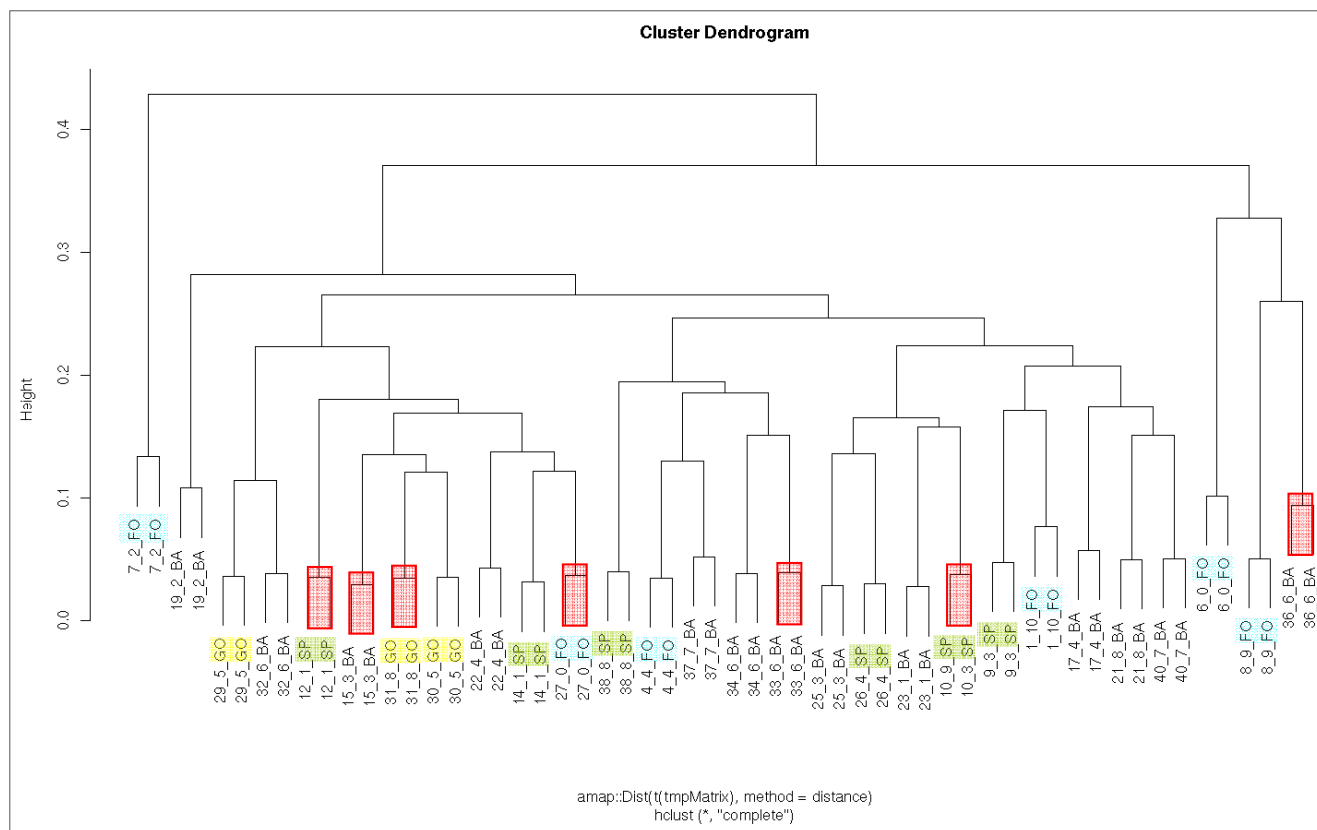
5.1.2 Análises dos dados gerados por *microarray* e análises matemáticas

Os resultados das análises foram iniciados com a geração de *cluster* hierárquico não supervisionado com distância de correlação, utilizando-se o método completo, o qual toma o máximo das distâncias entre todos os elementos que compõem um grupo (*cluster*). Na análise não supervisionada a busca de relação entre os genes ou entre as amostras é realizada sem informação adicional, ou seja, é baseada apenas na expressão gênica dos dados obtidos no experimento (TRICHE et al. 2001; RAMASWAMY e GOLUB 2002).

O agrupamento hierárquico se baseia em uma matriz de distância par a par, a partir do qual as amostras mais semelhantes, ou pacientes, são reunidos dois a dois, sucessivamente, formando uma única árvore hierárquica. As análises de *cluster* apresentam como objetivo encontrar grupos de genes, ou de amostras, que se comportem de modo semelhante nas condições experimentais estudadas (FREEMAN et al. 2000; ZHANG et al. 2001). O agrupamento de amostras complexas permite

identificar subgrupos que, em estudos de doenças, por exemplo, podem apresentar diferentes prognósticos ou respostas a tratamentos.

Após a normalização dos dados foi então gerado cluster hierárquico não supervisionado, ou dendrograma (Figura 9), no qual todas as amostras se parearam com sua duplicata, demonstrando a alta qualidade dos dados e a ausência da necessidade de se re-hibridizar alguma amostra. Além disso, é possível perceber que não existe agrupamento por cidade de origem da amostra e nem agrupamento pelo dia em que o experimento de hibridização foi realizado, o que corrobora com a alta qualidade dos dados gerados por *array*.



Legenda: Cluster hierárquico não supervisionado das 28 amostras *main* e *swap*, baseado no perfil de expressão de todos os genes da lâmina. O primeiro número é referente à amostra (1 a 40), o segundo é referente à hibridização (dia 0 a 8) e as letras são referentes à cidade de origem do paciente (BA: Barretos; FO: Fortaleza, em azul; GO: Goiânia, em amarelo e SP: São Paulo, em verde). Os pacientes NR estão representados pelos retângulos vermelhos.

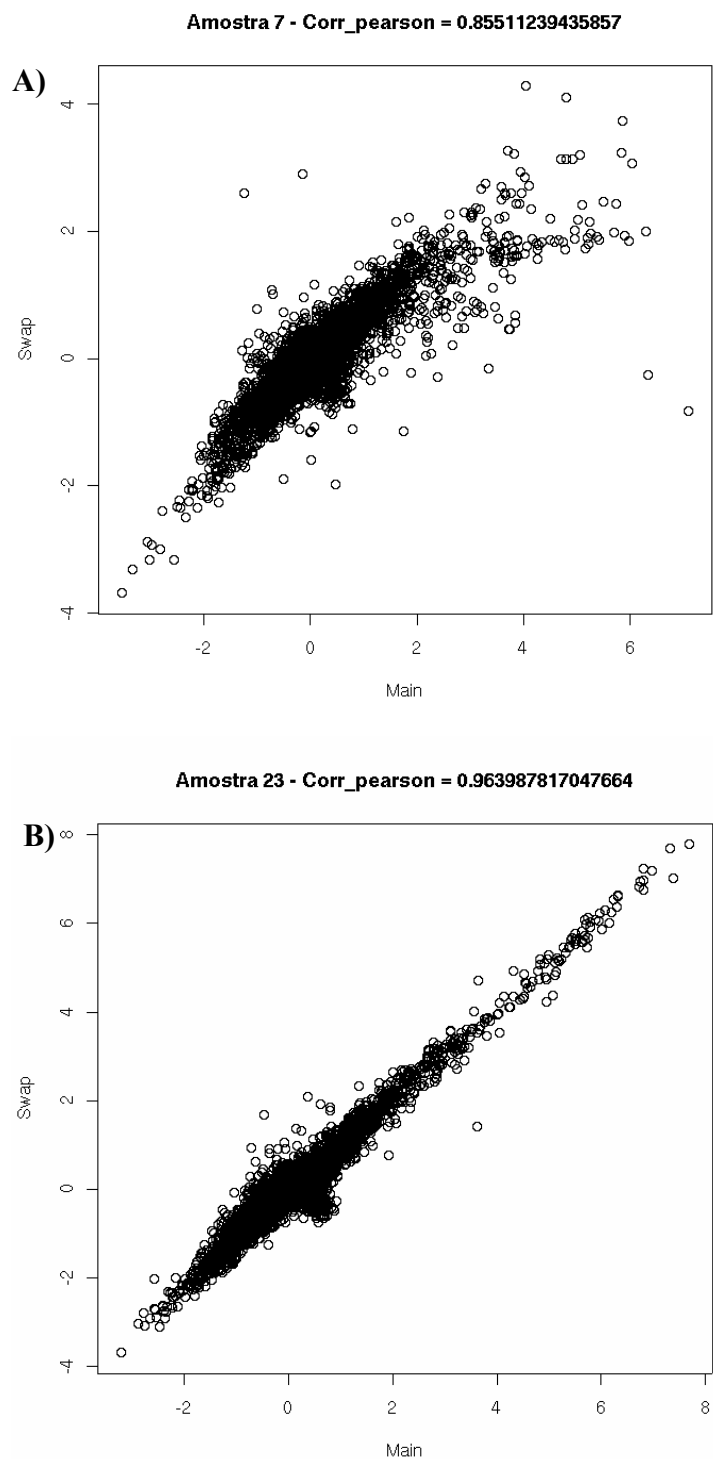
Figura 9 - Cluster hierárquico.

Outro dado importante de ser analisado, do ponto de vista da qualidade da lâmina de *array* gerada, é o número de *spots* utilizáveis nas análises seguintes em cada lâmina. Após a eliminação dos controles restam 4.600 *spots* e em nossas lâminas a variação de *spots* utilizáveis foi de 4.556 a 4.596, com média de 4.592 *spots*, o que se traduz em dados de ótima qualidade.

As análises de controle de qualidade também geraram gráficos do tipo MM *plot* (Figura 10), nos quais são calculadas as correlações de Pearson entre os *spots* da lâmina controle e os da lâmina *swap*. Quanto mais próximo de 1 for o resultado,

melhor é a correlação entre as lâminas. Neste estudo, os resultados variaram de 0,855 (Figura 10A) a 0,963 (Figura 10B), com média de 0,934, novamente demonstrando a ótima qualidade das lâminas e dos dados gerados, dispensando a necessidade de se refazer novas hibridizações.

Em seguida, foram analisados os genes mais diferencialmente expressos, corrigidos por FDR (*False Discovery Rate*), ordenados pelo menor p-valor obtido pelo teste não paramétrico de Wilcoxon. A correção do p-valor é necessária devido a realização de busca exaustiva. Esta tabela foi criada a partir da diferença de ordenação entre R e NR; assim sendo, os genes que apresentam valores de *fold mean* positivos são superexpressos nos pacientes R, enquanto que os genes com valores negativos de *fold mean* são superexpressos nos pacientes NR. Nenhum gene apresentou significância estatística (Tabela 15).



Legenda: Gráficos utilizados para controle de qualidade das amostras. O painel A se refere à amostra 7, que apresentou a menor correlação entre as duplicatas. O painel B se refere à amostra 23, que apresentou a maior correlação entre as duplicatas.

Figura 10 - Gráficos MM plot (marginal model).

Tabela 15 - Genes mais diferencialmente expressos, na comparação RxNR, ordenados em ordem decrescente de *fold*.

Official Symbol	Name	GenBank	Fold	P.value
KRT1	keratin 1	BF739130.1	-1.713	0.9723
COL17A1	Collagen, type XVII, alpha 1	BE703135.1	-1.5466	0.9574
KRT4	keratin 4	BF831397.1	1.4265	0.9552
IFI44L	interferon-induced protein 44-like	BE767619.1	-1.3105	0.9552
SERPINA3	serpin peptidase inhibitor, clade A, member 3	BF837767.1	-1.2646	0.9723
PPP2R2C	protein phosphatase 2	BE703712.1	-1.2647	0.9552
KLK7	kallikrein-related peptidase 7	BE713379.1	-1.1342	0.9552
IFI44	interferon-induced protein 44	AW381443.1	-1.1305	0.9552
LAMA3	laminin, alpha 3	BG945958.1	-1.1279	0.9723

O método de Discriminante Linear de Fisher foi executado utilizando a função *lda* (análise por discriminante linear) do pacote MASS do ambiente de análise estatística R; com o objetivo da busca por potenciais classificadores moleculares, realizada no trabalho de Mestrado. O Discriminante de Fisher leva em conta a informação de proporção entre R e NR, para predizer a classe das amostras. No caso dos classificadores previamente identificados, a proporção utilizada foi de 60% R e 40% NR.

Assim sendo, estudo de Mestrado anterior identificou os 100 melhores trios de genes classificadores e quadras, através do Discriminante Linear de Fisher; e, através de validação cruzada, identificou 4 trios (Tabela 2) e 27 quadras classificadoras (Tabela 3), com 100% de acerto (VC= 35).

Os classificadores se baseiam no perfil de expressão gênica combinada de cada trio ou quadra de genes para classificar dada amostra. Uma vez conhecida a regra de classificação, esta foi aplicada aos dados gerados por *microarray* das amostras do grupo da validação independente, a fim de analisar o comportamento dos classificadores neste novo banco de dados. Ou seja, a regra de classificação que

identificou anteriormente os 100 melhores trios e as 100 melhores quadras foi agora aplicada aos dados gerados pelas amostras do grupo da validação.

A utilização de RNAs referência diferentes em ambos os experimentos gerou um artefato, não sendo possível obter a sobreposição das amostras. Desta forma, foi necessário centralizar os níveis de expressão, utilizando a média não ponderada, para que as nuvens das amostras antigas e novas ficassem centralizadas a partir de seus centros de massa, e com isso pudessem ser sobrepostas e posteriormente comparadas.

Assim, após a centralização dos dados, a regra de classificação anterior foi aplicada às novas amostras e o comportamento das mesmas analisado. Como resultado, o melhor trio (Tabela 16) validado de genes, trio número 3, classificou corretamente 18 de 28 amostras, sendo que 14R e 4NR foram classificadas corretamente e, conseqüentemente, 7R e 3NR foram classificadas de forma incorreta. O valor estimado de sensibilidade do trio classificador é 66,6%, com (IC 95% = 43,1-84,5); enquanto que o valor estimado de especificidade é 57,14%, com (IC 95% = 20,23-88,2).

Tabela 16 - Trios de genes que classificaram os pacientes em respondedores ou não respondedores, com aplicação da regra de classificação em grupo independente de 28 amostras.

Trio	Gene 1	Gene 2	Gene 3	SVD_RvsNR ^(b)	VC ^(c)	Predição das 28	Acerto de NR	Acerto de R
3	ATP9B	KIAA1691	Hs.218008	10,926	35	18	4	14
5	KDEL3	KIAA1691	TRIM9	10,684	35	17	3	14
18	LAMA3	KIAA0819	TRIM2	9,984	35	16	3	13
4	FAD104	PPP1R9B	IFIT1	10,882	35	10	2	8

^(b)singular value decomposition ^(c)validação cruzada por *leave-one-out*

Já as quadras, 8 e 9, validadas de genes (Tabela 17) classificaram corretamente 19 de 28 amostras, sendo que 16R e 3NR foram classificadas corretamente e, conseqüentemente, 5R e 4NR foram classificadas de forma incorreta. O valor estimado de sensibilidade dessas quadras classificadoras são 76%, com (IC 95% = 52-90); enquanto que o valor estimado de especificidade são 43%, com (IC 95% = 12-79).

As quadras 6, 27 e 47, validadas de genes (Tabela 17) classificaram corretamente 18 de 28 amostras, sendo que 13R e 5NR foram classificadas corretamente e, conseqüentemente, 8R e 2NR foram classificadas de forma incorreta. O valor estimado de sensibilidade dessas quadras classificadoras são 62%, com (IC 95% = 38-81); enquanto que o valor estimado de especificidade são 71%, com (IC 95% = 30-95).

Todo classificador pode ser representado por uma equação que descreve a separação das amostras no espaço, o que facilitaria a implementação destes na rotina clínica. Para isso, caso o classificador seja linear, é necessário calcular a regra do plano, para trios, ou regra do hiperplano, para quadras. Caso o classificador não seja linear, é necessário calcular a equação da superfície.

Tabela 17 - Quadras de genes que classificaram os pacientes em respondedores ou não respondedores, com aplicação da regra de classificação em grupo independente de 28 amostras.

Quadra	Gene 1	Gene 2	Gene 3	Gene 4	SVD RvsNR ^(b)	VC ^(c)	Predição das 28	Acerto de NR	Acerto de R
8	DFFA	KIAA1691	TRIM9	CDCA4	14,327	35	19	3	16
9	KIAA0819	KIAA1691	TRIM37	HLA-DOB	14,300	35	19	3	16
6	ATP9B	KIAA1691	Hs.218008	GTF2H4	14,514	35	18	5	13
27	DKFZP566C134	RALA	ITGB5	KIAA0095	13,191	35	18	5	13
47	BBS2	KIAA1691	KIAA0335	ZDHHC6	12,520	35	18	5	13
1	KIAA1691	TRIM9	ATP9B	Hs.218008	16,176	35	18	4	14
4	Hs.218008	KIAA1691	FLJ10648	CGBP	14,825	35	17	4	13
30	CGBP	KIAA1691	KIAA0819	TRIM37	13,149	35	17	3	14
71	E2IG4	KIAA0819	KIAA1691	PP2135	11,690	35	17	3	14
18	KDEL3	KIAA1691	TRIM9	EIF2B3	13,661	35	16	3	13
38	APP	DKFZP566C134	RBM12	PIGL	12,795	35	16	4	12
55	AHCYL1	KIAA1691	TRIM9	DUSP18	12,210	35	16	2	14
5	ZNF151	KIAA1691	TRIM9	ATP9B	14,762	35	15	2	13
15	SMCY	TPT	C3AR1	C14orf130	13,830	35	15	3	12
40	ATP5L	SMCY	AKAP10	DBT	12,769	35	15	4	11
54	IAN4L1	KIAA1691	DKK4	ATP9B	12,233	35	15	2	13
56	ITGB5	SMCY	ATP9B	DKFZP434C171	12,197	35	15	2	13
62	DKFZP564O043	FAD104	KIAA1170	HMGCL	11,927	35	15	3	12
16	DKK4	KIAA1691	HLA-DOB	CLASP2	13,769	35	14	3	11
19	Hs.289123	TRIM9	KIAA1691	ATP9B	13,536	35	14	3	11
42	DMTF1	KIAA1691	TRIM9	ATP9B	12,712	35	14	2	12
53	Hs.379757	KIAA1691	TRIM9	KDEL3	12,257	35	14	2	12
13	NOS3	KIAA1691	TRIM9	ATP9B	14,121	35	13	2	11
29	Jade-1	KIAA1691	TRIM9	SCAMP-4	13,176	35	13	2	11
32	SDCCAG10	ATP9B	TRIM9	KIAA1691	13,075	35	12	0	12
100	FLJ12565	DKFZP434A1315	KLF8	WWP2	9,980	35	11	3	8
12	FAD104	PPP1R9B	IFIT1	FLJ20321	14,133	35	10	2	8

^(b)singular value decomposition ^(c)validação cruzada por *leave-one-out*

Apesar de validados classificadores com valores estimados de sensibilidade e especificidade de 62% e 71%, respectivamente, foi ainda aplicada a estratégia de árvore de decisão binária para classificação (DUDOIT et al. 2002a), com a finalidade de melhorar a técnica de classificação, aumentando a taxa de acerto.

Esta técnica não-linear e supervisionada consiste em agregar regras de classificação diversas para melhorar o desempenho do classificador. Assim, são realizadas sucessivas partições no espaço de genes, utilizando mais de um classificador, de modo a classificar melhor do que utilizando um classificador

isolado. Na busca da composição das árvores foram utilizados os 4 trios de genes e as 27 quadras classificadoras.

A árvore A de decisão (Figura 11) identificada utiliza inicialmente a regra da quadra 13, classificando as 28 amostras em dois braços, R ou NR. Estão classificadas como R as amostras 12, 15, 17, 1, 22, 23, 27, 6, 9, 31, 32, 33, 37, 38, 40 e 8. Já no braço NR estão as amostras 14, 19, 25, 26, 4, 7, 10, 21, 29, 30, 34 e 36.

O segundo passo é a aplicação da regra da quadra 6, nas amostras provenientes do braço R da regra 13. Da mesma forma, a regra 6 classifica as amostras em R ou NR. O braço NR classifica corretamente todas as amostras (27, 31 e 33); enquanto que o braço R acerta a classificação de 6, 17, 1, 22, 23, 9, 32, 37, 38, 40 e 8 e erra a de 12 e 15.

O passo final é a aplicação da regra da quadra 47, nas amostras provenientes do braço NR da regra 13. Do mesmo modo, a regra 47 classifica as amostras em R ou NR. O braço R classifica corretamente todas as amostras (14, 19, 25, 26, 4, 7, 30 e 34); enquanto que o braço NR acerta a classificação de 10 e 36 e erra a de 21 e 29.

Como resultado da aplicação da árvore A de decisão, tem-se que as amostras R 21 e 29 e as amostras NR 12 e 15, são as únicas classificadas de forma incorreta.

Ou seja, a partir da combinação das regras de classificação das quadra 13, 6 e 47 (Tabela 18) foi obtida árvore A de decisão (Figura 11) que classificou corretamente 24 das 28 amostras, sendo que 19 amostras R e 5 NR foram classificadas corretamente, e, conseqüentemente, 2 amostras R e 2 NR foram classificadas de forma incorreta. Assim, o valor estimado de sensibilidade é 90,4%,

Tabela 18 - Aplicação, nas 28 amostras clinicamente R ou NR, das regras de classificação das quadras de genes que compõe as árvores A e B de decisão. 0 significa que a regra da quadra erra a classificação da amostra e 1 que acerta.

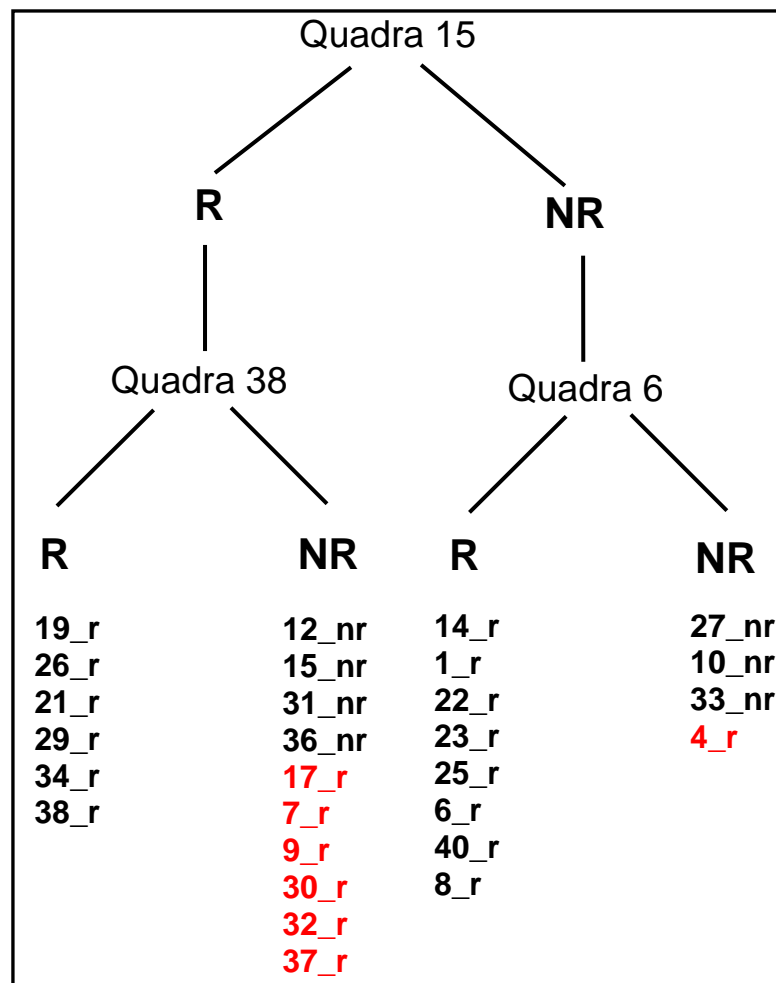
Amostra	12	14	15	17	19	1	22	23	25	26	27	4	6	7	9	10	21	29	30	31	32	33	34	36	37	38	40	8
			N																									
	NR	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
Quadra 13	0	0	0	1	0	1	1	1	0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	1	1	1	1	1
Quadra 6	0	1	0	1	0	1	1	1	1	0	1	0	1	0	1	1	0	0	0	1	1	1	0	1	1	1	1	1
Quadra 47	0	1	1	0	1	0	1	0	1	1	1	1	1	1	1	1	0	0	1	0	0	1	1	1	0	1	1	0
Quadra 15	0	0	0	1	1	0	0	0	0	1	1	0	0	1	1	1	1	1	1	0	1	1	1	0	1	1	0	0
Quadra 38	1	1	1	0	1	1	0	0	1	1	0	1	0	0	0	0	1	1	0	1	0	0	1	1	0	1	1	1

Foi também identificada árvore B de decisão (Figura 12), a qual utiliza inicialmente a regra da quadra 15 para classificar as 28 amostras em dois braços, R ou NR. Estão classificadas como R as amostras 12, 15, 17, 19, 26, 7, 9, 21, 29, 30, 31, 32, 34, 36, 37 e 38. Já no braço NR estão as amostras 14, 1, 22, 23, 25, 27, 4, 6, 10, 33, 40 e 8.

O segundo passo é a aplicação da regra da quadra 38, nas amostras provenientes do braço R da regra 15. Da mesma forma, a regra 38 classifica as amostras em R ou NR. O braço R classifica corretamente todas as amostras (19, 26, 21, 29 34 e 38); enquanto que o braço NR acerta a classificação de 12, 15, 31 e 36 e erra a de 17, 7, 9, 30, 32 e 37.

O passo final é a aplicação da regra da quadra 6, nas amostras provenientes do braço NR da regra 15. Do mesmo modo, a regra 6 classifica as amostras em R ou NR. O braço R classifica corretamente todas as amostras (14, 1, 22, 23, 25, 6, 40 e 8); enquanto que o braço NR acerta a classificação de 27, 10 e 33 e erra a da amostra 4.

Como resultado da aplicação da árvore de decisão, tem-se que as amostras 17, 7, 9, 30, 32, 37 e 4, todas R, são as únicas amostras classificadas de forma incorreta.



Legenda: Árvore de decisão que classifica as amostras utilizando as regras de classificação da quadra 15, e em seguida das quadras 38 e 6. A árvore erra a classificação das amostras em vermelho, e acerta a das amostras em preto.

Figura 12 - Árvore B de decisão.

A partir da combinação da regra de classificação da quadra 15, quadra 38 e quadra 6 (Tabela 18) foi obtida árvore B de decisão (Figura 12) que classificou corretamente 21 das 28 amostras, sendo que 14 amostras R e todas as 7 NR foram

classificadas corretamente, e, conseqüentemente, 7 R foram classificadas de forma incorreta. Como resultado, o valor estimado de sensibilidade é 66,6%, com (IC 95% = 43-84,5); enquanto que o valor estimado de especificidade é 100,0%, com (IC 95% = 56-100).

Deste modo, nota-se aumento nas taxas de acerto das árvores de decisão, quando comparadas ao trio de genes ou às quadras isoladas (Tabela 19).

Tabela 19 - Cálculo dos valores estimados de sensibilidade e especificidade, com seus respectivos intervalos de confiança para IC 95%, do melhor trio, melhores quadras e melhores árvores de decisão identificadas.

Classificador	Valor Estimado de Sensibilidade	IC 95% Sensibilidade	Valor Estimado de Especificidade	IC 95% Especificidade
Trio 3	66,6%	[43,1-84,5%]	57,14%	[20,2-88,2%]
Quadras 8 e 9	76,0%	[52,0-90,0%]	43,0%	[12,0-79,0%]
Quadras 6, 27 e 47	62,0%	[38,0-81,0%]	71,0%	[30,0-95,0%]
Árvore A de Decisão	90,4%	[68,1-98,3%]	71,4%	[30,2-94,8%]
Árvore B de Decisão	66,6%	[43,0-84,5%]	100%	[56,0-100%]

5.2 OBJETIVO SECUNDÁRIO: ESTUDO DE VIAS METABÓLICAS E GENES

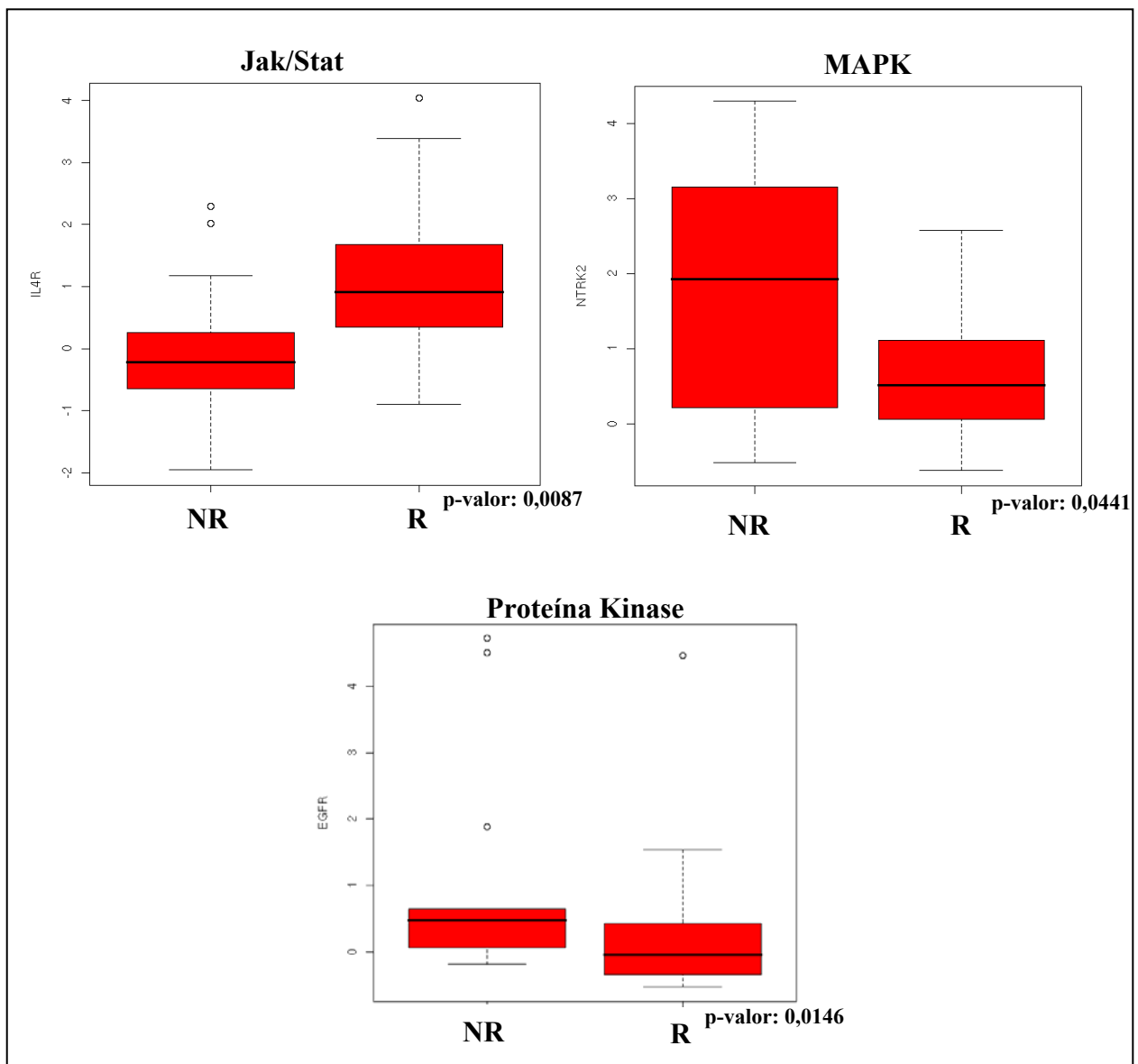
A fim de determinar quais genes estão diferencialmente expressos entre os respondedores e os não respondedores, no grupo das 35 amostras do estudo de Mestrado, foi utilizado o teste estatístico de Wilcoxon. Este teste não assume nenhuma distribuição dos dados, é não paramétrico e sua principal utilização é no ranqueamento de dados, analisando as diferenças entre grupos estudados, neste caso,

pacientes R e NR. É importante lembrar que, como explicado anteriormente, foram utilizadas 24 amostras (13 R e 11 NR) do grupo das 35 amostras.

Através da análise das três vias selecionadas e do Teste de Wilcoxon, considerando $|fold|$ de 0,51 a 1,09 e valor de p : $0,0062 \geq p \geq 0,0441$, foram identificados 6 genes, IL4R; NTRK2 ou TRKB; GNA12; EGFR; CMET e TGFA (Tabela 20) diferencialmente expressos entre as comparações R/NR no grupo das 35 amostras, a partir dos dados gerados por *microarray*. Estes genes apresentaram significância estatística e biológica. Os *boxplots* referentes ao gene mais diferencialmente expressos de cada via estão representados na Figura 13.

Tabela 20 - Lista dos 6 genes mais diferencialmente expressos nas 3 vias, na comparação R X NR no grupo das 35 amostras do projeto de Mestrado, pelo Teste de Wilcoxon. *Fold* positivo significa gene mais diferencialmente expresso nos pacientes respondedores e sinal negativo, nos não respondedores.

<i>Functional Modules</i>	<i>Name</i>	<i>Genbank</i>	<i>Unigene</i>	<i>Annotation</i>	<i>Fold</i>	<i>P.value</i>
Jak/Stat	IL4R	BE698417	Hs.75545	interleukin 4 receptor	1.09	0.0087
	NTRK2	BF828250	Hs.439109	neurotrophic tyrosine kinase, receptor, type 2	-1.02	0.0441
MAPK	EGFR	BG006329	Hs.77432	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	-0.73	0.0146
	GNA12	BI032676	Hs.182874	guanine nucleotide binding protein (G protein) alpha 12	-0.51	0.0062
Protein Kinase	EGFR	BG006329	Hs.77432	epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)	-0.73	0.0146
	MET	BQ316492	Hs.419124	met proto-oncogene (hepatocyte growth factor receptor)	0.7	0.0312
	TGFA	BC005308	Hs.170009	transforming growth factor, alpha	-0.54	0.007



Legenda: Distribuição do nível de expressão do gene mais diferencialmente expresso de cada uma das três vias analisadas, comparando os grupos dos pacientes R com NR, no grupo das 35 amostras do projeto de Mestrado.

Figura 13 - *Boxplot* ou gráfico-caixa.

Após as análises dos dados de expressão gênica gerados por *microarray*, foram realizados ensaios de imunistoquímica, a fim de comparar os dados identificados.

As 24 lâminas de imunistoquímica (13 R e 11 NR), provenientes das 35 amostras do estudo de Mestrado, foram analisadas com auxílio de microscópio e suporte do Departamento de Anatomia Patologia do Hospital AC Camargo.

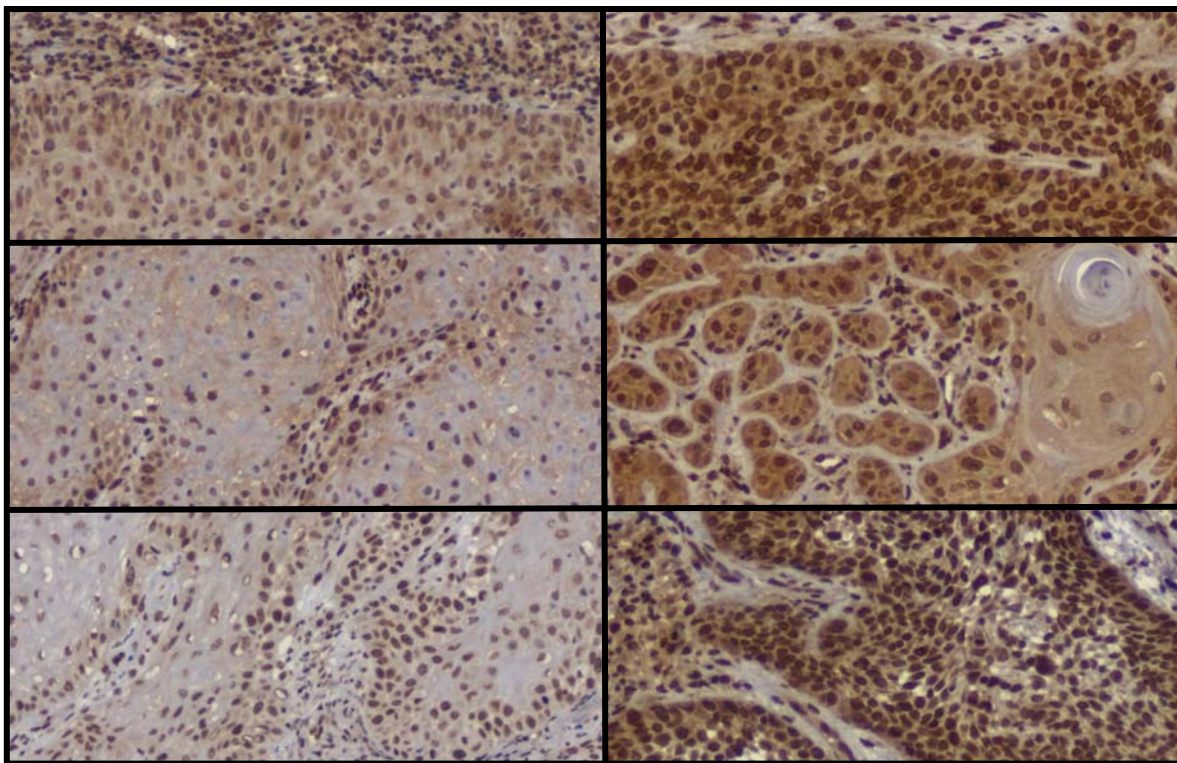
Cada anticorpo foi analisado de acordo com sua localização celular. Assim, foi atribuída escala que variou de negativo a 3 cruces, de acordo com a intensidade de coloração identificada para os anticorpos com localização citoplasmática e de membrana. Foi atribuído 1 para coloração fraca, 2 para moderada e 3 para forte. Já para os anticorpos com localização nuclear, foi atribuído valor que variou de negativo a 100%, de acordo com o número de núcleos corados em relação ao total (Tabela 21). A reação do anticorpo IL4R não funcionou.

Finalizada a atribuição de *scores*, foram realizadas análises estatísticas para comparação de médias para os diversos experimentos. Foram considerados dois grupos, R x NR, com amostras independentes e com hipótese de que as amostras não apresentam distribuição normal. Desta forma, foi realizado teste não paramétrico de Wilcoxon; com hipótese nula H₀ de que as duas médias são iguais, com coeficiente de confiança de 95% (Apêndice 2).

Tabela 21 - Resultados das reações de imunoistoquímica, realizadas com 5 anticorpos diferentes, em amostras de CE de laringe e hipofaringe, do grupo do projeto de Mestrado, em pacientes R e NR. C significa análise da marcação do anticorpo em citoplasma celular, M na membrana e N no núcleo.

		Anticorpos									
		TRKB	TRKB-p			c-MET		EGFR		TGFα	
Amostras	C	C	M	N	C	N	C	M	C	N	
R	LA27	1	neg.	3	30%	2	40%	neg.	3	2	90%
	LA30	1	3	3	neg.	2	neg.	neg.	3	1	10%
	LA26	1	3	3	10%	2	neg.	neg.	2	1	5%
	LA12	neg	1	1	neg.	2	80%	neg.	3	2	100%
	LA35	neg	neg.	neg.	neg.	3	90%	neg.	3	2	100%
	LA21	neg	2	2	50%	1	80%	neg.	3	2	100%
	LA22	neg	neg.	neg.	neg.	1	70%	neg.	2	1	100%
	LA38	neg	neg.	neg.	neg.	1	90%	neg.	3	2	100%
	LA37	1	1	1	neg.	3	80%	neg.	3	2	100%
	LA41	1	neg.	neg.	neg.	2	50%	neg.	3	2	100%
	LA40	neg	neg.	neg.	neg.	2	90%	neg.	3	3	100%
	LA39	neg	1	1	neg.	3	40%	neg.	3	3	100%
	LA24	neg	2	2	40%	2	80%	neg.	3	2	90%
	NR	LA28	2	1	neg.	50%	3	90%	neg.	3	3
LA25		neg.	2	neg.	95%	2	90%	neg.	3	3	90%
LA31		1	3	3	70%	3	70%	neg.	3	3	100%
LA04		3	3	3	neg.	3	neg.	neg.	3	2	90%
LA32		2	2	neg.	60%	2	50%	neg.	3	2	100%
LA11		neg.	neg.	neg.	neg.	2	90%	neg.	3	2	100%
LA16		1	1	1	neg.	2	70%	neg.	3	2	70%
LA17		neg.	1	1	neg.	2	neg.	neg.	3	2	100%
LA36		neg.	1	1	neg.	3	90%	neg.	3	3	100%
LA09		neg.	neg.	neg.	neg.	3	90%	neg.	3	3	100%
LA08	1	2	2	90%	2	40%	3	neg.	3	100%	

De acordo com as hipóteses acima, o único experimento que apresentou diferença significativa nas médias entre os grupos foi o do anticorpo TGF α , com localização citoplasmática (Figura 14 e Apêndice 2). O p-valor identificado na comparação de médias entre os dois grupos foi de 2,6% (p-valor = 0,02692), que é menor que 5%, assim, a hipótese nula é rejeitada, indicando a existência de diferença significativa entre as médias (Apêndice 2).



Legenda: Imunoistoquímica realizada com anticorpo TGFA. Na coluna da esquerda estão 3 fotomicrografias de 3 pacientes distintos do grupo dos R e na coluna da direita estão 3 de 3 pacientes distintos do grupo dos NR. Essas amostras são referentes ao grupo do projeto de Mestrado. Nota-se que a coloração citoplasmática é mais intensa no grupo dos NR (p -valor = 0,02692), à direita. As fotos foram tiradas em aumento de 200 vezes.

Figura 14 - Fotomicrografias de cortes histológicos de CE de laringe e hipofaringe.

Os dados de expressão gênica anteriormente identificados, através de *microarray* (Tabela 20), mostraram que TGFA estava significativamente superexpresso no grupo dos pacientes NR.

Os resultados de imunoistoquímica (Figura 14 e Apêndice 2) validaram este dado, já que foi observada maior abundância da proteína TGFA, com significância estatística (p -valor: 0,02692), no grupo dos NR. Este dado é certamente interessante e promissor, uma vez que a localização citoplasmática de TGFA poderia ser utilizada, juntamente com outros fatores, como marcador de resposta para a condição clínica estudada.

TGFA é um fator de crescimento com 33% de homologia ao EGF e que se liga ao EGFR, ativando a fosforilação da tirosina e também estimulando a proliferação celular. TGFA também apresenta papel importante na iniciação tumoral, induzindo fenótipo transformado reversível. A ligação de TGFA a EGFR resulta em mudança conformacional no receptor, o que promove homodimerização com outras moléculas de EGFR ou heterodimerização com outros membros da família HER (principalmente HER2); dimerização resulta em subsequente autoativação de tirosina quinase do domínio intracelular do receptor. Este processo irá ativar uma via de sinalização intracelular, levando à inibição da apoptose, ativação de proliferação celular e angiogênese, assim como um aumento no potencial metastático (ROSKOSKI 2004).

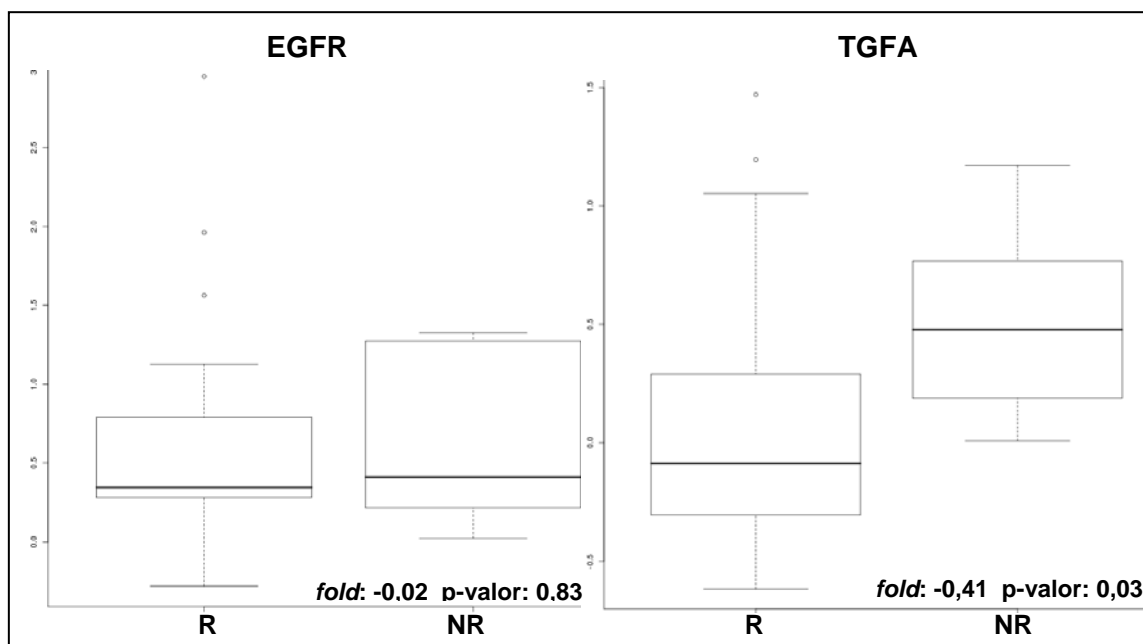
TGFA participa de uma via de sinalização autócrina em células epiteliais da mucosa transformada, mas o mesmo não ocorre em células normais. Tendo como alvo o RNA mensageiro de TGFA, com a técnica de oligonucleotídeos *antisense*, foi demonstrado decréscimo de até 93% da proteína TGFA e 76% de redução da proliferação celular em linhagens celulares humanas de CECP (GRANDIS et al. 1998). A expressão de TGFA também está fortemente correlacionada com pior prognóstico em câncer de mama (KENNY e BISSELL 2007).

Células tumorais podem gerar seus próprios sinais de crescimento, assim, reduzindo a dependência de estímulo do microambiente normal, ao adquirir a capacidade de se tornar autosuficiente. Um dos exemplos é a produção de TGFA (HANAHAN e WEINBERG 2000).

A superexpressão de TGFA no grupo dos NR deste estudo poderia ser responsável, entre outros fatores, pelo aumento da inibição de apoptose e ativação da

angiogênese, o que poderia explicar a não resposta à quimiorradioterapia. Já a ativação de proliferação celular poderia ser responsável por um fenótipo mais agressivo no grupo dos NR.

Em seguida, foi realizada análise da expressão gênica diferencial entre os pacientes R e NR, através dos dados gerados por *microarray*. Esta análise ocorreu no grupo das 28 amostras da validação e também apresentou TGFA superexpresso no grupo dos NR (Figura 15), com significância estatística (p-valor:0,03), corroborando mais uma vez este achado.



Legenda: Distribuição do nível de expressão dos genes EGFR e TGFA, comparando os grupos dos pacientes R com NR, referentes ao grupo da validação.

Figura 15 - *Boxplot* ou gráfico-caixa.

Os dados gerados por *array* das amostras do estudo de Mestrado também mostraram superexpressão do gene EGFR no grupo dos pacientes NR (Tabela 20), e a expressão de sua proteína, com localização citoplasmática, também estava aumentada no mesmo grupo, através das análises de imunistoquímica (Apêndice 2).

Apesar desses dados corroborarem os dados de *array*, a diferença identificada na imunohistoquímica não apresentou significância estatística (p-valor: 0,458, Apêndice 2).

O receptor do fator de crescimento epidermal (EGFR) é um receptor transmembrana de tirosina quinase da família ErbB, o qual se apresenta anormalmente ativado em muitos tumores epiteliais. Esta família é composta por 4 receptores parecidos entre si, EGFR/ErbB1/HER1; ErbB2/HER2; ErbB3/HER2 e ErbB4/HER4. Esses receptores ativam vias de sinalização *downstream* que não são lineares, mas consistem em ricas multicamadas e redes cruzadas conectadas, o que permite interações horizontais e respostas combinatoriais múltiplas. Isso poderia explicar a variedade de respostas biológicas, frente à ativação de um receptor específico em uma célula específica.

A ativação do receptor leva ao recrutamento e a fosforilação de inúmeros substratos intracelulares, os quais, por sua vez, ativam vias mitogênicas e outras atividades promotoras de tumor. A desregulação dessas vias de sinalização, que fortemente regulam o receptor ErbB, contribuem para a transformação maligna. Muitos mecanismos levam à ativação aberrante do receptor que é observada em cânceres, incluindo superexpressão do receptor, mutação, dimerização do receptor ligante-dependente e ativação ligante independente (KLAPPER et al. 2000; OLAYIOYE et al. 2000; YARDEN e SLIWKOWSKI 2001; MENDELSON e BASELGA 2003).

Além disso, existem estudos que demonstraram que altos níveis de EGFR foram associados com menor resposta tumoral à radioterapia. Desta forma, esses estudos pré-clínicos *in vivo* sugerem que a superexpressão de EGFR pode servir

como um preditor de resposta tumoral à radioterapia, e como alvo terapêutico para aumentar a eficiência da radioterapia (AKIMOTO et al. 1999; MILAS et al. 2004).

Assim, a resistência à quimioterapia nos pacientes NR pode ser devido à superexpressão de EGFR, entre outros fatores.

A análise da expressão gênica diferencial, realizada no grupo das 28 amostras da validação, também apresentou EGFR superexpresso no grupo dos NR (Figura 15). Embora não haja significância estatística (p-valor: 0,83), existe uma visível tendência de superexpressão deste gene nos pacientes NR.

Novamente, de acordo com os dados gerados por *microarray* das amostras do projeto de Mestrado, o gene NTRK2 se mostrou significativamente superexpresso no grupo dos NR (Tabela 20), e a superexpressão da sua proteína também foi identificada para o mesmo grupo. Este dado ocorreu com a proteína localizada no citoplasma, tanto para a versão fosforilada (p-valor:0,28) quanto para a não fosforilada (p-valor:0,22); e também para a proteína localizada no núcleo (p-valor:0,19). Mas, apesar dos dados de imunistoquímica corroborarem os dados de *array*, não houve significância estatística (Apêndice 2).

O gene receptor tirosina quinase neurotrófico tipo 2, NTRK2 ou TRKB, e seu ligante fator neurotrófico cérebro-derivado, BDNF, são essenciais para o desenvolvimento e função do sistema nervoso (KLEIN et al. 1993; ERNFORS et al. 1994; JONES et al. 1994). Por outro lado, altos níveis de expressão da proteína TRKB estão envolvidos em diversas neoplasias humanas (GEIGER e PEEPER 2007). Além disso, também foi identificada como oncoproteína associada à capacidade metastática, e sua superexpressão levou a resistência a anoikis, alta tumorigenicidade e a metástases altamente invasivas (DOUMA et al. 2004).

Anoikis é um tipo de apoptose resultante da perda de interação célula-matriz, e é sugerida ser uma barreira à metástase. Muitas células tumorais são anoikis resistentes; desta forma, esta supressão ajudaria na disseminação de células tumorais, facilitando a ocorrência de metástases (MEREDITH et al. 1993; FRISCH e FRANCIS 1994; LIOTTA e KOHN 2001).

Assim, a ausência de resposta à quimioterapia no grupo dos NR poderia ser explicada, entre outros fatores, pela superexpressão de NTRK2 neste grupo e pela resistência a anoikis.

Apesar disso, a análise da expressão gênica diferencial através dos dados gerados por *microarray*, realizada no grupo das amostras da validação, não apresentou NTRK2 mais expresso no grupo dos NR (*fold*= 0,33); embora o p-valor não seja significativo (p-valor:0,56) .

6 DISCUSSÃO

A análise do perfil de expressão gênica é um dos avanços mais significativos desta década, e o desenvolvimento de assinaturas gênicas e a identificação de alvos moleculares são objetivos recorrentes de inúmeros estudos (GÖNEN 2009). Atualmente, o maior desafio da metodologia de *array* é a validação dos dados identificados em conjunto independente de amostras (NTZANI e IOANNIDIS 2003; RANSOHOFF 2004; GÖNEN 2009).

Assinatura gênica é uma regra que prediz a resposta do paciente, normalmente sobrevida ou progressão, a partir da expressão de um número relativamente pequeno de genes. O desenvolvimento de assinaturas envolve “peneirar” entre milhares de genes e selecionar alguns poucos que irão predizer a resposta (GÖNEN 2009).

A disponibilidade de classificadores capazes de indicar, através de uma biópsia prévia, quais pacientes portadores de CECP responderiam à quimiorradioterapia seria de grande importância e interesse clínico. A possibilidade de discriminação prévia dos pacientes que se beneficiariam do tratamento conservador de preservação de órgão e dos pacientes que se beneficiariam da cirurgia é bastante importante, uma vez que seria evitado tratamento clínico ineficaz para pacientes NR, pois estes seriam encaminhados de imediato à cirurgia, evitando a progressão tumoral local.

Como resultado da utilização dos classificadores, ocorreria melhora na qualidade de vida advindo da preservação de órgão nos R; toxicidade desnecessária

seria evitada e ocorreriam reduções no custo do tratamento dos NR; e ainda poderia ocorrer aumento na sobrevida global.

Existem inúmeros trabalhos que identificaram marcadores moleculares em CECP, através da análise do perfil de expressão gênica pela técnica de *array*. Apesar dos resultados serem bastante instigantes, nenhum desses estudos foram suficientemente validados para permitir sua utilização na rotina clínica. Dessa forma, o valor preditivo dos marcadores moleculares individuais ou combinados permanece insuficiente (SINGH e PFISTER 2008).

Em nosso projeto prévio de Mestrado, foi identificado e validado matematicamente conjuntos de genes cujos níveis de expressão combinados podem ser utilizados como classificadores moleculares. Esses classificadores, 4 trios (Tabela 2) e 27 quadras de genes (Tabela 3), são capazes de agrupar com 100% de precisão pacientes R e NR ao tratamento clínico, realizado com quimioterapia concomitante à radioterapia, em CE de laringe e hipofaringe localmente avançado.

A validação independente é imprescindível para implementação dos classificadores no uso clínico, pois é considerada a forma mais robusta e confiável de validação de dados gerados por *array*, devido à grande possibilidade de ocorrência de *overfitting* nesses experimentos (NTZANI e IOANNIDIS 2003; RANSOHOFF 2004; GÖNEN 2009).

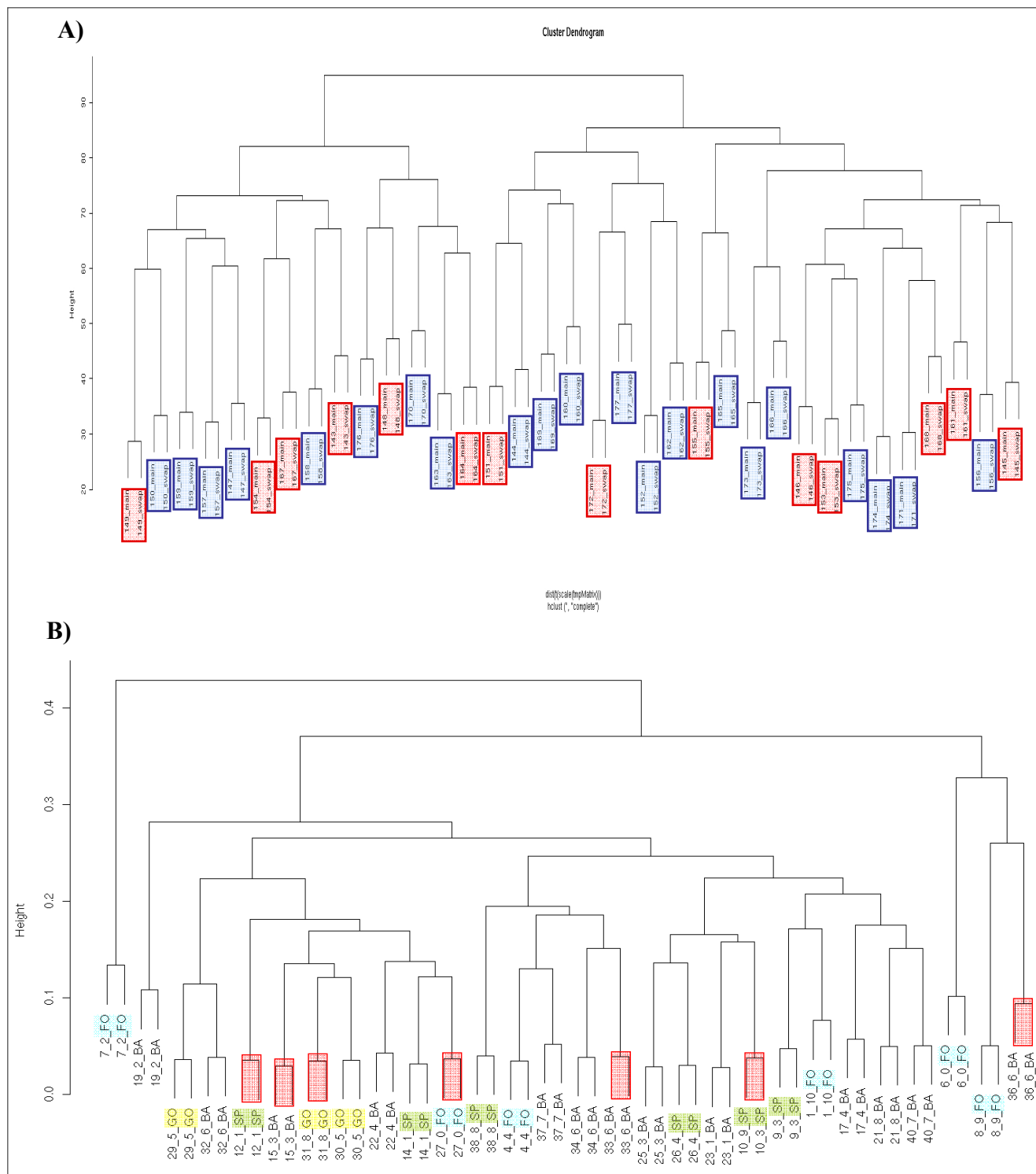
Assim, é de grande valia a identificação de classificadores de resposta ao tratamento clínico em CECP, bem como sua validação independente. Esse tipo de validação elimina a chance de identificação de padrão discriminatório devido ao acaso; e, conseqüentemente, eles podem ser introduzidos na rotina clínica; após validação em estudo prospectivo, com coorte maior.

Para tanto, foi determinado o perfil de expressão de 4.600 genes presentes em lâmina de *array*, de 28 amostras de biópsia coletadas previamente ao início do tratamento quimiorradioterápico.

Como demonstrado pela análise de agrupamento, ou cluster hierárquico (Figura 9 e 16B), o perfil de expressão global dos 4600 genes presentes na lâmina de *array* não são capazes de separar os pacientes dos diferentes grupos, R e NR ao tratamento clínico recebido.

Ou seja, não existem diferenças suficientes capazes de separar ambos os grupos. O mesmo dado foi demonstrado no conjunto das 35 amostras utilizadas no estudo prévio de Mestrado, sobre identificação dos classificadores (Figura 16A).

Os genes mais diferencialmente expressos (Tabela 15) entre os grupos R e NR também foram analisados, e nenhum apresentou p-valor significativo. Os genes mais diferencialmente expressos também não foram capazes de separar as amostras R e NR do grupo da identificação dos classificadores. Esses dados são novamente resultados da homogeneidade genética dos pacientes de CE de laringe e hipofaringe, que se mostrou bastante grande.



Legenda: *Cluster* hierárquico não supervisionado das 35 amostras, do grupo da identificação dos classificadores, em A; e em B das 28 amostras do grupo da validação. Os pacientes NR estão representados pelos retângulos vermelhos, em A e B; e os pacientes R, representados em azul, em A.

Figura 16 - Comparação dos *clusters* hierárquicos.

Dessa forma, esses dados sugerem que a busca de classificadores de resposta compostos por grupos de genes diferencialmente expressos capazes de definir *clusters* não é a melhor metodologia e ser empregada, como é realizado em muitos trabalhos com dados gerados por *microarray*.

Discriminante Linear de Fisher (HASTIE et al. 2001) é um método utilizado em estatística para identificar combinação linear de traços que melhor separem duas ou mais classes de objetos ou eventos. Esta técnica consegue manter uma boa relação entre facilidade e desempenho, tornando-se uma boa escolha para este tipo de estudo.

Além disso, existe grande vantagem na implementação desta metodologia, pois possibilita a criação de banco de dados, contra o qual uma amostra desconhecida pode ser testada e em seguida classificada, previamente ao início do tratamento, a partir de fragmento de biópsia. Deste modo, se torna possível direcionar cada paciente, individualmente, para o tratamento mais eficaz, de acordo com seu grupo de classificação, R ou NR.

O ideal é que os R realizem quimiorradioterapia, tratamento conservador e convencional, e obtenham a preservação do órgão. Já o ideal para os NR é que sejam encaminhados para cirurgia radical curativa, desta forma são evitadas progressão da doença e conseqüente ressecção mais invasiva, bem como toxicidade desnecessária de fármacos quimioterápicos.

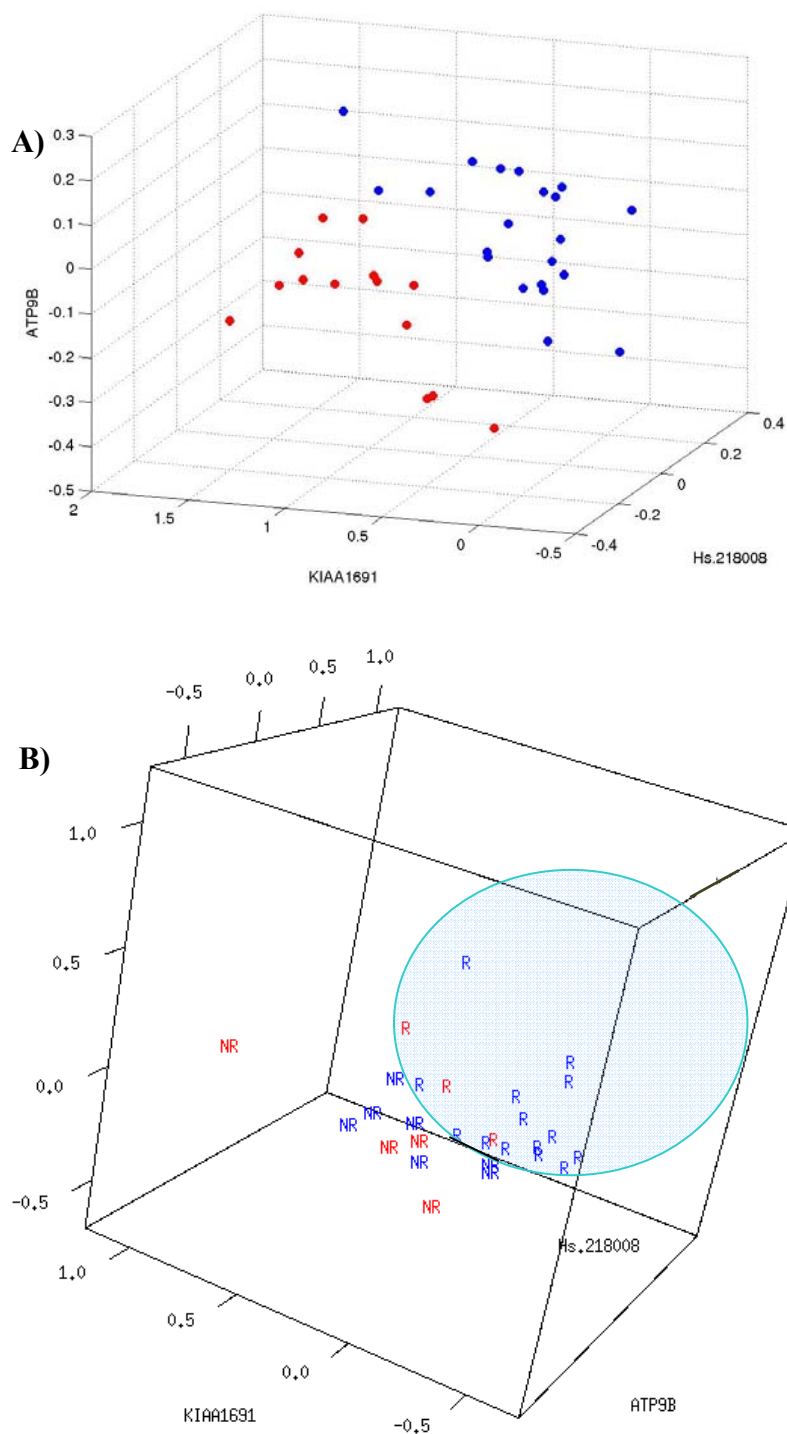
A regra de classificação que anteriormente identificou os 100 melhores trios e as 100 melhores quadras, dentre elas os 4 trios de genes classificadores (Tabela 2) e as 27 quadras (Tabela 3) com 100% de acerto na validação cruzada (VC= 35), foi agora aplicada aos dados gerados, novamente por *microarray*, nas amostras do grupo da validação (Tabelas 16 e 17).

Outra vantagem do método utilizado, Discriminante Linear de Fisher, é que o mesmo permite a utilização de normalizadores distintos, porém, permanecendo a capacidade de classificar. Assim sendo, quando foram utilizados RNAs referência diferentes, as nuvens de dados de ambos os experimentos não ficaram centralizadas, mas a capacidade de separar R e NR permaneceu em cada um dos experimentos isolados. Isto pode ser visualizado através de gráfico tri-dimensional (Figura 17), com a representação do trio de genes do grupo da identificação dos classificadores (Figura 17A) e do grupo da validação independente (Figura 17B). Não é possível visualizar as quadras de genes classificadoras, pois são necessárias 4 dimensões, uma vez que se tem 4 genes, e não mais apenas 3.

Assim, de acordo com o nível de expressão de cada um dos genes que compõe cada classificador, cada amostra isolada, ou paciente, assume determinada localização espacial, podendo ser representado por gráfico 3-D (Figura 17), no caso de trios.

Quando o trio 3 é aplicado ao grupo de dados da identificação de classificadores, existe um plano que separa ambos os grupos (Figura 17A); enquanto que quando o mesmo é aplicado ao grupo da validação isso não ocorre de forma linear, mas existe uma superfície que delimita os mesmos grupos (Figura 17B), R e NR. Essa separação pode ser descrita por uma equação da superfície.

Através da Figura 17B, pode-se visualizar que, o melhor trio validado de genes, o número 3 (Tabela 16), classificou incorretamente 7 amostras clinicamente R em NR (Tabela 18). O mesmo trio também classificou incorretamente 3 amostras clinicamente NR em R (Tabela 18).



Legenda: Aplicação do trio 3 de genes classificadores em dados da identificação, em A; e em B da validação independente. As amostras clinicamente R estão representadas em azul e as NR em vermelho, em A e B. Em B, a superfície não linear azul delimita a classificação da amostras, pelo trio de genes, em R. O restante das amostras são classificadas, pelo trio de genes, em NR.

Figura 17 - Comparação de *plots* 3D de trio de classificador molecular.

Comparando o melhor trio validado com as melhores quadras, o trio 3 de genes apresentou 66,6% de sensibilidade e 57,14% de especificidade; já as melhores quadras de genes, 8 e 9, apresentaram 76% de sensibilidade e 43% de especificidade e as quadras 6, 27 e 47 apresentaram 62% e 71% (Tabela 19), respectivamente.

Sensibilidade mede a taxa de acerto da classificação dos pacientes R, e especificidade a de acerto dos NR. O erro tipo I mede a taxa de erro da classificação dos NR, e o erro tipo II a de erro dos R. Embora as quadras 6, 27 e 47 apresentem sensibilidade e especificidade maiores que 60%, é sempre desejável diminuir o erro tipo II, responsável pela sensibilidade dos dados próximo de 62%, nas melhores quadras de genes.

Deste modo, foi utilizada a estratégia de árvore de decisão (DUDOIT et al. 2002a), técnica que consiste em agregar regras de classificação diversas para melhorar o desempenho do classificador. Assim, são utilizados mais de um classificador, de modo a classificar melhor do que utilizando um classificador isolado. Na busca pela identificação das árvores foram utilizados os 4 trios de genes e as 27 quadras classificadoras, com VC=35 (Tabelas 16 e 17).

As árvores foram criadas sob a perspectiva de otimizar a classificação dos pacientes NR, já que a classificação incorreta de um NR em R acarretaria maior prejuízo do que o inverso, uma vez que o NR receberia um tratamento ineficaz e a doença avançaria. Por fim, este NR seria encaminhado da mesma forma para a cirurgia, só que a doença local estaria em estágio avançado e ainda apresentaria maior chance de desenvolvimento de metástase a distância, o que, provavelmente, diminuiria a sobrevida global do NR.

Após as análises de árvores (Tabela 22), as A (Figura 11) e B (Figura 12) foram as melhores identificadas. A árvore B apresentou taxa de 66% de especificidade e 100% de sensibilidade, enquanto que a árvore A taxas de 90,4% e 71,4% (Tabela 19), respectivamente.

Tabela 22 - Acertos e erros das árvores A e B de decisão. Visualização do número de amostras classificadas em cada grupo.

	Árvore A		Árvore B	
Acerto	19 R	5 NR	14 R	7 NR
Erro	2 R	2 NR	7 R	0 NR

Embora a especificidade da árvore A não seja de 100% como na árvore B (Tabela 19), a A apresenta valor de erro tipo II razoavelmente menor que B, uma vez que sua sensibilidade é de 90,4%, maior que da B, que é de apenas 66%. Além disso, a árvore A apresenta erro tipo I pequeno, pois sua especificidade é de 71,4%.

A árvore A (Figura 14 e Tabela 21) erra a classificação de 2 amostras R e 2 NR; mas, o peso do erro das NR é maior, pois são 2 NR em 7 amostras; contra 2 R em 21 amostras. Dessa forma, provavelmente a especificidade da árvore A está sendo prejudicada pelo menor número de amostras NR no grupo. Assim, muito provavelmente o desempenho real da árvore A tende a melhorar, com a diminuição do desbalanceamento dos grupos de amostras. Como resultado, a árvore A composta pelas quadras 13, 6 e 47 (Figura 14 e Tabela 19), pode ser considerada o melhor classificador preditor de resposta identificado e validado deste estudo.

Uma vez que os classificadores testados apresentaram altas taxas de sensibilidade e especificidade, pode-se concluir que ocorreu a validação dos mesmos.

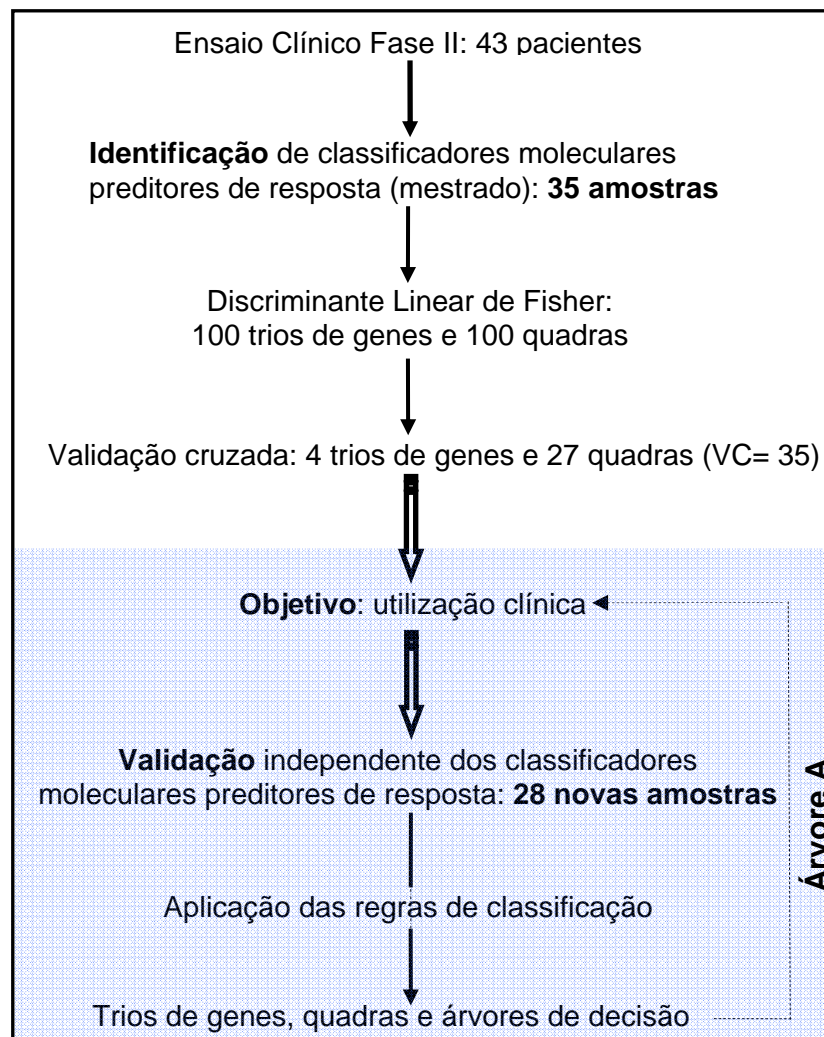
Além disso, as amostras são provenientes de 4 instituições e de 3 estados diferentes (Figura 9), o que torna o grupo mais heterogêneo e a validação realizada mais robusta.

Os valores de especificidade e sensibilidade, identificados no estudo de validação independente, demonstram que a utilização desses classificadores apresenta vantagem quanto ao método de seleção dos pacientes utilizado até o momento. Após a validação independente, é necessária a realização de estudo prospectivo com coorte maior, para que os classificadores possam ser utilizados na rotina clínica.

Outro fator que também torna os classificadores importantes é o fato de nenhuma variável clinicopatológica ser capaz de prever a resposta ao tratamento clínico (Tabela 4). A única variável que apresentou significância estatística (p-valor: 0,032; Apêndice 3) foi metástase em linfonodo regional, N₁; e, provavelmente, foi devido a erro amostral, já que a mesma variável clínica não foi significativa (p-valor: 0,32; Apêndice 3) no grupo das 35 amostras utilizadas na identificação dos classificadores (Tabela 1), em projeto de Mestrado.

Além disso, seria mais coerente em relação à biologia do tumor se pacientes com linfonodo acometido, ou em estágios mais avançados de acometimento, não respondessem ao tratamento; e o que ocorreu neste grupo foi exatamente o oposto, com 100% dos NR sendo clinicamente N0.

O objetivo principal deste estudo está esquematizado na Figura 18, assim como a visão global dos 3 estudos realizados no Hospital AC Camargo e suas conexões.



Legenda: Esquema detalhando o objetivo geral deste projeto, delimitado pelo retângulo azul, e dos 2 outros estudos relacionados, realizados previamente a este.

Figura 18 - Esquema que engloba o objetivo geral do projeto.

Uma vez que genes em uma assinatura predizem resposta, pode parecer natural concluir que eles serão alvos moleculares apropriados para terapia. Mas isso não é necessariamente verdadeiro, pois a forma como assinaturas gênicas são desenvolvidas não prioriza o mecanicismo nem as ligações causais entre genes e resposta. A falácia deste pensamento ocorre devido aos conceitos estatísticos de correlação e causalidade; desta forma, uma assinatura gênica não tem que necessariamente identificar todos (ou mesmo um) elos de mecanismo de uma via. Além disso, sem algum entendimento das vias, se torna impossível a identificação de alvos terapêuticos (GÖNEN 2009).

Assim, até mesmo a assinatura gênica mais criteriosamente validada não irá necessariamente revelar novos alvos moleculares. Isso ocorre devido à natureza do preditor, uma vez que um bom preditor precisa meramente estar correlacionado com resposta. Uma ligação de causa entra ambos é útil, mas não necessária (GÖNEN 2009).

Enquanto genes de uma assinatura bem validada podem ser pontos de partida para explorações futuras, a identificação de alvos moleculares terapêuticos irá requerer experimentos mais cuidadosos e focados para identificar relações de causa.

Células tumorais podem gerar seus próprios sinais de crescimento, reduzindo a dependência de estímulo do microambiente normal ao adquirir a capacidade de se tornar autosuficiente. Um dos exemplos é a produção de TGFA. A ligação de TGFA a EGFR ativa uma via de sinalização intracelular, levando à inibição da apoptose, ativação de proliferação celular e angiogênese, assim como um aumento no potencial metastático. A expressão de TGFA também está fortemente correlacionada com pior prognóstico em câncer de mama.

A superexpressão de TGFA foi identificada nos grupos dos NR deste estudo (Figura 15), e do estudo prévio responsável pela identificação de classificadores (Tabela 20), por *array*. Além disso, este dado foi ainda corroborado por imunohistoquímica (Tabela 21 e Figura 14). Todos os 3 dados de TGFA identificados apresentaram significância estatística.

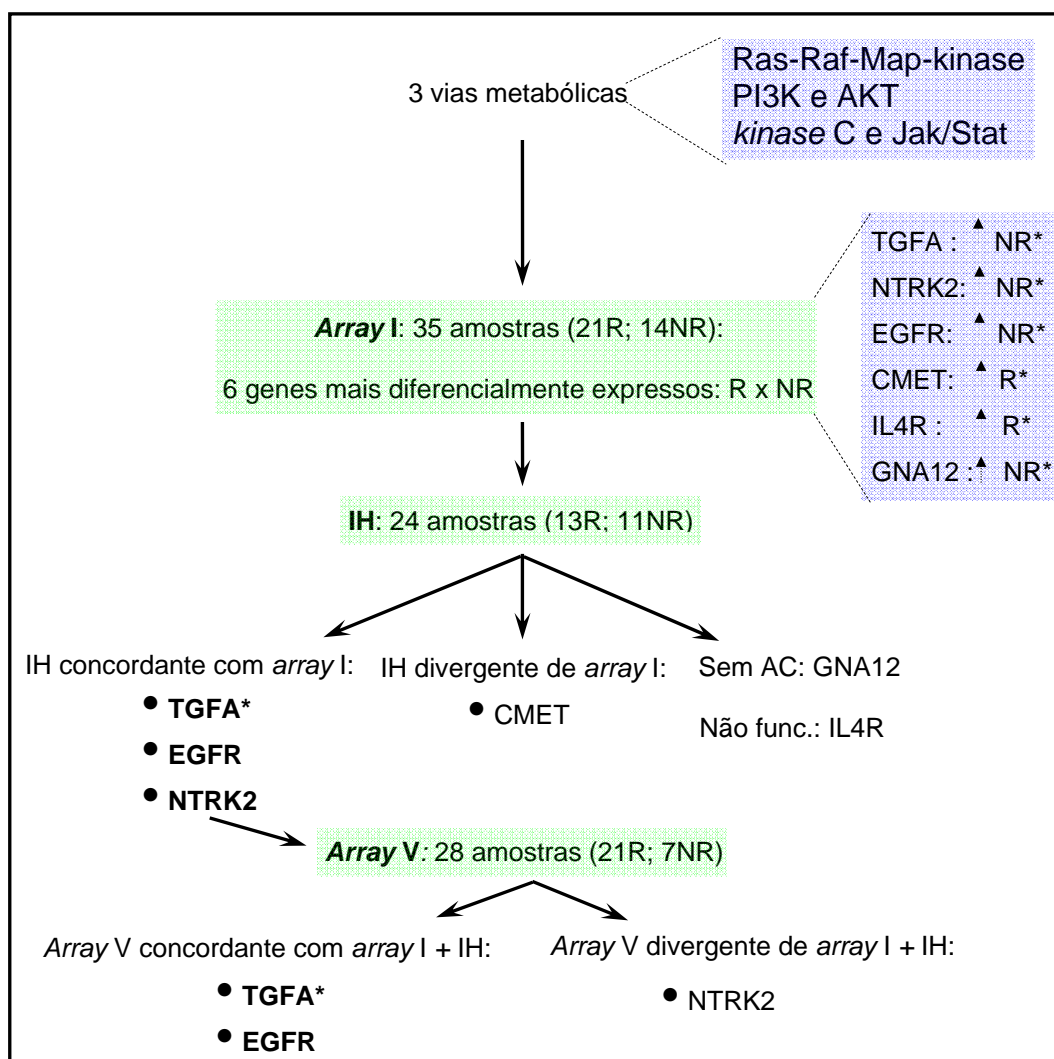
Este gene superexpresso pode ser um dos fatores responsáveis pela maior inibição de apoptose no grupo dos NR, o que pode ser um dos fatores responsáveis por promover maior resistência tumoral ao tratamento clínico. Além disso, o aumento de angiogênese e proliferação celular podem promover maior sobrevivência dessas células tumorais.

O gene EGFR também se mostrou superexpresso nos pacientes NR, em amostras de ambos os estudos, tanto nas análises de *array* (Tabela 20, Figura 13 e 15) como por imunohistoquímica (Tabela 21). Apesar disso, apenas o resultado de diferença de expressão gênica entre os grupos R e NR, das amostras do projeto de Mestrado, apresentou p-valor significativo.

A ativação deste receptor leva ao recrutamento e a fosforilação de inúmeros substratos intracelulares, os quais, por sua vez, ativam vias mitogênicas e outras atividades promotoras de tumor. Além disso, existem estudos que demonstraram que altos níveis de EGFR foram associados com menor resposta tumoral à radioterapia. Assim, a superexpressão de EGFR pode ser um dos fatores responsáveis por promover radiorresistência nos pacientes NR, já que essas células tumorais teriam maior capacidade de se multiplicar.

Como resultado do estudo de genes alvos diferencialmente alterados em vias relacionadas à resposta ao tratamento clínico (Tabela 20), foi possível identificar

alterações que podem estar relacionadas à resistência dos pacientes NR ao tratamento clínico. Toda a metodologia que compõe o objetivo secundário deste projeto está esquematizada na Figura 19.



Legenda: Esquema detalhando o objetivo secundário deste projeto. I significa grupo de amostras da identificação dos classificadores e V da validação. IH é ensaio de imunoistoquímica e o asterisco significa p-valor com significância estatística.

Figura 19 - Esquema global do objetivo secundário do projeto.

Estes genes identificados (Figura 15), principalmente TGFA devido à significância do p-valor, sugerem potenciais candidatos para explorações futuras mais detalhadas e focadas na função gênica, para que seja possível determinar se existe relação de causa entre a superexpressão do gene e a resistência do paciente ao tratamento clínico.

Caso exista relação de causa entre os fatores, e não apenas de correlação, os genes apresentam potencial para serem utilizados como alvos terapêuticos, uma vez que alvo molecular é um traço do microambiente de células tumorais, que contribui com o crescimento maligno e que pode ser externamente modificado com propósitos de tratamento.

7 CONCLUSÃO

O objetivo geral deste projeto foi atingido com a validação dos classificadores de resposta dos pacientes ao tratamento clínico, compostos por trios ou quadras de genes. Também foram identificadas árvores de decisão, que obtiveram maiores taxas de especificidade e sensibilidade, quando comparadas aos trios de genes ou quadras isoladas. A árvore A, a melhor árvore de decisão identificada, apresenta 90,4 % de sensibilidade e 71,4 % de especificidade, e provavelmente sua especificidade aumentaria com o aumento do número de amostras NR e conseqüente diminuição do desbalanceamento entre as amostras R e NR.

Sendo assim, sugerimos a árvore A como o melhor classificador identificado e validado em amostras independentes, para realizar a discriminação dos pacientes R e NR e conseqüente encaminhamento ao tratamento mais adequado. No futuro, é sugerido estudo prospectivo com coorte maior, para que o classificador possa ser utilizado na rotina clínica.

Para a realização do objetivo secundário deste projeto foram selecionados 6 genes, a partir do estudo de 3 vias metabólicas, a fim de identificar alterações gênicas possivelmente relacionadas com a resistência dos pacientes ao tratamento clínico. Após análise de expressão gênica diferencial dos dados gerados por *array* em 2 conjuntos de dados diferentes, entre os pacientes R e NR, e após ensaios de imunistoquímica, pode-se concluir que a superexpressão do gene TGFA no grupo dos NR foi validada e o objetivo secundário alcançado. TGFA é responsável por inibir apoptose, ativar proliferação celular e promover angiogênese, desta forma, sua

superexpressão poderia ser uma das causas, entre outras, da resistência dos pacientes ao tratamento; já que estes tumores teriam maior capacidade de sobrevivência, de formar novos vasos e novas células tumorais.

Estudos futuros focados na função gênica são sugeridos, para determinar se existe relação de causalidade entre a superexpressão do gene e a resistência do paciente ao tratamento clínico. Caso a relação entre os fatores seja de causa, e não apenas de correlação, o gene é um alvo terapêutico em potencial, existindo a possibilidade de ser externamente modificado com propósitos de tratamento

8 REFERÊNCIAS BIBLIOGRÁFICAS

[Anonymus] Induction chemotherapy plus radiation compared with surgery plus radiation in patients with advanced laryngeal cancer. The Department of Veterans Affairs Laryngeal Cancer Study Group. **N Engl J Med** 1991; 324:1685-90.

Akimoto T, Hunter NR, Buchmiller L, et al. Inverse relationship between epidermal growth factor receptor expression and radiocurability of murine carcinomas. **Clin Cancer Res** 1999; 5:2884-90.

Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. **Molecular biology of the cell**. 3rd ed. New York: Garland Publishing; 1994. Câncer; p.1255-91.

Alevizos I, Mahadevappa M, Zhang X, et al. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. **Oncogene** 2001; 20:6196-204.

Allison M. Is personalized medicine finally arriving. **Nat Biotechnol** 2008; 26:509-17.

Alroy I, Yarden Y. The ErbB signaling network in embryogenesis and oncogenesis: signal diversification through combinatorial ligand-receptor interactions. **FEBS Lett** 1997; 410:83-6.

Baker SG, Kramer BS, Srivastava S. Markers for early detection of cancer: statistical guidelines for nested case-control studies. **BMC Med Res Methodol** 2002; 2:4.

Balkwill F, Charles KA, Mantovani A. Smoldering and polarized inflammation in the initiation and promotion of malignant disease. **Cancer Cell** 2005; 7:211-7.

Balm AJ, Rasch CR, Schornagel JH, et al. High-dose superselective intra-arterial cisplatin and concomitant radiation (RADPLAT) for advanced head and neck cancer. **Head Neck** 2004; 26:485-93.

Barbosa JF. **Câncer da faringe**. São Paulo: Editora Manole; 1979. Câncer da hipofaringe; p.157-86.

Barreto BD. **Identificação de assinaturas moleculares preditoras de resposta à quimio e radioterapia combinada em tumores localmete avançados de laringe e hipofaringe**. São Paulo; 2005. [Dissertação de Mestrado-Fundação Antônio Prudente].

Bataini P, Brugere J, Bernier J, Jaulerry CH, Picot C, Ghossein NA. Results of radical radiotherapeutic treatment of carcinoma of the pyriform sinus: experience of the Institut Curie. **Int J Radiat Oncol Biol Phys** 1982; 8:1277-86.

Benjamin B, Bingham B, Hawke M, Stammberger H. **Atlas colorido de otorrinolaringologia**. São Paulo: Artes Médicas; 1996. A laringe; p.223-305.

Bergers G, Benjamin LE. Angiogenesis: tumorigenesis and the angiogenic switch. **Nat Rev Cancer** 2003; 3:401-10.

Bierie B, Stover DG, Abel TW, et al. Transforming growth factor-beta regulates mammary carcinoma cell survival and interaction with the adjacent microenvironment. **Cancer Res** 2008; 68:1809-19.

Bierie B, Chung CH, Parker JS, et al. Abrogation of TGF-beta signaling enhances chemokine production and correlates with prognosis in human breast cancer. **J Clin Invest** 2009; 119:1571-82.

Bissell MJ, Radisky D. Putting tumours in context. **Nat Rev Cancer** 2001; 1:46-54.

Blume-Jensen P, Hunter T. Oncogenetic kinase signaling. **Nature** 2001; 411:355-65.

Boffetta P, Hecht S, Gray N, et al. Smokeless tobacco and cancer. **Lancet Oncol** 2008; 9:667–75.

Braakhuis BJ, Senft A, de Bree R, et al. Expression profiling and prediction of distant metastases in head and neck squamous cell carcinoma. **J Clin Pathol** 2006; 59:1254-60.

Bray F, Moller B. Predicting the future burden of cancer. **Nat Rev Cancer** 2006; 6:63-74.

Byers RM, Wolf PF, Ballantyne AJ. Rationale for elective modified neck dissection. **Head Neck Surg** 1988; 10:160-7.

Carter SL, Eklund AC, Kohane IS, et al. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. **Nat Genet** 2006; 38:1043-8.

Chan TO, Rittenhouse SE, Tschlis PN. AKT/PKB and other D3 phosphoinositide-regulated kinases: kinase activation by phosphoinositide-dependent phosphorylation. **Annu Rev Biochem** 1999; 68:965-1014.

Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. **Proc Natl Acad Sci U S A** 2005; 102:3738-43.

Chang JC, Wooten EC, Tsimelzon A, et al. Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. **Lancet** 2003; 362:362-9.

Chi JT, Wang Z, Nuyten DS, et al. Gene expression programs in response to hypoxia: Cell type specificity and prognostic significance in human cancers. **PLoS Med** 2006; 3:e47.

Choi P, Chen C. Genetic expression profiles and biologic pathway alterations in head and neck squamous cell carcinoma. **Cancer** 2005; 104:1113-28. Erratum in: **Cancer** 2005; 104:2532.

Chung CH, Parker JS, Karaca G, et al. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. **Cancer Cell** 2004; 5:489-500.

Chung CH, Parker JS, Ely K, et al. Gene expression profiles identify epithelial-to-mesenchymal transition and activation of nuclear factor- κ B signaling as characteristics of a high risk head and neck squamous cell carcinoma. **Cancer Res** 2006; 66:8210-8.

Condeelis J, Pollard JW. Macrophages: obligate partners for tumor cell migration, invasion, and metastasis. **Cell** 2006; 124:263-6.

Cotran RS, Kumar V, Robbins SL. **Robbins pathologic basis of disease**. 6th ed. Philadelphia: W. B. Saunders; 1999. Neoplasia; p.260-327.

Coussens LM, Werb Z. Inflammation and cancer. **Nature** 2002; 420:860-7.

Cristo EB. **Métodos estatísticos na análise de experimentos de *microarray***. São Paulo; 2003. [Dissertação de Mestrado-Instituto de Matemática e Estatística-USP].

D'Souza G, Kreimer AR, Viscidi R, et al. Case-control study of human papillomavirus and oropharyngeal cancer. **N Engl J Med** 2007; 356:1944-56.

de Visser KE, Eichten A, Coussens, LM. Paradoxical roles of the immune system during cancer development. **Nat Rev Cancer** 2006; 6:24-37.

DeNardo DG, Johansson M, Coussens LM. Immune cells as mediators of solid tumor metastasis. **Cancer Metastasis Rev** 2008; 27:11-8.

DeRisi J, Penland L, Brown PO, et al. Use of a cDNA microarray to analyse gene expression patterns in human cancer. **Nat Genet** 1996; 14:457-60.

Dubben HH, Thames HD, Beck-Bornholdt HP. Tumor volume: a basic and specific response predictor in radiotherapy. **Radiother Oncol** 1998; 47:167-74.

Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. **J Natl Cancer Inst** 2007; 99:147-57.

Douma S, VanLaar T, Zevenhoven J, Meuwissen R, VanGarderen E, Peeper DS. Suppression of anoikis and induction of metastasis by the neurotrophic receptor TrkB. **Nature** 2004; 430:1034-9.

Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. **J Am Stat Assoc** 2002a; 457:77-87.

Dudoit S, Yang YH, Speed TP, Callow MJ. Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. **Statistica Sinica** 2002b; 12:111-39.

Dumont N, Tlsty TD. Reflections on miR-ing effects in metastasis. **Cancer Cell** 2009; 16:3-4.

Ernfors P, Lee KF, Jaenisch R. Mice lacking brain-derived neurotrophic factor develop with sensory deficits. **Nature** 1994; 368:147-50.

Esteves, Gustavo Henrique. **Métodos estatísticos para a análise de dados de cDNA microarray em um ambiente computacional integrado**. São Paulo; 2007. [Tese de Doutorado-Instituto de Matemática e Estatística-USP].

Etzioni R, Urban N, Ramsey S, et al. The case for early detection. **Nat Rev Cancer** 2003; 3:243-52.

Farinati F, Zhou Z, Bellah J, Lieber CS, Garro AJ. Effect of chronic ethanol consumption on activation of nitrosopyrrolidine to a mutagen by rat upper alimentary tract lung and hepatic tissue. **Drug Metab Dispos** 1985; 13:210-4.

Feher O, Martins SJ, Lima CA, Salvajoli JV, Simpson AJ, Kowalski LP. Pilot trial of concomitant chemotherapy with paclitaxel and split-course radiotherapy for very advanced squamous cell carcinoma of head and neck. **Head Neck** 2002a; 24:228-35.

Feher O, Valadares AD, Nicolau UR, et al. Paclitaxel and cisplatin concurrent with radiotherapy RT for larynx preservation in advanced resectable laryngeal and hypopharyngeal squamous cell carcinomas SCC: preliminary results of a phase II study. **Proceedings ASCO** 2002b; 21:A962.

Feldman AL, Costouros NG, Wang E, et al. Advantages of mRNA amplification for microarray analysis. **Biotechniques** 2002; 33:906-14.

Field JK, Malliri A, Butt SA. p53 overexpression in end stage squamous cell carcinoma of the head and neck: correlates with a very poor clinical out-come. **Int J Oncol** 1993a; 3:431-5.

Field JK, Pavelic ZP, Spandidos DA, Stambrook PJ, Jones AS, Gluckman JL. The role of the p53 tumor suppressor gene in squamous cell carcinoma of head and neck. **Arch Otolaryngol Head Neck Surg** 1993b; 119:1118-22.

Fonseca CA, Otto SS, Paumgartten FJ, Leitão AC. Non-toxic mutagenic and clastogenic activities of Mate-Chimarrão (*Ilex Paraguariensis*). **J Environ Pathol Toxicol Oncol** 2000; 19:333-46.

Forastiere AA, Leong T, Rowinsky E, et al. Phase III comparison of high-dose paclitaxel + cisplatin + granulocyte colony-stimulating factor versus low-dose paclitaxel + cisplatin in advanced head and neck cancer: Eastern Cooperative Oncology Group Study E1393. **J Clin Oncol** 2001; 19:1088-95.

Forastiere AA, Goepfert H, Maor M, et al. Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. **N Engl J Med** 2003; 349:2091-8.

Foulds L. The experimental study of tumor progression: a review. **Cancer Res** 1954; 14:327-39.

Frisch SM, Francis H. Disruption of epithelial cell-matrix interactions induces apoptosis. **J Cell Biol** 1994; 124:619-26.

Freeman WM, Robertson DJ, Vrana KE. Fundamentals of DNA hybridization arrays for gene expression analysis. **Biotechniques** 2000; 29:1042-55.

Garden AS, Harris J, Vokes EE, et al. results of Radiation Therapy Oncology Group 97-03: a randomized phase ii trial of concurrent radiation and chemotherapy for advanced squamous cell carcinomas of the head and neck. **J Clin Oncol** 2004; 22:2856-64.

Garro AJ, Seitz HK, Lieber CS. Enhancement of dimethylnitrosamine metabolism and activation to a mutagen following chronic ethanol consumption. **Cancer Res** 1981; 41:120-4.

Geiger TR, Peeper DS. Critical role for TrkB kinase function in anoikis suppression, tumorigenesis, and metastasis. **Cancer Res** 2007; 67:6221-9.

Gillison ML, Koch WM, Capone RB, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. **J Natl Cancer Inst** 2000; 92:709-20.

Gilman AG, Ruddon RW, Molinoff PB, Limbird LE, Hardman JG. Antineoplastic agents. In: Chabner B, Allegra C, Curt G, Calabresi P, editors. **Goodman e Gilman's: the pharmacological basis of therapeutics**. New York: McGraw-Hill; 1996. p.1233-87.

Ginsburg GS, McCarthy JJ. Personalized medicine: revolutionizing drug discovery and patient care. **Trends Biotechnol** 2001; 19:491-6.

Glinsky GV, Berezovska O, Glinskii AB. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. **J Clin Invest** 2005; 115:1503-21.

Goldenberg D, Golz A, Joachims HZ. The beverage maté: a risk factor for cancer of the head and neck. **Head Neck** 2003; 25:595-601.

Gomes LI, Silva RL, Stolf BS, et al. Comparative analysis of amplified and nonamplified RNA for hybridization in cDNA microarray. **Anal Biochem** 2003; 321:244-51.

Gönen M. Statistical aspects of gene signatures and molecular targets. **Gastrointest Cancer Res** 2009; 3:S19-21.

Goswami S, Sahai E, Wyckoff JB, et al. Macrophages promote the invasion of breast carcinoma cells via a colony-stimulating factor-1/epidermal growth factor paracrine loop. **Cancer Res** 2005; 65:5278-83.

Grandis JR, Chakraborty A, Zeng Q, Melhem MF, Tweardy DJ. Down modulation of TGF-alpha protein expression with antisense oligonucleotides inhibits proliferation of head and neck squamous carcinoma but not normal mucosal epithelial cells. **J Cell Biochem** 1998; 69:55-62.

Hanahan D, Weinberg RA. The hallmarks of cancer. **Cell** 2000; 100:57-70.

Hanna E, Shrieve DC, Ratanatharathorn V, et al. A novel alternative approach for prediction of radiation response of squamous cell carcinoma of head and neck. **Cancer Res** 2001; 61:2376-80.

Hanna E, Sherman A, Cash D, et al. Quality of life for patients following total laryngectomy vs chemoradiation for laryngeal preservation. **Arch Otolaryngol Head Neck Surg** 2004; 130:875-9.

Hastie T, Tibshirani R, Friedman J. **The elements of statistical learning**. New York: Springer-Verlag; 2001.

Hepper GH. Tumor heterogeneity. **Cancer Res** 1984; 44:2259-65.

Hoffmann D, Melkian A, Adams JD, Brunnemann ND, Haley ND. New aspects of tobacco carcinogenesis. **Carcinogenesis** 1985; 8:239-56.

Hsu TC. Alcohol: a cocarcinogen in head and neck malignancies. **Adv Exp Med Biol** 1992; 320:35-7.

Huang E, Cheng SH, Dressman H, et al. Gene expression predictors of breast cancer outcomes. **Lancet** 2003; 361:1590-6.

Hubank M, Schatz DG. Identifying differences in mRNA expression by representational difference analysis of cDNA. **Nucleic Acids Res** 1994; 22:5640-8.

Hurst DR, Edmonds MD, Welch DR. Metastamir: the field of metastasis-regulatory microRNA is spreading. **Cancer Res** 2009; 69:7495-8.

Jain KK. Personalized medicine. **Curr Opin Mol Ther** 2002; 4:548-58

Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2009. **CA Cancer J Clin** 2009; 59:225-49.

Jones KR., Lodge-Rigal RD, Reddick RL, Tudor GE, Shockley WW. Prognostic factors in the recurrence of stage I and II squamous cell cancer of the oral cavity. **Arch Otolaryngol Head Neck Surg.** 1992; 118:483-5.

Jones KR, Farinas I, Backus C, Reichardt LF. Targeted disruption of the BDNF gene perturbs brain and sensory neuron development but not motor neuron development. **Cell** 1994; 76:989-99.

Kearns CM, Gianni L, Egorim MJ. Paclitaxel pharmacokinetics and pharmacodynamics. **Semin Oncol** 1995; 22:16-23.

Kenny PA, Bissell, MJ. Targeting TACE-dependent EGFR ligand shedding in breast cancer. **J Clin Invest** 2007; 117:337-45.

Kissin B, Kaley MM. Alcohol and cancer. In: Kissin B, Begleiter H, editors. **The biology of alcoholism**. New York: Plenum Press; 1974. p.481-511.

Klapper L, Waterman H, Sela M, et al. Tumor-inhibitory antibodies to HER-2/ErbB-2 may act by recruiting c-Cbl and enhancing ubiquitination of HER-2. **Cancer Res** 2000; 60:3384-8.

Klein R, Smeyne RJ, Wurst W, et al. Targeted disruption of the *trkB* neurotrophin receptor gene results in nervous system lesions and neonatal death. **Cell** 1993; 75:113-22.

Kowalski LP, Batista MB, Santos CR, Scopel A, Salvajoli JV, Torloni H. Prognostic factors in T3,N0-1 glottic and transglottic carcinoma. a multifactorial study of 221 cases treated by surgery or radiotherapy. **Arch Otolaryngol Head Neck Surg** 1996; 122:77-82.

Kreimer AR, Clifford GM, Boyle P, et al. Human papillomavirus types in head and neck squamous cell carcinomas worldwide: a systematic review. **Cancer Epidemiol Biomarkers Prev** 2005; 14:467-75.

Kutler DI, Wreesmann VB, Goberdhan A, et al. Human papillomavirus DNA and p53 polymorphisms in squamous cell carcinomas from Fanconi anemia patients. **J Natl Cancer Inst** 2003; 95:1718-21.

Lakhani SR, Ashworth A. Microarray and histopathological analysis of tumours: the future and the past? **Nat Rev Cancer** 2001; 1:151-7.

Lakhani SR, O'Hare MJ, Asworth A. Profiling familial breast cancer. **Nat Med** 2001; 7:408-10.

Lee SW, Tomasetto C, Sanger R. Positive selection of candidate tumor-suppressor genes by subtractive hybridization. **Proc Natl Acad Sci U S A** 1991; 88:2825-9.

Leemans CR, Tiwari R, Nauta JJ, van der Waal I, Snow GB. Regional lymph node involvement and its significance in the development of distant metastases in head and neck carcinoma. **Cancer** 1993; 71:452-6.

Lefebvre JL, Chevalier D, Luboinski B, Kirkpatrick A, Collette L, Sahmoud T. Larynx preservation in pyriform sinus cancer: Preliminary results of a European Organization for Research and Treatment of Cancer phase III trial: EORTC Head and Neck Cancer Cooperative Group. **J Natl Cancer Inst** 1996; 88:890-9.

Lennon GG, Lehrach H. Hybridization analyses of arrayed cDNA libraries. **Trends Genet** 1991; 7:314-7.

Lewis TS, Shapiro PS, Ahn NG. Signal transduction through MAP kinase cascades. **Adv Cancer Res** 1998; 74:49-139.

Liang P, Pardee AB. Differential display of eukariotic messenger RNA by means of the polymerase chain reaction. **Science** 1992; 257:967-70.

Lindberg R. Distribution of cervical lymph node metastases from squamous cell carcinoma of the upper respiratory and digestive tracts. **Cancer** 1972; 29:1146-9.

Liotta L, Petricoin E. Molecular profiling of human cancer. **Nat Rev Genet** 2000; 1:48-56.

Liotta LA, Kohn EC. The microenvironment of the tumour-host interface. **Nature** 2001; 411:375-9.

Little PF, Williams RB, Wilkins MR. Inter-individual variation in expression: a missing link in biomarker biology. **Trends Biotechnol** 2008; 27:5-10.

Liu AY, Roudier MP, True LD. Heterogeneity in primary and metastatic prostate cancer as defined by cell surface CD profile. **Am J Pathol** 2004; 165:1543-56.

Löning T, Ikenberg H, Becker J, Gissman L, Hoepfer I, zur Hausen H. Analysis of oral papillomas leukoplakias and invasive carcinomas for human papillomavirus type related DNA. **Invest Dermatol** 1985; 84:417-20.

Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. **Nat Rev Cancer** 2005; 5:845-56.

Macoska JA. The progressing clinical utility of DNA microarrays. **CA Cancer J Clin** 2002; 52:50-9.

Maier H, Born IA, Veith S, Adler D, Seitz HK. The effect of chronic ethanol consumption on salivary gland morphology and function in the rat. **Alcoholism Clin Exp Res** 1986; 10:420-7.

Maier H, Born IA, Mall G. Effect of chronic ethanol and nicotine consumption on the function and morphology of the salivary glands. **Klin Wochenschr** 1988; 66:140-50.

Maier H, Dietz A, Gewelke U, Heller WD, Weidauer H. Tobacco and alcohol and the risk of head and neck cancer. **Clin Invest** 1992; 70:320-7.

Marx J. DNA arrays reveal cancer in its many forms. **Science** 2000; 289:1670-2.

Mancuso AA, Mukherji SK, Schmalfluss I, et al. Preradiotherapy computed tomography as a predictor of local control in supraglottic carcinoma. **J Clin Oncol** 1999; 17:631-37.

Mathé EA, Nguyen GH, Bowman ED, et al. MicroRNA expression in squamous cell carcinoma and adenocarcinoma of the esophagus: associations with survival. **Clin Cancer Res** 2009; 15:6192-200.

Matos de Sá G. Tratamento das cadeias linfáticas no câncer da laringe. In: Noronha MJR, Dias FL, editores. **Câncer da laringe: uma abordagem multidisciplinar**. Rio de Janeiro: Revinter; 1997. p.227-32.

Maula SM, Luukkaa M, Grenman R, Jackson D, Jalkanen S, Ristamaki R. Intratumoral lymphatics are essential for the metastatic spread and prognosis in squamous cell carcinomas of the head and neck region. **Cancer Res** 2003; 63:1920-6.

Meireles SI, Carvalho AF, Hirata R, Jr, et al. Differentially expressed genes in gastric tumors identified by cDNA array. **Cancer Lett** 2003; 190:199-211.

Meireles SI, Cristo EB, Carvalho AF, et al. Molecular classifiers for gastric cancer and nonmalignant diseases of the gastric mucosa. **Cancer Res** 2004; 64:1255-65.

Mendelsohn J, Baselga L. Status of epidermal growth factor receptor antagonists in the biology and treatment of cancer. **J Clin Oncol** 2003; 21:2787-99.

Mendenhall WM, Parsons JT, Devine JW, Cassisi NJ, Million RR. Squamous cell carcinoma of the pyriform sinus treated with surgery and/or radiotherapy. **Head Neck Surg** 1987; 10:88-92.

Meredith JE, Fazeli B, Schwartz MA. The extracellular matrix as a cell survival factor. **Mol Biol Cell** 1993; 4:953-61.

Milas L, Fan Z, Andratschke NH, Ang KK. Epidermal growth factor receptor and tumor response to radiation: in vivo preclinical studies. **Int J Radiat Oncol Biol Phys** 2004; 58:966-71.

Ministério da Saúde. Instituto Nacional de Câncer. **Estimativa/2008 incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2007.

Ministério da Saúde. Instituto Nacional de Câncer. **Câncer de laringe**. Disponível em: <URL:http://www.inca.gov.br/conteudo_view.asp?id=332> [2009 jun 12]

Mueller MM, Fusenig NE. Friends or foes: bipolar effects of the tumour stroma in cancer. **Nat Rev Cancer** 2004; 4:839-49.

Mufti SI, Salvagnini M, Lieber CS, Garro AJ. Chronic ethanol consumption inhibits repair of dimethylnitrosamine-induced DNA alkylation. **Biochem Biophys Res Commun** 1988; 152:423-31.

Muir C, Weiland L. Upper aerodigestive tract cancers. **Cancer** 1995; 75:147-53.

[NCCN] National Comprehensive Cancer Network Practice Guidelines. **Head and neck cancer** (v.1. 2007). http://www.nccn.org/professionals/physician_gls/pdf/head-and-neck.pdf> [2009 set 12]

Negm RS, Verma M, Srivastava S. The promise of biomarkers in cancer screening and detection. **Trends Mol Med** 2002; 8:288-93.

Nilsson R, Björkegren J, Tegnér J. On reliable discovery of molecular signatures. **BMC Bioinformatics** 2009; 29:10-38.

Nowell PC. The clonal evolution of tumor cell populations. **Science** 1976; 194:23-8.

Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. **Lancet** 2003; 362:1439-44.

O'Donnell RK, Kupferman M, Wei SJ, et al. Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. **Oncogene** 2005; 24:1244-51.

Obermueller E, Vosseler S, Fusenig NE, Mueller MM. Cooperative autocrine and paracrine functions of granulocyte colony-stimulating factor and granulocyte-macrophage colony-stimulating factor in the progression of skin carcinoma cells. **Cancer Res** 2004; 64:7801-12.

Olayioye M, Neve R, Lane H, et al. the ErbB signaling network: Receptor heterodimerization in development and cancer. **EMBO J** 2000; 19:3159-67.

Oliveira CAB, Valle HA, Monteiro GBM, et al. Patologia das neoplasias da laringe. In: Noronha MJR, Dias FL, editores. **Câncer da laringe: uma abordagem multidisciplinar**. Rio de Janeiro: Revinter; 1997. p.8-14.

Pantel K, Brakenhoff RH. Dissecting the metastatic cascade. **Nat Rev Cancer** 2004; 4:448-56.

Parkin DM, Pisani P, Ferlay J. Global cancer statistics. **CA Cancer J Clin** 1990; 49:33-64.

Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. **Proc Natl Acad Sci U S A** 1994; 91:5022-6.

Perou CM, Jeffrey SS, van de Rijn M, Rees CA, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. **Proc Natl Acad Sci U S A** 1999; 96:9212-7.

Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. **Nature** 2000; 406:747-52.

Petricoin EF, Zoon KC, Kohn EC, Barrett JC, Liotta LA. Clinical proteomics: translating benchside promise into bedside reality. **Nat Rev Drug Discov** 2002; 1:683-95.

Pfister DG, Laurie SA, Weinstein GS, et al. American Society of Clinical Oncology clinical practice guideline for the use of larynx-preservation strategies in the treatment of laryngeal cancer. **J Clin Oncol** 2006; 24:3693-704.

Phan JH, Moffitt RA, Stokes TH, Liu J, Young AN, Nie S, Wang MD. Convergence of biomarkers, bioinformatics and nanotechnology for individualized cancer treatment. **Trends Biotechnol** 2009; 27:350-8.

Pignon JP, Bourhis J, Domenge C, Designe L. Chemotherapy added to locoregional treatment for head and neck squamous-cell carcinoma: Three meta-analyses of updated individual data: MACH-NC Collaborative Group: Meta-Analysis of Chemotherapy on Head and Neck Cancer. **Lancet** 2000; 355:949-55.

Pollack JR. Parallel analysis of gene copy number and expression using cDNA microarrays. **Methods Mol Biol** 2003; 224:89-97.

Pramana J, Van den Brekel MW, van Velthuisen ML, et al. Gene expression profiling to predict outcome after chemoradiation in head and neck cancer. **Int J Radiat Oncol Biol Phys** 2007; 69:1544-52.

Quackenbush J. Computational analysis of microarray data. **Nat Rev Genet** 2001; 2:418-27.

Ramaswamy S, Golub TR. DNA microarrays in clinical oncology. **J Clin Oncol** 2002; 20:1932-41.

Ransohoff DF. Developing molecular biomarkers for cancer. **Science** 2003; 299:1679-80.

Ransohoff DF. Rules of evidence for cancer molecular-marker discovery and validation. **Nat Rev Cancer** 2004; 4:309-14.

Raybaud-Diogenè H, Fortin A, Morency R, Roy J, Monteil RA, Têtu B. Markers of radioresistance in squamous cell carcinomas of the head and neck: a clinicopathologic and immunohistochemical study. **J Clin Oncol** 1997; 15:1030-8.

Reid BC, Winn DM, Morse DE, Pendrys DG. Head and neck in situ carcinoma: incidence, trends, and survival. **Oral Oncol** 2000; 36:414-20.

Roepman P, Wessels LF, Kettelarij N, et al. An expression profile for diagnosis of lymph node metastases from primary head and neck squamous cell carcinomas. **Nat Genet** 2005; 37:182-6.

Roepman P, de Jager A, Groot Koerkamp MJ, Kummer JA, Slootweg PJ, Holstege FC. Maintenance of head and neck tumor gene expression profiles upon lymph node metastasis. **Cancer Res** 2006a; 66:11110-4.

Roepman P, de Koning E, van Leenen D, et al. Dissection of a metastatic gene expression signature into distinct components. **Genome Biol** 2006b; 7:R117.

Rose WC. Taxol: a review of its preclinical in vivo antitumor activity. **Anticancer Drugs** 1992; 3:311-21.

Rosenberg B. Platinum coordination complexes in cancer chemotherapy. **Naturwissenschaften** 1973; 60:399-406.

Roskoski R, Jr. The ErbB/HER receptor protein-tyrosine kinases and cancer. **Biochem Biophys Res Commun** 2004; 319:1-11.

Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. **Nat Genet** 2000; 24:227-35.

Rowinsky EK, Onetto N, Canetta RM, Arbuck SG. Taxol: the first of the taxanes an important new class of antitumor agents. **Semin Oncol** 1992; 19:646-62.

Rowinsky EK, McGuire WP, Donehower RC. The current status of Taxol. **Print Pract Gynecol Oncol Updates** 1993; 11:1-16.

Rowinsky EK, Donehower RC. Paclitaxel taxol. **N Engl J Med** 1995; 332:1004-14.

Santos CR, Kowalski LP, Magrin J, et al. Prognostic factors in supraglottic carcinoma patients treated by surgery or radiotherapy. **Ann Otol Rhinol Laryngol** 1998; 107:697-702.

Schena M, Shalon D, Davis R.W, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. **Science** 1995; 270:467-70.

Schroeder A, Mueller O, Stocker S, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. **BMC Mol Biol** 2006; 31;7:3.

Seiwert TY, Jagadeeswaran R, Faoro L, et al. The MET receptor tyrosine kinase is a potential novel therapeutic target for head and neck squamous cell carcinoma. **Cancer Res** 2009; 69:3021-31.

Shiga H, Heath EI, Rasmussen AA, et al. Prognostic value of p53 glutathione S-transferase π and thymidylate synthase for neoadjuvant cisplatin-based chemotherapy in head and neck cancer. **Clin Cancer Res** 1999; 5:4097-104.

Shin DM, Glisson BS, Khuri FR, Hong WK, Lippman SM. Role of paclitaxel, ifosfamide, and cisplatin in patients with recurrent or metastatic squamous cell carcinoma of the head and neck. **Semin Oncol** 1998; 25(2 Suppl 4):40-4; discussion 45-8.

Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. **J. Natl Cancer Inst.** 2003; 95:14-8.

Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. **J Clin Oncol** 2005; 23:7332-41.

Singh B, Pfister DG. Individualized treatment selection in patients with head and neck cancer: do molecular markers meet the challenge? **J Clin Oncol** 2008; 26:3114-6.

Singh D, Febbo PG, Ross K, et al. Gene expression correlates of clinical prostate cancer behavior. **Cancer Cell** 2002; 1:203-9.

Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. **J Mol Biol** 1975; 98:503-17.

Southern EM, Mir K, Shchepinov M. Molecular interactions on microarrays. **Nat Genet** 1999; 21:5-9.

Stears RL, Martinsky T, Schena M. Trends in microarray analysis. **Nat Med** 2003; 9:140-5.

Steeg PS. Heterogeneity of drug target expression among metastatic lesions: lessons from a breast cancer autopsy program. **Clin Cancer Res** 2008; 14:3643-5.

Stolf BS, Carvalho AF, Martins WK, et al. Differential expression of IGFBP-5 and two human ESTs in thyroid glands with goiter adenoma and papillary or follicular carcinomas. **Cancer Lett** 2003; 191:193-202.

Stolf BS, Abreu CM, Mahler-Araújo MB, et al. Expression profile of malignant and non-malignant diseases of the thyroid gland reveals altered expression of a common set of genes in goiter and papillary carcinomas. **Cancer Lett** 2005; 227:59-73.

Stolf BS, Santos MM, Simao DF, et al. Class distinction between follicular adenomas and follicular carcinomas of the thyroid gland on the basis of their signature expression. **Cancer** 2006; 106:1891-900.

Takes RP, Baatenburg de Jong RJ, Wijffels K, et al. Expression of genetic markers in lymph node metastases compared with their primary tumours in head and neck cancer. **J Pathol** 2001; 194:298-302.

Theoharides TC, Conti P. Mast cells: The Jekyll and Hyde of tumor growth. **Trends Immunol** 2004; 25:235-41.

Tillonen J, Homman N, Rautio M, Jousimies-Somer H, Salaspuro M. Role of yeasts in the salivary acetaldehyde production from ethanol among risk groups for ethanol-associated oral cavity cancer. **Clin Exp Res** 1999; 23:1409-15.

Triche TJ, Schofield D, Buckley J. DNA microarrays in pediatric cancer. **Cancer J** 2001; 7:2-15.

Valk PJ, Verhaak RG, Beijnen MA, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. **N Engl J Med** 2004; 350:1617-28.

van de Vijver MJ, He YD, Van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. **N Engl J Med** 2002; 347:1999-2009.

van den Broek GB, Rasch CR, Pameijer FA, et al. Pretreatment probability model for predicting outcome after intraarterial chemoradiation for advanced head and neck carcinoma. **Cancer** 2004; 101:1809-17.

Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. **Proc Natl Acad Sci U S A** 1990; 87:1663-7.

van Houten V, Snijders P, van den Brekel M, et al. Biological evidence that human papillomaviruses are etiologically involved in a subgroup of head and neck squamous cell carcinomas. **Int J Cancer** 2001; 93:232-5.

van't Veer LJ, Linn SC, Keijzer R, et al. Gene expression profiling predicts clinical outcome of breast cancer. **Nature** 2002; 415:530-6.

Vandenbrouck C, Eschwege F, de la Rochefordiere A, et al. Squamous cell carcinoma of the pyriform sinus: retrospective study of 351 cases treated at the Institut Gustave-Roussy. **Head Neck Surg** 1987; 10:4-13.

Vartanian JG, Carvalho AL, Yueh B, et al. Long-term quality-of-life evaluation after head and neck cancer treatment in a developing country. **Arch Otolaryngol Head Neck Surg** 2004; 130:1209-13.

Vassallo A, Correa P, De Stefani E, et al. Esophageal cancer in Uruguay: a case-control study. **J Natl Cancer Inst** 1985; 75:1005-9.

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. **Science** 1995; 270:484-7.

Vermund H. Role of radiotherapy in cancer of the larynx as related to the TNM system of staging. **Cancer** 1970; 25:485-504.

Vivanco I, Sawyers CL. The phosphatidylinositol 3-kinase-Akt pathway in human cancer. **Nat Rev Genet** 2002; 2:489-501.

Wang E, Miller LD, Ohnmatcht GA, Liu ET, Marincola FM. High-fidelity mRNA amplification for gene profiling. **Nat Biotechnol** 2000; 18:457-9.

Wang J, Hu L, Hamilton SR, Coombes KR, Zhang W. RNA amplification strategies for cDNA microarray experiments. **Biotechniques** 2003; 34:394-400.

Warner GC, Reis PP, Jurisica I, et al. Molecular classification of oral cancer by cDNA microarrays identifies overexpressed genes correlated with nodal metastasis. **Int J Cancer** 2004; 110:857-68.

Weber RS, Forastiere A, Rosenthal DI, Laccourreye O. Controversies in the management of advanced laryngeal squamous cell carcinoma. **Cancer** 2004; 101:211-9.

Welsh J, Chada K, Dalal SS, Cheng R, Ralph D, McClelland M. Arbitrality primed PCR fingerprinting of RNA. **Nucleic Acid Res** 1992; 20:4965-70.

Weymuller EA, Yueh B, Deleyiannis FW, Kuntz AL, Alsarraf R, Coltrera MD. Quality of life in patients with head and neck cancer: lessons learned from 549 prospectively evaluated patients. **Arch Otolaryngol Head Neck Surg** 2000; 126:329-35.

Wildsmith SE, Archer GE, Winkley AJ, Lane PW, Bugelski PJ. Maximization of signal derived from cDNA microarrays. **Biotechniques** 2001; 30:202-6, 208.

Wolf GT, Carey TE, Schmaltz SP, et al. Altered antigen expression predicts outcome in squamous cell carcinoma of the head and neck. **J Natl Cancer Inst** 1990; 82:1566-72.

Wolf GT, Carey TE. Tumor antigen phenotype, biologic staging, and prognosis in head and neck squamous carcinoma. **J Natl Cancer Inst Monogr** 1992; 13:67-74.

Wolf GT, Fisher SG: Effectiveness of salvage neck dissection for advanced regional metastases when induction chemotherapy and radiation are used for organ preservation. **Laryngoscope** 1992; 102:934-9.

[WHO] World Health Organization. **World Cancer Day: Global action to avert 8 million cancer-related deaths by 2015.** 2006. Available from: URL:<http://www.who.int/mediacentre/news/releases/2006/pr06/en/index.html>>

[2009 set 12]

Wu JM, Fackler MJ, Halushka MK, et al. Heterogeneity of breast cancer metastasis: comparison of therapeutic target expression and promoter methylation between primary tumors and their multifocal metastases. **Clin Cancer Res** 2008; 14:1938-46.

Wyckoff J, Wang W, Lin EY, Wang Y, Pixley F, Stanley ER, et al. A paracrine loop between tumor cells and macrophages is required for tumor cell migration in mammary tumors. **Cancer Res** 2004; 64:7022-9.

Wyckoff JB, Wang Y, Lin EY, Li JF, Goswami S, Stanley ER, et al. Direct visualization of macrophage-assisted tumor cell intravasation in mammary tumors. **Cancer Res** 2007; 67:2649-56.

Yamabuki T, Daigo Y, Kato T, et al. Genome-wide gene expression profile analysis of esophageal squamous cell carcinomas. **Int J Oncol** 2006; 28:1375-84.

Yang L, Huang J, Ren X, et al. Abrogation of TGF beta signaling in mammary carcinomas recruits Gr-1+CD11b+ myeloid cells that promote metastasis. **Cancer Cell** 2008; 13:23-35.

Yang YH, Speed T. Design issues for cDNA microarray experiments. **Nat Rev Genet** 2002; 3:579-88.

Yarden Y, Sliwkowski M. Untangling the ErbB signaling network. **Nat Rev Mol Cell Biol** 2001; 2:127-37.

Zhang W, Laborde PM, Coombes KR, Berry DA, Hamilton SR. Cancer genomics: promises and complexities. **Clin Cancer Res** 2001; 7:2159-67.

zur Hausen H. Papillomaviruses causing cancer: evasion from host-cell control in early events in carcinogenesis. **J Natl Cancer Inst** 2000; 92:690-8.

Apêndice 1 - Tampão 5X *Second Strand Buffer*.

[Final] no tampão	Volume adicionado	[Solução estoque]
100 mM Tris pH 6,9	100 μ L	1M
23 mM MgCl ₂	23 μ L	1M
450 mM KCl	225 μ L	2M
0,75 mM β -NAD	7,5 μ L	100 mM
50 mM (NH ₄) SO ₄	50 μ L	1M
H ₂ O	594,5 μ L	
Vol. final	1000 μ L	

Apêndice 2 - Cálculo dos p-valores para a diferença identificada entre pacientes respondedores e não respondedores por imunistoquímica.

TRKB citoplasma: R = c(1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0); NR = c(2, 0, 1, 3, 2, 0, 1, 0, 0, 0, 1)

Wilcoxon exact(resp, nresp)

Exact Wilcoxon rank sum test

data: R and NR

W = 52.5, **P-value = 0.2191**

alternative hypothesis: true μ is not equal to 0

TRKB-p citoplasma: R = c(0, 3, 3, 1, 0, 2, 0, 0, 1, 0, 0, 1, 2); NR = c(1, 2, 3, 3, 2, 0, 1, 1, 1, 0, 2)

Wilcoxon exact(resp, nresp)

Exact Wilcoxon rank sum test

data: resp and nresp

W = 53, **P-value = 0.287**

alternative hypothesis: true μ is not equal to 0

TRKB-p membrana: R = c(3, 3, 3, 1, 0, 2, 0, 0, 1, 0, 0, 1, 2); NR = c(0, 0, 3, 3, 0, 0, 1, 1, 1, 0, 2)

Wilcoxon exact(resp,nresp)

Exact Wilcoxon rank sum test

data: resp and nresp

W = 79, **P-value = 0.6879**

alternative hypothesis: true μ is not equal to 0

TRKB-p núcleo: R = c(0.3, 0, 0.1, 0, 0, 0.5, 0, 0, 0, 0, 0, 0, 0.4); NR = c(0.5, 0.95, 0.7, 0, 0.6, 0, 0, 0, 0, 0, 0.9)

Wilcoxon exact(resp, nresp)

Exact Wilcoxon rank sum test

data: resp and nresp

W = 51.5, **P-value = 0.1924**

alternative hypothesis: true μ is not equal to 0

c-MET cytoplasm R = c(2,2,2,2,3,1,1,1,3,2,2,3,2); NR = c(3,2,3,3,2,2,2,2,3,3,2)

Wilcoxon exact(resp, nresp)

data: resp and nresp

W = 46.5, **p-value = 0.1288**

alternative hypothesis: true μ is not equal to 0

c-MET nucleo: R = c(0.40,0,0,0.80,0.90,0.80,0.70,0.90,0.80,0.50,0.90,0.40,0.80);

NR = c(0.90,0.90,0.70,0,0.50,0.90,0.70,0,0.90,0.90,0.40)

Wilcoxon exact(resp, nresp)

data: resp and nresp

W = 65, **P-value = 0.7264**

alternative hypothesis: true μ is not equal to 0

EGFR citoplasma: R = rep(0, 13); NR = c(rep(0, 10),3)

Wilcoxon exact(resp, nresp)

Exact Wilcoxon rank sum test

data: resp and nresp

W = 65, **P-value = 0.4583**

alternative hypothesis: true μ is not equal to 0

EGFR membrane: R = c(3,3,2,3,3,3,2,3,3,3,3,3); NR = c(3,3,3,3,3,3,3,3,3,0)

Wilcoxon exact(resp, nresp)

Exact Wilcoxon rank sum test

data: resp and nresp

W = 68, **P-value = 1**

alternative hypothesis: true μ is not equal to 0

TGF α cytoplasm: R = c(2,1,1,2,2,2,1,2,2,2,3,3,2); NR = c(3,3,3,2,2,2,2,2,3,3,3)

Wilcoxon exact(resp, nresp)

Exact Wilcoxon rank sum test

data: resp and nresp

W = 36, **P-value = 0.02692**

alternative hypothesis: true μ is not equal to 0

TGF α nucleo: R = c(90,10,5,100,100,100,100,100,100,100,100,100,90); NR =
c(10,90,100,90,100,100,70,100,100,100,100)

Wilcoxon exact(resp, nresp)

Exact Wilcoxon rank sum test

data: resp and nresp

W = 74, **P-value = 0.9112**

alternative hypothesis: true μ is not equal to 0

Apêndice 3 - Cálculo dos p-valores para a correlação entre variáveis clínicopatológicas e resposta clínica dos pacientes.

- Resp Idade = tabela Contig (classif, as.factor(idade))

[,1] [,2]

[1,] 2 5

[2,] 8 13

- ✓ `chisq.test (Resp Idade)`

Pearson's Chi-squared test with Yates' continuity correction

data: Resp Idade

X-squared = 0, df = 1, p-value = 1

Warning message: In `chisq.test (Resp Idade)`: Chi-squared approximation may be incorrect

- ✓ **`fisher.test (Resp Idade)`**

Fisher's Exact Test for Count Data

data: Resp Idade

p-value = 1

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval: 0.0511701 5.3544706

sample estimates:

odds ratio: 0.6597596

- Resp Genero= tabela Contig (classif, idade)

[,1] [,2]

[1,] 2 5

[2,] 8 13

- ✓ `chisq.test (Resp Genero)`

Pearson's Chi-squared test with Yates' continuity correction

data: Resp Genero

X-squared = 0, df = 1, p-value = 1

Warning message: In `chisq.test (Resp Genero)`: Chi-squared approximation may be incorrect

✓ **fisher.test (Resp Genero)**

Fisher's Exact Test for Count Data

data: Resp Genero

p-value = 1

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval: 0.0511701 5.3544706

sample estimates:

odds ratio: 0.6597596

- Resp Fuma= tabela Contig (classif, fuma)

[,1]

[1,] 7

[2,] 21

✓ **chisq.test (Resp Fuma)**

Chi-squared test for given probabilities

data: Resp Fuma

X-squared = 7, df = 1, p-value = 0.008151

✓ **fisher.test (Resp Fuma)**

Error in fisher.test(Resp Fuma) :

'x' must have at least 2 rows and columns

- Resp Bebe= tabela Contig (classif, bebe)

[,1] [,2]

[1,] 0 7

[2,] 4 17

✓ **chisq.test (Resp Bebe)**

Pearson's Chi-squared test with Yates' continuity correction

data: Resp Bebe

X-squared = 0.3889, df = 1, p-value = 0.5329

Warning message: In chisq.test(Resp Bebe) : Chi-squared approximation may be incorrect

✓ **fisher.test (Resp Bebe)**

Fisher's Exact Test for Count Data

data: Resp Bebe

p-value = 0.5453

alternative hypothesis: true odds ratio is not equal to 1

95 percent confidence interval: 0.000000 4.690284

sample estimates:

odds ratio: 0

- Resp T= tabela Contig (classif, as.factor(T))

[,1] [,2] [,3]

[1,] 0 6 1

[2,] 1 15 5

✓ **chisq.test (Resp T)**

Pearson's Chi-squared test

data: Resp T

X-squared = 0.6984, df = 2, p-value = 0.7052

Warning message: In chisq.test(Resp T) : Chi-squared approximation may be incorrect

✓ **fisher.test (Resp T)**

Fisher's Exact Test for Count Data

data: Resp T

p-value = 1

alternative hypothesis: two.sided

- Resp N= tabela Contig (classif, as.factor(N))

[,1] [,2] [,3]

[1,] 7 0 0

[2,] 9 9 3

✓ `chisq.test (Resp N)`

Pearson's Chi-squared test

data: Resp N

X-squared = 7, df = 2, p-value = 0.0302

Warning message: In `chisq.test(Resp N)` : Chi-squared approximation may be incorrect

✓ **`fisher.test (Resp N)`**

Fisher's Exact Test for Count Data

data: Resp N

p-value = 0.03226

alternative hypothesis: two.sided

• Resp Diag= tabela Contig (`classif, diagnostico_histologico`)

[,1] [,2] [,3]

[1,] 2 4 1

[2,] 2 13 6

✓ `chisq.test (Resp Diag)`

Pearson's Chi-squared test

data: Resp Diag

X-squared = 1.7815, df = 2, p-value = 0.4103

Warning message: In `chisq.test(Resp Diag)` : Chi-squared approximation may be incorrect

✓ **`fisher.test (Resp Diag)`**

Fisher's Exact Test for Count Data

data: Resp Diag

p-value = 0.5751

alternative hypothesis: two.sided

• Resp Sitio= tabela Contig (`classif, as.factor(sitio)`)

[,1] [,2] [,3] [,4] [,5] [,6]

[1,] 1 4 0 1 1 0

[2,] 1 4 2 6 4 4

✓ `chisq.test (Resp Sitio)`

Pearson's Chi-squared test

data: Resp Sitio

X-squared = 5.8286, df = 5, p-value = 0.3233

Warning message: In `chisq.test (Resp Sitio)`: Chi-squared approximation may be incorrect

✓ **`fisher.test (Resp Sitio)`**

Fisher's Exact Test for Count Data

data: Resp Sitio

p-value = 0.3979

alternative hypothesis: two.sided

• Tabela cidade:

	BA	FO	GO	SP
nr	3	1	1	2
r	10	5	2	4

✓ X-squared: 0.9007

✓ **Fisher exact test: p-valor= 1**

• `tabela Contig(as.vector(no_respN), as.vector(no_resp Resposta))`

Tab N0 Resp Mestrado = `tabelaContig(as.vector(no_resp N), as.vector(no_resp Resposta))`

NR	R
N0	6 12
N1	5 2
N2	2 5
N3	1 2

✓ `chisq.test(tab N0 Resp Mestrado)`

Pearson's Chi-squared test

data: tab N0 Resp Mestrado

X-squared = 3.6508, df = 3, p-value = 0.3017

Warning message:

In `chisq.test(tab N0 Resp Mestrado)` :

Chi-squared approximation may be incorrect

✓ **`fisher.test(tabN0 Resp Mestrado)`**

Fisher's Exact Test for Count Data

data: tab N0 Resp Mestrado

p-value = 0.327

alternative hypothesis: two.sided