

**CLASSIFICADORES MOLECULARES DO
CARCINOMA FOLICULAR DA TIREOIDE
BASEADOS EM ALTERAÇÕES DA
EXPRESSÃO GÊNICA**

ANA HELENA PAGOTTO

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

Área de concentração: Oncologia

Orientador: Dr. Luiz Fernando Lima Reis

São Paulo

2010

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Pagotto, Ana Helena

Classificadores moleculares do carcinoma folicular da tireoide baseado em alterações da expressão gênica / Ana Helena Pagotto – São Paulo, 2010.

103p.

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. BIOPSIA ASPIRATIVA POR AGULHA FINA. 4. CÂNCER DE TIREOIDE. 5. ADENOMA FOLICULAR. 6. ADENOCARCINOMA FOLICULAR. 7. TÉCNICAS DE AMPLIFICAÇÃO DE RNA.

“Cada pessoa deve trabalhar para o seu aperfeiçoamento e, ao mesmo tempo, participar da responsabilidade coletiva por toda a humanidade”.

Marie Curie (1867-1934)

“Quem quer que se tenha ocupado seriamente de trabalho científico de qualquer tipo percebe que sobre os portões de entrada do templo da ciência estão escritas as palavras: Deveis ter fé. É uma qualidade que o cientista não pode dispensar”.

Max Karl Ernst Ludwig Planck (1858-1947)

"Se a vida lhe der um limão, faça dele uma limonada."

Provérbio popular brasileiro

DEDICATÓRIA

À minha mãe Edi, pelo amor incondicional, pelo esforço em prol da minha educação e dos meus estudos desde cedo, pelo colo amigo de todas as horas.

Ao meu padrasto Antonio, por compartilhar das minhas alegrias, pelo apoio nos momentos difíceis e, principalmente, por ter me dado o privilégio de fazer parte de uma família tão maravilhosa.

Ao meu pai Clóvis (IN MEMORIAN). Mesmo o destino tendo nos separado antes mesmo de termos nos olhado, sei que estou sendo bem cuidada aí de cima.

À minha avó Isabel (IN MEMORIAN), meu anjo bom.

*Ao Daniel Stuginski, pela imensa ajuda, paciência, abdicção,
sensibilidade e amor.*

*Finalmente, dedico esta tese à Tatiana Stuginski, que além do
exemplo de força e superação, tornou-se um estímulo direto e
diário à realização deste trabalho, para que eu pudesse contribuir
ao menos um pouco com a ciência contra o câncer.*

AGRADECIMENTOS

Ao meu orientador Dr. Luiz Fernando Lima Reis, pela oportunidade, confiança e paciência em tempos tão complicados.

Ao CEPID e à FAPESP, pelo suporte financeiro e críticas ao projeto.

Ao doutor Paulo Grimaldi, pela boa vontade, iniciativa e trabalho de coleta sem os quais este trabalho definitivamente não seria possível. Sou imensamente grata.

Aos membros da Banca de Qualificação e da Banca Examinadora, pela atenção dispensada e críticas ao projeto.

Ao Prof. Dr. Luiz Paulo Kowalski e Departamento de Cabeça e Pescoço do Hospital A.C. Camargo.

À todos do Departamento de Imagem do Hospital A.C. Camargo, pelo esforço coletivo na coleta de amostras.

Ao Departamento de Anatomia Patológica, em especial ao Dr. Victor Piana de Andrade, por elaborar a folha de trabalho para coleta e pelo esforço na coleta de amostras pós-cirúrgicas. A todos os médicos que participaram da coleta de amostras na sala de congelação.

Aos membros e ex-membros do Laboratório BioInfo, do Instituto de Matemática e Estatística da USP, especialmente Dr. Eduardo Jordão, Roberto Hirata e Gustavo Esteves, pelas análises matemáticas e pelo trabalho exaustivo em períodos pré-relatório.

Ao Laboratório de Bioinformática do Hospital A.C. Camargo, em especial à Dra. Helena Brentani, Renato Puga, pela ajuda com a pré-análise das lâminas.

Aos membros do Comitê de Ética em Pesquisa do Hospital A.C. Camargo.

Aos funcionários da Biblioteca do Hospital A.C. Camargo, pelo impecável serviço. À Suely, por me ajudar com a formatação da tese e pelo excelente atendimento.

À Secretaria da Pós Graduação, pela organização, ajuda e cuidados tanto com os trabalhos como também com as disciplinas. Ao diretor da pós graduação, Prof. Dr. Fernando Soares, por contribuir com a qualidade do curso e por ter dado condições para a conclusão deste trabalho.

Às secretárias Aline Pacífico, Haydée Espinosa e Cláudia Novaes, pelo excelente trabalho e pela ajuda em momentos de pane total.

Ao Dr. Juan Postigo Diaz e esposa, por me receberem tão bem em Lima. Por toda a equipe e amigos que fiz no INEN, pelos ensinamentos que me passaram e pela oportunidade acompanhar tantos procedimentos. À Margiory, Rafael, Gisella e Dra. Romy, por toda a assistência e bons momentos.

À Beatriz Stolf, por toda a ajuda quando cheguei ao Ludwig e por me ensinar tantas coisas. É um orgulho dar continuidade ao seu trabalho.

A todos os alunos, ex-alunos e funcionários do laboratório (em tempos longínquos chamado de Laboratório da Inflamação, depois Laboratório de Genômica Funcional, depois de “Terceiro andar da Casinha” e, finalmente, de “as meninas lá de cima”). Ao Alex, pela confecção das lâminas e por toda

a ajuda. À Bianca e Coló, pelo trabalho em equipe com as linhagens e discussões de projetos.

Às “minhas meninas” Nair, Camila, Letícia, pelo convívio, amizade, gulodices e risadas, seja para comemorar em tempos áureos, seja para fazer piada das próprias desgraças. E também por gostarem de cozinhar mais do que eu. À Ju, Marina, Vladimir, Grá e todos que se juntaram (e se apertaram, se espremeram) no novo laboratório. Vocês todos são incríveis.

À Dra. Adriana Abalen, Dra. Silvia Rogatto e todos os seus alunos.

À Dra. Dirce Carraro, por ceder o RNA das linhagens celulares.

Ao Rodrigo Piologo pela ajuda com os pôsteres de congressos.

Aos meus alunos do programa de Ensino de Jovens e Adultos do Colégio Lourenço Castanho e seu coordenador Miguel Castilho, por terem me proporcionado uma das experiências mais ricas da minha vida.

À Shirley, por ter me dado a oportunidade de trabalho após o encerramento da minha bolsa, permitindo a minha permanência em São Paulo e conclusão do meu projeto. A todos os funcionários da Wizard Liberdade e a todos os meus alunos, que me apóiam e me animam tanto, mesmo sem entender direito o que a *teacher* faz.

À Renata Scavone, pela ajuda quando precisei de trabalho. Espero poder retribuir da melhor maneira possível.

Ao Philipp e Ricardo, por terem sido tão solícitos.

À Daniela Filipini e Fernanda Molo, amigas queridas, de conversas nas quais sempre se misturam amizade, preocupações, desabafos, comida e biologia molecular.

À Dra. Cristina, por toda a atenção e ajuda no meu tratamento. Sem esta ajuda, tenho dúvidas se a conclusão deste trabalho seria possível.

À Tamara Prior, pelo incentivo, amizade e inspiração devido à sua paixão pela vida acadêmica.

Ao Erivelto, por todo o incentivo e torcida.

Ao Eduardo Deberaldini e Daniel Conti, grandes amigos de infância, presentes e companheiros desde o tempo dos trabalhos para as feiras de ciências da escola.

À todos os amigos da Borracharia 64, querida turma que permanece unida desde a faculdade, ainda que com alguns membros morando tão longe; por fazerem com que eu me sinta em família, pelo contato diário (sem falta) por e-mail e pelo esforço de todos para pegarem a estrada quando necessário.

Ao Dr. Arthur Gruber, meu orientador na iniciação científica, que com seu gênio de sair fásca e sua impressionante habilidade em lecionar fez valer a pena cada dia da minha primeira incursão como aspirante a cientista. Muito obrigada pela oportunidade e por toda a atenção.

À minha “sogra” Bel, ao Jaber, à Tati e “Tia” Goldi, ao Cris e ao Ozzy, por me acolherem tão bem, pelas conversas, risos e lágrimas, pelo exemplo de união em situações tão difíceis que enfrentaram neste último ano.

Ao meu namorado Daniel Stuginski, por ter me ajudado tanto, por ter se preocupado e participado de tudo de forma tão direta e comprometida, por me trazer paz, confiança e muito amor.

À toda a minha família, por todo o apoio e torcida. Por todas as orações e palavras de carinho. À minha tia Vilma, por sempre se lembrar de mim e ser tão sensível. Em especial, quero agradecer ao meu padrasto Antonio por sempre me dizer o quão importante é a nossa família. À minha mãe, por tudo o que faz por mim, por tudo o que fazemos juntas e por me fazer sentir um amor que transcende qualquer coisa explicável (e obrigada por ter aprendido a usar a webcam e o messenger quando eu fui para o Peru!).

Finalmente, agradeço a todos os pacientes que cederam amostras para este estudo. Que possamos um dia transformar o conhecimento gerado graças a estas amostras em melhores oportunidades para os que futuramente precisarem. Muito obrigada.

RESUMO

Pagotto AH. **Classificadores moleculares do carcinoma folicular da tireoide baseado em alterações da expressão gênica**. São Paulo; 2010. [Tese de Doutorado-Fundação Antônio Prudente].

A tireoidectomia é o método de escolha para o tratamento dos tumores malignos da tireoide. Considerando que a maioria dos nódulos tireoideanos é benigna (95% dos casos), a precisão no diagnóstico é fundamental para evitar cirurgias diagnósticas. O principal método utilizado na análise pré-operatória de nódulos tireoideanos é a citologia de material obtido por PAAF (Punção Aspirativa por Agulha Fina), que apresenta excelente precisão no diagnóstico de carcinomas anaplásicos, medulares e papilíferos. No entanto, a citologia de material obtido por PAAF não é capaz de diferenciar o adenoma (nódulo benigno) do carcinoma folicular, uma vez que é a invasão da cápsula pelas células tumorais que os distingue. Para detectar a invasão capsular, é necessário realizar cortes histológicos de todo o nódulo, o que só pode ser realizado após remoção total da lesão. Como medida diagnóstica, quando o diagnóstico por PAAF é inconclusivo e, diante da possibilidade de diagnóstico de um carcinoma folicular e consequentes metástases, o paciente geralmente é submetido à tireoidectomia total. Como apenas 30% das lesões foliculares com diagnóstico inconclusivo representam lesões malignas, um método com maior precisão no diagnóstico diferencial de adenoma e carcinoma foliculares poderia evitar 70% das tireoidectomias nestes pacientes. Em estudo anterior, utilizando a técnica de microarranjos de cDNA, nosso grupo identificou 14 trios de genes que foram capazes de distinguir precisamente amostras de adenoma e carcinoma folicular. O presente trabalho teve como objetivo principal a validação destes classificadores, através da mesma técnica, para que possam ser empregados no diagnóstico diferencial destas lesões. Foram testadas 14 amostras de RNA extraídas por punção pré ou pós cirúrgicas de nódulos

com diagnóstico histopatológico de adenoma ou carcinoma folicular. Com estas amostras foi realizada a otimização do protocolo de amplificação de RNA para quantidades iniciais inferiores a 30ng de RNA total, visando o uso de RNA extraído das células obtidas por PAAF, já que a maior limitação no uso de material colhido por esta técnica consiste no baixo número de células obtido, resultando em quantidade insuficiente de RNA para a realização dos experimentos de microarranjos de cDNA. Os cinco melhores trios aqui validados apresentam sensibilidade de 100% e especificidade que varia entre 77 e 44%. Com estes trios, foram construídas árvores de decisão para aumentar a taxa de acerto do teste.

SUMMARY

Pagotto AH. **[Molecular classifiers for follicular carcinoma of the thyroid gland based on alterations of gene expression]**. São Paulo; 2010. [Tese de Doutorado-Fundação Antônio Prudente].

Thyroidectomy is the treatment of choice in thyroid malign tumors cases. Considering that the great part of all thyroid nodules is benign (95%), diagnostic precision is essential to avoid diagnostic surgeries. The main technique used in the preoperative analysis of thyroid nodules is Fine-Needle Aspiration Biopsy (FNAB), that shows great diagnostic precision of anaplastic, medullary and papillary carcinomas. However, citologic examination of FNAB-collected samples doesn't allow the distinction between adenoma (benign) and follicular carcinoma (malign), as distinction is made on basis of capsule invasion by tumor cells. Histological examination of of the entire capsule is necessary to detect this invasion and this procedure can only be performed after total removal of the lesion. As a diagnostic measure, when facing an inconclusive diagnosis by FNAB, the patient undergoes total thyroidectomy, not taking the risk of preserving a carcinoma and its metastasis. Considering that only 30% of all inconclusive follicular lesions represent malign ones, it would be very useful to develop a diagnostic tool with better precision to distinguish adenoma from follicular carcinomas, as this could avoid 70% of thyroidectomies in this group of patients. Using cDNA microarrays, our group identified 14 trios of genes that could precisely classify adenoma and follicular carcinoma samples. Through the same technique, the present study aimed to validate these classifiers, so they could be used in the differential diagnosis of these pathologies. We tested 14 RNA samples from adenoma and follicular carcinoma obtained by pre and post operative FNAB. The RNA amplification protocol was optimized using the collected samples, so it can be performed from minute amounts of total RNA (<30 ng), since the major limitation of using PAAF collected samples is their

low cellularity. This restrict number of cells per sample results in insufficient mass of RNA to perform cDNA microarray assays. The five best trios validated show 100% sensibility and specificity ranging between 77 e 44%. Aiming to raise the test's hit rate, decision trees were constructed.

LISTA DE FIGURAS

Figura 1	Cortes histológicos de adenoma e carcinoma foliculares.....	6
Figura 2	Representação tridimensional de trios classificadores.....	18
Figura 3	Fluxograma representando o desenho experimental do presente trabalho.....	25
Figura 4	Representação gráfica da reflexão provocada pela lâmina espelhada e pela lâmina comum.....	32
Figura 5	Comparação entre fluorescência emitida com a utilização lâmina convencional e lâmina espelhada.....	33
Figura 6	Qualidade dos RNAs totais que compõem o pool de referência.....	64
Figura 7	Interface do <i>software</i> ND-1000 V3.3.0, do espectrofotômetro NanoDrop®, representando o resultado de leitura da amostra AH18713007.....	66
Figura 8	Espectro de absorbância do RNA total de amostras de punção de peça cirúrgica extraído com o <i>kit</i> Pico Pure.....	68
Figura 9	Perfil do RNA da amostra AHF002.....	70
Figura 10	Análise da qualidade de RNA amplificado, através de fracionamento em gel de agarose.....	71
Figura 11	Análise da qualidade do RNA amplificado de amostras pós-cirúrgicas de interesse, através de fracionamento em gel de agarose.....	73

Figura 12	Imagem da porção de uma lâmina após digitalização.....	75
Figura 13	MM plot da amostra g12, antes e após a normalização, com o respectivo coeficiente de Pearson.....	76
Figura 14	MA plot da amostra g12.....	76
Figura 15	Análise da uniformidade dos spots depositados por cada agulha.....	77
Figura 16	Dendrograma mostrando o agrupamento das amostras testadas.....	78
Figura 17	Curva ROC para o Trio 18	81
Figura 18	Árvores de decisão estabelecidas com os melhores trios classificadores	83

LISTA DE QUADROS

Quadro 1	Trios classificadores e seus respectivos genes.....	17
Quadro 2	Amostras de adenoma e carcinoma foliculares a utilizadas na validação dos trios e quartetos classificadores.....	27
Quadro 3	Linhagens celulares selecionadas para compor o RNA referência.....	42
Quadro 4	Concentração dos nucleotídeos na mistura utilizada para marcação indireta do cDNA.....	50
Quadro 5	Amostras pós cirúrgicas coletadas no primeiro semestre deste trabalho.....	61
Quadro 6	Massa de RNA total das linhagens do RNA referência após extração.....	62
Quadro 7	Rendimento da extração do RNA total das amostras de carcinoma e adenoma foliculares.....	65
Quadro 8	Rendimento do primeiro ciclo de amplificação do RNA referência.....	67
Quadro 9	Rendimento do segundo ciclo de amplificação do RNA referência.....	67
Quadro 10	Rendimento de RNA das amostras após dois ciclos de amplificação.....	72
Quadro 11	Melhores trios classificadores, erros e acertos de cada classificador.....	80

Quadro 12	Sensibilidade, especificidade , acurácia e SVD dos melhores trios classificadores.....	80
Quadro 13	Identificação das amostras classificadas corretamente e erroneamente por cada um dos melhores trios.....	82

LISTA DE ABREVIATURAS

µg	Micrograma
µL	Microlitro
18S	RNA ribossômico 18S
28S	RNA ribossômico 28S
aa-dUTP	<i>aminoalil-dUTP</i>
ANOVA	análise de variância, do inglês <i>analysis of variance</i>
aRNA	RNA amplificado
BSA	albumina sérica bovina, do inglês: <i>bovine serum albumine</i>
CA	Carcinoma
CB	tampão condicionante (<i>Conditioning Buffer, kit Pico Pure®</i>)
cDNA	DNA complementar
Cm	Centímetro
cm²	centímetro quadrado
CONEP	Comissão Nacional de Ética em Pesquisa
Cy3	cianina 3
Cy5	cianina 5
dATP	<i>desoxiadenosina trifosfato</i>
dCTP	desoxicitosina trifosfato
DEPC	Dietilpirocarbonato
dGTP	desoxiguanina trifosfato
DMSO	Dimetilsulfóxido
DNA	ácido desoxiribonucleico, do inglês: <i>desoxiribonucleic acid</i>
dNTPs	mistura de desoxinucleotídeos trifosfato
dTTP	desoxitimina trifosfato
dUTP	desoxiuridina trifosfato
EB	tampão de eluição (<i>Elution Buffer, kit Pico Pure®</i>)
EDTA	ácido etilenodiaminotetracético, do inglês: <i>ethylenediamine tetraacetic acid</i>
g	unidade de medida da força centrífuga relativa
g	Gramma
IME	Instituto de Matemática e Estatística

INEN	<i>Instituto Nacional de Enfermedades Neoplasicas</i>
LGEA	<i>Laboratory of Gene Expression Analysis</i>
log₂	Logaritmo na base 2
M	Molar
mg	Miligrama
mL	Mililitro
mM	Milimolar
mRNA	RNA mensageiro
n	número amostral
ng	Nanograma
NH₄Ac	acetato de amônio
nm	Nanômetro
°C	grau Celsius
ORESTES	<i>Open reading frame ESTs</i>
PAAF	punção aspirativa por agulha fina
PB	par de bases
PBS	tampão fosfato em solução salina, do inglês <i>phosphate saline buffer</i>
PCR	reação em cadeia da polimerase, do inglês: <i>polymerase chain reaction</i>
pH	potencial hidrogeniônico
PoliA	cauda poliadenilada do RNA mensageiro
qsp	quantidade suficiente para
R	linguagem de programação "R"
RNA	ácido ribonucléico, do inglês : <i>ribonucleic acid</i>
RNAse	Ribonuclease
rNTPs	mistura de ribonucleotídeos trifosfato
ROC	<i>Receiver Operating Characteristic</i>
rpm	rotações por minuto
SDS	dodecil sulfato de sódio , do inglês: <i>sodium dodecyl sulfate</i>
SSC	solução salina com citrato de sódio, do inglês: <i>saline-sodium citrate</i>
Tris	Tris (hidroximetil) aminometano
U	Unidade
USP	Universidade de São Paulo

UV	Ultravioleta
WB1	tampão de lavagem 1 (<i>Washing Buffer 1, kit Pico Pure®</i>)
WB2	tampão de lavagem 2 (<i>Washing Buffer 2, kit Pico Pure®</i>)
XB	tampão de extração (<i>Extraction Buffer, kit Pico Pure®</i>)

ÍNDICE

1	INTRODUÇÃO	1
1.1	Informações gerais	1
1.2	Classificação	2
1.3	Diagnóstico por punção aspirativa por agulha fina (PAAF)	4
1.4	Carcinoma folicular x adenoma	5
1.5	Abordagem metodológica: microarranjos de cDNA	8
1.5.1	Informações gerais	8
1.5.2	Variabilidade nos experimentos de microarranjos	10
1.5.3	Análise estatística dos dados de microarranjos.....	12
1.5.4	A técnica de <i>microarray</i> de cDNA e a distinção molecular entre carcinoma folicular e adenoma.....	13
1.6	Dados preliminares.....	14
2	JUSTIFICATIVA.....	19
3	OBJETIVOS	21
3.1	Objetivo geral	21
3.2	Objetivos específicos.....	21
3	DESENHO EXPERIMENTAL.....	23
4	CASUÍSTICA.....	26
5	MATERIAIS E MÉTODOS	28
5.1	Reagentes	28
5.2	A lâmina de microarranjos utilizada.....	30
5.2.1	Substrato	30
5.2.2	Conjunto de sondas depositadas nas lâminas	33
5.3	Coleta e Preservação das amostras.....	34

5.3.1	Coleta e preservação das amostras coletadas por punção de peça cirúrgica (punção pós-operatória)	34
5.3.1	Coleta e preservação das amostras coletadas por PAAF diagnóstica (punção pré-operatória)	35
5.3.2	Aspectos éticos envolvidos na coleta e preservação das amostras	36
5.4	Extração do RNA total	36
5.4.1	Extração do RNA total de amostras pós-cirúrgicas	37
5.4.2	RNA total de amostras coletadas por punção aspirativa pré-cirúrgica ..	39
5.5	Preparação do RNA referência.....	39
5.6	Amplificação do RNA total	43
5.6.1	Amplificação do RNA referência	44
5.6.2	Amplificação das amostras (pré e pós-cirúrgicas)	47
5.7	Marcação indireta das amostras e referência	48
5.8	Hibridização em lâmina espelhada	52
5.9	Captura das imagens	54
5.10	Análise da qualidade das lâminas	55
5.11	Análise matemática e estatística dos dados.....	58
6	RESULTADOS	61
6.1	Coleta e preservação de amostras	61
6.2	Extração do RNA total	63
6.2.1	Rendimento do RNA referência	63
6.2.2	Rendimento do RNA total das amostras de adenoma e carcinomas foliculares	64
6.3	Amplificação do RNA total	66
6.3.1	Rendimento do RNA referência	66
6.3.2	Rendimento da amplificação do RNA total de amostras colhidas por PAAF	68
6.4	Marcação e hibridização das amostras	73
6.5	Análise da qualidade das lâminas	75
6.6	Análise matemática e estatística dos dados.....	79

7	DISCUSSÃO	84
8	CONCLUSÕES	93
9	REFERÊNCIAS BIBLIOGRÁFICAS	95

APÊNDICES

Apêndice 1 Folha de trabalho na coleta por PAAF

Apêndice 2 Termo de Consentimento Livre e Informado

Apêndice 3 Trios utilizados na validação independente

1 INTRODUÇÃO

1.1 INFORMAÇÕES GERAIS

Nódulos tireoideanos são comuns na população geral: estima-se que a prevalência clínica seja de 2 a 7% (YAMASHITA 2000; MACKENZIE e MORTIMER 2004). Esta prevalência varia muito em função do método de detecção; exames ultrassonográficos e autópsias revelam prevalências de 27 e 50%, respectivamente (MACKENZIE e MORTIMER 2004). A maior parte dos nódulos é benigna: adenomas foliculares, bóciós multinodulares, cistos e tireoidites crônicas.

O câncer de tireoide é o tumor endócrino mais comum, excetuando-se os tumores de ovário (CLARK e NOGUCHI 2000). Os tumores clinicamente evidentes são raros, com incidência anual de dois a quatro novos casos por 100.000 indivíduos nos países com ingestão adequada de iodo. No entanto, carcinomas microscópicos são extremamente prevalentes, sendo que a maior taxa de prevalência já observada foi de 35,6%, em um estudo envolvendo 101 autópsias, publicado por HARACH et al. (1985). Estima-se que um a cada 10.000 destes tumores microscópicos podem, eventualmente, manifestar-se clinicamente (NOGUCHI 2000). A frequência de malignidades em nódulos examinados em cirurgias varia de 3 a 31%. A menor proporção de malignidade é observada em pacientes entre 31 e 40

anos. Esta proporção dobra em pacientes jovens (até 20 anos) e é quase seis vezes maior em pacientes com mais de 70 anos (YAMASHITA 2000).

Na maior parte dos casos de câncer de tireoide é difícil apontar um fator etiológico específico: fatores genéticos participam da etiologia de alguns tipos tumorais (NOGUCHI 2000), como por exemplo, a translocação cromossômica t(2;3)(q13;p25), que resulta na fusão dos genes PAX8 e PPARG. Outros exemplos são os genes LGALS3, TPO, TP53, hTERT e HMBE-1, cujo aumento da expressão parece estar relacionada aos cânceres de tireoide. No entanto, a utilização destes genes como marcadores moleculares não se mostrou útil no diagnóstico destes tumores (CERUTTI et al. 2004). Também são conhecidos alguns fatores ambientais, como a exposição à radiação ionizante, principalmente na infância e adolescência, que aumenta o risco do desenvolvimento tanto de nódulos benignos quanto de carcinomas (MACKENZIE e MORTIMER 2004). Também, a deficiência ou excesso de iodo na dieta podem aumentar a incidência de câncer de tireoide (KANNO et al. 1992).

1.2 CLASSIFICAÇÃO

O câncer da tireoide pode ser classificado histologicamente em: carcinoma papilífero, carcinoma folicular, carcinoma indiferenciado (ou anaplásico), carcinoma medular e linfoma originado na tireoide.

O carcinoma papilífero é o mais comum (50-80% dos casos de câncer de tireoide), exceto em populações com dieta pobre em iodo. Embora

ocorram frequentemente em associação com micrometástases, o prognóstico é excelente em cerca de 80% dos casos, sendo variável nos 20% restantes (NOGUCHI 2000).

O carcinoma folicular (10-40% dos casos) é um tumor epitelial que apresenta evidências de diferenciação de células foliculares, porém sem os elementos diagnósticos observados no carcinoma papilífero. Podem ser classificados em *minimamente invasivos* e *muito invasivos*. Embora o carcinoma folicular minimamente invasivo tenha um prognóstico muito bom, não apresenta elementos típicos de malignidade e, por este motivo, a distinção entre este tipo de tumor e um tumor folicular benigno é complicada (NOGUCHI 2000).

Os carcinomas indiferenciados ou anaplásicos (5-10% dos casos) é provavelmente o tumor sólido mais agressivo com taxa de mortalidade próxima a 100% (MARTINS et al. 2004). É um tumor raro, mais frequente em áreas de bócio endêmico, com incidência acentuadamente decrescente nos últimos anos (NOGUCHI 2000).

O carcinoma medular (3-12% dos casos) é originado das células C ou parafoliculares, produtoras de calcitonina. Podem ser esporádicos ou familiares (NOGUCHI 2000).

O linfoma maligno originado na tireoide (2% dos casos) desenvolve-se na glândula acometida por longos períodos de tireoidite de Hashimoto e geralmente é do tipo célula T (NOGUCHI 2000).

1.3 DIAGNÓSTICO POR PUNÇÃO ASPIRATIVA POR AGULHA FINA (PAAF)

Atualmente, diversos métodos podem ser empregados na análise de nódulos tireoideanos, como a ultra-sonografia, cintilografia e punção aspirativa por agulha fina (PAAF), entre outros (INOHARA et al. 1999).

Nos últimos 20 anos, a PAAF de nódulos tireoideanos tem se mostrado muito mais vantajosa que qualquer outro procedimento diagnóstico, com taxas de sensibilidade e especificidade de 80% a maiores que 90% (HAMBURGER 1994; KEBEBEW et al. 2005). É um método de coleta de células do interior do nódulo para a realização de estudo citológico (SANTOS 1999). O procedimento consiste em remover uma pequena amostra de células ou fluído por aspiração através de uma agulha inserida no nódulo. Com o material aspirado é feito uma espécie de esfregaço, que é então corado e examinado ao microscópio. Desta forma, os tipos celulares que compõe a amostra podem ser avaliados. Trata-se de uma técnica minimamente invasiva, bem tolerada pelo paciente e que não requer anestesia. Complicações são raras e os efeitos colaterais são geralmente limitados a um pequeno desconforto ao deglutir por algumas horas (CHAPLIN 2005). Tal segurança, no entanto, não exclui a importância da experiência do profissional que realiza a punção, uma vez que este fator contribui muito para a qualidade do procedimento.

A PAAF de nódulos tireoideanos tem sido proposta como método de triagem pré-operatória, reduzindo o número de pacientes encaminhados

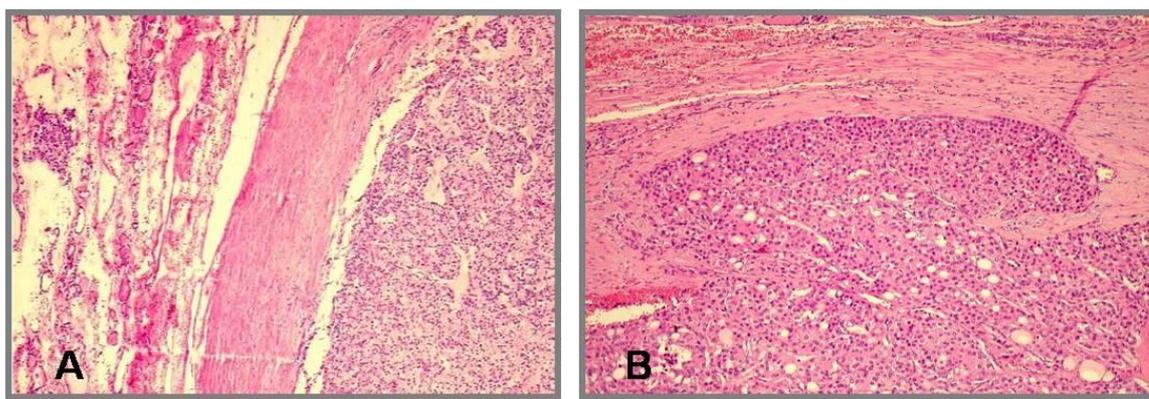
desnecessariamente para a cirurgia. Também é utilizada para o acompanhamento da evolução de nódulos benignos e de pacientes operados (SANTOS 1999).

Este método apresenta excelente precisão no diagnóstico da maior parte das malignidades da tireoide (GREBE e HAY 1999). O estudo citológico do material colhido por PAAF levou a uma importante diminuição do número de tireoidectomias realizadas para a remoção tumores benignos e um aumento concomitante do número de casos de câncer encontrados em amostras excisadas cirurgicamente. No entanto, o diagnóstico do carcinoma folicular através da PAAF é muito complicado devido à semelhança citológica deste tumor com o adenoma folicular (BARTOLAZZI et al. 2001).

1.4 CARCINOMA FOLICULAR X ADENOMA

Uma das distinções mais difíceis na citologia da tireoide é entre carcinomas foliculares minimamente invasivos e adenomas foliculares. Adenomas foliculares são tumores tireoideanos benignos, mais frequentes que os carcinomas e que geralmente acometem indivíduos de menor faixa etária (LEMOINE et al. 1989). Ambos são encapsulados por definição e o principal elemento que os distingue é a invasão capsular nos casos de carcinoma folicular (**Figura 1**). Para a invasão capsular ser considerada inequívoca, o tumor deve penetrar toda a espessura da cápsula, sendo que irregularidades no contorno da borda interna da cápsula ou grupos de células foliculares situadas dentro da cápsula não são suficientes. Até hoje,

o diagnóstico de carcinoma folicular da tireoide não pode ser feito apenas com base na citologia, pois neste caso tanto atipias nucleares, celulares e mitoses não são significantes (YAMASHITA 2000). Assim, o diagnóstico definitivo só pode ser dado após o exame minucioso de cortes histológicos de todo o tumor para avaliação da integridade capsular.



Fonte: FCM-UNICAMP Faculdade de Ciências Médicas (2010)

Figura 1 - Cortes histológicos de adenoma e carcinoma foliculares. **1A**: corte histológico de um adenoma folicular, evidenciando a integridade da cápsula; **1B**: carcinoma folicular, em porção da cápsula invadida por células tumorais.

Quando, na análise pré-operatória, há dúvida quanto ao diagnóstico de carcinoma folicular ou adenoma, o paciente é geralmente submetido à tireoidectomia total, sendo que apenas 30% dos pacientes submetidos à cirurgia apresentam tumor maligno, conforme atestado em análises pós-cirúrgicas de fragmentos tumorais incluídos em parafina. Como resultado desta imprecisão no diagnóstico, muitos pacientes diagnosticados com lesão folicular pela citologia passam pela excisão cirúrgica da tireoide e, após a avaliação do nódulo retirado, recebem o diagnóstico de tumor benigno.

Nódulos benignos da tireoide geralmente não implicam na realização

de tireoidectomia total, sendo necessária apenas a retirada da lesão. Em alguns casos, o nódulo benigno sequer exigiria procedimento cirúrgico. No entanto, após a remoção total da glândula, todos os pacientes são submetidos à reposição de hormônios tireoideanos exógenos por toda a vida.

Atualmente, diversos estudos têm como objetivo minimizar o número de intervenções cirúrgicas não-terapêuticas. Por exemplo: a padronização da terminologia patológica e estabelecimento de critérios para o diagnóstico de lesões foliculares; a avaliação de parâmetros clínicos e sua associação com o risco de malignidade; a análise de aspectos radiológicos, etc. Finalmente, a identificação de novos marcadores moleculares promete aumentar a acurácia da citologia de amostras colhidas por PAAF. Alguns estudos sugeriram que a expressão da proteína galectina 3 em aspirados poderia ser utilizada como método diagnóstico de lesão folicular maligna, porém alguns trabalhos, através de ensaios imunocitoquímicos de células obtidas por PAAF (MEHROTRA et al. 2004) e imunohistoquímica de fragmentos emblocados em parafina (MILLS et al. 2005) demonstraram que a galectina 3 não é um marcador confiável para distinguir o carcinoma folicular do adenoma. Além disso, muitos aspirados apresentam celularidade muito baixa, não sendo apropriados para este tipo de ensaio (MILLS et al. 2005). CERUTTI et al. (2004) propôs a combinação dos marcadores DDIT3, ARG2 e ITM1 para a distinção molecular destes dois tumores. No entanto, a acurácia do método foi estimada em 83%, sendo que as amostras classificadas erroneamente como adenoma incluíram um caso de carcinoma

folicular com invasão capsular e vascular, mostrando que a combinação destes marcadores não é adequada para a utilização na rotina clínica. Neste sentido, estudos sobre a expressão gênica podem contribuir para a identificação de novos marcadores, permitindo o desenvolvimento de novas ferramentas para o diagnóstico pré-operatório correto destes tumores.

1.5 ABORDAGEM METODOLÓGICA: MICROARRANJOS DE cDNA

1.5.1 Informações gerais

Uma das técnicas mais utilizadas atualmente no estudo da expressão gênica diferencial é a técnica de microarranjos de cDNA. A técnica foi primeiramente descrita em 1995 por SCHENA et al. e é utilizada principalmente para a análise da expressão gênica diferencial entre duas amostras distintas (ex: células normais *versus* células tumorais; diferenças de expressão gênica entre dois tipos tumorais, etc.). Uma nova classificação dos tumores, baseada na expressão gênica e não no perfil histológico, está sendo desenvolvida a partir de dados resultantes das pesquisas com microarranjos. Esta classificação promete trazer grande progresso no desenvolvimento de novos tratamentos para o câncer, uma vez que tumores de aparência histológica semelhante podem ter comportamento clínico diferente e mostrar diferenças na resposta à quimioterapia (GOLUB et al. 1999; RAMASWAMY e GOLUB 2002).

O sistema de arranjos é composto de um substrato, que pode ser uma lâmina de vidro (convencionalmente chamada de microarranjo ou, em inglês, *microarray*) ou membrana de nitrocelulose (chamada de macroarranjo ou *macroarray*), no qual milhares de clones de cDNAs ou oligonucleotídeos são depositados mecanicamente em forma de pontos (“*spots*”). A este substrato são adicionadas as sondas de cDNAs marcados com fluorocromos (para uso em lâminas de vidro) ou isótopos radioativos (para uso em membranas de nitrocelulose) sintetizadas por transcrição reversa a partir do mRNA extraído das células que se deseja estudar. Após a hibridização da sonda ao substrato (aqui, os termos são colocados de acordo com a denominação utilizada no primeiro trabalho sobre microarranjos, de SCHENA et al. (1995). No entanto, em muitos trabalhos denomina-se “sonda” o material fixado na lâmina ou membrana e “substrato” o material marcado com fluorocromos ou isótopo radioativo), a imagem dos arranjos é digitalizada e passa por uma análise computacional para a quantificação da expressão de cada uma das sequências impressas.

Estes ensaios fornecem uma quantidade muito grande de dados. Transformar estes dados em informações relevantes é um grande desafio, em grande parte porque os métodos matemáticos e estatísticos utilizados não são padronizados ou então inadequados a certos desenhos experimentais.

Embora a técnica de microarranjos se embase em conceitos simples, ela oculta uma série de problemas. É uma tecnologia ainda sob desenvolvimento e padronização.

1.5.2 Variabilidade nos experimentos de microarranjos

A maior parte dos laboratórios utiliza diferentes plataformas e parâmetros de análise na realização dos experimentos, o que torna difícil a comparação entre os trabalhos. Muitos problemas envolvendo o desenho e variabilidade experimentais devem ainda ser estudados e resolvidos, uma vez que cada etapa do processo é uma potencial fonte de erro (MURPHY 2002).

Assim, a análise dos dados passa por diversas etapas que visam corrigir os erros que possam ter ocorrido. Uma limitação importante à acurácia da análise dos microarranjos é o erro causado por diferenças na incorporação e emissão de fluorescência entre os corantes utilizados. Muitos estudos mostram que diferentes corantes se ligam de forma diferente à amostra devido às suas características químicas. Além disso, a configuração do leitor (*scanner*) também pode afetar a mensuração da intensidade dos sinais de cada fluorocromo (LYNG et al. 2004), dificultando a distinção entre efeitos biológicos verdadeiros e artefatos. Este viés pode introduzir um grande número de sinais falso-positivos e falso-negativos, caso esta diferença não seja corrigida (ROSENZWEIG et al. 2004). Dentre as estratégias utilizadas para obter esta correção (chamada de “normalização”), a mais comum é a replicata com inversão de corantes (“*dye-swap*”). Embora este recurso seja, na maioria das vezes, suficiente para corrigir o erro induzido pela incorporação dos fluoróforos, existem críticas ao seu uso, embasadas no aumento no custo do experimento e na quantidade de RNA necessária (ROSENZWEIG et al. 2004). Outra alternativa, também cara, é a

repetição de hibridizações (replicatas) de um mesmo conjunto de amostras, que pode aumentar a confiança na mensuração da expressão de cada gene (ROSENZWEIG et al. 2004).

Múltiplos níveis de replicata são comuns em experimentos de microarranjos que utilizam dois corantes. As *replicatas biológicas* consistem de RNAs extraídos de múltiplos indivíduos de uma população e são importantes na redução da variância da estimativa da diferença média e no aumento do poder de detecção de diferenças biologicamente importantes. Em um estudo que busca diferenças entre um tratamento e um grupo-controle não tratado, as replicatas biológicas são muito importantes, já que a estimativa da variância dentro de cada grupo é essencial para determinar se há diferenças significantes entre os diferentes grupos. As *replicatas técnicas*, que podem ser, por exemplo, replicatas de um mesmo gene dentro de um *array* ou presença de genes-controle, aumentam a precisão e a confiança da medida de intensidade da fluorescência (WEI et al. 2004). A normalização baseada em genes-controle considera que as duas amostras hibridizadas contêm a mesma quantidade de moléculas do gene controle. Assim, a razão das intensidades dos dois canais pode ser ajustada para um e então são calculados limites de confiança para a identificação de genes diferencialmente expressos (QUACKENBUSH 2001).

Muitos estudos realizam a normalização intra-lâmina de forma global, assumindo que os efeitos dos erros são os mesmos ao longo de toda a plataforma. Porém, a deposição robótica das sequências de cDNA nas lâminas exerce influência na qualidade do sinal fluorescente, que pode variar

nos diferentes subarranjos devido a diferenças na deposição realizada por agulhas distintas (MARY-HUARD et al. 2004). Este efeito pode ser explicado de diversas formas: pelo tempo em que a agulha fica mergulhada na placa contendo as amostras de cDNA, o tempo em que as agulhas ficam em contato com a lâmina, a velocidade com que as agulhas se movem, concentração e salinidade das soluções de cDNA, temperatura e umidade relativa na cabine do robô, características físico-químicas das agulhas e da lâmina. Assim, por provocar alterações estatisticamente significantes, este fator deve ser considerado nos processos de normalização (MARY-HUARD et al. 2004). O processo de determinação dos valores do fator de calibração que é utilizado na calibração da razão de expressão para cada gene depende das características dos dados para cada agulha. Por exemplo, algumas agulhas apresentam efeitos não-lineares, enquanto outras agulhas do mesmo experimento se comportam de maneira diferente. Assim, para corrigir estes efeitos, pode ser feita uma normalização que considera as diferentes subregiões das lâminas. Geralmente é utilizado um método de regressão local (ex: LOWESS ou LOESS) que partem de um gráfico de dispersão (*scatter-plot*) das intensidades dos sinais dos dois corantes para obter um fator de calibração. Comparada com outras técnicas, a normalização baseada no *scatterplot* é mais robusta (BERGER et al. 2004).

1.5.3 Análise estatística dos dados de microarranjos

Após a normalização dos dados, é feita a análise estatística, que é, provavelmente, o problema mais difícil do uso da tecnologia de

microarranjos, uma vez que estes ensaios, ao invés de envolver poucas variáveis e muitas amostras, envolvem muitas variáveis e geralmente poucas amostras. Para extrair informações biológicas relevantes de um apanhado de variabilidade e “ruído” podem ser utilizados diversos métodos como “*fold change*”, estatística univariada (ex: teste-*t*), análise de variância (ANOVA), entre outros (BERGER et al. 2004).

1.5.4 A técnica de microarranjos de cDNA e a distinção molecular entre carcinoma folicular e adenoma

Diversos estudos mostram que é possível a distinção molecular entre adenoma e carcinoma folicular com base no perfil de expressão gênica e/ou proteica destas lesões (BARDEN et al. 2003; CERUTTI et al. 2004; WEBER et al. 2005). No entanto, estes trabalhos ainda necessitam de validação e, além disso, utilizam amostras de fragmentos tumorais pós cirúrgicos, o que inviabiliza sua utilização para fins diagnósticos pré-operatórios. A aplicação de um teste molecular em amostras pré-cirúrgicas tem como principal fator limitante a pequena quantidade de material obtido por punção da lesão.

O RNA necessário para os ensaios de microarranjos pode ser extraído de células prontamente isoladas de tecidos frescos (cirurgias, biópsias), punção por agulha fina ou de culturas de linhagens celulares de interesse. Há vantagens e desvantagens para cada uma destas fontes celulares. A análise de tumores excisados, embora teoricamente permita uma comparação fiel em relação às características deste tumor *in vivo*, pode estar sujeita a erros devido à contaminação da amostra com tipos celulares

diferentes, dificultando o estudo do perfil de expressão gênica de um tipo celular específico. A técnica de aspiração por agulha fina, além de ser minimamente invasiva, possibilita também o estudo do curso natural da doença já que o mesmo tumor de um mesmo indivíduo pode ser analisado diversas vezes em diferentes espaços de tempo. Esta técnica tem a desvantagem de fornecer uma quantidade muito limitada de material, sendo necessária a amplificação do RNA *in vitro* (WANG et al. 2000), já que os experimentos de microarranjos requerem quantidades relativamente grandes de RNA total, geralmente maiores que 20 µg (PETALIDIS et al. 2003) e o rendimento de RNA extraído de amostras de PAAF é da ordem de nanogramas. O ideal é comparar resultados de ensaios realizados com células obtidas por diferentes métodos, estabelecendo assim as semelhanças e diferenças entre estas abordagens.

1.6 DADOS PRELIMINARES

A partir do RNA extraído de fragmentos de peças cirúrgicas e utilizando a técnica de microarranjos de cDNA, nosso grupo determinou o perfil de expressão de 4600 genes em amostras de carcinoma folicular de adenoma da tireoide (trabalho de doutorado da aluna Beatriz Simonsen Stolf, processo Fapesp nº 99/05299-5) no intuito de identificar as diferenças na expressão gênica das duas patologias.

Devido à grande similaridade entre os perfis de expressão dos dois grupos de amostras e do número de genes da plataforma utilizada, não foi

possível classificar as amostras por agrupamento hierárquico, tanto com base na expressão global quanto no perfil de expressão de genes selecionados em função de sua expressão diferencial. Então, foi utilizado um procedimento de aprendizagem supervisionado para identificar assinaturas moleculares de amostras individuais e, com base nestas assinaturas, realizar a classificação das amostras. Talvez a maior vantagem do uso de algoritmos de aprendizado supervisionado seja a possibilidade de criação de um banco de dados contra o qual a amostra-teste poderia ser classificada, conforme demonstrado por RAMASWAMY et al. (2001). Um grupo conhecido de amostra é utilizado para o treinamento e com o classificador resultante é então feita a predição de classe de uma amostra desconhecida, conforme descrito por GOLUB et al. (1999). Assim, foi utilizada a análise discriminante de Fisher e, por busca sequencial, foi possível identificar mais de 100 trios de genes capazes de separar as amostras de carcinoma folicular e adenoma. Destes, 14 trios foram capazes de distinguir 100% das amostras. O **Quadro 1** mostra os trios de classificadores identificados neste trabalho realizado previamente por nosso grupo.

Outra vantagem do uso de classificadores baseados em assinaturas de expressão é que estes, ao contrário daqueles baseados na análise comparativa, não variam em sua composição caso as amostras em análise sejam alteradas, embora possam ter sua habilidade de predição afetada quando a amostra difere na expressão coordenada dos genes. Este problema pode ser contornado através da utilização de um número maior de classificadores compostos por genes diferentes, reduzindo assim a

probabilidade de uma amostra escapar de todos os testes. É importante salientar que apenas uma fração dos genes que compõem os trios fazem parte do conjunto de genes identificados como diferencialmente expressos (dado não mostrado), apoiando a idéia de que, especialmente para amostras biologicamente semelhantes, a busca por diferenças no perfil de expressão pode não ser a estratégia adequada.

A significância dos achados aqui descritos foi avaliada através de diversos métodos matemáticos, uma vez que não foi possível obter um número de amostras satisfatório para a validação biológica.

Primeiramente, todos os 14 trios testados obtiveram 100% de eficiência na análise do tipo “*leave-one-out*”, tanto para encontrar o mesmo trio classificador quando $n-1$ amostras foram utilizadas quanto para classificar corretamente a amostra deixada de fora. Conforme já mencionado, este tipo de classificador pode ter sua performance comprometida pela variação na expressão gênica observada em diferentes amostras. Então, foi feita a combinação dos trios de modo a construir classificadores com dois ou mais trios e a eficiência em classificar corretamente as amostras foi maior que 90% quando combinados 11 ou mais trios. Finalmente, a identificação destes trios foi também validada por *bootstrap*, indicando um p-valor igual a 0,04, sugerindo que estes achados não ocorreram ao acaso e realmente refletem informações significantes de diferenças na assinatura de expressão gênica.

Quadro 1 - Trios classificadores e seus respectivos genes.

Trios	Gene1	Gene2	Gene3
1	FLJ10305 ^a	AXL	JPH3
2	GRCC9	BBS2	RBM9
3	DUSP22	SLC5A6	H41
4	UBE1C	SMCR8	HERC4
5	AXL	ZFYVE21	KIAA1244
6	MYH10	GRCC9	KIAA1463
7	SLC5A6	ZFYVE21	HK3
8	OLFML3	DUSP22	CLDN2
9	LZTS1	SLC5A6	DUSP22
10	RIT2	DUSP22	CLDN2
11	SYCP1	MEP50	INSIG2
12	SNAP25	SNAI2	NDRG2
13	PPP3CA	SLC5A6	DUSP22
14	FSTL1	COL15A1	FLJ12572

^a) Genes em negrito são aqueles que fazem parte do conjunto de genes identificados como diferencialmente expressos e tem p-valor <0.01 para a expressão diferencial em adenomas e carcinomas foliculares.

Os classificadores identificados, se validados biologicamente, podem ser utilizados para a discriminação dos dois tipos tumorais utilizando o RNA extraído de células colhidas por PAAF, evitando que pacientes com diagnóstico citológico inconclusivo sejam submetidos à tireoidectomia para a obtenção do diagnóstico anátomo-patológico definitivo. Deste modo, a cirurgia seria realizada apenas quando a presença do tumor maligno estivesse confirmada, ou seja, apenas na minoria dos casos. A **Figura 2**

mostra dois trios que separam perfeitamente as amostras de adenoma e carcinoma folicular.

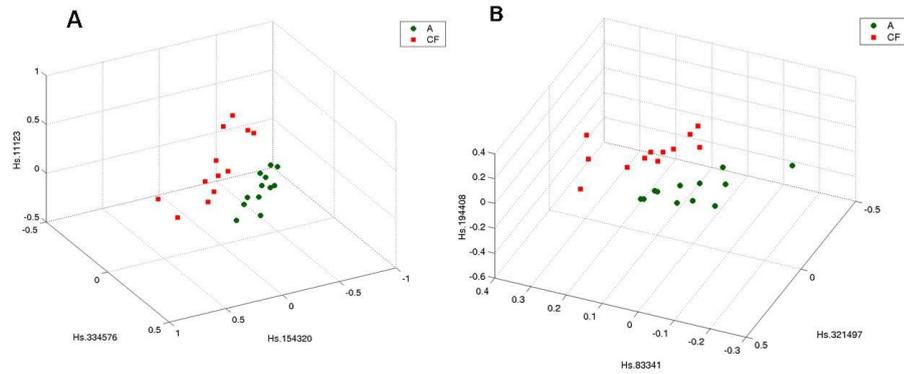


Figura 2 - Representação tridimensional de trios classificadores. Nas figuras 1A e 1B estão representados dois trios classificadores baseados na assinatura de expressão gênica de amostras individuais. Adenomas estão representados por círculos verdes e carcinomas foliculares por quadrados vermelhos.

2 JUSTIFICATIVA

O carcinoma folicular de tireoide é um tumor endócrino com prognóstico favorável se o tratamento for realizado em uma fase não muito avançada da doença. No entanto a distinção clínica entre o carcinoma folicular da tireoide e o adenoma ainda é um desafio. Embora a excisão cirúrgica seja atualmente recomendada em todos os casos onde há dúvida quanto ao diagnóstico de adenoma e carcinoma folicular, apenas 30% dos pacientes submetidos à cirurgia apresenta o tumor maligno, indicando que esta abordagem pode ser desnecessária em grande parte dos casos, já que a maior parte dos adenomas não requer tratamento cirúrgico, salvo em casos em que causam compressão, ocasionando problemas respiratórios e de deglutição. Mais importante, os resultados citológicos falso-negativos podem levar a um atraso no tratamento do carcinoma folicular, com consequências potencialmente graves ao paciente (WEBER et al. 2005).

Atualmente, o principal método de avaliação dos nódulos tireoideanos é a PAAF. Conforme mencionado, embora este método apresente excelente precisão no diagnóstico de carcinomas anaplásicos, medulares e papilíferos, não é capaz de distinguir entre o carcinoma folicular e o adenoma folicular. Assim, o valor preditivo da análise citológica de amostras colhidas por PAAF não é alto o bastante para determinar se um paciente deve ou não ser submetido à cirurgia (SMITH et al. 2005). A alta prevalência de nódulos, a baixa frequência de malignidade nos mesmos, a considerável morbidade

pós-operatória e o custo significativo da cirurgia demonstram a urgente necessidade da identificação pré-cirúrgica precisa dos tumores malignos (WEISS et al. 1996), o que justifica o presente projeto, que tem como objetivo principal o desenvolvimento de um método eficaz de distinção molecular destes dois tumores que possa ser realizado a partir de células colhidas por PAAF, diminuindo o número de tireoidectomias diagnósticas, com a enorme vantagem de ser minimamente invasivo, não oferecendo riscos ao paciente e podendo ser implementado sem maiores problemas na rotina clínica.

3 OBJETIVOS

3.1 OBJETIVO GERAL

A partir do RNA de células colhidas por PAAF de nódulos com suspeita de adenoma ou carcinoma folicular, validar trios de genes selecionados em estudo prévio capazes de distinguir os dois tumores através da análise da expressão destes genes nestas amostras, utilizando a técnica de microarranjos de cDNA, no intuito de desenvolver uma ferramenta diagnóstica pré-operatória eficaz e minimamente invasiva.

3.2 OBJETIVOS ESPECÍFICOS

a. Aperfeiçoar o protocolo de amplificação de RNA de modo a amplificar quantidades iniciais de 300ng e 30 ng de RNA total extraído de pequenas amostras aspiradas de nódulos tireoideanos com suspeita diagnóstica de adenoma ou carcinoma folicular, coletadas antes ou após a cirurgia.

b. Estabelecer a metodologia para utilização de lâminas espelhadas, aparentemente mais sensíveis quando a massa de aRNA disponível é baixa.

c. Validar os trios de genes identificados no trabalho de doutorado da aluna Beatriz Stolf (proc FAPESP no. 99/05299-5) que foram capazes de separar molecularmente amostras de carcinoma folicular de amostras de adenoma. Esta validação será realizada com amostras independentes através da hibridização do RNA extraído (a) de células obtidas por punção de peças cirúrgicas com diagnóstico anátomo-patológico de adenoma ou carcinoma folicular e (b) de amostras de carcinoma folicular e adenomas colhidas por PAAF pré-operatória, também através da técnica de microarranjos de cDNA.

3 DESENHO EXPERIMENTAL

Através de análises de microarranjos de cDNA, STOLF et al. (2006) identificaram 65 trios de genes que separaram precisamente amostras de adenoma e carcinoma folicular. Para validar de trios de genes que possam ser usados no diagnóstico molecular destes tumores, foi utilizado um microarranjo de cDNA contendo os trios de interesse, além de sequências necessárias para normalização dos dados.

Na primeira etapa deste trabalho, foram testadas amostras de RNA extraídas de fragmentos de peças cirúrgicas de pacientes submetidos à tireoidectomia com diagnóstico histopatológico de adenoma ou carcinoma folicular. No entanto, para a elaboração de um teste molecular que possa ser usado na rotina, seria necessário o uso de RNA extraído das células obtidas por PAAF. Neste caso, a maior limitação consiste no baixo número de células obtido por esta técnica, implicando em quantidade insuficiente de RNA para a realização dos testes. Uma alternativa é o uso da metodologia de amplificação de RNA, que permite a obtenção de massa satisfatória de mRNA a partir de quantidades mínimas de RNA total. No entanto, a massa de RNA total comumente utilizada no protocolo de amplificação (3µg) não é obtida em amostras colhidas por PAAF devido ao baixo número de células presentes no aspirado. Para permitir a utilização do RNA extraído das células colhidas por PAAF, o protocolo de amplificação de RNA utilizado em nosso laboratório foi otimizado de forma a partir de quantidades iniciais

iguais a 300ng e 30ng de RNA total. Esta padronização possibilitou então a realização de testes com o material obtido por aspiração.

As amostras foram obtidas de duas formas: após a cirurgia, puncionando-se o nódulo presente na peça cirúrgica, mimetizando a PAAF diagnóstica pré-operatória; e através da punção aspirativa diagnóstica propriamente dita, realizada quando o paciente é atendido no Departamento de Imagem desta instituição. Após a extração do RNA total de cada amostra, este foi amplificado, marcado com fluorocromos e hibridizado em uma lâmina de microarranjos de cDNA. Após a hibridização das amostras, foi realizada captura das imagens e quantificação do sinal fluorescente. Com estes dados, foi feita a análise matemática e estatística dos dados, de modo a verificar se os trios de genes classificadores selecionados eram capazes de classificar corretamente cada amostra em adenoma ou carcinoma folicular.

A **Figura 3** ilustra o fluxograma do desenho experimental.

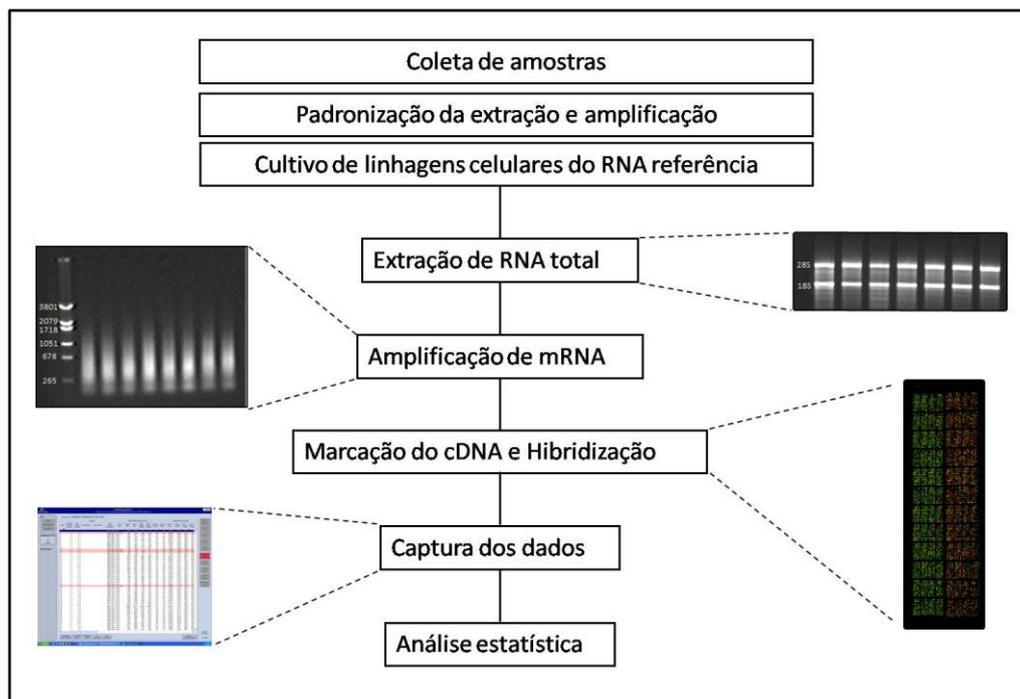


Figura 3 - Fluxograma representando o desenho experimental do presente trabalho.

4 CASUÍSTICA

Neste trabalho foram analisadas 14 amostras, das quais nove amostras tiveram diagnóstico anátomo-patológico de adenoma folicular e cinco amostras de carcinoma folicular. Destas amostras, duas foram coletadas no Hospital A.C. Camargo, dez no Hospital Brasil (Santo André) e duas no Instituto de Enfermedades Neoplasicas (Lima-Peru).

Dados clínicos foram obtidos através de informações fornecidas pela equipe médica responsável pelos pacientes e pela revisão dos respectivos prontuários, assim como dados do laudo anátomo-patológico. No caso de amostras colhidas por PAAF, o diagnóstico final só foi estabelecido após remoção cirúrgica do nódulo.

O **Quadro 2** mostra as amostras de adenoma e carcinoma foliculares utilizadas neste projeto.

Quadro 2 - Amostras de adenoma e carcinoma foliculares a utilizadas na validação dos trios e quartetos classificadores.

Identificação da amostra	Instituição de coleta	Sexo do paciente	Tipo de PAAF	Método de Extração do RNA total	Diagnóstico Anátomo-patológico
AHF012	Hospital Brasil	F	Pós-cirúrgica	PicoPure	Adenoma folicular
AHF020	Hospital Brasil	F	Pós-cirúrgica	PicoPure	Adenoma folicular
AH02947706	A.C. Camargo	M	Pós-cirúrgica	Trizol	Adenoma folicular
AHF004	Hospital Brasil	M	Pós-cirúrgica	PicoPure	Adenoma folicular
AHF21	Hospital Brasil	F	Pós-cirúrgica	PicoPure	Adenoma folicular com células de Hurtle
AHF28	Hospital Brasil	M	Pós-cirúrgica	PicoPure	Adenoma folicular
AH162234007	A.C. Camargo	F	Pós-cirúrgica	PicoPure	Adenoma folicular
JP20	INEN	M	Pós-cirúrgica	PicoPure	Adenoma folicular
JP10	INEN	F	Pré-cirúrgica	Trizol	Adenoma folicular
AHF16	Hospital Brasil	F	Pós-cirúrgica	PicoPure	Carcinoma folicular
AHF19	Hospital Brasil	F	Pós-cirúrgica	PicoPure	Carcinoma folicular
AHF30	Hospital Brasil	M	Pós-cirúrgica	PicoPure	Carcinoma folicular
AHF31	Hospital Brasil	F	Pós-cirúrgica	PicoPure	Carcinoma folicular
AHF32	Hospital Brasil	F	Pós-cirúrgica	PicoPure	Carcinoma folicular

Uma vez que a amostra deve ser coletada antes do exame anátomo-patológico confirmativo, foi necessária a coleta de um grande número de amostras de diagnóstico indeterminado para a obtenção dos casos de interesse, já que a grande maioria das lesões encontradas não correspondeu a casos de adenoma ou carcinoma folicular, embora ambas as enfermidades fossem consideradas como suspeita diagnóstica prévia à cirurgia. Tal fato será discutido com maiores detalhes adiante.

5 MATERIAIS E MÉTODOS

5.1 REAGENTES E EQUIPAMENTOS ESPECÍFICOS

Coleta de amostras e extração do RNA total:

RNA later[®] (Ambion[®] Inc)

PicoPure[®] RNA Extraction kit (Arcturus[®])

TRIzol[®] reagent (Gibco[®])

DEPC (Invitrogen[®])

Clorofórmio (Merck[®])

Etanol (Merck[®])

Isopropanol (Merck[®])

Seringa 1ml (BD Becton Dickinson[®])

Agulha 20G (BD Becton Dickinson[®])

Amplificação:

RNAasin (Promega[®])

dNTP (Eppendorf[®])

Ribomax[®] Large Scale RNA Production (Promega[®])

Transcriptase reversa IMPROM II RT (Promega[®])

Tampão da enzima Improm II a 5x (Promega[®])

RNAse H (USB[®]),

E. coli DNA ligase (Invitrogen[®])

E. Coli DNA Polimerase I (Promega[®])

T4 DNA polimerase (Invitrogen[®])

T7 RNA polimerase (Promega[®])

Fenol (Merck[®])

Álcool isoamílico (Merck[®])

Glicogênio (EMD Chemicals[®])

Acrilamida (Ambion[®])

Oligo comercial: oligo pd(N)6 (Amersham[®])

Oligo sintetizado: oligo dT (24)-T7: (5'-GGC CAG TGA ATT GTA ATACGA CTC ACT ATA GGG AGG CGG TTT TTT TTT TTT TTT TTT TTT TTT-3').

Marcação e hibridização:

Alexa 555 (Invitrogen)

Alexa 647 (Invitrogen)

Cy3-dCTP (GE Healthcare)

Formamida deionizada (Applied Biosystems)

DNA de espermatozoides de salmão a 5 µg/µl (Invitrogen)

Cot 1 DNA(Gibco[®])

Poli(A) DNA (Amersham[®])

QIAQuick[®] PCR Purification kit (Qiagen[®])

DMSO (Invitrogen[®])

RNeasy[®] Midi Kit (Qiagen[®])

Membrana Filtrante GD/X Nylon 0.45 μ m (Whatman[®])

Lâminas HiSens[®] A Nexterion (Schott[®])

Estação de hibridização Gene Tac[®] (Genomic Solutions[®])

leitor confocal a laser ScanArray[®] (Packard Bioscience[®])

Software ScanArray Express[®] (Packard BioScience[®]).

Quantificação e análise de qualidade do RNA:

Espectrofotômetro NanoDrop ND1000 (NanoDrop[®])

Software ND-1000 V3.3.0 (NanoDrop[®])

Marcador de peso molecular 100 pb – 100bp DNA ladder (Invitrogen[®])

Marcador de peso molecular λ *Hind* DNA Ladder (Biolabs[®])

Polaroid[®] GelCam[®] camera system (Polaroid[®])

Agarose (Life Technologies[®])

Brometo de etídeo (Sigma[®])

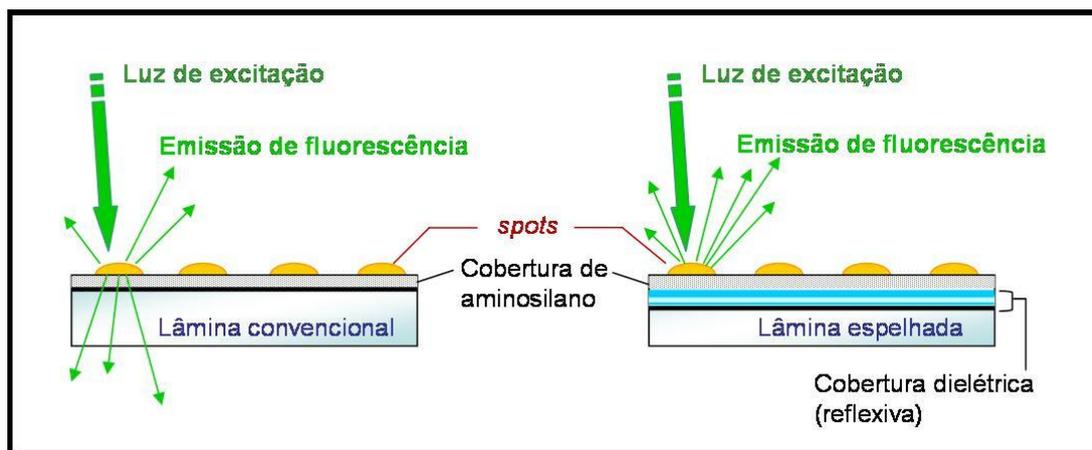
5.2 A LÂMINA DE MICROARRANJOS UTILIZADA

5.2.1 Substrato

As lâminas de microarranjos foram previamente confeccionadas pela *Microarray Facility* do Instituto Ludwig de Pesquisa sobre o Câncer e são constituídas de vidro silanizado com uma camada dielétrica reflexiva (HiSens A, Schott-Nexterion[®]).

A escolha por uma lâmina reflexiva (espelhada) ao invés das tradicionais lâminas de vidro silanizado transparentes deve-se ao fato que, considerando as condições experimentais padronizadas em nosso laboratório, é necessária uma massa inicial de RNA menor do que aquela necessária para a utilização da lâmina de vidro transparente. A celularidade do material colhido por PAAF é muito baixa e, não raro, o rendimento de RNA, mesmo após dois ciclos de amplificação, é insuficiente.

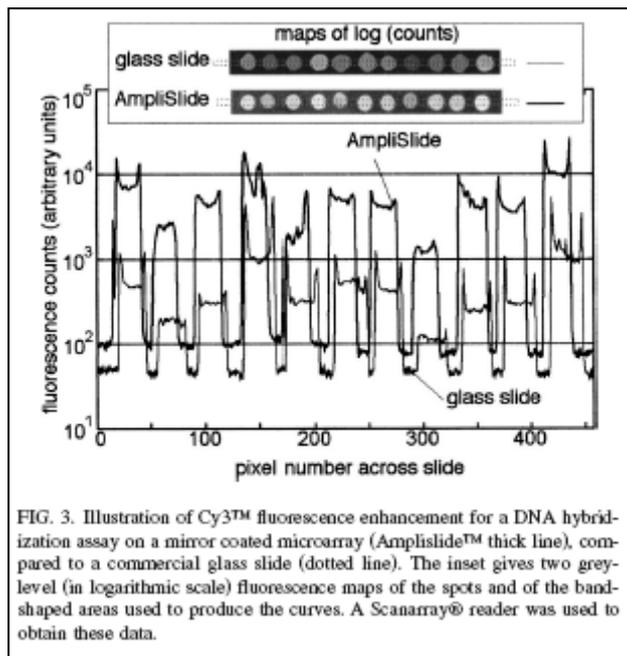
A detecção da expressão gênica é prejudicada quando uma massa inferior a 3 µg de RNA é utilizada porque o sinal de fluorescência torna-se muito baixo. Infelizmente, os *scanners* utilizados não operam em sua máxima eficiência, pois os detectores captam apenas uma pequena fração da fluorescência emitida. Isto acontece porque grande parte da luz utilizada na excitação não é absorvida pelos *spots*, passando através da lâmina de vidro. Além disso, a luz emitida pelos corantes fluoróforos ligados às moléculas de cDNA não é direcionada e parte é perdida pelo verso da lâmina transparente. Recentemente, novas tecnologias tornaram possível aumentar a sensibilidade do ensaio de fluorescência, aumentando o número de fótons detectados, sem aumentar a massa de RNA utilizada. Neste projeto, optamos por empregar uma nova lâmina para microarranjos que contém finas camadas dielétricas (isolantes) imediatamente abaixo da cobertura de aminosilano onde são imobilizados os *spots*. Estas camadas reflexivas fazem com que a fluorescência que seria perdida seja refletida de volta aos *spots* (**Figura 4**), aumentando a excitação luminosa em mais de 6 vezes, preservando ainda a razão entre os corantes fluoróforos.



Fonte: Adaptado de SCHOTT North América (2010)

Figura 4 - Representação gráfica da reflexão provocada pela lâmina espelhada e pela lâmina comum.

CHOUMANE et al. (2005) comparou uma lâmina comercial comum de microarranjos com a versão espelhada. Em ambas, foi depositado o mesmo número de *spots*, em condições idênticas de marcação de DNA, hibridização e lavagem das lâminas. Na **Figura 5** é mostrada a mesma área de duas lâminas, contendo os mesmos *spots* e um gráfico ilustrando a fluorescência emitida por cada *spot* e captada pelo leitor a laser. Observa-se que a fluorescência emitida pelos *spots* na lâmina espelhada é sempre maior do que a emitida pelo *spot* correspondente na lâmina comum.



Fonte: CHOUMANE et al. (2005)

Figura 5 - Comparação entre fluorescência emitida com a utilização lâmina convencional e lâmina espelhada. A figura representa os sinais emitidos por 11 spots da lâmina convencional (linha inferior) e lâmina espelhada (linha superior).

5.2.2 Conjunto de sondas depositadas nas lâminas

A lâmina aqui utilizada carrega 4800 sondas de cDNA, depositadas em sua superfície com o auxílio do robô Flexys® (Genomic Solutions®). As sequências gênicas representadas nesta lâmina foram escolhidas a partir de sequências ORESTES, geradas pelo Projeto Genoma do Câncer (<http://www.compbio.ludwig.org.br/ORESTES>). Estas sequências foram selecionadas de acordo com os seguintes critérios: tamanho entre 300 e 800 pares de bases; ausência de sequência repetitiva; homologia com um único gene no genoma humano; não mais do que 85% de homologia com nenhuma sequência do genoma humano em qualquer segmento de 100 pares de base; o mais 3' possível mas sempre 5' em relação ao primeiro

sítio de poliadenilação; e 100% de homologia com um gene humano de sequência conhecida (BRENTANI et al. 2005).

As sequências imobilizadas na lâmina são as mesmas utilizadas no projeto da aluna Beatriz Stolf (proc. FAPESP 99/05299-5), que identificou os trios classificadores que separam as amostras de adenoma das de carcinoma folicular que são utilizados neste trabalho de validação. Devido a problemas técnicos na deposição das sequência nas lâminas, alguns clones não puderam ser empregados na análise matemática, fato que será discutido adiante.

5.3 COLETA E PRESERVAÇÃO DAS AMOSTRAS

Conforme mencionado, foram obtidas amostras por dois procedimentos distintos: punção aspirativa da peça cirúrgica e punção aspirativa propriamente dita, realizada no paciente durante os procedimentos iniciais de diagnóstico (amostra pré-cirúrgica).

5.3.1 Coleta e preservação das amostras coletadas por punção de peça cirúrgica (punção pós-operatória)

Nos casos de peças cirúrgicas, foi realizada uma punção dos nódulos suspeitos na peça cirúrgica, quando esta era encaminhada à sala de congelação do centro cirúrgico. Logo após a tireoidectomia, o patologista realizava a punção dos nódulos tireoideanos, de modo a mimetizar a PAAF realizada nos exames pré-operatórios. Cada amostra aspirada

(aproximadamente 50µl) foi preservada em tubos contendo 400 µl de RNAlater®, incubada à temperatura ambiente por 24 horas e depois armazenada em freezer a -20°C. Para cada amostra, o patologista responsável pela punção preencheu uma *folha de trabalho* (**Apêndice 1**) com as informações sobre o nódulo puncionado. O restante da peça foi submetida à congelação e exame anátomo-patológico. Após a confirmação do diagnóstico de adenoma ou carcinoma folicular, as amostras armazenadas em RNAlater® foram utilizadas para a extração do RNA com o *kit* PicoPure®. Também foi extraído o RNA total de amostras sem o diagnóstico de interesse, para ser utilizado nas etapas de padronização, poupando as raras amostras de adenoma ou carcinoma folicular.

5.3.2 Coleta e preservação das amostras coletadas por PAAF diagnóstica (punção pré-operatória)

Nas amostras colhidas por PAAF guiada por ultrassom, a coleta foi realizada por médicos do Departamento de Imagem do Hospital A.C. Camargo e todo o aspirado foi encaminhado para exame citológico de modo a não comprometer o diagnóstico do paciente. O material utilizado no projeto foi aquele obtido pela lavagem das agulhas utilizadas na punção após a obtenção do material para diagnóstico. Somente foi utilizado o material que seria descartado pelo médico durante o procedimento normal da PAAF diagnóstica. A lavagem das agulhas foi feita com TRIZOL® e o RNA destas amostras foi imediatamente extraído pelo método deste reagente, uma vez que a preservação em RNAlater® e extração com o *kit* PicoPure® não

mostrou-se eficiente neste tipo de amostra pela presença de grande volume relativo de sangue. A padronização dos métodos de preservação e extração é mostrada ao final deste trabalho, em seção específica.

5.3.3 Aspectos éticos envolvidos na coleta e preservação das amostras

O presente projeto conta com um termo de consentimento livre e esclarecido dos sujeitos da pesquisa (**Apêndice 2**) aprovado pela CONEP. A identidade destes é mantida em sigilo através de codificação.

A escolha dos métodos para a obtenção de material foi feita de modo a não interferir nos procedimentos diagnósticos e de tratamento, sendo utilizado apenas material que, na rotina, seria descartado. Portanto, a coleta de amostras utilizando os procedimentos descritos no item anterior não oferece riscos aos pacientes.

As amostras coletadas foram utilizadas exclusivamente neste trabalho. É previsto que nenhuma amostra ou produto desta será armazenado em período posterior ao término do estudo. Os resultados da pesquisa serão tornados públicos, sejam eles favoráveis ou não.

5.4 EXTRAÇÃO DO RNA TOTAL

A utilização de amostras coletadas por PAAF implicou no uso de métodos distintos de extração do RNA total. Isto porque, em amostras coletadas por PAAF pós-cirúrgica, foi possível utilizar o *kit* PicoPure[®], que

garante maior rendimento e melhor qualidade do RNA obtido. Em amostras pré-cirúrgicas, devido à grande presença de sangue e consequente incompatibilidade com o *kit* PicoPure[®], foi utilizado o método do TRIzol[®], o mesmo empregado no estudo original, porém com algumas alterações, já que as amostras coletadas neste estudo são predominantemente líquidas, enquanto no trabalho de Stolf eram constituídas de fragmentos tumorais sólidos.

5.4.1 Extração do RNA total de amostras pós-cirúrgicas

Foi utilizado o kit de extração de RNA total PicoPure[®]. Este *kit* é recomendado na extração de RNA quando a celularidade da amostra é muito baixa. De acordo com o fabricante, permite a extração de RNA total de boa qualidade de uma quantidade inferior a 10 células. Considerando que as amostras de lesões foliculares contêm um número muito pequeno de células, um método eficiente de obtenção de RNA destas células é muito importante neste projeto.

O protocolo utilizado foi adaptado do protocolo de isolamento de RNA de *pellet* de células descrito no manual do produto. Este protocolo refere-se a células em cultura e sugere o *pellet* de células seja lavado com 1 ml de meio de cultura e centrifugadas. Neste projeto, as células estão preservadas em RNAlater[®]. Foi verificado que, ao centrifugar as células para formar o *pellet*, a amostra não apresentava uma sedimentação completa e algumas células permaneciam aderidas à parede do tubo ou em suspensão. Assim, foi adicionado 1 ml de PBS à amostra, que proporcionou maior

sedimentação da amostra por diminuir a densidade do meio, além de dissolver cristais de RNAlater[®] que formavam-se eventualmente.

Após a adição do PBS, a amostra foi homogeneizada delicadamente por inversão e centrifugada a 3000x g por 5 minutos. O sobrenadante foi descartado, as células foram adicionadas de 100 µl de *Extraction Buffer* (XB) e homogeneizadas por pipetagem, com o cuidado de não formar espuma. Foi feita incubação a 42° C por 30 minutos e centrifugação a 3000x g por dois minutos para sedimentar debris celulares. O sobrenadante foi então transferido para um tubo limpo e foram pipetados 100µl de etanol 70%. O extrato foi então pipetado sobre a coluna do *kit*, acoplada a um tubo de coleta e pré-condicionada com 250 µl de *Conditioning Buffer* (CB). Para que o RNA se ligasse na coluna, esta foi a centrifugada por dois minutos a 100x g e, imediatamente em seguida, centrifugada a 16000x g por 30 segundos para remover o restante da solução. Foram pipetados 100 µl de *Wash Buffer 1* (WB1) na coluna de purificação seguido de centrifugação a 8000x g por um minuto. Em seguida foram feitas mais duas lavagens com 100 µl de *Wash Buffer 2* (WB2) cada, sendo a primeira centrifugação a 8000x g por um minuto e a segunda a 16000x g por um minuto para remover qualquer resíduo do tampão de lavagem. Para eluir o RNA, a coluna foi acoplada em um tubo limpo e foram pipetados 11µl de *Elution Buffer* (EB). Este é o menor volume recomendado pelo fabricante e foi escolhido porque a concentração de RNA esperada é baixa. A coluna foi incubada por um minuto à temperatura ambiente, centrifugada por um minuto a 1000xg para distribuir o

tampão de eluição pela coluna e depois por um minuto a 16000x g para eluir o RNA.

Dos 11 microlitros eluídos, um foi utilizado para quantificação no espectrofotômetro NanoDrop[®], lembrando que a amostra de referência (“branco”) utilizado na calibração deve ser a solução de eluição pura (*Elution Buffer*).

5.4.2 RNA total de amostras coletadas por punção aspirativa pré-cirúrgica

A extração do RNA total das amostras pré-cirúrgicas foi realizada imediatamente após a coleta, uma vez que o uso de solução conservante (RNAlater[®]) não mostrou-se benéfico. Foi utilizado o método do TRIzol[®], seguindo o protocolo do fabricante, com uma pequena modificação, que consistiu na homogeneização da amostra com uma seringa e agulha de calibre 20.

5.5 PREPARAÇÃO DO RNA REFERÊNCIA

O RNA referência é uma mistura de moléculas utilizada como amostra-controle para gerar medidas de razão para diferentes amostras testes, permitindo a subsequente análise das informações obtidas nas hibridizações.

Neste projeto, o RNA referência a ser utilizado nas hibridizações foi um *pool* de RNA extraído de tecido tumoral de peças cirúrgicas de tireoide.

Segundo JEFFREY e POLLACK (2003), a referência ideal é aquela que produz um sinal mínimo em todos os elementos do arranjo, além de ser facilmente produzida e reprodutível, não só no tempo mas também em diferentes laboratórios. Dessa forma, o que mais se aproxima da referência ideal seria um conjunto de linhagens celulares como o utilizado na Universidade de Stanford (PEROU et al. 1999 e 2000, ROSS et al. 2000). Entretanto, as 11 linhagens citadas nestes trabalhos não estão disponíveis no Instituto Ludwig e, portanto, um novo conjunto, composto por 15 linhagens celulares, foi escolhido, tomando o cuidado de selecionar tipos celulares equivalentes. As linhagens escolhidas foram: Daudi (linfoma de Burkitt), DLD-1 (adenocarcinoma de cólon), DU 145 (carcinoma de próstata), FaDU (carcinoma epidermoide de faringe), GM 637 (fibroblasto), H 1080 (fibrossarcoma), H 146 (carcinoma de pulmão), HB4a (célula luminal de mama), HEK 293 (linhagem epitelial de rim embrionário humano), Jurkat (Leucemia Aguda de Células T), Saos-2 (Osteossarcoma), SK-BR-3 (Adenocarcinoma de Mama), SK-MEL-28 (Melanoma), T24 (Carcinoma de Bexiga), T98G (Glioblastoma).

Foi feito o cultivo e extração de RNA de 13 das 15 linhagens citadas. O RNA das 2 linhagens restantes foi gentilmente cedido pela Dra Dirce Carraro, do Laboratório de Análise da Expressão Gênica do Hospital do Câncer- A.C. Camargo.

Uma alíquota de cada linhagem foi descongelada e transferida para um tubo plástico cônico contendo 15 ml de meio de cultura. Os tubos foram centrifugados por 5 minutos a 3000 rpm (centrífuga CELM[®], modelo LS II). O

meio de cultura foi removido, o *pellet* ressuspendido em 10 ml de novo meio de cultura e cada amostra foi transferida para frascos de cultura pequenos (25 cm²) e incubada em estufa a 37°C e 5% CO₂. As células foram examinadas diariamente ao microscópio para verificação da confluência, aparência morfológica e qualquer sinal de contaminação microbiológica até que a cultura atingisse 90% de confluência, momento em que era feita a primeira passagem. A passagem foi feita da seguinte maneira: o meio de cultura foi removido com pipeta Pasteur e bomba de vácuo e as células aderidas ao fundo do frasco foram lavadas com PBS previamente aquecido a 37°C. O PBS foi removido e foi adicionada tripsina. A tripsina foi espalhada sobre a superfície da monocamada celular e imediatamente removida por aspiração. Os frascos foram incubados em estufa a 37°C por um minuto, tempo suficiente para que houvesse o descolamento das células do fundo do frasco. Após o descolamento, foi adicionado meio de cultura previamente aquecido a 37°C, o suficiente para preencher o fundo de 2 frascos de cultura maiores. Após a distribuição da suspensão de células entre os frascos, estes foram incubados em estufa até que a cultura novamente atingisse 90% de confluência, momento no qual era feita nova passagem. No período de incubação das células, o meio de cultura foi renovado em dias alternados.

O RNA foi extraído das culturas celulares (4 frascos de 162 cm² de cada linhagem, com 90% de confluência) utilizando-se o método do TRIzol[®] de acordo com as instruções do fabricante e ressuspendido em H₂O tratada com DEPC. Os RNAs extraídos foram quantificados e depois avaliados através de fracionamento eletroforético em gel de agarose a 1% em

condições desnaturantes. Nos casos onde o RNA não apresentou qualidade satisfatória, a respectiva linhagem celular foi cultivada novamente e o RNA total foi extraído com o RNeasy[®] Midi Kit, já que o tampão de lise do *kit* tem a guanidina em concentração duas vezes maior em relação ao TRIzol[®], podendo degradar RNAses endógenas com maior eficiência. O **Quadro 3** mostra as linhagens celulares e o método de extração de RNA empregado para cada uma.

Quadro 3 - Linhagens celulares selecionadas para compor o RNA referência.

Linhagem	Número ATCC ou outra referência	Descrição	Método de extração de RNA
Daudi	CCL-213	Linfoma de Burkitt	TRIzol [®]
DLD-1	CCL-221	Adenocarcinoma de Cólon	TRIzol [®]
DU 145	HTB-81	Carcinoma de Próstata	TRIzol [®]
FaDU	HTB-43	Ca. Epidermóide de Faringe	TRIzol [®]
GM 637	Kucherlapati et al. (1978)	Fibroblasto	RNeasy [®] Midi kit
H 1080	CCL-121	Fibrossarcoma	TRIzol [®]
H 146	HTB-173	Carcinoma de Pulmão	RNeasy [®] Midi kit
HB4a	Stamps et al. 1994	Célula Luminal de Mama	Cloreto de Césio
HEK 293	CRL-1573	Célula epitelial de rim embrionário humano	RNeasy [®] Midi kit
Jurkat	TIB-152	Leucemia Aguda de Células T	TRIzol [®]
Saos-2	HTB-85	Osteossarcoma	TRIzol [®]
SK-BR-3	HTB-30	Adenocarcinoma de Mama	TRIzol [®]
SK-MEL-28	HTB-72	Melanoma	TRIzol [®]
T24	HTB-4	Carcinoma de Bexiga	TRIzol [®]
T98G	CRL-1690	Glioblastoma	TRIzol [®]

Para montar o *pool*, foram misturados 45 µg de cada RNA. Essa quantidade foi baseada na massa obtida na linhagem com menor rendimento (linhagem H146). Foram feitas alíquotas do *pool* de referência e o RNA armazenado em etanol 100% a -60°C. Este *pool* passou a ser usado na maioria dos projetos de análise expressão diferencial por microarranjos de nossa instituição.

5.6 AMPLIFICAÇÃO DO RNA TOTAL

A amplificação das amostras de RNA deste projeto seria originalmente feita com base no protocolo descrito por GOMES et al. (2003). Este protocolo utiliza a técnica de “*template-switch*” e amplificação guiada pela T7 RNA polimerase. No entanto, neste projeto, o protocolo de GOMES et al. (2003) foi substituído pelo protocolo descrito por VAN GELDER et al. (1990), com algumas modificações. O protocolo descrito por VAN GELDER et al. (1990) não baseia-se na técnica de “*template-switch*” e realiza uma síntese convencional de segunda fita de cDNA. WANG et al. (2003) realizaram uma comparação entre o protocolo baseado em “*template-switch*” 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 recomendado quando a amplificação parte de uma massa muito pequena de RNA. Embora a massa inicial de RNA referência utilizada neste experimento não seja um fator crítico, as amostras colhidas por PAAF apresentam celularidade muito baixa

e, conseqüentemente, massa insuficiente de RNA total, necessitando de um protocolo mais eficiente. Para evitar a inserção de vieses experimentais, será utilizado o mesmo protocolo e o mesmo número de ciclos de amplificação tanto para as amostras quanto para a referência. Foram realizados dois ciclos de amplificação.

É importante salientar que, antes da amplificação das amostras de interesse, o protocolo foi testado e padronizado utilizando as amostras correspondentes a outras patologias, também coletadas por PAAF e com baixa celularidade.

5.6.1 Amplificação do RNA referência

O RNA utilizado como referência nos experimentos de microarranjos é um conjunto de RNAs totais de 15 linhagens celulares. Foram feitas alíquotas do RNA referência na concentração de 450ng/µl armazenadas a -60°C em 3 volumes de etanol 100%.

A Primeiro ciclo de amplificação

Foi amplificada massa inicial de 18 µg de RNA total. Inicialmente, o RNA em etanol foi precipitado através da adição de 1µl de NaOAc (3 M, pH 5,4) e resfriamento a -20°C por 12 horas. Após a centrifugação e lavagem do *pellet* de RNA com etanol a 70%, o *pellet* foi ressuspendido em 12 µl de água tratada com DEPC e este volume foi então dividido igualmente entre seis tubos, para que cada tubo contivesse aproximadamente 3µg de RNA total, quantidade ideal para realização do protocolo de amplificação.

Foi realizada a transcrição reversa do RNA utilizando-se oligo dT (24)-T7, um iniciador que contém a sequência 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 TTT TTT-3'). O RNA total (3µg) foi misturado com 0,5 µg de oligo d(T)24-T7 em um volume final de 11 µl e o RNA foi denaturado a 70°C por 10 minutos e em seguida colocado no gelo. A transcrição reversa foi feita com a adição de 6,65 µl de água tratada com DEPC, 6,4µl de MgCl₂ a 25mM, 5,35µl do tampão da enzima Improm a 5x, 2,6µl de dNTPs a 10mM, 0,6µl de RNaseIn[®] e 1,3µl da transcriptase reversa IMPROM II[®] RT e a reação ocorreu a 42°C por duas horas. A síntese da segunda fita de cDNA foi feita com a adição de 10 µl de *Second Strand Buffer* (ver composição no final deste item) a 5x, 1 µl de dNTPs a 10mM, 0,12 µl de RNaseH a 5U/ µl, 0,5 µl de *E. coli* DNA ligase a 10 U/ µl, 1,3 µl de *E. coli* DNA Polimerase I a 10 U/ µl, seguida de incubação a 16°C por duas horas. Em seguida, foi adicionado 1 µl da enzima T4 DNA polimerase a 5U/ µl e a amostra incubada a 16° C por mais 10 minutos. O cDNA dupla-fita foi então purificado pelo método fenol-clorofórmio e a precipitação foi realizada na presença de 1 µl de acrilamida (5 mg/ml), 10% de NaOAc (3M, pH 5,2) e três volumes de etanol absoluto gelado, por , no mínimo , 12 horas a -20° C. Após a centrifugação e duas lavagens do *pellet* com etanol a 70%, o cDNA foi ressuscitado em 10 µl de água tratada com DEPC. Foi feita então a transcrição *in vitro* utilizando o *kit* RiboMAX[®] Large Scale RNA Production e seguindo as recomendações do fabricante. O tempo de incubação utilizado foi de oito horas a 37° C. O RNA amplificado

foi purificado através do método do TRIzol[®] e ressuspendido em 20 µl de água tratada com DEPC. Deste volume, 2 µl foram utilizados para a quantificação no espectrofotômetro NanoDrop[®].

Composição do *Second Strand Buffer (5x)*: 100 mM Tris pH 6,9, 23 mM MgCl₂, 450 mM KCl, 0,75 mM β-NAD, 50 mM (NH₄) SO₄

B Segundo ciclo de amplificação:

Apesar da massa de RNA referência obtida no primeiro ciclo de amplificação ser suficiente para a realização das hibridizações, é necessária a realização de um segundo ciclo de amplificação para limitar a introdução de vieses em relação às amostras utilizadas, que obrigatoriamente terão que passar por dois ciclos de amplificação em vista da baixa celularidade das mesmas.

Apenas parte do material obtido após o primeiro ciclo de amplificação foi utilizado para a realização do segundo ciclo, pois esta massa de RNA é suficiente para a obtenção da massa de RNA desejada ao final do segundo ciclo. Foram feitas nove reações. Em cada uma delas, a massa inicial de RNA amplificado (um ciclo) foi igual a 800 ng.

Em cada tubo, foi adicionado 1,3 µl de dN6 *random hexamer* (2 µg/µl) e água tratada com DEPC q.s.p. 9 µl. O RNA foi denaturado a 70°C por 10 minutos e em seguida colocado no gelo por dois minutos. A transcrição reversa foi feita com a adição de 6,4µl de MgCl₂ a 25mM, 5,35µl do tampão da enzima IMPROM[®] a 5x, 2,6µl de dNTP a 10mM, 0,6µl de RNaseIn[®] e

1,3µl da transcriptase reversa IMPROM[®] II RT e a reação ocorreu a 42°C por 90 minutos.

A síntese da segunda fita de cDNA foi feita da seguinte forma: inicialmente, foi utilizado 0,2 µl de RNase H (5U/µl), incubação a 37°C por 10 minutos, desnaturação a 95°C por dois minutos e resfriamento em gelo. Em seguida, foi adicionado 0,3 µl de Oligo dT24-T7 (0,5 µg/µl), 42°C por 10 minutos. Foram então adicionados 10 µl de *Second Strand Buffer* a 5x, 1 µl de dNTP a 10mM, 0,5 µl de *E. coli* DNA ligase a 10 U/ µl, 1,3 µl de *E. coli* DNA Polimerase I a 10 U/ µl, seguida de incubação a 16°C por 2 horas. Imediatamente foi pipetado 1 µl da enzima T4 DNA polimerase a 5U/ µl e a amostra novamente incubada a 16° C por 10 minutos. As etapas de purificação do cDNA e transcrição *in vitro* e purificação do aRNA foram feitas conforme descrito no protocolo do primeiro ciclo de amplificação. No entanto, a amostra foi ressuspensa em 30 µl de água tratada com DEPC e 2 µl foram utilizados para a quantificação no espectrofotômetro NanoDrop[®].

5.6.2 Amplificação das amostras (pré e pós-cirúrgicas)

A amplificação das amostras coletadas por punção de peça cirúrgica e preservadas em RNAlater[®] foi realizada seguindo o mesmo protocolo descrito anteriormente, porém partindo de quantidades menores de RNA total.

A massa de RNA total mínima foi igual a 20ng para amostras pós-cirúrgicas e 40ng para amostras pré-cirúrgicas.

5.7 MARCAÇÃO INDIRETA DAS AMOSTRAS E REFERÊNCIA

A marcação do cDNA pode ser direta ou indireta. Na marcação direta, são utilizados desoxinucleotídeos modificados com fluoróforos, como por exemplo, dCTP marcado com Cy3 ou Cy5. Estes nucleotídeos modificados são incorporados ao cDNA durante a síntese da primeira fita, a partir da transcrição reversa de uma fita-molde de RNA.

Na marcação indireta do cDNA, são utilizados nucleotídeos modificados com aminoalil, como 5-(3-aminoalil)-dUTP (aa-A-dUTP), incorporados ao cDNA na reação de transcrição reversa do RNA. Somente então as amostras marcadas com aminoalil são acopladas ao fluoróforo (ex: Cy3, Cy5, Alexa 555, Alexa 647 etc.). Por ser menor, a molécula de aa-dUTP é incorporada muito mais facilmente do que os nucleotídeos que já são associados ao fluoróforo, tornando a marcação mais uniforme, maximizando a intensidade de sinal.

Embora ligeiramente mais complexa, a marcação indireta com aminoalil tem apresentado maior eficiência e consistência, a custos menores (t'HOEN et al. 2003), sendo o método aqui empregado.

A primeira etapa no processo de marcação do cDNA é a transcrição reversa do RNA amplificado. Para isso, foi utilizada massa de 11,5 µg de RNA poli-A amplificado de cada amostra e 9 µg de RNA-referência. Como cada amostra foi posteriormente hibridizada com o RNA-referência, em duplicata com inversão de corantes (*dye-swap*), para cada amostra foram feitas duas lâminas, *main* e *swap*. Na lâmina *main*, a amostra foi marcada

com o fluoróforo Alexa 555 e a referência com o Alexa 647. Na lâmina *swap* ocorre o inverso.

Devido à duplicata, cada lâmina recebe metade do material transcrito, ou seja, cada lâmina será hibridizada com o produto da transcrição de 5,75 µg de RNA poli-A de uma amostra (metade da massa inicial de 11,5µg) e com 4,5 µg do produto da transcrição do RNA poli-A referência. Desta forma, através de uma mesma reação de transcrição reversa, é obtido material para as duas lâminas, reduzindo os vieses que poderiam ser introduzidos entre as duplicatas, caso as amostras fossem transcritas separadamente. A massa de RNA das amostras foi escolhida baseada no rendimento total da amostra de menor rendimento.

Inicialmente, foi prevista a utilização de massa de RNA amplificado em torno de 1 a 3 µg. No entanto, após uma série de testes, foi constatado que estas amostras não apresentam qualidade satisfatória para uso em quantidade tão pequena em experimentos de microarranjos. Adicionalmente, devido a problemas relacionados à logística da *Microarray Facility* do Instituto Ludwig de Pesquisa sobre o Câncer, foi impossível a confecção das lâminas em período próximo à etapa de hibridização, diminuindo a eficiência das lâminas, havendo a necessidade do emprego de maior massa de cDNA, visando à hibridização no maior número de *spots* possível. A transcrição reversa foi feita da seguinte forma: o RNA (11,5 µg para amostras ou 9 µg para referência) foi misturado com 0,5 µg de oligo pd(N)6 a 5 µg/ µl em um volume final de 15,4 µl, o RNA foi denaturado a 70°C por 10 minutos e em seguida colocado no gelo. A transcrição reversa

foi feita com a adição de 9,6 µl de MgCl₂ a 25mM (concentração final: 6mM), 8µl do tampão da enzima IMPROM[®] II a 5x, 4 µl de uma mistura de dNTPs contendo aminoalil-dUTP e de baixa concentração de timina (concentração dos nucleotídeos mostrada na **Quadro 4**), 1 µl de RNasin[®] e 2 µl da transcriptase reversa IMPROM[®] II RT e a reação ocorreu a 42°C por duas horas.

Quadro 4 - Concentração dos nucleotídeos na mistura utilizada para marcação indireta do cDNA.

Nucleotídeo	Concentração final
dATP	5 mM
dCTP	5 mM
dGTP	5 mM
dTTP	2 mM
Aminoallyl dUTP	3mM

Após a síntese do cDNA, a amostra foi incubada a 70° C por 20 minutos na presença de 20 µl de solução de NaOH 1M e EDTA 0,25M para a degradação do RNA-molde. A reação foi neutralizada com 20 µl de HCl 1M. Foi feita precipitação do cDNA na presença de 300 µl de isopropanol, 2 µl de glicogênio a 20 ng/µl, 2 µl de acrilamida a 100 ng/µl e 40 µl de NH₄Ac a 7,5 M por 20 minutos à temperatura ambiente. A amostra foi centrifugada a 12000x g por 30 minutos para a formação do *pellet* e o sobrenadante foi descartado. O *pellet* foi lavado duas vezes em etanol a 70% e uma vez em etanol absoluto para facilitar a secagem do pellet, que foi ressuspensionado em 16 µl de tampão de bicarbonato de sódio a 0,0625 M, pH 9,0, diluído no momento do uso. A amostra foi então dividida em duas alíquotas de 8 µl

cada uma (*main* e *swap*). O Alexa 555 e Alexa 647 foram então preparados com a adição de 5 µl de DMSO a cada tubo de fluorocromo (tomando o cuidado para protegê-los da luz), seguido de forte agitação (vortex) durante 15 segundos. Os 5 µl de corante foram imediatamente adicionados às amostras (uma alíquota de amostra foi marcada com Alexa 555 e a outra com o Alexa 647) e feita incubação em temperatura ambiente por duas horas no escuro. De modo a remover fluorocromos não incorporados e fragmentos muito pequenos de cDNA, a reação foi purificada com o QIAQuick[®] PCR Purification kit, podendo ser substituído pelo PureLink[®] PCR Purification Kit (Invitrogen). Na purificação, a amostra foi adicionada de 40 µl de NaOAc a 100 mM e pH 5,2 e 500 µl de *Binding Buffer* (PBI), misturando-se delicadamente. O cDNA marcado foi aplicado no centro da coluna de purificação do *kit*, seguido de centrifugação a 10000x g por um minuto. O filtrado foi descartado e foram adicionados 750 µl de etanol 80% à coluna. Após nova centrifugação a 10000x g por um minuto, o filtrado foi descartado e foi feita mais uma lavagem com 750 µl de etanol 80%. Após descartar o filtrado, a coluna foi centrifugada novamente para remover qualquer resíduo de etanol e transferida para um tubo limpo de 2 ml. Foram adicionados 50 µl de *Elution Buffer* (EB) aquecido a 70° C ao centro da coluna e após dois minutos foi feita a centrifugação a 10000x g por um minuto para eluir o cDNA. A eluição foi repetida com mais 50 µl de EB e a amostra final (100 µl) marcada com Alexa 555 foi misturada com a referência marcada com Alexa 647, assim como a amostra marcada com Alexa 647 foi misturada com a referência marcada com Alexa 555. Após a mistura, foram adicionados 2 µl

de *poly A DNA* (2 µg/µl), 2 µl de *Cot 1 DNA* (1 µg/µl) e o volume da amostra foi reduzido por centrifugação a vácuo até o volume de 12,5 µl. Foram adicionados 40 µl de solução de hibridização a 2X (SSC 1X, SDS 0,2%, solução de Denhardt 10X) aquecida a 70°C, 50 µl de formamida deionizada e 2 µl de DNA de esperma de salmão a 5 µg/µl e a amostra foi incubada a 45°C até o momento do uso.

5.8 HIBRIDIZAÇÃO EM LÂMINA ESPELHADA

Com os cDNAs já sintetizados, iniciou-se a etapa de hibridização das amostras. Quatro horas antes da hibridização, a lâmina de microarranjos foi incubada em solução contendo SSC 5x, SDS 2%, BSA 1% e solução de Denhardt a 10% (tudo filtrado em membrana de 0,45 µm e pré aquecido a 42°C). A solução de pré-hibridização contém uma grande quantidade de albumina sérica bovina, que se liga fisicamente a grupos reativos presentes na superfície da lâmina, promovendo o bloqueio físico destes componentes. Do contrário, haveria *background* elevado e/ou manchas em forma de cauda de cometa. Antes de realizar a pré-hibridização das amostras, a etiqueta com o código de barras de cada lâmina foi parcialmente removida com o auxílio de um bisturi, deixando-se apenas a identificação numérica, que se encontra na extremidade da lâmina. Este procedimento foi imprescindível para evitar vazamentos na hibridização, já que a etiqueta impedia a vedação da câmara. Após o período de pré-hibridização, as lâminas foram lavadas com auxílio de pinça, mergulhando-as dez vezes em água deionizada

aquecida a 42°C. Após a lavagem, cada lâmina foi colocada em um tubo cônico de 50 ml limpo e seco e centrifugada a 1000x g por três minutos para secagem e evitar marcas de água na lâmina.

A lâmina foi então colocada na estação de hibridização, aquecida a 70°C por dois minutos e esta temperatura foi mantida até a aplicação da amostra de cDNA marcado.

Antes de ser colocado na câmara de hibridização da estação, o cDNA marcado foi aquecido a 95°C por cinco minutos, centrifugado brevemente e mantido a 70°C. A mistura de cDNA foi então aplicada sobre a lâmina na câmara de hibridização e mantida a 42°C por 20 horas sob agitação e ao abrigo da luz. Após este período o cassete de hibridização foi desmontado e foi constatado que houve vazamento em dois deles. Após desmontar os cassetes, as lâminas foram imediatamente imersas em solução SSC 2x a 42°C.

As lâminas foram então encaixadas em estante de vidro próprias para a lavagem, submersas em solução SSC 2x sob agitação por cinco minutos. Em seguida foram feitas duas lavagens de 10 minutos cada em solução contendo SSC 0,1x e SDS 0,1% e mais uma lavagem em SSC 0,1x, tudo a 42°C. Os recipientes utilizados para lavagens com SDS não podem ser usados para a última lavagem com SSC, pois resíduos de SDS podem interferir na qualidade da imagem captada pelo *scanner*.

Após estas lavagens, a lâmina foi colocada em tubo cônico de 50 ml e centrifugada por dois minutos a 1000x g. A lâmina foi retirada imediatamente do tubo cônico após a centrifugação. Este detalhe é especialmente

importante para evitar manchas causadas pelo vapor que se forma quando o tubo permanece fechado, ainda que por alguns instantes. Outro detalhe que minimiza as manchas de vapor, muito comuns em lâminas espelhadas, é a centrifugação em centrífuga refrigerada a 4°C, pois a baixa temperatura implica em menor formação de manchas causadas pelo vapor na superfície da lâmina imediatamente após a centrifugação.

As lâminas foram mantidas ao abrigo da luz até o momento da captura das imagens.

5.9 CAPTURA DAS IMAGENS

Depois de secas, as lâminas passaram pelo processo de captação de imagens. Foi utilizado o leitor confocal a *laser* ScanArray® para captar os sinais de intensidade de fluorescência emitidos pelas amostras hibridizadas em cada *spot*. São usados filtros específicos para captar separadamente o sinal de cada corante (550 nm para Alexa 555 e 650 nm para Alexa 645). Para efeito de visualização, os sinais de intensidade são convertidos em uma escala de cor representados por verde para o canal 1 (555) e vermelho para o canal 2 (645). As duas imagens (uma correspondente ao canal 1 e a outra ao canal 2) são sobrepostas, tornando possível a observação da expressão diferencial entre elas. A quantificação da intensidade do sinal de luz emitido de cada *spot* foi feita utilizando-se o *software* ScanArray Express®.

As quantificações das marcações de *spot* e *background* foram obtidas pelo método histograma, utilizando a mediana dos valores dos pixels e percentis de 80 a 95 para o sinal e 5 a 20 para o *background*. Este método é simples, estável e exclui pixels de características extremas do cálculo das intensidades do *spot* e do *background*.

5.10 ANÁLISE DA QUALIDADE DAS LÂMINAS

As tabelas de quantificação foram enviadas para o Laboratório de Bioinformática do Hospital do Câncer (<http://bioinfo.lbhc.hcancer.org.br>), sob coordenação da Dra Helena Brentani, que utiliza a ferramenta ArrayManager (desenvolvida no próprio laboratório) para análise da qualidade das lâminas.

Primeiramente foi realizada a correção dos valores de intensidade dos *spots*, através da subtração do valor do *background* de cada *spot* do respectivo sinal de intensidade deste. Em seguida, foram excluídos da análise os genes utilizados como controle e também os *spots* que não emitiram sinal, resultando em um conjunto de 4600 genes para o prosseguimento da análise. Foi então calculada a razão entre os dois canais (Alexa 555 / Alexa 647) para todos os genes em todas as lâminas e realizada a transformação destes valores aplicando-se o logaritmo na base 2 (\log_2) e seguinte normalização das replicatas que, neste caso, constitui da média aritmética (para todos os genes) entre os valores os valores das razões supracitadas ((lâmina principal + replicata com inversão de corantes)/2). A grande vantagem do uso do logaritmo da razão de expressão

pode ser exemplificada da seguinte forma: genes que são superexpressos por um fator igual a 2, possuem uma razão de expressão igual a 2, enquanto aqueles subexpressos pelo mesmo fator possuirão a razão de expressão igual a 0,5. Utilizando os logaritmos destes valores, um gene superexpresso por um fator igual a 2 possui \log_2 da razão igual a 1, enquanto um gene subexpresso pelo mesmo fator possui \log_2 da razão igual a -1. Já os genes que não são diferencialmente expressos (razão igual a 1) possuem \log_2 da razão igual a zero (QUACKENBUSH 2001).

De forma geral, na avaliação da qualidade das lâminas, é feito primeiramente um gráfico do tipo MA plot, no qual cada ponto representa um *spot*. O eixo x corresponde à intensidade do sinal e o eixo y corresponde à razão entre os sinais do canal 1 e 2. Em seguida, é feita a normalização dos dados por LOWESS, um método de regressão não-linear localmente ponderado, utilizando-se o parâmetro de suavização (*span*) igual a 0,4 e grau 2. Na análise, foram considerados apenas *spots* com relação *sinal/background* >1,05.

As diferentes propriedades de incorporação dos corantes e suas diferentes características físicas produzem viéses de marcação, que é a fonte de erro sistemático mais importante em microarranjos de duas cores. A normalização visa igualar esta intensidade relativa dos dois corantes. Adicionalmente, as replicatas do tipo “dye-swap”, nas quais as condições de hibridização são repetidas, com a inversão dos corantes na segunda hibridização, são muito importantes na redução destes viéses (QUACKENBUSH 2001; YANG et al. 2002).

Para verificar se houve algum viés introduzido por diferenças na deposição das sondas durante a confecção da lâmina, é avaliada a uniformidade dos *spots* feitos por cada agulha do robô que imprime os microarranjos nas lâminas. Ao total, são 48 agulhas. Os dados são representados em *boxplots*, sendo que cada *boxplot* representa as amostras depositadas por cada agulha.

Para avaliar a reprodutibilidade das hibridizações, foi calculado o coeficiente de correlação de Pearson entre cada lâmina e sua replicata. Este coeficiente é uma medida do grau de relação linear entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1. O valor 0 (zero) significa que não há relação linear, o valor 1 indica uma relação linear perfeita e o valor -1 também indica uma relação linear perfeita mas inversa, ou seja quando uma das variáveis aumenta a outra diminui. Quanto mais próximo estiver de 1 ou -1, mais forte é a associação linear entre as duas variáveis. Quando este coeficiente é alto e positivo, significa que duas lâminas produzem perfis de expressão gênica similares, o que é extremamente importante para a utilização de replicatas, especialmente em trabalhos onde o número de amostras é limitado. Foram gerados gráficos do tipo MM plot, nos quais são calculadas as correlações entre os *spots* do par de lâminas (*main* e *swap*). Após o cálculo do coeficiente, os dados são normalizados para corrigir as diferenças inerentes aos dois fluoróforos.

Também foi feito o grupamento hierárquico das amostras pela correlação de distancia euclidiana. Nesta análise, foram incluídas três amostras de carcinoma papilífero e duas de bócio, visando determinar se as

patologias malignas eram classificadas de forma próxima umas das outras e se o mesmo aconteceria com as amostras benignas.

5.11 ANÁLISE MATEMÁTICA E ESTATÍSTICA DOS DADOS

As tabelas com os dados de quantificação foram analisadas pelo laboratório BioInfo do IME-USP para a busca dos trios classificadores de interesse para validação.

A análise dos dados foi realizada através do pacote MaigesPack (ESTEVES 2007), do projeto Bioconductor (<http://www.bioconductor.org>), que utilizam o ambiente computacional R (<http://www.r-project.org>). Para a normalização dos dados, foi empregado o método de LOESS, utilizando o parâmetro de suavização (*span*) igual a 0,4 e grau 2, conforme implementado nos pacotes do R e Bioconductor.

Para a utilização dos dados obtidos no trabalho de identificação dos classificadores (STOLF et al. 2006) e o presente trabalho de validação, foi realizada a centralização dos dados de expressão gênica obtidos em ambos os estudos. Esta centralização é necessária devido à diferença entre o RNA-referência utilizado no trabalho de identificação e o utilizado no projeto de validação dos classificadores. Enquanto neste trabalho foi utilizada uma referência composta de um conjunto de RNAs provenientes de linhagens celulares de diferentes tipos tumorais, a referência do trabalho original era composta por um conjunto de RNAs provenientes de tecido tireoideano normal de uma série independente de cinco pacientes. A utilização de

referências distintas impede a sobreposição das amostras. Assim, torna-se necessário centralizar os níveis de expressão pela média não-ponderada, de modo que as nuvens das amostras antigas e novas fiquem centralizadas a partir de seus centros de massa. Após esta etapa, as amostras podem ser sobrepostas e então comparadas.

Na construção dos classificadores moleculares a serem validados, foi empregado o Discriminante Linear de Fisher, que apresenta boa relação entre facilidade e desempenho. Este método possibilita a criação de um banco de dados, contra o qual é possível testar uma amostra desconhecida e então classificá-la. Os classificadores se baseiam no perfil de expressão gênica combinada de cada trio de genes para classificar dada amostra. Uma vez conhecida a regra de classificação, esta foi aplicada aos dados de microarranjos das amostras do grupo de validação independente, ou seja, a mesma regra de classificação utilizada na identificação dos trios classificadores (STOLF et al. 2006) foi aplicada às amostras independentes coletadas neste projeto. Após a seleção dos melhores trios classificadores validados, foram montadas árvores de decisão para aumentar a taxa de acerto do teste. Em geral, esta estratégia consiste em ordenar os classificadores e seus possíveis resultados, dando origem a uma representação gráfica em forma de árvore, na qual cada nó interno representa um classificador e cada galho representa um dos possíveis resultados do teste. Em cada divisão, são testados novos classificadores no intuito de aumentar a eficiência do teste (GE e WONG 2008).

Neste trabalho, a regra de decisão estabelecida para a árvore de decisão foi a seguinte:

- Dada uma amostra, ela é classificada como resultado que pode ser A (adenoma folicular) ou C (carcinoma folicular).

1) aplica-se trio 1

1.1) se classificado como A -> então RESULTADO = A

caso contrário:

2) aplica-se trio 2

2.1) se classificado como A -> então RESULTADO = A

2.2) se classificado como C (carcinoma folicular) -> então RESULTADO =
C

6 RESULTADOS

6.1 COLETA E PRESERVAÇÃO DE AMOSTRAS

Número de amostras coletadas: 136, representando tecido normal, adenomas, carcinomas foliculares, carcinomas papilíferos, bócio, bócio adenomatoso e um caso de carcinoma medular. A coleta foi realizada em colaboração com o Departamento de Imagem do Hospital A.C. Camargo (coleta de amostras pré-cirúrgicas); Departamento de Cirurgia e Pescoço do Hospital A.C. Camargo, INEN (Lima, Peru) e Hospital Brasil (punções pós-cirúrgicas). Entre todas estas amostras, foram encontrados apenas dez casos de adenoma e seis casos de carcinoma folicular. Destas dez amostras, somente duas foram provenientes de punção pré-cirúrgica, uma de adenoma e uma de carcinoma folicular. Devido a problemas durante a hibridização, algumas amostras foram perdidas, restando para análise nove amostras de adenoma e cinco amostras de carcinoma folicular. O **Quadro 2** (seção 5, Casuística) mostra as amostras analisadas.

Para exemplificar a dificuldade na obtenção de material para este estudo, o **Quadro 5** mostra as amostras coletadas com suspeita pré-operatória de lesão folicular, no período de um semestre no Hospital A.C. Camargo, cujos pacientes foram encaminhados para a cirurgia. Em seguida, é mostrada a conclusão do laudo anátomo-patológico pós-operatório. No campo referente a suspeita pré-operatória (resultado da PAAF), foram

colocadas as informações fornecidas pelo patologista responsável através do preenchimento da folha de trabalho descrita anteriormente (informações transcritas *ipsis litteris*). Estes dados evidenciam a dificuldade do diagnóstico pré-operatório das lesões foliculares e justificam o desenvolvimento de um método de classificação mais eficaz.

Quadro 5 - Amostras pós cirúrgicas coletadas no primeiro semestre deste trabalho.

Amostra	PAAF	Congelação
AH01786006	Folicular	Bócio colóide
AH02477706	Folicular	Bócio colóide
AH02754706	Lesão folicular	Ca papilífero
AH02733406	Lesão folicular	Microcarcinoma papilífero
AH02656706	Padrão folicular, adenoma	Bócio colóide
AH14728007	Células foliculares	Bócio adenomatoso
AH03031906	Lesão folicular	Bócio multinodular
AH03062906	Padrão folicular	Bócio colóide
AH11906007	Lesão folicular	Ca papilífero
AH05420496	Lesão folicular	Bócio adenomatoso
AH02925606	Lesão folicular com aspecto oncocítico	Bócio nodular
AH01786006	Folicular	Bócio colóide
AH02477706	Folicular	Bócio colóide
AH02754706	Lesão folicular	Ca papilífero
AH02733406	Lesão folicular	Microcarcinoma papilífero
AH02656706	Padrão folicular, adenoma	Bócio colóide
AH14728007	Células foliculares	Bócio adenomatoso
AH03031906	Lesão folicular	Bócio multinodular
AH03062906	Padrão folicular	Bócio colóide
AH11906007	Lesão folicular	Ca papilífero
AH05420496	Lesão folicular	Bócio adenomatoso
AH02925606	Lesão folicular com aspecto oncocítico	Bócio nodular

6.2 EXTRAÇÃO DO RNA TOTAL

6.2.1 Rendimento do RNA referência

Após a extração dos RNAs, foi feita a quantificação das amostras em espectrofotômetro NanoDrop® ND1000. O rendimento para cada linhagem (extração do conteúdo de quatro frascos de cultura de 162 cm² com aproximadamente 90% de confluência) é mostrado no **Quadro 6**.

Quadro 6 - Massa de RNA total das linhagens do RNA referência após extração.

Linhagem	Massa de RNA total	Linhagem	Massa de RNA total
Daudi	314,4 µg	HEK293	200,97 µg
DLD-1	413,28 µg	Jurkat	772,8 µg
DU145	1128 µg	Saos2	124,8 µg
FaDu	526,8 µg	SK BR3	377,6 µg
GM637	485,79 µg	SK MEL28	340,76 µg
H1080	288 µg	T24	234,85 µg
H146	45 µg	T98G	631,85 µg
HB4a	201,64 µg	-	-

Após a quantificação, uma amostra de RNA total de cada linhagem celular foi analisada através de fracionamento em gel de agarose a 1% corado com brometo de etídeo. Foram aplicados 2 µg de cada amostra, diluídas em 5 µl de tampão desnaturante (2x TAE, 30% glicerol e uréia 7M) 2x e H₂O tratada com DEPC (q.s.p. 10 µl). Os resultados são mostrados na **Figura 6**. A intensidade da banda 28S está sempre maior do que a banda 18S, indício de que os RNAs estão com boa qualidade.

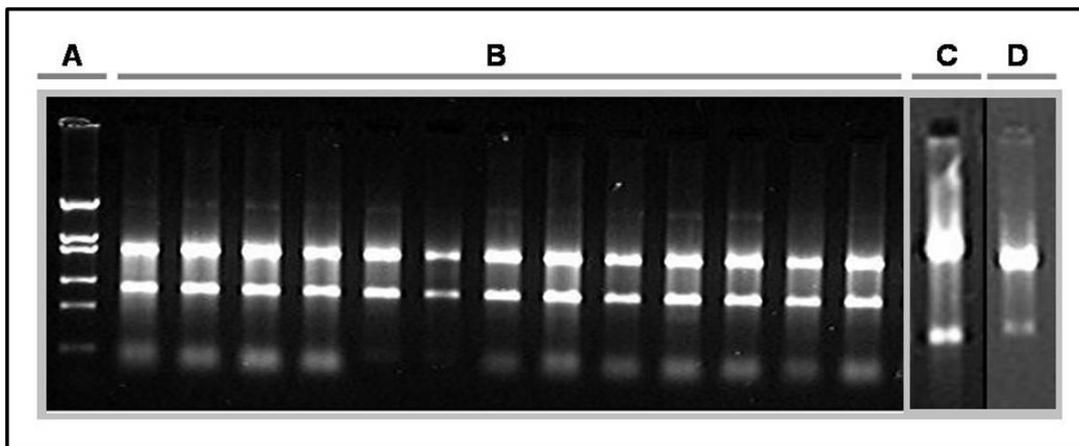


Figura 6 - Qualidade dos RNAs totais que compõem o pool de referência. Análise por fracionamento em gel de agarose a 1% corado com brometo de etídeo. Foram aplicados 2 μ g de cada amostra, diluídas em 5 μ l de tampão desnaturante (2x TAE, 30% glicerol e uréia 7M) 2x e H₂O tratada com DEPC (q.s.p. 10 μ l). A: marcador de peso molecular. B: qualidade dos RNAs das 11 linhagens extraídos em nosso laboratório; C: qualidade do RNA da linhagem HEK 293 e D: HB4a, cedidos pela Dra. Dirce Carraro (LGEA – Hospital A.C. Camargo).

6.2.2 Rendimento do RNA total das amostras de adenoma e carcinomas foliculares

O **Quadro 7** abaixo mostra o rendimento da extração do RNA total das amostras de carcinoma e adenoma foliculares.

Quadro 7 - Rendimento da extração do RNA total das amostras de carcinoma e adenoma foliculares.

Amostra	Diagnóstico final	RNA total (ng)	260/280	260/230
AHF004	Adenoma Folicular	1292	1,93	2,15
AHF006*	Adenoma Folicular	10	0,024	0,03
AHF012	Adenoma Folicular	1096	2,03	2,09
AHF020	Adenoma Folicular	575	1,87	0,56
AHF021	Adenoma Folicular	287,5	2,08	1,84
AHF028	Adenoma Folicular	900	2,23	2,12
JP10	Adenoma Folicular	560	1,92	1,33
JP20	Adenoma Folicular	1288	2,03	1,47
AHF016	Carcinoma Folicular	239	1,43	0,64
AHF019	Carcinoma Folicular	708	1,74	1,37
AHF030	Carcinoma Folicular	211	2,18	1,97
AHF031	Carcinoma Folicular	2320	1,89	1,84
AHF32	Carcinoma Folicular	1710	2,07	2,21
JP14*	Carcinoma Folicular	1784	2,23	0,53

* Amostras não utilizadas na validação dos classificadores devido a problemas na hibridização, com massa insuficiente para reibridização.

A **Figura 7** mostra o gráfico de leitura da amostra AH18713007 no espectrofotômetro NanoDrop[®], exemplificando o padrão observado em uma amostra de RNA total sem evidências de contaminação por DNA genômico ou excesso de sais resultantes dos processos de preservação e extração.

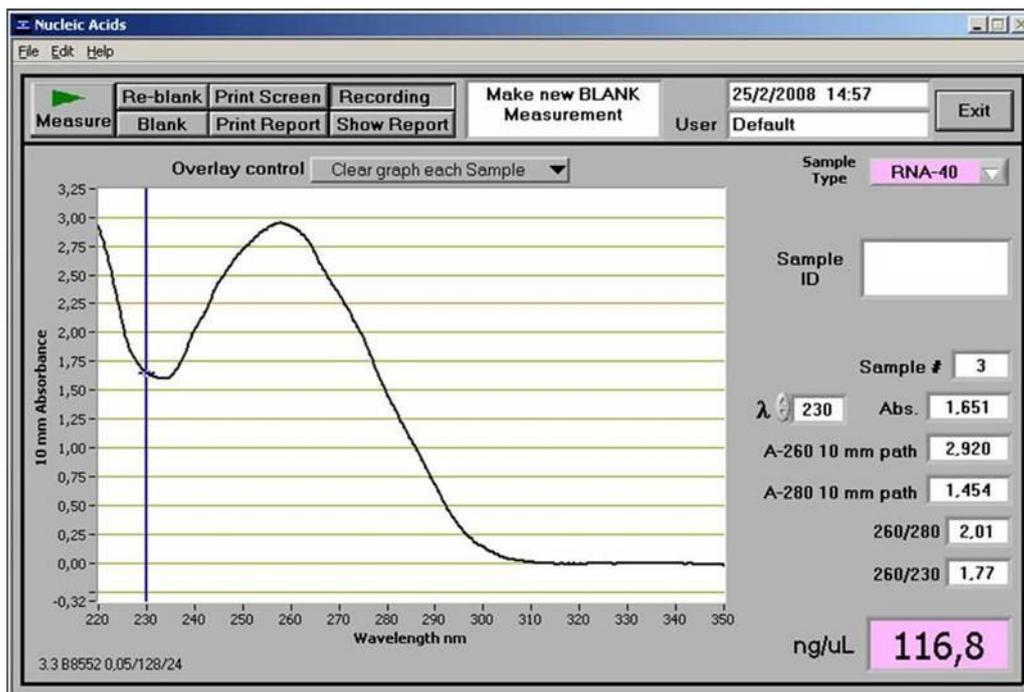


Figura 7 - Interface do *software* ND-1000[®] V3.3.0, do espectrofotômetro NanoDrop[®], representando o resultado de leitura da amostra AH18713007

6.3 AMPLIFICAÇÃO DO RNA TOTAL

6.3.1 Rendimento do RNA referência

No primeiro ciclo de amplificação, a massa inicial e RNA-referência foi igual a 18µg de RNA total. O rendimento do primeiro ciclo de amplificação do RNA-referência é mostrado na **Quadro 8**.

Quadro 8 - Rendimento do primeiro ciclo de amplificação do RNA referência.

Tubo	Razão 260/280	Razão 260/230	Concentração de RNA	Massa total de RNA (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

No segundo ciclo de amplificação, foram feitas nove reações. Em cada uma delas, a massa inicial de RNA amplificado resultante do primeiro ciclo foi igual a 800 ng. O **Quadro 9** mostra o rendimento destas reações.

Quadro 9 - Rendimento do segundo ciclo de amplificação do RNA referência.

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

Para a utilização do RNA referência amplificado, foi feito um *pool* de todo o material obtido após os dois ciclos de amplificação.

6.3.2 Rendimento da amplificação do RNA total de amostras colhidas por PAAF

A **Figura 8** mostra o espectro de absorvância de quatro amostras. A maior parte das amostras apresentou RNA total com alto grau de pureza, como exemplificado na Figura 8A e 8B. Algumas amostras, no entanto, apresentaram baixa concentração de RNA, assim como a presença de contaminantes. Proteínas apresentam um pico de absorvância no comprimento de onda de 280 nm e outros contaminantes tais como sais e compostos orgânicos apresentam pico de absorvância em torno de 230 nm. Amostras de baixa pureza são mostradas nas Figuras 8C e 8D.

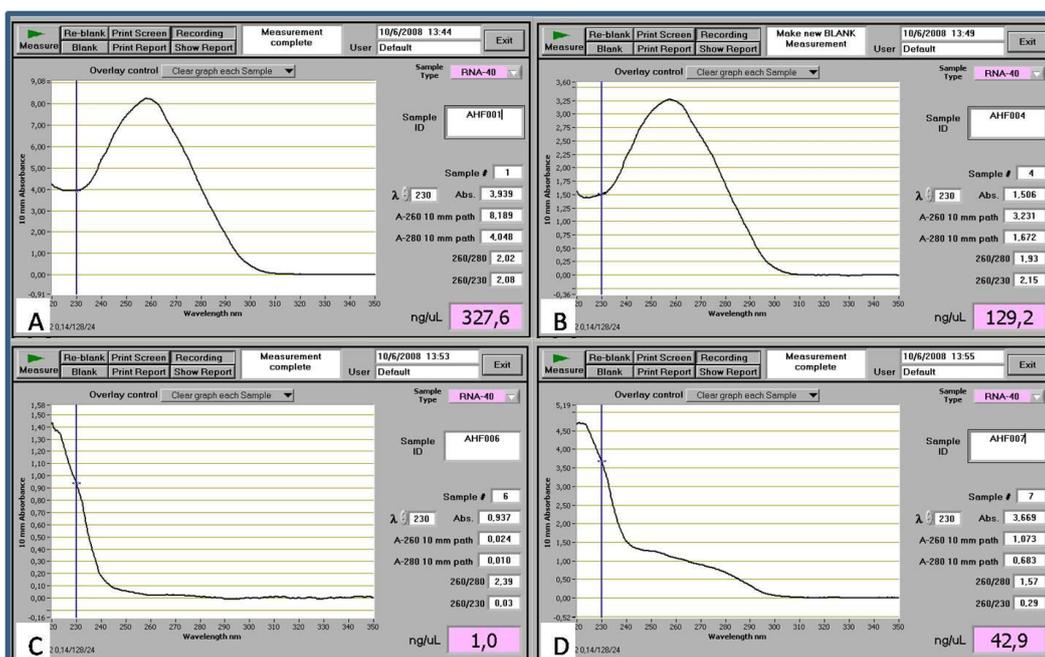


Figura 8 - espectro de absorvância do RNA total de amostras de punção de peça cirúrgica extraído com o *kit* PicoPure[®]. A) amostra AHF001; B) amostra AHF004; C) amostra AHF006, D) amostra AHF007. A e B apresentam alto grau de pureza, conforme pode ser observado pelas razões 260/280 e 260/230 altas. Já C e D representam amostras com contaminantes.

É observado que, de acordo com as razões 260/280 e 260/230, o RNA amplificado tem maior qualidade em relação ao RNA total inicial. A **Figura 9** mostra o perfil do RNA da amostra AH002 em três momentos: a) extração; b) após um ciclo de amplificação; c) após dois ciclos de amplificação. Entre as amostras do amplificadas nesta série, esta foi a que apresentou menor rendimento. A amplificação partiu de 333 ng de RNA total. Após o primeiro ciclo de amplificação, a massa gerada foi de 189 ng de RNA mensageiro. Considerando que cerca de 3% do RNA total é composto por RNA *poli A*, houve amplificação na ordem de 18 vezes em relação à massa original. Após o segundo ciclo, o rendimento foi de 121 vezes em relação ao primeiro ciclo e 2300 vezes em relação ao RNA mensageiro inicial. Nota-se que, no caso apresentado, houve um aumento de 1,47 vezes na razão 260/280 e de 2,08 vezes na razão 260/230 entre o RNA total e o RNA amplificado.

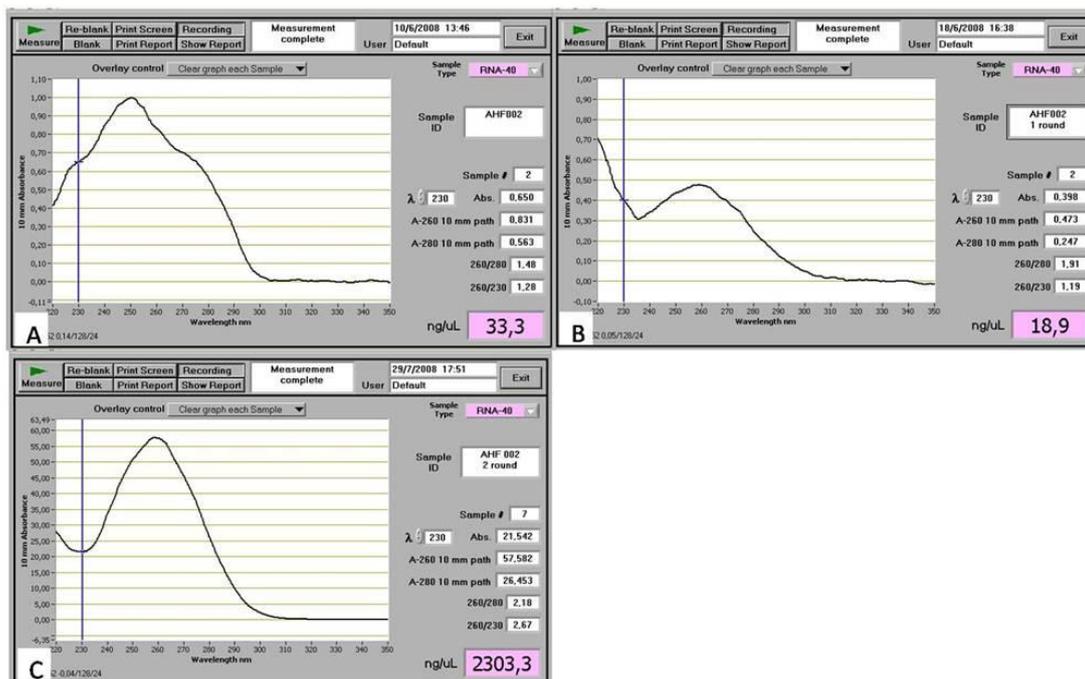


Figura 9 - Perfil do RNA da amostra AHF002. A) após extração; B) após um ciclo de amplificação; C) após dois ciclos de amplificação.

A qualidade do RNA amplificado das amostras AHF001 a AHF005 pode ser observada na **Figura 10**, onde 1 μg de cada amostra foi fracionada em gel de agarose em condições desnaturantes e coradas pelo brometo de etídeo.

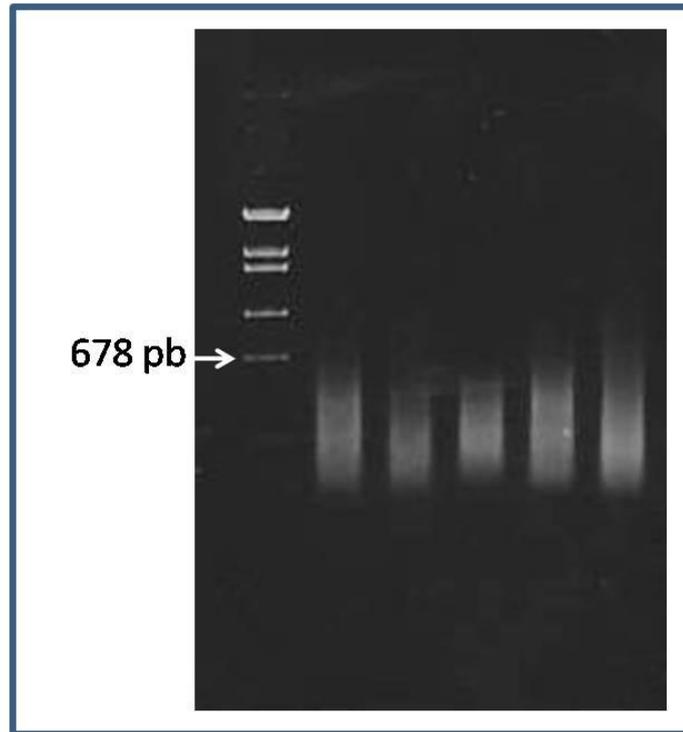


Figura 10 - Análise da qualidade do RNA amplificado através de fracionamento em gel de agarose. Fracionamento dos aRNAs das amostras AHF001 a AHF005 após dois ciclos de amplificação em gel de agarose a 1,5 % em condições desnaturantes e corados pelo brometo de etídeo. Foi utilizado 1 μ g de aRNA de cada amostra.

O **Quadro 10** mostra o rendimento de RNA das amostras de *adenoma e carcinoma foliculares* após dois ciclos de amplificação.

Quadro 10 - Rendimento de RNA das amostras após dois ciclos de amplificação.

Amostra	Diagnóstico final	RNA total (ng)	aRNA após 2 ciclos (µg)	Razão 260/280	Razão 260/230
AHF004	Adenoma	1292	63,8	2,05	2,47
AHF006*	Adenoma	10	15,558	2,45	2,91
AHF012	Adenoma	1096	81,544	2,33	2,88
AHF020	Adenoma	575	129,28	2,31	2,77
AHF021	Adenoma	287,5	86,56	2,33	2,76
AHF028	Adenoma	900	179	2,17	2,67
AH02947706	Adenoma	931	65,73	1,98	1,66
AH162234007	Adenoma	710	21,24	1,82	1,46
JP10	Adenoma	560	93,12	2,06	2,54
JP20	Adenoma	1288	71,11	2,24	2,77
AHF030	Ca. Folicular	1600	36,25	2,16	2,88
AHF031	Ca. Folicular	2320	81,44	2,19	2,77
AHF32	Ca. Folicular	1710	85,74	2,16	2,76
AHF016	Ca. Folicular	239	89,85	2,29	2,83
AHF019	Ca. Folicular	708	101,65	2,29	2,79
JP14*	Ca. Folicular	1784	72,25	2,33	3,2

* Amostras não utilizadas na análise final por apresentarem problemas na marcação ou hibridização.

A qualidade do RNA amplificado pode ser observada na **Figura 11**, onde 1 µg de cada amostra foi fracionada em gel de agarose em condições desnaturantes e coradas pelo brometo de etídeo.

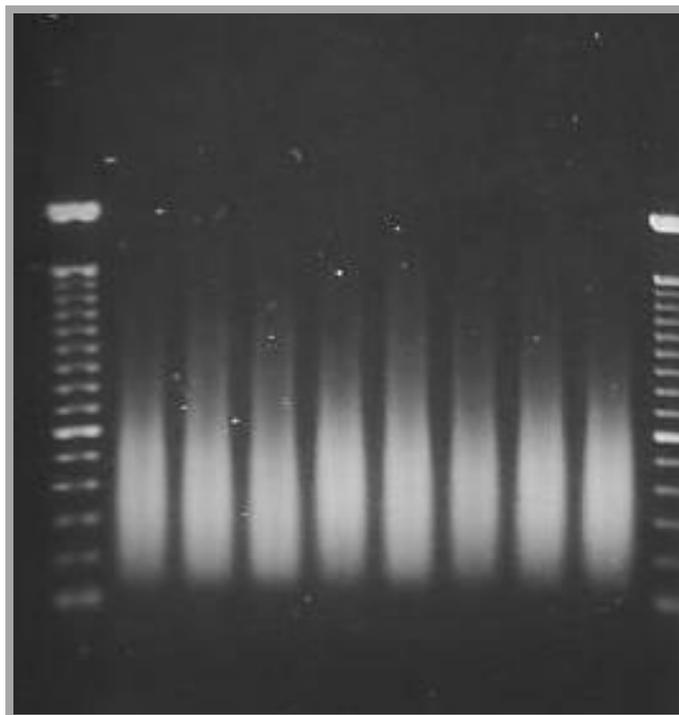


Figura 11 - Análise da qualidade do RNA amplificado de amostras pós cirúrgicas de interesse, através de fracionamento em gel de agarose. Fracionamento dos aRNAs das amostras pós-cirúrgicas (RNA total extraído com o *kit* PicoPure) após dois ciclos de amplificação em gel de agarose a 1,5 % em condições desnaturantes e corados pelo brometo de etídeo. Foi utilizado 1 μ g de aRNA de cada amostra. O marcador de peso molecular utilizado é de 100 pares de bases.

6.4 MARCAÇÃO E HIBRIDIZAÇÃO DAS AMOSTRAS

Devido a problemas na estação de hibridização, algumas lâminas não apresentaram qualidade satisfatória e foram refeitas. No entanto, não houve RNA suficiente da amostra AHF006, que teve que ser descartada. Foi necessário o descarte de mais uma amostra, jp14, pois esta apresentou baixíssima eficiência na marcação, em diversas tentativas. Apesar de os experimentos na maioria das vezes terem ocorridos de forma satisfatória,

enfrentamos alguns problemas, como falha na manutenção do vácuo dois dos módulos da estação de hibridização que levaram à perda de algumas lâminas, que tiveram que ser refeitas. Algumas lâminas também quebraram dentro do cassete de hibridização no momento do aquecimento e, para evitar este problema, o cassete de hibridização foi fixado na estação com um pouco menos de pressão de modo a evitar as rachaduras. Para compensar a menor pressão exercida sobre a lâmina e evitar o comprometimento da vedação da mesma, parte da etiqueta de identificação da lâmina foi removida, na região onde a etiqueta entrava em contato com a borracha de vedação. A remoção desta parte da etiqueta permitiu que a borracha de vedação mantivesse contato direto com o vidro da lâmina em toda a sua extensão.

Após as hibridizações, foi feita a captura e determinação dos valores dos sinais de fluorescência dos *spots* de cada lâmina, através do leitor confocal a *laser ScanArray Express[®] Microarray Scanner* e do *software ScanArray Express[®]*. A **Figura 12** mostra um exemplo de lâmina hibridizada, após captura dos sinais fluorescentes e digitalização da imagem.

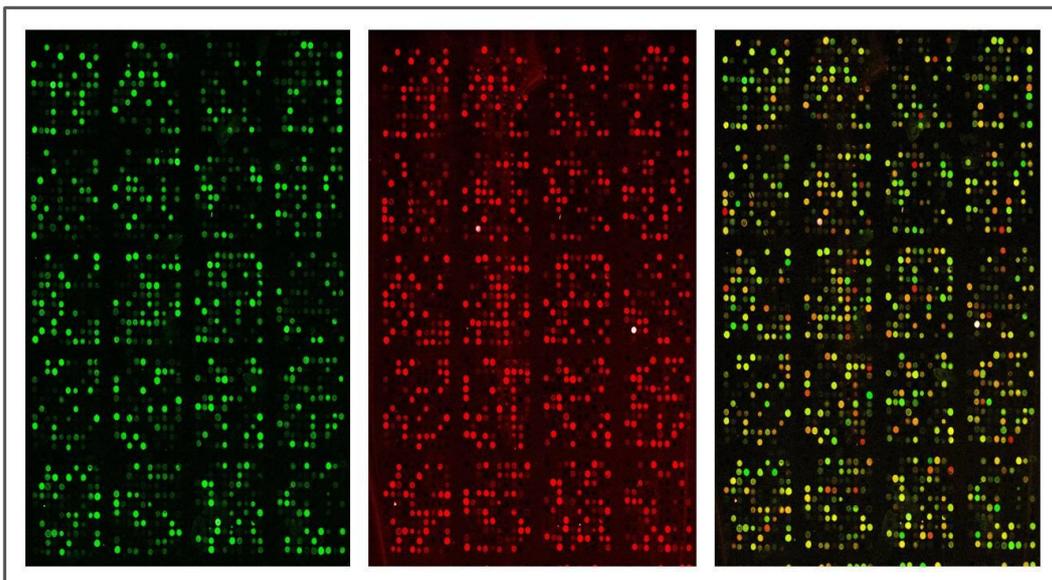


Figura 12 - Imagem da porção de uma lâmina após digitalização. Porção de uma lâmina de microarranjos após hibridização e digitalização da imagem, mostrando cada canal isoladamente e, por último, a sobreposição das imagens.

Após quantificação dos *spots* de uma lâmina, o programa fornece uma tabela contendo todos os valores de sinais fluorescentes gerados pela lâmina. Com estes dados é realizada as etapas posteriores de análise de qualidade das hibridizações e análises matemáticas e estatísticas de validação dos classificadores.

6.5 ANÁLISE DA QUALIDADE DAS LÂMINAS

Na pré-análise realizada pelo Laboratório de Bioinformática do Hospital do Câncer, foi feito um MM plot para cada par de lâminas (cada amostra com seu *swap*). A **Figura 13** mostra um MM plot gerado para a amostra G12 mostrando o índice de correlação de Pearson para este par de lâminas.

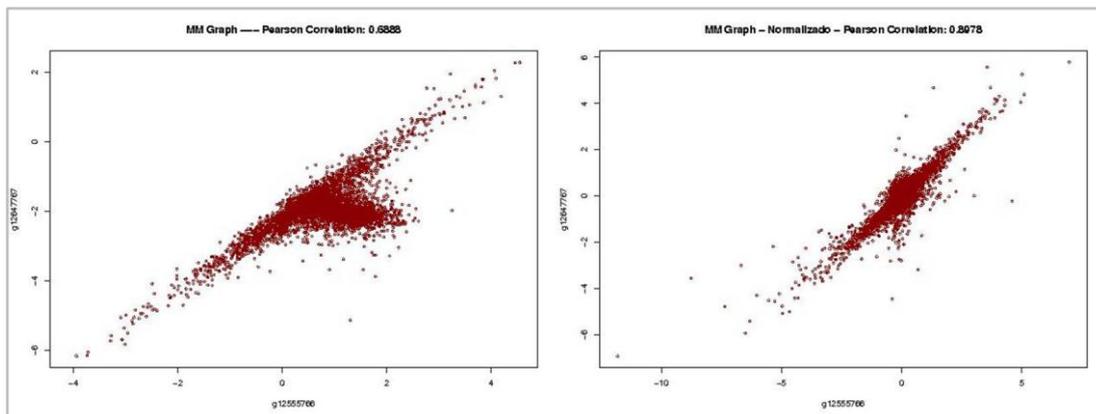


Figura 13 - MM plot da amostra g12, antes e após a normalização, com o respectivo coeficiente de Pearson

Também foi feito um MA plot (**Figura 14**), onde cada ponto é um *spot*. O eixo x corresponde à intensidade do sinal e eixo y corresponde à razão entre o sinal do canal 1 e do canal 2 (valores em \log_2).

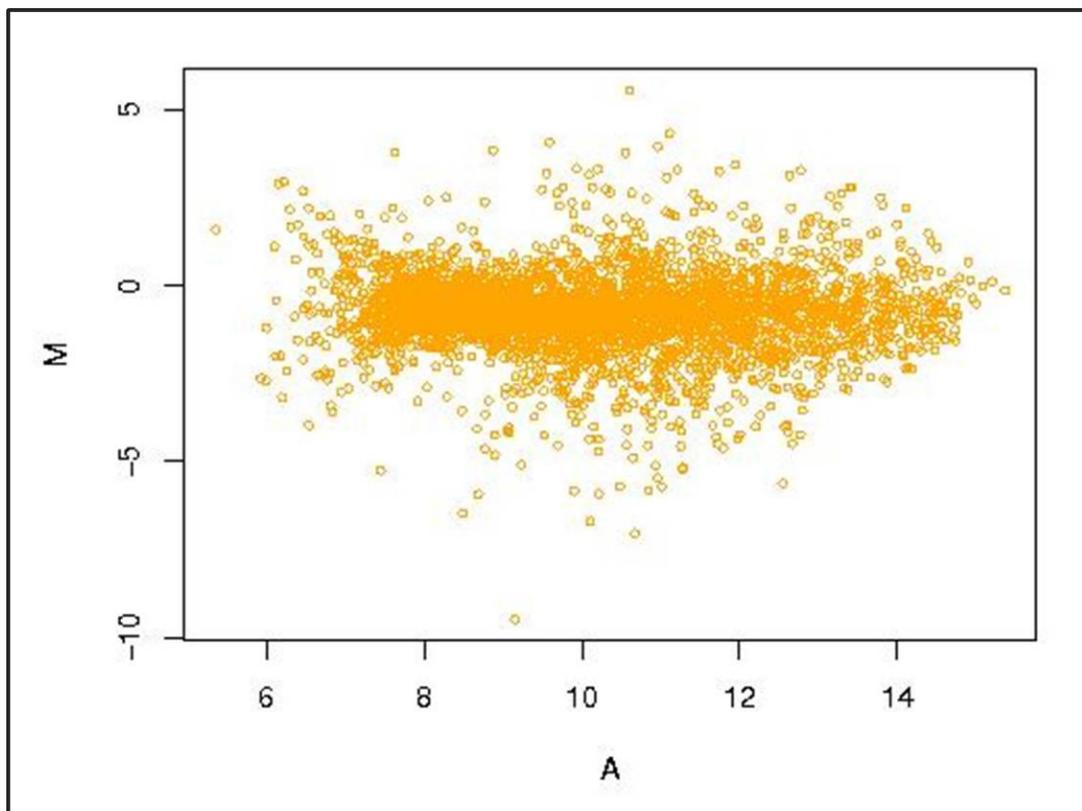


Figura 14 - MA plot da amostra g12.

Para avaliar se houve algum viés induzido na deposição das sondas durante a confecção da lâmina, foi feita uma análise da uniformidade dos *spots* depositados por cada agulha. Ao total são 48 agulhas. Na **Figura 15** são mostrados os 48 *boxplots* correspondentes às agulhas. Cada *boxplot* representa as amostras depositadas por cada agulha.

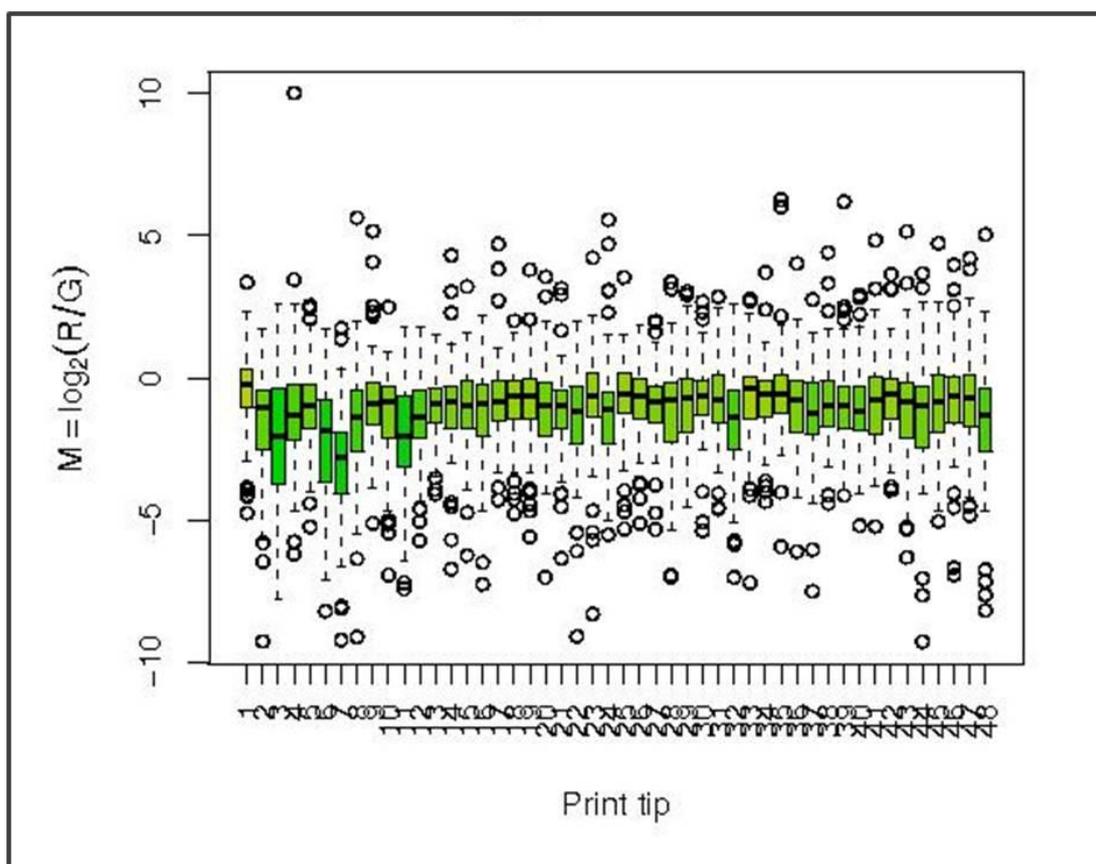


Figura 15 - Análise da uniformidade dos spots depositados por cada agulha. *Boxplot* representando a análise de variação de intensidade de sinal segundo a deposição dos clones da lâmina por cada agulha. Amostra g12, lâmina *main*.

Conforme mostrado na **Figura 16**, o agrupamento hierárquico não é capaz de separar as amostras de adenoma e carcinoma folicular. Em muitos casos, o agrupamento hierárquico pode permitir a distinção das amostras,

mas neste caso parece não ser a abordagem adequada. Este dado está de acordo com a publicação original de STOLF et al. (2006), confirmando que dados baseados somente na expressão diferencial não são capazes de permitir a distinção dos tumores, ao menos na plataforma utilizada.

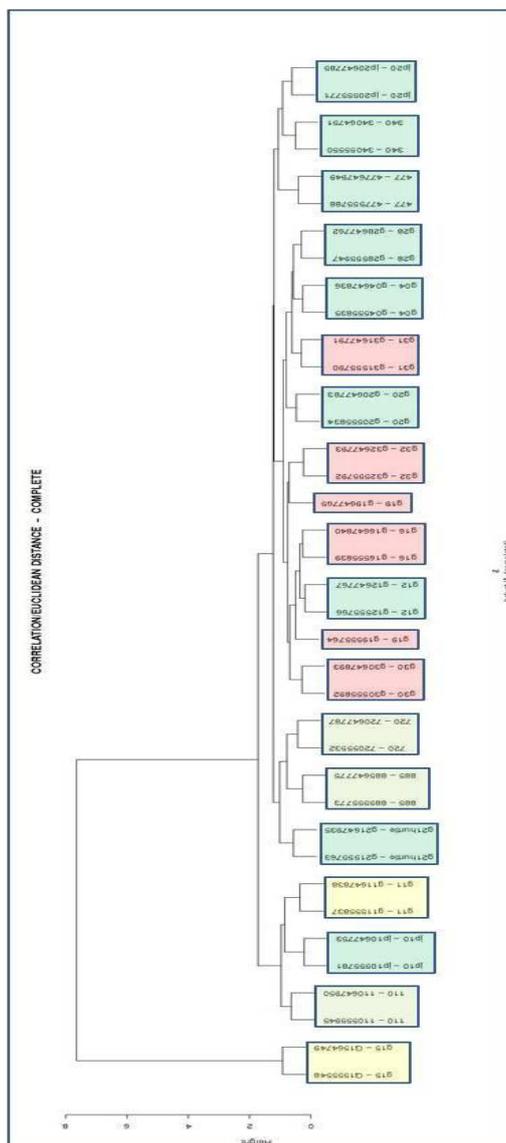


Figura 16 - Dendrograma mostrando o agrupamento das amostras testadas. Em azul, os adenomas folliculares; vermelho: carcinomas folliculares; verde-claro: carcinomas papilíferos; amarelo: bócios.

6.6 ANÁLISE MATEMÁTICA E ESTATÍSTICA DOS DADOS

No trabalho original de STOLF et al. (2006), foram escolhidos 100 trios de genes classificadores, sendo que 14 foram publicados devido à sua maior capacidade de discriminação. Dos 14 trios publicados, dez não puderam ser analisados devido a problemas na impressão dos genes na nova plataforma. Devido a este problema, foi realizada uma etapa de busca pelos 100 melhores trios classificadores obtidos no primeiro trabalho na lâmina utilizada neste trabalho. Dos 100 melhores trios do trabalho original, 31 estão presentes também na lâmina confeccionada para este trabalho. Destes 31 trios, apenas quatro fazem parte do conjunto de trios publicados.

Utilizando os 31 trios encontrados na nova lâmina, foi feita a análise de quais trios possibilitam a melhor distinção entre as duas patologias em questão.

Os genes que compõem cada trio são apresentados no **Apêndice 3**.

Baseado na *sensibilidade* de cada trio é possível estabelecer um *ranking* dos melhores trios classificadores aqui encontrados (**Quadro 11**). A sensibilidade é a fração dos que obtiveram resposta positiva no teste entre aqueles que possuem a doença, medindo a capacidade do teste em identificar corretamente a doença entre aqueles que a possuem. A particular importância da sensibilidade do teste neste trabalho será discutida mais adiante.

Quadro 11 - Melhores trios classificadores, erros e acertos de cada classificador.

Trio	Total de acertos	Amostras de Carcinoma Folicular classificadas corretamente	Amostras de Adenoma Folicular classificadas erroneamente
18	12	5/5	2/9
7	11	5/5	3/9
24	11	5/5	3/9
15	10	5/5	4/9
9	9	5/5	5/9

Quadro 12 – Sensibilidade, especificidade , acurácia e SVD dos melhores trios classificadores.

Trio	Sensibilidade	Especificidade	Acurácia	SVD
18	1	0,777778	0,857143	8,127
7	1	0,666667	0,785714	9,9142
24	1	0,666667	0,785714	7,8567
15	1	0,555556	0,714286	8,3951
9	1	0,444444	0,785714	9,1399

A análise ROC (*Receiver Operating Characteristic*) é uma ferramenta poderosa para medir e especificar problemas no desempenho do diagnóstico em medicina. Esta análise, por meio de um método gráfico simples e robusto, permite estudar a variação da *sensibilidade* e *especificidade*, para diferentes valores de corte.

A *área abaixo da curva ROC* está associada ao poder discriminante de um teste diagnóstico.

Geometricamente, a curva ROC é um gráfico de pares "x" e "y", que correspondem a (*1-especificidade*) e à *sensibilidade*, respectivamente, em um plano designado por plano ROC unitário. A designação de plano ROC

unitário se deve ao fato de as coordenadas deste gráfico representarem medidas de probabilidade e, por conseguinte, variarem entre zero e um (BRAGA 2000).

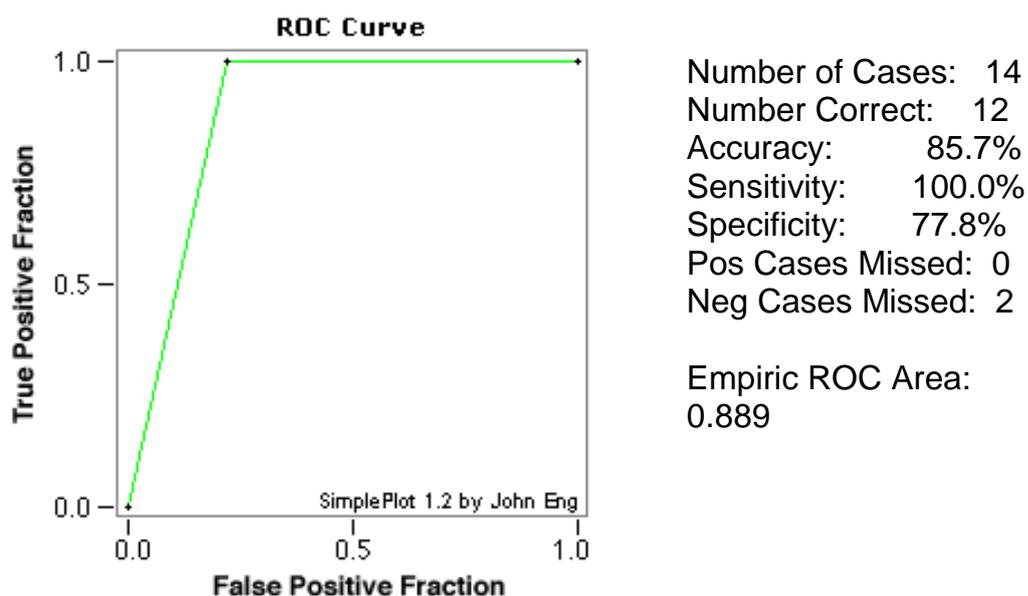


Figura 17 - Curva ROC para o Trio 18.

O Quadro 13 contém a comparação entre o SVD obtido no trabalho de Stolf e após a validação, mostrando que os resultados foram reproduzíveis, não havendo variação significativa do SVD.

Quadro 13 - Comparação entre os resultados obtidos no trabalho de Stolf e após a validação

Trio	Gene 1	Gene 2	Gene 3	SVD Identificação	SVD Validação	acertos	erros
18	FUBP1	GRCC9	UBE1C	8.2693	8,127	12	2
7	SNAI2	UBE1C	SMCR8	9.6997	9,9142	11	3
24	TRIM2	POLD3	TSHR	7.8567	7,8567	11	3
15	SNAI2	NDRG2	SNAP25	8.3951	8,3951	10	4
9	PLVAP	MCSP	POMT1	9.1399	9,1399	9	5

Podemos combinar estes trios de forma a re-classificar as amostras de carcinoma folicular classificadas como adenoma por outro classificador e fazer o mesmo com as amostras de adenoma que foram classificadas como carcinoma folicular.

Com estes dados, é possível utilizar a estratégia de árvore de decisão, que consiste em utilizar sucessivamente mais de uma regra de discriminação (mais de um classificador), aumentando a taxa de acerto do teste (DUDOIT et al. 2003). A Figura 18 mostra as árvores de decisão estabelecidas com os melhores trios classificadores:

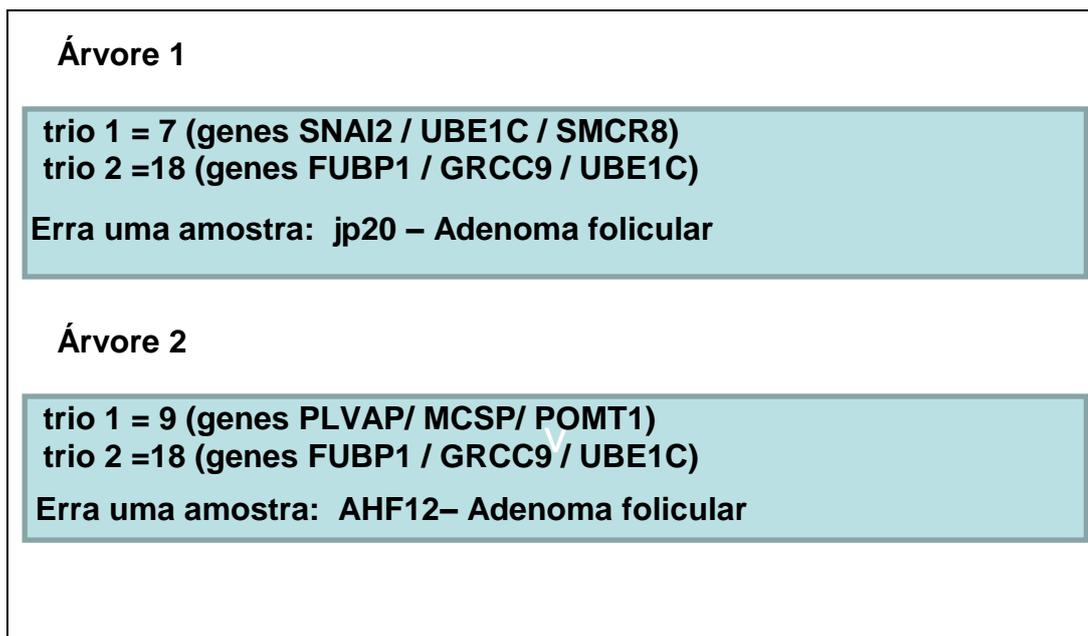


Figura 18: Árvores de decisão estabelecidas com os melhores trios classificadores

7 DISCUSSÃO

Entre os tumores de tireoide, o diagnóstico dos carcinomas foliculares através da PAAF pré-operatória é um grande desafio. Para um diagnóstico preciso deste tumor, é necessária a inspeção anatomo-patológica de toda a cápsula uma vez que o diagnóstico diferencial para os adenomas foliculares é a invasão capsular. Desta forma, um grande número de pacientes é submetido à tireoidectomias desnecessárias.

Através de dados da expressão gênica, nosso grupo utilizou a Análise Discriminante Linear de Fisher para identificar assinaturas gênicas, construindo trios de genes (chamados de preditores ou classificadores) para a distinção entre as lesões malignas e benignas (trabalho publicado por STOLF et al. 2006). Na plataforma de 4600 sequências expressas utilizada por nosso grupo, as duas patologias não são distinguíveis molecularmente pela análise da expressão global de genes seguida por abordagens de agrupamento (*clustering*).

Ainda que a distinção baseada no perfil de expressão de um conjunto de genes diferencialmente expressos fosse possível, esta abordagem pressupõe, para cada amostra a ser estudada, a obtenção dos dados dos dois grupos de amostras. Este passo é necessário para que se possa determinar a correlação entre cada uma das amostras, o que, na rotina diagnóstica, poderia ser uma impossibilidade (MEIRELES et al. 2004).

Um dos primeiros estudos relacionados à seleção de classificadores com base no perfil de expressão de um dado conjunto de genes foi o trabalho de VAN'T VEER et al. (2002). Este grupo identificou uma assinatura gênica, composta de 70 genes que, segundo os dados, indicam a necessidade de adjuvância em câncer de mama.

Uma das maiores vantagens da utilização da classificação baseada em assinaturas gênicas ao invés do perfil de expressão comparativa é a possibilidade de criar um banco de dados contra o qual uma amostra-teste possa ser classificada (RAMASWAMY et al. 2001). Em outras palavras, uma vez que 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. 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 (BARRETO 2009). A respeito do problema da determinação do número suficiente de genes para uma classificação acurada, trabalhos mostram que isto depende da acurácia desejada (ANCONA et al. 2006; ZENG et al. 2008). ANCONA et al. (2006) apresenta que, na classificação de câncer de cólon, a taxa de erro de um classificador com 16384 genes é de 11% ($p= 0,025$), enquanto para um classificador composto por apenas dois genes, esta taxa é de 16% ($p= 0,05$), mostrando que é possível realizar a classificação com base na expressão de um pequeno número de genes.

O valor destas assinaturas obtidas de forma experimental frequentemente se estendem além de sua publicação original. Em câncer de mama, por exemplo, diversas assinaturas gênicas, incluindo a Mammaprint e Oncotype DX vêm sendo investigadas para testes diagnósticos comerciais (ROSS 2009) e sendo então validadas em ensaios clínicos em larga escala (SPARANO e PAIK 2008; CARDOSO et al. 2008).

Um dos mais importantes requisitos para uma maior aceitação de um classificador é a validação do desempenho do mesmo em amostras independentes (FOEKENS et al. 2006).

O problema de um classificador estar adaptado de forma muito estreita ao conjunto de amostras utilizadas no treinamento (“*overfitting*”) não é algo restrito aos dados gerados pela técnica de microarranjos, porém pode ser especialmente importante devido à alta dimensionalidade deste tipo de estudo: com um enorme número de variáveis para grupos de observações muito reduzidos, é possível obter apenas uma fração muito pequena de elementos de predição. Limitar o número de características analisadas diminui a ocorrência de “*overfitting*” nas regras de predição. A inclusão de elementos irrelevantes pode também prejudicar substancialmente o desempenho de alguns classificadores (SONTROP et al. 2009).

Em diversos trabalhos, é possível encontrar uma regra de predição tal que todas ou quase todas as observações a partir dos dados de treinamento possam ser classificadas corretamente. Entretanto, isto não significa que a regra de predição que esteja altamente adaptada aos dados de treinamento

será capaz de prever corretamente novas observações independentes. (BOULESTEIX et al. 2008).

Por este motivo, é de suma importância que classificadores moleculares sejam validados em um conjunto de amostras independentes, que é o propósito deste trabalho. Em outras palavras: a predição pode ser feita de forma confiável em *novos* pacientes? Tal pergunta só pode ser respondida com segurança através de muitos (ou um grande) conjuntos de validação. O restrito número de amostras, especialmente de carcinomas foliculares, somado à dificuldade de obtenção de RNA destas amostras, constituem os principais obstáculos à realização de um estudo em larga escala em um curto espaço de tempo. No entanto, diante destas dificuldades e da forma com que foram propostas alternativas, é possível obter informações importantes para reforçar a validade deste estudo e ajudar no delineamento de futuros trabalhos de validação nos quais a obtenção de amostras seja um fator limitante. Um exemplo disto foi o estabelecimento de colaborações para a coleta de amostras em diferentes hospitais, incluindo um centro internacional. A importante consequência destas colaborações, não previstas no projeto original, é de que o impacto da acurácia do classificador na comunidade médica aumenta de acordo com as diferenças entre o conjunto utilizado na obtenção dos classificadores e o conjunto independente utilizado na validação. Isto porque é muito mais difícil encontrar resultados similares em um conjunto de validação cujas amostras foram coletadas em laboratórios diferentes, em épocas diferentes e utilizando pacientes de diferentes etnias, condições sociais e localização

geográfica, do que em um conjunto de validação mais homogêneo ou então um subconjunto daquele empregado na análise original (FOEKENS et al. 2006).

Outra diferença importante entre as amostras do trabalho de STOLF et al. (2006) e o presente trabalho é que aquele utilizou fragmentos dos nódulos tireoideanos obtidos na peça cirúrgica. O inconveniente do uso deste tipo de amostra é que não há meios de obtê-la antes da cirurgia. Assim, embora este tipo de amostra seja uma importante fonte para pesquisa, não apresenta aplicabilidade diagnóstica pré-cirúrgica e não promoveria a diminuição do número de tireoidectomias desnecessárias. Desta forma, é necessário, além de validar os classificadores em um conjunto de amostras independentes, verificar se estes classificadores também são capazes de distinguir amostras coletadas por punção aspirativa por agulha fina (PAAF).

A PAAF é um dos principais métodos diagnósticos de nódulos tireoideanos, porém a análise citológica do material coletado não permite a distinção entre adenomas e carcinomas foliculares. Fragmentos de tecido e aspirados tireoideanos fornecem amostras de composição diferentes: a coleta de fragmentos da peça cirúrgica pode ser direcionada para a porção de tecido suspeito no interior do nódulo, através de análise macro ou microscópica; este tipo de amostra concentra um número maior de células patológicas (sejam benignas ou malignas) e apresenta celularidade satisfatória. Já a coleta por punção aspirativa do nódulo fornece amostras de baixa celularidade, com presença de colóide e sangue, dificultando ainda

mais a obtenção de RNA em quantidade e qualidade suficientes para ensaios de microarranjos de cDNA.

Outra dificuldade relacionada à utilização de amostras obtidas por PAAF é a espera pelo diagnóstico final. Na PAAF pré-cirúrgica, há diversas possibilidades diagnósticas, sendo que a maioria não é constituída de lesões foliculares (bócio, Ca papilífero, tireoidite de Hashimoto). Assim, é necessária a coleta de um grande número de amostras para a obtenção de algumas poucas amostras com suspeita de lesão folicular, que só terão o diagnóstico confirmado após a cirurgia. Na coleta por PAAF pós-cirúrgica, é mais frequente a obtenção de uma amostra de interesse, pois só são coletadas amostras com indicação cirúrgica por suspeita de carcinoma ou adenoma foliculares. No entanto, foi constatado que, mesmo as amostras com suspeita pré-operatória de lesão folicular podem tratar-se de outras patologias, após diagnóstico anátomo-patológico da peça cirúrgica, conforme constatado na etapa de coleta de amostras deste trabalho, evidenciando a dificuldade do diagnóstico pré-operatório das lesões foliculares e justificando o desenvolvimento de um método de classificação mais eficaz.

Voltando à análise dos classificadores aqui validados, a escolha do melhor classificador envolve uma série de fatores. Um critério de decisão pelo melhor teste depende, em grande parte, dos benefícios associados aos resultados corretos e dos “custos” associados aos incorretos. Por exemplo, a previsão de uma tempestade que acaba por não ocorrer (*falso positivo*) tem um custo menor em relação à falha na previsão de uma tempestade que

ocorre (*falso negativo*) (BRAGA 2000). No contexto do presente trabalho, os “custos” não se referem ao significado econômico, e sim ao sentido de *consequência* à saúde do paciente.

Em um teste diagnóstico, há dois tipos de erro que podem ocorrer na decisão: a escolha de uma *falha* (no sentido de declarar um doente como são) ou a escolha de um *falso alarme* (declarar uma pessoa são como doente). Nossos resultados contêm três classificadores com alta porcentagem de acertos, porém foram considerados como melhores classificadores não aqueles que classificaram corretamente o maior número de amostras, mas aqueles que, *também*, classificaram as amostras da forma mais adequada ao problema apresentado. Ao optar por um classificador para adenomas e carcinomas foliculares, é preferível optar por um *falso alarme* a uma *falha*, ou seja, optar por “um mal menor” em termos de diagnóstico, considerando as consequências de cada opção. Aqui, é considerado um “mal menor” diagnosticar um paciente com carcinoma folicular, quando, na verdade, trata-se de um adenoma, do que classificar um paciente como portador de uma patologia benigna, quando na verdade trata-se de uma malignidade que deve abordada de forma mais radical. Em relação às consequências de um falso positivo e um falso negativo, primeiramente deve-se levar em consideração que, atualmente, diante de uma suspeita de adenoma ou carcinoma folicular, o paciente é submetido à tireoidectomia total para diagnóstico pós-operatório. Considerando que a prevalência do carcinoma folicular é relativamente baixa, cerca de 70% destas tireoidectomias poderiam ser evitadas. Visando reduzir o número de

cirurgias desnecessárias sem comprometer o diagnóstico correto daqueles que *realmente* necessitam de tratamento cirúrgico, é de suma importância que o teste apresente baixo número de falso-negativos. Em outras palavras, o teste deve, imprescindivelmente, ter alta sensibilidade.

Um bom teste possui um alto valor para a sensibilidade e para a especificidade, pois ele identificará corretamente aqueles que têm a doença e aqueles que não têm. No entanto, conforme mencionado anteriormente, alguns aspectos devem ser levados em consideração para ressaltar a especial importância da sensibilidade do teste em relação à sua especificidade neste trabalho em particular, no qual não foi possível encontrar valores ideais de sensibilidade e especificidade em um mesmo trio classificador. Para nosso melhor trio classificador, a curva ROC mostra uma sensibilidade de 100% e uma especificidade de 66,7%.

É importante ressaltar que a forma como as assinaturas gênicas são desenvolvidas não prioriza mecanismos biológicos e ligações causais entre os genes e a classificação do tumor. Enquanto o foco das assinaturas gênicas é a discriminação, os aspectos funcionais que resultam em diferentes fenótipos não estão diretamente acessíveis a partir das informações provenientes da assinatura gênica (HUMMEL et al. 2008). Isto significa que os genes selecionados como classificadores moleculares não estão, necessariamente, envolvidos nos mecanismos da patogênese da doença (SIMON et al. 2003; RANSOHOFF 2005).

Desta forma, até mesmo uma assinatura gênica validada de forma extremamente criteriosa não irá necessariamente revelar mecanismos da

patogênese ou novos alvos moleculares. Uma ligação de causa entre os genes dos classificadores e as diferenças na gênese das duas patologias seria muito útil, porém não é necessária à classificação molecular (GÖNEN 2009).

É de suma importância que os classificadores aqui validados sejam testados em um conjunto maior de amostras. Devido à dificuldade na obtenção dos tipos de amostras necessárias, seria necessária a realização de um estudo multicêntrico. Com base na heterogeneidade das amostras, tanto do conjunto de treinamento como no conjunto de validação independente, poderiam ser coletadas amostras de pacientes de ambos os sexos, de todas as idades e de nódulos de diferentes tamanhos, ao exemplo do estudo de FOEKENS et al. (2006). O mesmo trabalho sugere que, como critérios de exclusão, devam ser considerados: a realização de terapia sistêmica adjuvante e neoadjuvante, histórico de outra malignidade primária e baixa qualidade do RNA ou da hibridização. Uma vez que a variabilidade de dados gerados por diferentes laboratórios é um obstáculo inerente à técnica de microarranjos de cDNA, seria interessante realizar a análise de variância nos dados de expressão gênica. Pelo mesmo motivo, é imprescindível que haja uma padronização das técnicas utilizadas e rigorosa observação dos detalhes estabelecidos por tal. Assim é possível assegurar a reprodutibilidade requerida para um estudo multicêntrico e, potencialmente, para a aplicação na rotina clínica.

8 CONCLUSÕES

O objetivo geral deste projeto foi atingido com a validação dos trios classificadores capazes de realizar a distinção entre carcinomas foliculares e adenomas foliculares da tireoide, cujo diagnóstico pré-cirúrgico não pode ser realizado através das técnicas convencionais devido às semelhanças entre estas duas patologias.

De acordo com os dados aqui obtidos, o melhor trio classificador foi o trio número 18, composto pelos genes FUBP1, GRCC9 e UBE1C, seguido dos trios 7 (SNAI2, UBE1C e SMCR8) e 24 (TRIM2, POLD3 e TSHR). Com estes dados, é possível que, no futuro, possa haver um esforço concentrado em reunir amostras para avaliação destes classificadores em um maior número de amostras, visando o estabelecimento de um método diagnóstico que possa ser implementado na rotina clínica, cuja prática poderia diminuir significativamente o número de tireoidectomias totais realizadas devido às patologias aqui descritas.

O protocolo de amplificação foi otimizado de modo a amplificar amostras partindo de massa de RNA total inferior a 30ng. Tal fato tem especial importância no presente estudo, que empregou amostras de aspirados tireoideanos, sabidamente de baixa celularidade. O processo de amplificação proporcionou, inclusive, a utilização de amostras remanescentes de PAAFs diagnósticas pré-cirúrgicas, coletadas por lavagem da agulha utilizada com TRIzol[®], sem necessidade de punções

extras e sem comprometimento do exame citológico de rotina. Outra vantagem da padronização da extração e amplificação do RNA das amostras aqui descritas, coletadas por diferentes técnicas, é que isto pode proporcionar aos pesquisadores uma nova fonte de estudos, uma vez que diferentes informações podem ser obtidas de acordo com a natureza da amostra utilizada.

Foi estabelecida a metodologia para utilização de lâminas espelhadas, mais sensíveis que as lâminas de microarranjos não-reflexivas. No entanto, mesmo sendo melhores que as lâminas usuais, ambas sofreram a influência do período entre a impressão e a hibridização, que foi maior do que especificados pelo fabricante.

9 REFERÊNCIAS BIBLIOGRÁFICAS

Ancona N, Maglietta R, Piepoli A, et al. On the statistical assessment of classifiers using DNA microarray data. **BMC Bioinformatics** 2006; 7:387-401.

Barden CB, Shister KW, Zhu B, et al. Classification of follicular thyroid tumors by molecular signature: results of gene profiling. **Clin Cancer Res** 2003; 9:1792-800.

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 Antonio Prudente]

Bartolazzi A, Gasharri A, Papotti M, et al. Application of an immunodiagnostic method for improving preoperative diagnosis of nodular thyroid lesions. **Lancet** 2001; 357:1644-50.

Berger JA, Hautaniemi S, Järvinen AK, Edgren H, Mitra SK, Astola J. Optimized LOWESS normalization parameter selection for DNA microarray data. **BMC Bioinformatics** 2004; 5:194.

Braga ACS. **Curvas ROC: aspectos funcionais e aplicações**. Braga; 2000. [Tese de Doutorado-Universidade do Minho].

Boulesteix AL, Strobl C, Augustin T, Daumer M. Evaluating Microarray-based Classifiers: An Overview. **Cancer Informatics** 2008; 6:77-97.

Brentani RR, Carraro DM, Verjovski-Almeida S, et al. Gene expression arrays in cancer research: methods and applications. **Crit Rev Oncol Hematol** 2005; 54:95-105.

Cardoso F, Van't Veer L, Rutgers E, Loi S, Mook S, Piccart-Gebhart MJ. Clinical application of the 70-gene profile: the MINDACT trial. **J Clin Oncol** 2008; 26:729-35.

Cerutti JM, Delcelo R, Amadei MJ, et al. A preoperative diagnostic test that distinguishes benign from malignant thyroid carcinoma based on gene expression. **J Clin Invest** 2004; 113:1234-42.

Chaplin J. **Thyroid lumps and nodules: a patient's guide**. Medic8@Family Health Guide, 2005. Available from: <URL:<http://www.medic8.com/healthguide/articles/thyroidlumps.html>> [2010 jan 4]

Choumane H, Ha N, Nelep C, et al. Double interference fluorescence enhancement from reflective slides: application to bicolor microarray. **Appl Phys Lett** 2005; 87:031102.

Clark OH, Noguchi S. Preface. In: Clark OH, Noguchi S, editors. **Thyroid cancer: diagnosis and treatment**. St Louis, Missouri: Quality Medical Publishing; 2000; p. ix.

Dudoit S, Fridlyand J, Speed T. Comparison of discrimination methods for the classification of tumors using gene expression data. **J Am Stat Assoc** 2003; 97:77-87.

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

[FCM-UNICAMP] Faculdade de Ciências Médicas. Departamento de Anatomia Patológica. Técnicas Histológicas Empregadas. Anatomia Patológica Especial-módulo-**Patologia endócrina: adenoma folicular da tireoide/Infiltração da cápsula e vasos**. Disponível em: <URL:<http://www.fcm.unicamp.br/deptos/anatomia/lamendo13.html>> <URL:<http://www.fcm.unicamp.br/deptos/anatomia/lamendo15.html>> [2010 maio 12]

Foekens JA, Atkins D, Zhang Y, et al. Multicenter validation of a gene expression–based prognostic signature in lymph node-negative primary breast cancer. **J Clin Oncol** 2006; 24:1665-71.

Ge G, Wong GW Classification of premalignant pancreatic cancer mass-spectrometry data using decision tree ensembles. **BMC Bioinformatics** 2008, 9:275-87.

Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class prediction by gene expression monitoring. **Science** 1999; 286:531-7.

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(2 Suppl):S19-21.

Grebe SKG, Hay ID. Clinical evaluation of thyroid tumors. In: Thawley SE, Panje WR, Batsakis JG, et al., editors. **Comprehensive management of head and neck tumors**. 2nd ed. Philadelphia: WB Saunders; 1999. p.1694-709.

Hamburger JI. Diagnosis of thyroid nodules by fine needle biopsy: use and abuse. **J Clin Endocrinol Metab** 1994; 79:335-9.

Harach HR, Fransiella KO, Wasenius VM. Occult papillary carcinoma of the thyroid. a "normal" finding in Finland. a systematic autopsy study. **Cancer** 1985; 56:531-8.

Hummel M, Metzeler KH, Buske C, Bohlander SK, Mansmann U. Association between a prognostic gene signature and functional gene sets. **Bioinform Biol Insights** 2008; 22:329-41.

Inohara H, Honjo Y, Yoshii T, et al. Expression of galectin-3 in fine-needle aspirates as a diagnostic marker differentiating benign from malignant thyroid neoplasms. **Cancer** 1999; 85:2475-84.

Jeffrey SS, Pollack JR. The diagnosis and management of pre-invasive breast disease: promise of new technologies in understanding pre-invasive breast lesions. **Breast Cancer Res** 2003; 5:320-8.

Kanno J, Onodera H, Furuta K, et al. Tumor-promoting effects of both iodine deficiency and iodine excess in the rat thyroid. **Toxicol Pathol** 1992; 20:226-35.

Kebebew E, Peng M,Reiff E, et al. ECM1 and TMPRSS4 are diagnostic markers of malignant thyroid neoplasms and improve the accuracy of fine needle aspiration biopsy. **Ann Surg** 2005; 242:353-64.

Lemoine LR, Mayall ES, Willie FS, et al. High frequency of ras oncogene activation in all stages of human thyroid tumorigenesis. **Oncogene** 1989; 4:159-64.

Lyng H, Badiee A, Svendsrud DH, Hovig E, Myklebost O, Stokke T. Profound influence of microarray scanner characteristics on gene expression ratios: analysis and procedure for correction. **BMC Genomics** 2004; 5:10.

Mackenzie EJ, Mortimer RH. Thyroid nodules and thyroid cancer. **Med J Aust** 2004; 180:242-7.

Martins TD, Carrilho F, Leitão P, et al. Anaplastic thyroid tumor: retrospective analysis of 12 cases. **Acta Med Port** 2004; 17:349-51

Mary-Huard T, Daudin J, Robin S, Bitton F, Cabannes E, Hilson P. Spotting effect in microarray experiments. **BMC Bioinformatics** 2004; 5:63.

Mehrotra P, Okpokam A, Bouhaidar R, et al. Galectin-3 does not reliably distinguish benign from malignant thyroid neoplasms. **Histopathology** 2004; 45:493-500.

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.

Mills LJ, Poller DN, Yiangou C. Galectin-3 is not useful in thyroid FNA. **Cytopathology** 2005; 16:132-8.

Murphy D. Gene expression studies using microarrays: principles, problems, and prospects. **Adv Physiol Edu** 2002; 26:256-70.

Noguchi S. Overview. In: Clark OH, Noguchi S, editors. **Thyroid cancer: diagnosis and treatment**. St Louis: Quality Medical Pu; 2000. p.3-20.

Perou CM, Jeffrey SS, van de Rijn M, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. **PNAS** 1999; 96:9212-7.

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

Petalidis L, Bhattacharyya S, Morris GA, Collins VP, Freeman TC, Lyons PA. Global amplification of mRNA by template-switching PCR: linearity and application to microarray analysis. **Nucleic Acids Res** 2003; 31:e142.

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

Ramaswamy S, Tamayo P, Rifkin R. Multiclass cancer diagnosis using tumor gene expression signatures. **PNAS** 2001; 98:15149-54.

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

Rosenzweig BA, Pine PS, Domon OE, Morris SM, Chen JJ, Sistare FD. Dye bias correction in dual-labeled cDNA microarray gene expression measurements. **Environ Health Perspect** 2004; 112:480-7.

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.

Ross JS. Multigene classifiers, prognostic factors, and predictors of breast cancer clinical outcome. **Adv Anat Pathol** 2009; 16:204-15.

Ransohoff DF. Bias as a threat to the validity of cancer molecular-marker research. **Nat Rev Cancer** 2005; 5:142-9.

Santos GS. Punção aspirativa por agulha fina. In: Dedivitis RA, Guimarães AV, editores. **Patologia cirúrgica da tireoide**. São Paulo: Frôntis Editorial, 1999. p.67-74.

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

SCHOTT North America. Reflective optical coating: Nexterion®HiSens slides. Available from: <URL:http://www.us.schott.com/nexterion/english/products/coated_slides/reflective/reflective.html> [2010 jun 4].

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

Smith J, Cheifetz RE, Schneiderei N, Berean K, Thomson T. Can cytology accurately predict benign follicular nodules? **Am J Surg**. 2005; 189:592-5.

Sontrop HMJ, Moerland PD, Ham RVD, et al. A comprehensive sensitivity analysis of microarray breast cancer classification under feature variability. **BMC Bioinformatics** 2009, 10:389-411.

Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. **J Clin Oncol** 2008; 26:721-8.

Stolf BS, Santos MMS, 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: 891-900.

't Hoen PA, de Kort F, van Ommen GJ, den Dunnen JT. Fluorescent labelling of cRNA for microarray applications. **Nucleic Acids Res** 2003; 31:e20.

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

Van't Veer L, Dai H, Vijver M van de, et al. Gene expression profiling predicts clinical outcome of breast cancer. **Nature** 2002; 415:530-6.

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

Wang X, Hessner MJ, Wu Y, Pati N, Ghosh S. Quantitative quality control in microarray experiments and the application in data filtering, normalization and false positive rate prediction. **Bioinformatics** 2003; 19:1341-7.

Weber F, Shen L, Aldred MA, et al. Genetic classification of benign and malignant thyroid follicular neoplasia based on a three-gene combination. **J Clin Endocrinol Metab** 2005; 90:2512-21.

Wei C, Li J, Bumgarner RE. Sample size for detecting differentially expressed genes in microarray experiments. **BMC Genomics** 2004; 5:87.

Weiss M, Baruch A, Keydar I, Wreschner DH. Preoperative diagnosis of thyroid papillary carcinoma by reverse transcriptase polymerase chain reaction of the MUC1 gene. **Int J Cancer** 1996; 66:55-9.

Yamashita H. Evaluation of thyroid nodules. In: Clark OH, Noguchi S, editors. **Thyroid cancer: diagnosis and treatment**. St Louis, Missouri: Quality Medical Pu; 2000. p.21-38.

Yang YH, Dudoit S, Luu P, et al. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. **Nucleic Acids Res** 2002; 30:e15.

Zeng X, Li G, Yang JY, Yang MQ, Wu G. Dimension reduction with redundant gene elimination for tumor classification. **BMC Bioinformatics** 2008; 9(Suppl 6):S8.

Apêndice 1 - Folha de trabalho na coleta por PAAF

Projeto de Extração de RNA
PAAF Tireoide

Dr. Luis Fernando Lima Reis
Ana Helena Pagotto

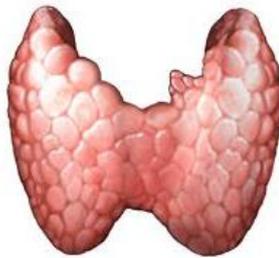
Identificação do Paciente:

RGH:

Procedimento:

Tireoidectomia Total Lobo Direito Istmo Lobo Esquerdo

Nódulos puncionados:



- | | | |
|----|--------------|--------------------|
| 1. | Localização: | Número de Lâminas: |
| 2. | Localização: | Número de Lâminas: |
| 3. | Localização: | Número de Lâminas: |
| 4. | Localização: | Número de Lâminas: |
| 5. | Localização: | Número de Lâminas: |

PAAF prévia:

Data:

Topografia:

Tamanho:

Diagnóstico:

Data:

Topografia:

Tamanho:

Diagnóstico

Contato: 3388-3236 – Laboratório de Genômica Funcional

Apêndice 2 - Termo de Consentimento Livre e Informado

Você está sendo admitido neste Serviço para estabelecer diagnóstico e/ou tratamento de sua lesão de tireoide. Para fins de diagnóstico, fator prognóstico e/ou como parte do tratamento, poderá haver necessidade de realizarmos remoção de material biológico relacionado à sua doença. Esta remoção poderá ser realizada por cirurgia ou por biópsia da lesão. Parte deste material retirado é utilizado para os exames clínicos e/ou laboratoriais, necessários para um diagnóstico definitivo.

Para obter um maior conhecimento clínico e científico do câncer, o corpo clínico do Hospital A.C. Camargo (médicos e pesquisadores) desenvolve pesquisa clínica científica. Através desta pesquisa é possível conhecer melhor os mecanismos da doença e, portanto, oferecer novas possibilidades de diagnóstico e tratamento. Grande parte desses trabalhos de pesquisa requer a utilização de fragmentos de tecidos, necessários para a busca de novos genes ou de alterações em genes já conhecidos. Esses fragmentos de tecidos correspondem ao excedente de material retirado de pacientes para a realização do diagnóstico, conforme mencionado acima.

Caso você concorde em ceder o excedente do seu material para as atividades de pesquisa, o fragmento de tecido ou líquido aspirado por agulha não utilizado será identificado no laboratório por um código formado por números e letras e portanto sua privacidade e identidade serão preservadas. A eventual inclusão dos dados em publicação científica será feita de modo a manter o anonimato do paciente.

É muito importante esclarecer que só será considerado excesso, o material restante após o estabelecimento final de seu diagnóstico e que nenhum fragmento de tecido será utilizado para atividades de pesquisa antes do seu diagnóstico definitivo. Concordando com a retirada e o uso deste fragmento excedente, como descrito acima, é necessário esclarecer que você não terá quaisquer benefícios ou direitos financeiros sobre os eventuais resultados decorrentes da pesquisa. Caso não concorde em doar

o fragmento de biópsia para pesquisa, sua decisão não influenciará, de nenhum modo, o seu tratamento.

Você receberá uma cópia deste documento e o original será arquivado em seu prontuário. Caso você tenha questões a fazer sobre este acordo ou alguma dúvida que não tenha sido esclarecida pelo seu médico, por gentileza, entre em contato com o coordenador do Comitê de Ética do Hospital A.C. Camargo, pelo telefone 2189-5083.

NOME DOS PESQUISADORES RESPONSÁVEIS PELO PROJETO:

Ana Helena Pagotto, Pós-Graduação – Fundação Antonio Prudente

Dr. Luiz Fernando Lima Reis – Fundação Antonio Prudente

RG do prontuário médico:

Médico responsável:

Data:

Nome do(a) paciente ou representante legal

Assinatura

Apêndice 3 - Trios utilizados na validação independente

Trio 1:

ugc|Hs.334697.2|Hs.334697 GENE=GRCC9 PROTSIM=pir:T26185 likely ortholog of mouse gene rich cluster, C9 gene

ugc|Hs.332633.3|Hs.332633 GENE=BBS2 PROTSIM=ref:NP_501325.1 Bardet-Biedl syndrome 2

ugc|Hs.433574.5|Hs.433574 GENE= PROTSIM= Homo sapiens cDNA FLJ32493 fis, clone SKNSH2000051.

SVD: 11.5424

Número de acertos: 7

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	2	1	2	1	1	1	2	2	2	2	1	1	2
Teste	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE

Trio 2:

ugc|Hs.154320.5|Hs.154320 GENE=UBE1C PROTSIM=ref:NP_568370.1 ubiquitin-activating enzyme E1C (UBA3 homolog, yeast)

ugc|Hs.11123.1|Hs.11123 GENE=DKFZP564G092 PROTSIM=ref:NP_490834.1 DKFZP564G092 protein

ugc|Hs.334576.1|Hs.334576 GENE=SMCR8 PROTSIM= Smith-Magenis syndrome chromosome region, candidate 8

SVD: 13.3317

Número de acertos: 9

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	2	2	1	1	1	1	1	2	2	1	1	2	2
Teste	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE

Trio 3:

ugc|Hs.334697.2|Hs.334697 GENE=GRCC9 PROTSIM=pir:T26185 likely ortholog of mouse gene rich cluster, C9 gene

ugc|Hs.21104.4|Hs.21104 GENE=KIAA1463 PROTSIM=pir:T34061 KIAA1463 protein

ugc|Hs.296842.7|Hs.296842 GENE=MYH10 PROTSIM=ref:NP_175858.1 myosin, heavy polypeptide 10, non-muscle

SVD: 11.0657

Número de acertos: 8

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	1	1	1	1	2	1	2	2	1	2	2	1	1
Teste	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE

Trio 4:

ugc|Hs.321497.1|Hs.321497 GENE= PROTSIM=ref:NP_060312.1 Homo sapiens cDNA FLJ31347 fis, clone MESAN2000023.

ugc|Hs.321579.4|Hs.321579 GENE=SLC5A6 PROTSIM=pir:T28017 solute carrier family 5 (sodium-dependent vitamin transporter), member 6

ugc|Hs.159237.2|Hs.159237 GENE=HK3 PROTSIM=ref:NP_175220.1 hexokinase 3 (white cell)

SVD: 11.0423

Número de acertos: 8

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	2	1	2	1	2	1	1	2	2	1	2	1	2
Teste	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE

Trio 7:

ugc|Hs.93005.1|Hs.93005 GENE=SNAI2 PROTSIM=ref:NP_492338.1 snail homolog 2 (Drosophila)

ugc|Hs.154320.5|Hs.154320 GENE=UBE1C PROTSIM=ref:NP_568370.1 ubiquitin-activating enzyme E1C (UBA3 homolog, yeast)

ugc|Hs.334576.1|Hs.334576 GENE=SMCR8 PROTSIM= Smith-Magenis syndrome chromosome region, candidate 8

SVD: 9.9142

Número de acertos: 11

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	2	1	1	1	1	1	1	2	2	2	2	2	2
Teste	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE

Trio 8:

ugc|Hs.72660.4|Hs.72660 GENE=PTDSR PROTSIM=ref:NP_177951.1 phosphatidylserine receptor

ugc|Hs.24950.6|Hs.24950 GENE=RGS5 PROTSIM=sp:P34295 regulator of G-protein signalling 5

ugc|Hs.321579.4|Hs.321579 GENE=SLC5A6 PROTSIM=pir:T28017 solute carrier family 5 (sodium-dependent vitamin transporter), member 6

SVD: 9.4584

Número de acertos: 6

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	2	2	2	1	1	2	1	2	1	1	2	2	2
Teste	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE

Trio 9:

ugc|Hs.107125.2|Hs.107125 GENE=PLVAP PROTSIM=pir:T31611 plasmalemma vesicle associated protein

ugc|Hs.111850.2|Hs.111850 GENE=MCSP PROTSIM=ref:NP_508659.1 mitochondrial capsule selenoprotein

ugc|Hs.99654.10|Hs.99654 GENE=POMT1 PROTSIM=ref:NP_524025.1 protein-O-mannosyltransferase 1

SVD: 9.1399

Número de acertos: 9

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	2	2	2	1	1	1	2	1	2	2	2	2	2
Teste	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Trio 10:

ugc|Hs.243960.6|Hs.243960 GENE=NDRG2 PROTSIM=pir:T48538 NDRG family member 2

ugc|Hs.154978.4|Hs.154978 GENE=FOE PROTSIM=pir:T29661 friend of EBNA2

ugc|Hs.24950.6|Hs.24950 GENE=RGS5 PROTSIM=sp:P34295 regulator of G-protein signalling 5

SVD: 8.9112

Número de acertos: 8

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	2	2	1	1	1	1	1	1	1	2	2	1	1
Teste	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE

Trio 11:

ugc|Hs.33829.1|Hs.33829 GENE=BHLHB3 PROTSIM=pir:T22341 basic helix-loop-helix domain containing, class B, 3

ugc|Hs.58382.1|Hs.58382 GENE=FLJ11101 PROTSIM=ref:NP_504829.1 hypothetical protein FLJ11101

ugc|Hs.182528.5|Hs.182528 GENE=ZNF263 PROTSIM=ref:NP_477243.1 zinc finger protein 263

SVD: 8.7094

Número de acertos: 7

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	2	2	2	1	1	1	2	2	2	2	2	1	1
Teste	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE

Trio 12:

ugc|Hs.112743.1|Hs.112743 GENE=SYCP1 PROTSIM=ref:NP_198994.1 synaptonemal complex protein 1

ugc|Hs.11039.6|Hs.11039 GENE=MEP50 PROTSIM=pir:T06120 MEP50 protein

ugc|Hs.7089.4|Hs.7089 GENE=INSIG2 PROTSIM=ref:NP_005533.1 insulin induced gene 2

SVD: 8.6048

Número de acertos: 8

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	2	1	1	1	2	1	2	2	1	2	2	2	2
Teste	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE

Trio 13:

ugc|Hs.70823.4|Hs.70823 GENE=SULF1 PROTSIM=ref:NP_508560.1 sulfatase 1

ugc|Hs.278507.1|Hs.278507 GENE=HARSL PROTSIM=ref:NP_186925.1 histidyl-tRNA synthetase-like

ugc|Hs.433559.1|Hs.433559 GENE=PSMD10 PROTSIM=ref:NP_178442.1 proteasome (prosome, macropain) 26S subunit, non-ATPase, 10

SVD: 8.5759

Número de acertos: 8

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	1	2	1	1	2	2	1	2	1	2	2	2	1
Teste	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE

Trio 14:

ugc|Hs.194728.1|Hs.194728 GENE=H6PD PROTSIM=ref:NP_173838.1 hexose-6-phosphate dehydrogenase (glucose 1-dehydrogenase)

ugc|Hs.215725.1|Hs.215725 GENE=FLJ25429 PROTSIM= hypothetical protein FLJ25429

ugc|Hs.154978.4|Hs.154978 GENE=FOE PROTSIM=pir:T29661 friend of EBNA2

SVD: 8.5494

Número de acertos: 9

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	2	1	1	1	1	1	2	1	1	2	1	2	1
Teste	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE

Trio 15:

ugc|Hs.93005.1|Hs.93005 GENE=SNAI2 PROTSIM=ref:NP_492338.1 snail homolog 2 (Drosophila)

ugc|Hs.243960.6|Hs.243960 GENE=NDRG2 PROTSIM=pir:T48538 NDRG family member 2

ugc|Hs.84389.10|Hs.84389 GENE=SNAP25 PROTSIM=ref:NP_172842.1 synaptosomal-associated protein, 25kDa

SVD: 8.3951

Numero de acertos: 10

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	1	1	2	1	2	1	1	2	2	2	2	2	2
Teste	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE

Trio 16:

ugc|Hs.5212.9|Hs.5212 GENE=SMUG1 PROTSIM=ref:NP_055126.1 single-strand selective monofunctional uracil DNA glycosylase

ugc|Hs.2437.2|Hs.2437 GENE=EIF2B5 PROTSIM=pir:T00478 eukaryotic translation initiation factor 2B, subunit 5 epsilon, 82kDa

ugc|Hs.332633.3|Hs.332633 GENE=BBS2 PROTSIM=ref:NP_501325.1 Bardet-Biedl syndrome 2

SVD: 8.3841

Numero de acertos: 6

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	2	1	2	1	1	1	2	2	2	2	1	1	1
Teste	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE

Trio 17:

ugc|Hs.14912.2|Hs.14912 GENE=KIAA0286 PROTSIM=pir:T16022 KIAA0286 protein

ugc|Hs.251383.1|Hs.251383 GENE=CHST4 PROTSIM=ref:NP_005760.1 carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 4

ugc|Hs.335918.5|Hs.335918 GENE=FDPS PROTSIM=sp:Q43315 farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase)

SVD: 7.1195

Número de acertos: 5

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	2	2	2	2	1	2	2	1	2	1	2	1	1
Teste	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE

Trio 18:

ugc|Hs.118962.7|Hs.118962 GENE=FUBP1 PROTSIM=ref:NP_180167.1 far upstream element (FUSE) binding protein 1

ugc|Hs.334697.2|Hs.334697 GENE=GRCC9 PROTSIM=pir:T26185 likely ortholog of mouse gene rich cluster, C9 gene

ugc|Hs.154320.5|Hs.154320 GENE=UBE1C PROTSIM=ref:NP_568370.1 ubiquitin-activating enzyme E1C (UBA3 homolog, yeast)

SVD: 8.127

Número de acertos: 12

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	1	1	2	1	1	1	1	2	2	2	2	2	2
Teste	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE

Trio 19:

ugc|Hs.13872.7|Hs.13872 GENE=C17 PROTSIM=ref:NP_061129.1 cytokine-like protein C17

ugc|Hs.119129.8|Hs.119129 GENE=COL4A1 PROTSIM=ref:NP_510664.1 collagen, type IV, alpha 1

ugc|Hs.334576.1|Hs.334576 GENE=SMCR8 PROTSIM= Smith-Magenis syndrome chromosome region, candidate 8

SVD: 7.7746

Número de acertos: 7

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32	
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF	
LDA	2	2	2	1	2	1	1	1	2	1	2	2	1	2	
Teste		FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE

Trio 20:

ugc|Hs.8361.2|Hs.8361 GENE=E2IG4 PROTSIM=pir:T46070 hypothetical protein, estradiol-induced

ugc|Hs.93605.3|Hs.93605 GENE=LZTS1 PROTSIM=pir:T47424 leucine zipper, putative tumor suppressor 1

ugc|Hs.5894.5|Hs.5894 GENE=FLJ10305 PROTSIM=ref:NP_565275.1 hypothetical protein FLJ10305

SVD: 8.0243

Número de acertos: 9

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32	
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF	
LDA	1	2	2	2	1	1	1	2	1	2	2	1	2	2	
Teste		TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE

Trio 21:

ugc|Hs.343551.1|Hs.343551 GENE=FLJ22584 PROTSIM=ref:NP_195462.1
hypothetical protein FLJ22584

ugc|Hs.194728.1|Hs.194728 GENE=H6PD PROTSIM=ref:NP_173838.1 hexose-6-
phosphate dehydrogenase (glucose 1-dehydrogenase)

ugc|Hs.321579.4|Hs.321579 GENE=SLC5A6 PROTSIM=pir:T28017 solute carrier
family 5 (sodium-dependent vitamin transporter), member 6

SVD: 8.0055

Número de acertos: 7

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	2	2	2	1	2	1	1	2	1	1	2	2	2
Teste	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE

Trio 22:

ugc|Hs.77929.5|Hs.77929 GENE=ERCC3 PROTSIM=ref:NP_568591.1 excision
repair cross-complementing rodent repair deficiency, complementation group 3
(xeroderma pigmentosum group B complementing)

ugc|Hs.251383.1|Hs.251383 GENE=CHST4 PROTSIM=ref:NP_005760.1
carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 4

ugc|Hs.149957.4|Hs.149957 GENE=RPS6KA1 PROTSIM=ref:NP_187484.1
ribosomal protein S6 kinase, 90kDa, polypeptide 1

SVD: 6.5554

Número de acertos: 9

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	2	1	2	1	1	1	1	1	2	1	2	1	2
Teste	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE

Trio 23:

ugc|Hs.18759.3|Hs.18759 GENE=C21orf108 PROTSIM=pir:T04824 chromosome 21 open reading frame 108

ugc|Hs.168073.1|Hs.168073 GENE=C20orf188 PROTSIM=ref:NP_056453.1 chromosome 20 open reading frame 188

ugc|Hs.9605.1|Hs.9605 GENE=CPSF5 PROTSIM=ref:NP_194285.1 cleavage and polyadenylation specific factor 5, 25 kDa

SVD: 7.8572

Número de acertos: 8

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	1	2	2	1	1	2	1	2	2	2	1	1	2
Teste	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE

Trio 24:

ugc|Hs.12372.3|Hs.12372 GENE=TRIM2 PROTSIM=pir:S44890 tripartite motif-containing 2

ugc|Hs.82502.2|Hs.82502 GENE=POLD3 PROTSIM=pir:T34036 polymerase (DNA directed), delta 3

ugc|Hs.123078.2|Hs.123078 GENE=TSHR PROTSIM=pir:T20123 thyroid stimulating hormone receptor

SVD: 7.8567

Número de acertos: 11

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	2	1	2	1	1	1	1	2	2	2	2	2	2
Teste	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE

Trio 25:

ugc|Hs.379090.1|Hs.379090 GENE= PROTSIM=sp:Q13491 Homo sapiens cDNA FLJ38338 fis, clone FCBBF3027104, highly similar to Mus musculus proteolipid M6B isoform alpha-beta-TMD-omega (M6B) mRNA.

ugc|Hs.11039.6|Hs.11039 GENE=MEP50 PROTSIM=pir:T06120 MEP50 protein

ugc|Hs.89472.1|Hs.89472 GENE=AGTR1 PROTSIM=pir:JC7319 angiotensin II receptor, type 1

SVD: 7.8165

Número de acertos: 9

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	2	1	1	1	2	1	2	2	2	2	1	2	2
Teste	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE

Trio 26:

ugc|Hs.94761.2|Hs.94761 GENE=KIAA1691 PROTSIM=ref:NP_174461.1 KIAA1691 protein

ugc|Hs.183105.1|Hs.183105 GENE=STRN3 PROTSIM=pir:T46032 striatin, calmodulin binding protein 3

ugc|Hs.24879.1|Hs.24879 GENE=PPAP2C PROTSIM=ref:NP_565255.1 phosphatidic acid phosphatase type 2C

SVD: 7.4962

Número de acertos: 4

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	1	1	2	2	2	1	2	2	1	2	1	1	1
Teste	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE

Trio 27:

ugc|Hs.24879.1|Hs.24879 GENE=PPAP2C PROTSIM=ref:NP_565255.1
phosphatidic acid phosphatase type 2C

ugc|Hs.134342.1|Hs.134342 GENE=LANCL2 PROTSIM=ref:NP_175700.1 LanC
lantibiotic synthetase component C-like 2 (bacterial)

ugc|Hs.8663.1|Hs.8663 GENE=KIAA0321 PROTSIM=pir:T49237 KIAA0321
protein

SVD: 7.4273

Número de acertos: 6

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32	
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF	
LDA	2	2	1	1	2	1	1	2	2	2	2	1	1	1	
Teste		FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE

Trio 28:

ugc|Hs.13872.7|Hs.13872 GENE=C17 PROTSIM=ref:NP_061129.1 cytokine-like
protein C17

ugc|Hs.406790.2|Hs.406790 GENE= PROTSIM= Homo sapiens mRNA full length
insert cDNA clone EUROIMAGE 2005635.

ugc|Hs.193143.3|Hs.193143 GENE=KIAA1069 PROTSIM=ref:NP_568881.1
KIAA1069 protein

SVD: 7.1723

Número de acertos: 9

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32	
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF	
LDA	1	2	2	1	1	1	1	2	2	1	2	2	2	2	
Teste		TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE

Trio 29:

ugc|Hs.103915.1|Hs.103915 GENE=KIAA0346 PROTSIM=ref:NP_509450.1
KIAA0346 protein

ugc|Hs.268281.2|Hs.268281 GENE=CRNKL1 PROTSIM=ref:NP_199411.1 Crn,
crooked neck-like 1 (Drosophila)

ugc|Hs.335918.5|Hs.335918 GENE=FDPS PROTSIM=sp:Q43315 farnesyl
diphosphate synthase (farnesyl pyrophosphate synthetase,
dimethylallyltranstransferase, geranyltranstransferase)

SVD: 7.1054

Número de acertos: 4

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	1	2	1	2	2	1	2	2	2	1	1	2	1	1
Teste	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE

Trio 30:

ugc|Hs.44208.2|Hs.44208 GENE=FLJ23153 PROTSIM=sp:Q9UHE8 likely ortholog
of mouse tumor necrosis-alpha-induced adipose-related protein

ugc|Hs.103291.5|Hs.103291 GENE=NRN1 PROTSIM=ref:NP_057672.1 neuritin 1

ugc|Hs.406790.2|Hs.406790 GENE= PROTSIM= Homo sapiens mRNA full length
insert cDNA clone EUROIMAGE 2005635.

SVD: 7.0533

Número de acertos: 9

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	2	1	1	1	1	1	2	2	2	2	2	1	2
Teste	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE

Trio 31:

ugc|Hs.99962.5|Hs.99962 GENE=PRG2 PROTSIM=ref:NP_501369.1 proteoglycan 2, bone marrow (natural killer cell activator, eosinophil granule major basic protein)

ugc|Hs.288658.2|Hs.288658 GENE=ZNF35 PROTSIM=ref:NP_477243.1 zinc finger protein 35 (clone HF.10)

ugc|Hs.124969.1|Hs.124969 GENE= PROTSIM=pir:T45715 Homo sapiens clone 24707 mRNA sequence

SVD: 6.7023

Número de acertos: 7

Amostras	340	477	g04	g12	g20	g21	g28	jp10	jp20	g16	g19	g30	g31	g32
Patologista	A	A	A	A	A	A	A	A	A	CF	CF	CF	CF	CF
LDA	2	2	2	1	1	2	1	1	2	2	1	2	1	2
Teste	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE