UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE QUÍMICA

INSTITUTO LUDWIG DE PESQUISA SOBRE O CÂNCER

IDENTIFICAÇÃO DE MARCADORES MOLECULARES PARA O CÂNCER DE
TIREÓIDE UTILIZANDO cDNA MICROARRAYS

BEATRIZ SIMONSEN STOLF

TESE DE DOUTORADO

ORIENTADOR: PROFA. DRA. LUISA LINA VILLA

CO-ORIENTADOR: PROF. DR. LUIZ FERNANDO LIMA REIS

SÃO PAULO

2003

"Identificação de marcadores moleculares para o câncer de tireóide por cDNA microarrays"

BEATRIZ SIMONSEN STOLF

Tese de Doutorado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Doutor em Ciências – Área: Bioquímica.

Aprovada por:

Prof. Dr. LUIZ FERNANDO LIMA REIS
(Presidente)
Ludwig

Prof. Dr. SERGIO VERJOVSKI DE ALMEIDA
IQ – USP

Profa. Dra. MARI CLEIDE SOGAYAR
IQ – USP

Profa. Dra. EDNA TERUKO KIMURA
ICB – USP

Prof. Dr. LUIZ ROBERTO NUNES
UMC

SÃO PAULO
29 DE SETEMBRO 2003

"Microarrays are one of the first of what will be a series of new technologies that provide a closer look at the workings of cells at a molecular level"-
Richard Simon, NCI

"It's too easy to find seemingly significant patterns that aren't really there. It's like staring at stars in the night sky: with enough desire and imagination, one can find just about any pattern one wants to see."- Tom Downey, Partek (statistical analysis provider)

A meus pais, por acreditarem em mim e me ajudarem a fazer o mesmo nas horas mais difíceis.

Ao Marcelo, por ser tão especial e pela surpreendente paciência
dos últimos tempos.

A Deus, por me dar curiosidade pelos mistérios da vida e vontade
de buscar explicações.

AGRADECIMENTOS

Ao meu orientador, Luiz Fernando, por conduzir o laboratório de uma forma tão pessoal e agradável. Agradeço pelas várias oportunidades em palestras e congressos, pela confiança em mim e no meu trabalho e por ter me ensinado a ser independente e profissional.

A minha orientadora Luisa Villa, pelo carinho e confiança. Pelas conversas profundas de algumas horas ou alguns minutos, que valeram por toda uma vida.

Ao Dr. Luiz Paulo Kowalski, que levantou o problema que se tornou o foco deste projeto. Um exemplo de cirurgião que se mantém atualizado e que enxerga muito além do que faz.

Ao nosso laboratório, que embora dividido em quatro salas é indivisível. Aos amigos, mais do que colegas que fiz dentro desse grupo, com quem partilhei, ao longo desses anos, muito mais do que experimentos de biologia molecular. Aqui vão algumas poucas palavras que exprimem parte do que tenho a agradecer a cada um de vocês : Lu e Wal, grandes companheiras de todas as horas, amigas e torcedoras: vocês não vão ficar livres de mim tão fácil! Alex, “nosso Magaiver”, com sua competência e capacidade de fazer coisas muito complexas de um jeito simples: obrigada pela ajuda, disposição e interesse! Susana, pela alegria e sensibilidade e pela companhia em momentos de descontração tão necessários. Abrantes, pela preocupação silenciosa, pela ajuda sempre que precisei; um exemplo de que alguns gestos falam mais do que mil palavras! Adriana, sincera e forte, que nessa fase final complicada me disse tantas coisas importantes... Lara, pelas conversas sérias e pelas piadas, com suas histórias mais engraçadas do mundo! Regina e Sibebe, pelas gentilezas e quitutes. Gustavo, sempre bem-humorado, pelas ajudas com explicações e gráficos. Fábio, meu ex-vizinho, pela paciência com meus desabafos e pelos aprimoramentos das listas e do site. Ana e Bianca, companheiras de intermináveis e sensacionais almoços e parceiras de “empreendimentos artísticos e estratégicos”- difícil encontrar um trio tão criativo! A Ana, também pela ajuda incansável nos infindáveis alinhamentos de spots...Cham, sempre dedicado, sorridente e calmo, disposto a ajudar e fazer tudo-que-for-preciso. Mariana, um exemplo de força de vontade, de interesse e de competência, sempre com um sorriso no rosto. Sarah, pela ajuda com os clones, pela simpatia e pela “boutique”! Ana Coló, Emerson e Vlad, exemplos de disposição e companheirismo mesmo com uma rotina tão corrida! Ao Álvaro (o “Aluno”) pelo

entusiasmo, interesse e alegria durante suas passagens pelo nosso laboratório. A Aline, que além de muito competente e cheia de iniciativa, é uma simpatia!

Aos colegas de outros laboratórios, difíceis de enumerar, pela amizade, torcida, pelo clima gostoso que criam no Instituto. Com certeza sentirei saudades...Em especial ao laboratório de Virologia e a Paty e Katy, que já fazem parte do nosso laboratório, pelo carinho e convívio. Ao Enríque, por conversas bastante importantes e pelo interesse e estímulo. A Angelita, pela amizade e companhia tão gostosa.

A Ana Cristina e Elis, pela ajuda com os sequenciamentos, unindo profissionalismo e amizade.

Aos funcionários do Ludwig, pela simpatia, atenção e ajuda. A Stella, pela competência, sensibilidade e carinho.

A Helena, Artur e Luiz Paulo, pelas análises de BLAST dos clones sequenciados. Acredito muito nessa nova parceria!

A nossos colaboradores do IME, Eduardo Jordão Neves e Roberto Hirata Jr., pela análise matemática dos dados de array, impossíveis aos “biólogos de bancada”. Jordão, obrigada pelas aulas e pela paciência com minha falta de conhecimento sobre o assunto!

Ao Dr. Fernando Soares, pela análise patológica de todas as lâminas do projeto e pela semi-microdissecção dos tecidos cirúrgicos.

Ao Carlinhos e à Myiuki, pelas lâminas para a análise patológica: obrigada pela eficiência, boa vontade e competência!

Aos colaboradores do projeto sul-americano de tireóide: Dr Juan Postigo, Maria Paula Curado e Marcos Brasilino, pelas amostras cirúrgicas e relatórios. Ao Dr. Juan, em especial, pela competência e cuidado, e pela hospitalidade quando de minha estadia no Peru.

A meus familiares: pais, irmãos, avós, tios e primos, pela torcida constante e por formarem uma família tão especial. A meus pais, por todas as oportunidades que me deram e continuam me dando na vida e pelo exemplo em tantos aspectos. Me sinto privilegiada. Ao Marcelo, meu namorado e grande amigo.

A meus amigos de faculdade, que compreendem essa vida meio maluca de pós graduando. Pelo convívio e amizade que se mantêm muito intensos mesmo com o passar dos anos.

A minhas queridas amigas Ciça, Dea, Eleine, Fla, Gê, Kiki, Mô, Rê e Sil, pelo carinho e torcida e por tentarem entender minha rotina e meus problemas mesmo não tendo a menor idéia do que faço: isso é uma das grandezas da amizade!

Aos meus amigos da Santa Teresa, pela torcida, carinho e orações. Muito obrigada!

Aos meus ex-colegas de lab: Márcia, Nancy, Clara (in memorian), Aurélio, Marcelo e Ricardo, pela torcida à distância. À minha ex-orientadora Bianca Zingales, por ter me ensinado a ter autonomia, precisão nos experimentos e na forma de expressão.

À secretaria de pós-graduação do IQ-USP, em especial ao Milton, pelas informações e pela atenção sempre que precisei.

Ao IQ-USP, seus docentes e funcionários, pelo excelente nível da pós-graduação, da pesquisa e dos cursos. Obrigada pela oportunidade de mais uma vez ser aluna desse curso.

Ao Instituto Ludwig, pelo apoio financeiro e infra-estrutura que permitiram o desenvolvimento de um trabalho tão custoso em um país tão sem recursos. Ao Prof. Ricardo Brentani.

Às bibliotecárias do Hospital, pela eficiência e gentileza.

À FAPESP pela bolsa de estudos.

ÍNDICE

1. INTRODUÇÃO

1.1. O câncer	1
1.2. O câncer de tireóide	2
1.3. Expressão diferencial em câncer de tireóide	3
1.3.1. Fatores de transcrição, fatores de crescimento, e seus receptores	4
1.3.2. Moléculas com função de adesão	7
1.3.3. Enzimas proteolíticas e moléculas que regulam sua função	9
1.3.4. Moléculas que regulam proliferação e morte celular	10
1.3.5. Moléculas com outras funções	11
1.3.6. Busca de marcadores em câncer de tireóide	13
1.4. Diagnóstico	15

2. SOBRE A ABORDAGEM METODOLÓGICA

2.1. Expressão diferencial	17
2.2. DNA microarray	20
A escolha da técnica de DNA microarray	
As amostras utilizadas como alvos para microarray	
A elaboração dos experimentos de microarray	
Determinação da intensidade dos spots de microarrays	
2.3. Métodos de análise	30
2.4. Amplificação de RNA	31

3. OBJETIVOS

4. MATERIAIS E MÉTODOS

4.1. Nucleotídeos radioativos e fluorescentes	35
4.2. Enzimas	35
4.3. Vetores	35
4.4. Oligonucleotídeos	35
4.5. Coleta das amostras	37
4.6. Extração de RNA total, tratamento com DNase I	38

4.7. Avaliação da qualidade do RNA: Northern blot e gel de agarose	38
Northern blot	
Gel de agarose com brometo de etídio	
4.8. Differential Display RT-PCR (DDRT-PCR)	41
4.9. Recuperação das bandas do gel de DDRT-PCR, reamplificação, clonagem	42
Preparação de bactérias competentes e transformação	
Seleção de clones positivos	
Preparação de DNA plasmidial recombinante em pequena escala	
4.10. Sequenciamento	44
4.11. Real-Time RT-PCR	44
4.12. RT-PCR semi-quantitativo	45
4.13. Southern blot	45
Hibridização de Southern blot com sonda de oligonucleotídio	
Hibridização de Southern blot com sonda de fragmento de DNA	
Quantificação das bandas do Southern blot	
4.14. Métodos de classificação	47
Análise de cluster	
Árvore de decisão	
4.15. Construção de arrays	
4.15.1. Membranas de nylon	48
Hibridização com cDNA radioativo	
Digitalização da imagem e quantificação do sinal dos spots	
Análise dos dados	
4.15.2. Lâminas de vidro	49
Hibridização com cDNA fluorescente	
Digitalização das imagens e análise da qualidade da hibridização	
Gel de agarose para análise da qualidade do cDNA fluorescente	
Quantificação da intensidade dos spots	
Análise dos dados de microarray	
4.16. Amplificação do RNA	52
 5. RESULTADOS	
5.1. DDRT-PCR e cDNA array em membrana de nylon	55
5.2. Identidade dos clones 18, 44, 199, 234	59

5.3. Real-Time RT-PCR	60
5.4. RT-PCR semi-quantitativo	63
5.5. Análise de cluster	67
5.6. Árvore de decisão	69
5.7. Os experimentos de microarray	
5.7.1. O uso de amostra de referência e a inversão de corantes	69
5.7.2. Normalização das intensidades dos dois canais	70
5.7.3. “Validação” da quantificação pelo programa Quantarray	73
5.7.4. Comparações A x B, ranking dos valores de p	75
5.7.5. Comparações AB x CD	86
5.8. Amplificação de RNA	100
 6. DISCUSSÃO	
6.1. Sobre os resultados obtidos	
6.1.1. DDRT-PCR e cDNA array em membrana de nylon	104
6.1.2. IGFBP-5	104
6.1.3. Real-Time RT-PCR	104
6.1.4. RT-PCR, cluster, árvore de decisão	105
6.1.5. Microrrays	
6.1.5.1. Os genes encontrados	105
6.1.5.2. Quantarray	107
6.1.5.3. Normalização	107
6.2. Sobre os problemas dos resultados obtidos	108
6.2.1. Câmera CCD x laser	109
6.2.2. Fixação do DNA na lâmina	110
6.2.3. Qualidade do RNA de tireóide	111
 7. CONCLUSÕES E PERSPECTIVAS	114
8. RESUMO	115
9. ABSTRACT	117
10. BIBLIOGRAFIA	119
CURRICULUM VITAE	130
ANEXO	

ABREVIATURAS

ATP	trifosfato de adenosina
BLAST	ferramenta básica de busca de alinhamentos locais
BSA	albumina sérica bovina
cDNA	DNA complementar, obtido por transcrição reversa de RNA
CTP	trifosfato de citidina
DDRT-PCR	“ differential display RT-PCR”
DEP	dietilpirocarbonato
DNA	ácido desoxirribonucleico
DNase	desoxirribonuclease
dNTPs	mistura de desoxinucleosídeos trifosfato
DO	densidade óptica
DTT	ditiotretol
EDTA	ácido etilenodiaminotetracético
EST	“expressed sequence tags”
GTP	trifosfato de guanidina
IGF	“insulin-like growth factor”
IGFBP	“ insulin-like growth factor binding protein”
IPTG	isopropil- β -D-tiogalactopiranosídeo
mRNA	RNA mensageiro
nt	nucleotídeo
pb	pares de bases
PAAF	punção aspirativa por agulha fina
PCR	reação em cadeia da polimerase
qsp	quantidade suficiente para
RNA	ácido ribonucleico
RNase	ribonuclease
rpm	rotações por minuto
RT-PCR	PCR a partir de reação de transcrição reversa
SDS	lauril sulfato de sódio
SSC	solução salina com citrato de sódio
TAE	solução de Tris, acetato e EDTA

TBE	solução de Tris, borato e EDTA
TE	solução de Tris e EDTA
T _m	temperatura de fusão (metade das moléculas na forma de fita simples)
Tris	Tris (hidroximetil) aminometano
UV	radiação ultravioleta
V	volts
X-gal	5-bromo-4-cloro-3-indolil- β -D-galactopiranosídeo

1. INTRODUÇÃO

1.1. O câncer

O câncer ou tumor maligno é uma das maiores causas de morte em países desenvolvidos. Nos Estados Unidos é a maior causa de morte entre mulheres, e dentro de alguns anos poderá ser a causa predominante na população como um todo, já que a mortalidade por doenças cardíacas tem diminuído consideravelmente nos últimos 40 anos (Henderson et al., 1991).

O tumor ou neoplasma, segundo a definição de Sir Rupert Willis (1952), é uma massa anormal de tecido cujo crescimento excede e não é coordenado pelo tecido normal, e cuja expansão persiste com igual intensidade mesmo após a interrupção do estímulo que a induziu (Cotran et al., 1994). O crescimento tumoral resulta do desbalanço entre o número de células se multiplicando e morrendo, geralmente pelo aumento da longevidade celular (Augustin-voss and Pauli, 1994).

O câncer surge como consequência do acúmulo de mutações num processo de múltiplas etapas denominado carcinogênese. Com poucas exceções, o câncer deriva de uma única célula que acumula alterações genéticas e epigenéticas (Ponder, 2001), que levam a mudanças qualitativas e quantitativas no padrão de expressão gênica, resultando em vantagens seletivas para o crescimento celular. Três principais grupos de genes sofrem alterações ao longo da progressão tumoral: oncogenes, que estimulam o crescimento celular, genes supressores de tumor, que inibem o crescimento, e genes que controlam a apoptose (Cotran et al., 1994). Os oncogenes são gerados ou “ativados” por mutações de efeito dominante, ou ganho de função (Alberts et al., 1994). Essa ativação pode se dar por mutações pontuais, por amplificações gênicas, por rearranjos cromossômicos ou pela perda da regulação de um fator inibitório. Seu alelo normal é chamado proto-oncogene, e geralmente corresponde a fator ou receptor de fator de crescimento, proteína quinase, proteína ligante de GTP ou proteínas nucleares (fatores de transcrição, por exemplo) (Augustin-voss and Pauli, 1994). Genes supressores de tumor codificam fatores inibitórios, e em tumores são inativados por mutações geralmente de efeito recessivo (Alberts et al., 1994), que levam a uma perda de função. São genes que geralmente regulam o crescimento celular ou vias de reparo de DNA e, por consequência, a perda de suas funções determina vantagens seletivas ou permite maior acúmulo de mutações ou aberrações cromossômicas. A sobrevivência de células somáticas necessita de fatores de

crescimento e sobrevivência que suprimam a apoptose, cujas principais vias são as de caspases e a mitocondrial (Evan and Vousden, 2001). Células tumorais apresentam mutações que desregulam a expressão de fatores de sobrevivência, de fatores anti-apoptóticos e fatores pró-apoptóticos (Evan and Vousden, 2001). Além desses três grupos de genes, é comum encontrar em tecidos com câncer mutações em genes que controlam a estabilidade do genoma. O resultado disso é o aumento das taxas de mutação gênica, dos erros na replicação do DNA e dos erros na segregação cromossômica (Loeb, 2001).

A caracterização dos tumores baseia-se em seu comportamento clínico (benigno ou maligno), no aspecto microscópico (histomorfológico) e em sua origem (histogenético) (Brasileiro Fo et al., 1993). Tumores benignos e malignos diferem quanto ao grau de diferenciação, taxa de crescimento, invasão local e presença de metástase (Cotran et al., 1994). De modo geral, tumores benignos são bem diferenciados, enquanto que os malignos podem ser tanto diferenciados quanto indiferenciados (anaplásicos). Tecidos indiferenciados apresentam grandes variações no tamanho e forma das células e núcleos, e núcleos hipercromáticos devido à abundância de DNA (Cotran et al., 1994). A taxa de crescimento dos tumores geralmente correlaciona-se inversamente com seu nível de diferenciação: tumores malignos crescem mais rapidamente do que os benignos (Cotran et al., 1994). A maior parte dos tumores benignos desenvolvem uma cápsula fibrosa e permanecem em seu local de origem, não apresentando capacidade de infiltrar, invadir ou metastatizar. Os tumores malignos nem sempre formam cápsula. Eles infiltram, invadem e destroem os tecidos vizinhos, além de penetrar em vasos sanguíneos e linfáticos e cavidades corpóreas, através dos quais se espalham pelo organismo formando as metástases (Cotran et al., 1994). Além disso, tumores malignos apresentam degenerações, o que não é comum nos benignos.

Tumores malignos são classificados como sarcomas ou carcinomas, caso o tecido afetado seja de origem mesodérmica ou epitelial, respectivamente. A grande maioria dos cânceres têm origem epitelial e apresenta aneuploidias, amplificações e deleções gênicas e grande instabilidade genética (Ramaswamy and Golub, 2002).

1.2. O câncer de tireóide

O câncer de tireóide é o tumor endócrino mais comum (Karayan et al., 2002) e corresponde a 1,1% do total de cânceres, segundo estatísticas mundiais (OncoLink). Ele pode ser classificado histologicamente como carcinoma papilífero, folicular,

indiferenciado (ou anaplásico) ou medular. Os carcinomas papilíferos correspondem a 50-80% dos casos de câncer da glândula, com incidência entre mulheres: homens de 3:1 e predominância em indivíduos de 30-40 anos. Os foliculares compreendem 10-40% dos cânceres tireoideanos, apresentam-se numa proporção entre mulheres: homens de 3:1 e atingem indivíduos na faixa dos 50 anos. Os indiferenciados (anaplásicos) compreendem 15% dos casos, com proporção mulheres: homens 1,5:1, em idade média superior a 50 anos. Os medulares correspondem a 3-12% dos casos, com igual incidência em mulheres e homens, podendo ser esporádicos (neste caso atingem principalmente indivíduos com mais de 50 anos) ou familiares (de efeito autossômico dominante) (Harach and Williams, 1994).

Os adenomas foliculares são tumores tireoideanos benignos, mais frequentes do que os carcinomas e que atingem indivíduos de menor faixa etária. De acordo com sua aparência microscópica, são classificados em micro ou macrofoliculares (Lemoine et al., 1989).

Observações clínicas indicativas de maior grau de malignidade como crescimento rápido, extensão extra-glandular, paralisia de cordas vocais, invasão local ou metástases distantes são pouco frequentes em pacientes portadores de carcinomas diferenciados de tireóide. Diagnósticos com base em dados sobre sexo, idade, firmeza do nódulo, normalidade do tecido extranodular e resposta do nódulo a hormônio tireoideano exógeno têm baixa sensibilidade e são pouco específicos, o que dificulta o diagnóstico preciso dos tumores tireoideanos (Hamburger and Kaplan, 1996).

1.3. Expressão diferencial em câncer de tireóide

O câncer de tireóide tem sido relativamente pouco estudado por biologia molecular quando comparado a outros tipos de câncer. De fato, até junho de 2003, em comparação com 13.033 e 6.371 trabalhos em genética de câncer de mama e pulmão, respectivamente, apenas 1.743 trabalhos foram publicados em câncer de tireóide (fonte= Pubmed; palavras-chave= câncer e gene). Além disso, esses trabalhos resultam da aplicação de um número restrito de técnicas. De fato, a maioria utilizou imunohistoquímica e RT-PCR, parte utilizou Southern e Northern blot e alguns realizaram análises de perda de heterozigose (LOH) ou metilação. No entanto, dificilmente encontramos, no mesmo trabalho, resultados obtidos com três ou mais métodos. Nos parágrafos que seguem, serão descritos alguns dos genes estudados em tumores de tireóide. A seleção da inclusão dos mesmos foi

o número de trabalhos que os relatam, a potencialidade de sua aplicação em diagnóstico ou a relação estreita entre sua alteração e algum tipo de tumor tireoideano.

1.3.1. Fatores de transcrição, fatores de crescimento, e seus receptores

Como mencionado no item 1.1, proto-oncogenes são genes que, quando mutados, apresentam um ganho de função qualitativo ou quantitativo, trazendo vantagens seletivas às células portadoras. Os genes mutados são denominados oncogenes. De modo geral, são genes envolvidos direta ou indiretamente no controle do crescimento celular. De fato, a expressão aumentada de fatores de crescimento está freqüentemente associada a tumorigênese. Em tumores malignos, a produção endógena de fatores de crescimento pelas células tumorais pode levar à progressão da malignidade de maneira autócrina (Kato et al., 2000).

As proteínas denominadas RAS participam da via de sinalização de fatores de crescimento que usam receptores tirosina quinase. Na maioria dos tumores humanos, a ativação de *Ras* ocorre por mutações que aumentam a atividade da proteína (Learoyd et al., 2000); (Fagin, 1994). Mutações nos genes *Ras* foram identificadas tanto em adenomas quanto em carcinomas tireoideanos, com prevalências de até 50%, sugerindo que são eventos relativamente comuns e precoces na progressão tumoral neste órgão (Learoyd et al., 2000); (Fagin, 1994). Alguns estudos descrevem maior prevalência de mutações em carcinomas tireoideanos do que em adenomas (Lemoine et al., 1989), sugerindo acúmulo destas mutações ao longo da progressão de tumor benigno para maligno.

O gene *c-Met* codifica para o receptor do fator de crescimento de hepatócitos (HGF), um potente mitogênico para células epiteliais, também promotor de motilidade e adesão (Di Renzo et al., 1992). A proteína MET está qualitativamente aumentada em vários tumores epiteliais (Ivan et al., 1997). Em tireóide ela está qualitativamente aumentada em 74% dos carcinomas papilíferos e 22% dos carcinomas foliculares, especialmente nos tumores de comportamento mais agressivo. Nenhuma das amostras de tecido normal, adenoma, carcinoma medular e anaplásico apresenta positividade (Di Renzo et al., 1992). A proteína MET está também aumentada em células transformadas que expressam os genes *Ret* ou *Ras* mutados, sugerindo uma relação entre os níveis de expressão dos três genes (Ivan et al., 1997).

O complexo AP-1, ativador de diversos genes envolvidos na proliferação celular, tumorigênese e metástase, é formado por três membros da família Jun (*c-Jun*, *JunB* e

JunD) e quatro da família Fos (*c-Fos*, *FosB*, *Fra-1* e *Fra-2*), que formam homo e heterodímeros (Curran and Franza, Jr., 1988). Linhagens de células de tireóide transformadas com diversos oncogenes apresentam aumento na atividade AP-1 e nos mRNAs e proteínas FRA-1 e JUNB. A atividade AP-1 e os níveis de FRA-1, JUND e C-JUN estão aumentados em linhagens celulares humanas de carcinomas tireoideanos em relação a células tireoideanas normais (Battista et al., 1998). O mRNA e a proteína de *Fra-1* foram analisados por RT-PCR e imunohistoquímica, respectivamente, e são detectados em 36% dos bócios, 75-88% dos adenomas e 100% dos carcinomas (foliculares, papilíferos e anaplásicos), mas não em tecido normal (Chiappetta et al., 2000).

Os genes *c-Myc* e *c-Fos*, que participam do controle da proliferação celular, estão alterados em alguns tipos de câncer e também estão mais expressos nos tumores tireoideanos (Terrier et al., 1988). Northern blots indicam aumento dos níveis de RNA de *c-Fos* em 61% e 91% dos carcinomas e adenomas, respectivamente (Terrier et al., 1988), sugerindo que essa ativação se dá no início do processo de tumorigênese. A ativação de *c-Myc* deve ocorrer ao longo da transição de adenoma para carcinoma, já que o aumento do RNA é observado predominantemente em carcinomas (57% dos casos).

O gene *Trk* codifica o receptor do fator de crescimento de neurônios (NGF) e sua expressão é restrita aos gânglios nervosos periféricos (Said et al., 1994). Os oncogenes *Trk* são criados pelo rearranjo da região 3' do gene com um de três genes diferentes, tornando o domínio tirosina quinase de *Trk* constitutivamente ativo. Um deles é o gene da tropomiosina (*Tpr*), cuja extremidade 5' está envolvida na formação do oncogene *Trk-T1*. O gene *Trk* está ativado em aproximadamente 20% dos carcinomas papilíferos (Said et al., 1994); (Farid et al., 1995).

O proto-oncogene *Ret* codifica para um trecho do receptor de GDNF (fator neurotrófico derivado de glia) (Blume-Jensen and Hunter, 2001). Ele apresenta atividade tirosina quinase e desempenha um papel importante na diferenciação neuronal e no desenvolvimento renal e intestinal (Blume-Jensen and Hunter, 2001). Na tireóide normal o gene é expresso apenas nas células C (parafoliculares), mas alterações em seu funcionamento são observadas em carcinomas papilíferos (Karayan et al., 2002); (Learoyd et al., 2000). Em carcinomas papilíferos o gene *Ret* é freqüentemente rearranjado, havendo fusão de seu domínio tirosina quinase com a sequência N-terminal de outros genes. Os rearranjos mais comuns são *Ret/Ptc1*, que envolve o gene *H4*, *Ret/Ptc2*, que envolve o gene da subunidade R1 α da proteína quinase A e *Ret/Ptc3*, que envolve o gene *ELE1*. Os rearranjos *Ret/Ptc4* (também gene *Ele1*), *Ret/Ptc5* (gene *Rfg5*), *Ret/Ptc6* (gene *Htif1*),

Ret/Ptc7 (gene relacionado a *Htif1*) e *Ret/Ptc8* (gene quinetina) são identificados em carcinomas papilíferos induzidos por radiação (Karayan et al., 2002). Os rearranjos *Ret/Ptc* são considerados eventos precoces na formação dos carcinomas papilíferos, e sua prevalência nesse tipo de tumor depende do país onde o estudo é realizado e da exposição a radiações, variando de 3 a 85% (Learoyd et al., 2000). Um trabalho recente realizado no Canadá com 73 amostras de PAAF conclui que a análise de rearranjos *Ret/Ptc* por RT-PCR pode auxiliar no diagnóstico de carcinomas papilíferos, sobretudo para as punções de resultado insuficiente e indeterminado (Cheung et al., 2001b). Rearranjos foram observados em 51% das amostras de carcinoma papilífero, e em nenhuma de carcinoma anaplásico, hiperplasia, adenoma e tireoidite. O método de PAAF identificou 36% dos carcinomas papilíferos, RT-PCR 51% e a combinação de ambos os métodos 70% (Cheung et al., 2001b). Em outro estudo, a proteína RET foi detectada por imunohistoquímica em 70% dos carcinomas papilíferos (78% dos clássicos e 63% dos de variante folicular), 50% dos carcinomas de células de Hurthle e carcinomas insulares, mas em nenhum adenoma, tecido hiperplásico, carcinoma folicular ou anaplásico (Cheung et al., 2001a).

O gene *Ret* também está envolvido na patogênese dos carcinomas medulares. Mutações pontuais no gene são freqüentemente observadas tanto nos casos familiares quanto nos esporádicos (Learoyd et al., 2000). Os casos familiares englobam as síndromes neuroendócrinas múltiplas MEN-IIA e MEN-IIB, nos quais a prevalência de mutações de *Ret* é de 98 e 95%, respectivamente, e o FMTC (carcinoma medular familiar), no qual a prevalência de mutações de *Ret* é de aproximadamente 88%. Os carcinomas medulares esporádicos não são tão estudados quanto os familiares, mas mutações em um códon específico, o 918, são encontradas em 28-86% dos indivíduos afetados, sendo a prevalência variável de acordo com o local e a amostragem do estudo, muitas vezes reduzida (Learoyd et al., 2000).

As proteínas da família ETS são fatores de transcrição envolvidos em transdução de sinal, progressão do ciclo celular e diferenciação, bem como em processos invasivos de carcinomas. Ensaio de imunohistoquímica indicam aumento da proteína ETS-1 em 98% dos carcinomas papilíferos e 87% dos foliculares, 50% dos anaplásicos e 60% dos medulares, comparados com 40% dos adenomas e 17% dos bóciós (Nakayama et al., 1999). A proteína e o mRNA de vários membros da família ETS também são mais abundantes em linhagens celulares e tecidos de carcinoma em comparação com amostras normais (de Nigris et al., 2001). O aumento da atividade de ETS-1 como fator de transcrição de alguns genes-alvo é observado nas linhagens de carcinoma e o efeito do

bloqueio da atividade ETS nessas linhagens suprime a proliferação e induz morte celular (de Nigris et al., 2001).

A família de receptores relacionados a EGFR (receptor de fator de crescimento epidérmico) inclui os genes *Egfr*, *c-ErbB2*, *c-ErbB3* e *c-ErbB4*. Carcinomas papilíferos apresentam expressão aumentada de todos esses receptores (Haugen et al., 1996). Dados de imunohistoquímica mostram marcação fraca de ERBB4 na maioria dos tecidos e média ou forte apenas em carcinomas, especialmente os papilíferos e anaplásicos. A marcação de ERBB3 é negativa em todos os tecidos normais e bócios e em 78% dos adenomas, e positiva na maior parte dos carcinomas, em especial os papilíferos e anaplásicos (91 e 100% das amostras positivas, respectivamente) (Haugen et al., 1996).

MK (midkine) é um fator de crescimento ligante de heparina cujo mRNA está aumentado em diversos carcinomas. Dos carcinomas papilíferos analisados por imunohistoquímica, 85% expressam a proteína MK, enquanto que os tecidos normais adjacentes expressam muito pouco ou nada (Kato et al., 2000).

1.3.2. Moléculas com função de adesão

A capacidade metastática está diretamente ligada a alterações na expressão de moléculas de adesão e de enzimas proteolíticas envolvidas na degradação tecidual. Neste item serão abordadas moléculas relacionadas com adesão célula-célula ou célula-matriz, e no próximo item enzimas proteolíticas.

Um dos marcadores mais estudados em tumores de tireóide e com grande potencial de aplicação é a galectina-3. Galectina-3 é uma lectina ligante de β -galactosídeos envolvida em diversos processos fisiológicos e patológicos como processamento de pré-mRNA (RNA splicing), adesão célula-célula, adesão célula-matriz, crescimento celular, transformação neoplásica, metástase e resposta imune. Sua expressão está aumentada em diversas linhagens celulares tumorais humanas e murinas bem como em tumores de estômago, cólon, sistema nervoso central e tireóide (Inohara et al., 1999). Diversos estudos mostram o aumento do mRNA e da proteína galectina-3 em carcinomas de tireóide em comparação com tumores benignos e tecido normal, tanto em PAAF quanto em amostras pós-cirúrgicas. De modo geral, a positividade em imunohistoquímica para carcinomas papilíferos, foliculares e anaplásicos é de 90-100%, para medulares 30-40%, para adenomas 0-10% e para tecidos normais e bócio 0% (Inohara et al., 1999); (Bartolazzi et al., 2001). Um trabalho com grande amostragem relatou sensibilidade,

especificidade e precisão de diagnóstico de 100, 98 e 99% respectivamente, para a diferenciação entre lesões tireoideanas benignas e malignas em PAAF (Bartolazzi et al., 2001), mostrando a eficácia desse marcador para distinção entre adenoma e carcinoma folicular.

Outra proteína analisada em alguns trabalhos, às vezes em associação com galectina-3 é CD44v6, uma das isoformas da família CD44. As CD44 são glicoproteínas que participam de processos fisiológicos e patológicos regulando adesão célula-célula, célula-matriz, ativação de linfócitos, migração, crescimento e progressão tumoral (Bartolazzi et al., 2001). Segundo dados de imunohistoquímica, a expressão da isoforma CD44v6 está aumentada em tumores de tireóide, especialmente nos carcinomas (Bartolazzi et al., 2001), mas a alta positividade para adenomas desaconselha o uso dessa proteína para diagnóstico.

Um processo de subtração foi utilizado para identificar RNAs presentes em uma linhagem celular de carcinoma papilífero mas não nas células normais. Um dos RNAs isolados foi o de HIP, uma proteína que modula interações célula-célula e célula-matriz, correlacionada com potencial metastático em câncer de fígado (de Nigris et al., 1998). Os níveis de mRNA de HIP estão aumentados em linhagens celulares de carcinoma de tireóide e em 80-100% dos tecidos com carcinoma, em comparação com 0% e 6% dos tecidos normais e com adenomas, respectivamente (de Nigris et al., 1998). O bloqueio da síntese da proteína HIP em células de carcinoma papilífero causou atraso do ciclo celular, mostrando o papel de HIP na proliferação celular (de Nigris et al., 1998).

As mucinas humanas são proteínas produzidas a partir de uma família de genes denominados *Muc* 1-6. Alterações em mucinas contribuem para mudanças no crescimento de células tumorais, reconhecimento imunológico e adesão celular, o que interfere no potencial invasivo e metastático do tumor. O gene *Muc-1* está superexpresso em vários tumores epiteliais. Reações de RT-PCR sob condições específicas geram 2 bandas de *Muc-1* em 87% dos carcinomas papilíferos mas não em amostras de bócio ou adenoma, indicando a ocorrência de um processamento alternativo do transcrito nesse tipo de carcinoma (Weiss et al., 1996).

E-caderina é uma molécula de adesão dependente de cálcio que medeia interações célula-célula. A expressão de E-caderina é normal em tumores diferenciados mas está reduzida nos indiferenciados e agressivos, o que sugere uma relação inversa entre E-caderina e potencial invasivo (Brabant et al., 1993); (Serini et al., 1996). Os níveis de mRNA e proteína estão reduzidos em carcinomas papilíferos, foliculares e anaplásicos em relação a tecidos normais, como visto em Northern blot e imunofluorescência,

respectivamente (Brabant et al., 1993). Em um outro trabalho não foi observada redução nos níveis da proteína E-caderina em carcinomas, mas sim diminuição em seu grau de fosforilação. Sua distribuição tornou-se pericelular ao invés de lateral (como em tecidos normais e adenomas) e desconectada dos filamentos de actina (Serini et al., 1996). Além das alterações na distribuição e funcionalidade de E-caderina, carcinomas apresentaram mudanças na expressão de integrinas, as principais responsáveis pela adesão da célula à membrana basal. Enquanto tecidos normais e adenomas expressaram apenas integrinas $\alpha 3 \beta 1$ nas regiões basais, carcinomas expressam $\alpha 6 \beta 4$, conforme observado em Northern blot, Western blot e imunohistoquímica (Serini et al., 1996). A diferença entre os dois artigos quanto à abundância da proteína E-caderina nos tumores poderia ser explicada pelas amostras analisadas nos dois trabalhos, pela reduzida amostragem do segundo trabalho e/ou pelos diferentes métodos utilizados.

1.3.3. Enzimas proteolíticas e moléculas que regulam sua função

Como descrito no item anterior, enzimas proteolíticas e seus reguladores têm importante papel na capacidade metastática dos tumores. As maiores famílias destas proteínas são: metaloproteinases de matriz, inibidores de metaloproteinases, serina, cisteína e asparagina proteinases e heparanases.

A proteína CD26, inicialmente identificada como um antígeno de linfócitos T ativados, corresponde à dipeptidil peptidase (DPP) IV, uma serina protease (Tanaka et al., 1995b). A quantificação da proteína DPP IV por imunohistoquímica e a medida de sua atividade mostraram alta positividade para carcinomas papilíferos (99-100% das amostras) e carcinomas foliculares (100% dos casos), certa positividade para adenomas foliculares (12-17%) e rara positividade para bóciós (2-9%) e doença de Graves (0-4%). Nenhum dos carcinomas anaplásicos e medulares analisados apresentou positividade por nenhuma das técnicas (Tanaka et al., 1995b). Northern blots mostraram aumento do RNA de *Dpp IV* em carcinomas foliculares e papilíferos em relação a tecidos benignos, e esse aumento foi mais significativo do que o observado para os transcritos dos genes *c-Met*, *c-ErbB2* e *Egfr* (Tanaka et al., 1995b).

Catepsinas B (CB) e L (CL) são cisteína endopeptidases envolvidas na degradação intra-lisossômica não-específica de proteínas celulares e endocitadas. As duas enzimas são capazes de degradar componentes da membrana basal, estão relacionadas com invasão tumoral e metástase e apresentam níveis aumentados em diversos carcinomas (Shuja and

Murnane, 1996). Na tireóide normal estas proteínas estão envolvidas no processamento da tireoglobulina. Ensaio de imunohistoquímica indicam aumento da marcação de CB e CL em uma amostra de carcinoma papilífero em comparação com o tecido normal adjacente. A atividade destas enzimas estava aumentada aproximadamente 15 vezes em carcinomas papilíferos e em uma amostra de carcinoma folicular em relação aos tecidos normais, mas estava praticamente inalterada em bócios, adenomas, carcinomas medulares e tecidos com tireoidite de Hashimoto (Shuja and Murnane, 1996).

A α 1-antitripsina controla a destruição tecidual e a invasão tumoral, regulando a atividade de proteases secretadas pelas células tumorais (Poblete et al., 1996). Na tireóide, expressão de α 1-antitripsina foi observada, por imunohistoquímica, em 9 dos 10 carcinomas papilíferos mas não no tecido normal adjacente. A proteína estava aparentemente mais abundante nos extratos provenientes de carcinomas papilíferos do que nos de tecidos normais, conforme observado por Western blot (Poblete et al., 1996).

1.3.4. Moléculas que regulam proliferação e morte celular

O balanço entre proliferação e morte celular é essencial para o desenvolvimento e manutenção de tecidos normais, e está freqüentemente alterado em tumores. A morte celular por apoptose é regulada por genes específicos, como o supressor de tumor *P53*. Um dos principais efetores induzidos por *P53* é o gene *Bax*, cujo produto aumenta a permeabilidade da mitocôndria, ativando a via que aciona a cascata das proteases caspases e leva à morte celular (Hermann et al., 2001). A expressão de *P53*, *BAX* e *P21* (outro gene ativado por *p53*) foi analisada por imunohistoquímica em amostras de bócio, adenoma e carcinoma (em sua maior parte foliculares). As proteínas *BAX*, *P53* citoplasmática (mas não nuclear) e *P21* estão aumentadas em carcinomas, e têm níveis semelhantes nos outros dois tecidos. Neste estudo observa-se acúmulo de *P53* em 49% dos carcinomas, considerando diversos trabalhos essa proporção varia de 23-61% (Hermann et al., 2001). Mutações no gene *P53* são raramente observadas em tumores tireoideanos diferenciados, sendo encontradas em até 70% dos carcinomas anaplásicos (Farid et al., 1995). A reintrodução de *P53* selvagem em linhagens celulares anaplásicas com *P53* mutado reduz sua proliferação e leva à reexpressão de marcadores de diferenciação tireoideana como tireoperoxidase, tireoglobulina e receptor de TSH, enfatizando a importância da inativação

do gene na progressão de carcinomas diferenciados para indiferenciados (Karayan et al., 2002).

A ceramidase ácida (CA) é a enzima responsável pela hidrólise de ceramida. A ceramida participa do metabolismo de esfingolipídios e da manutenção da membrana celular, bem como da regulação da proliferação e morte celulares e da indução de apoptose em células tumorais (Maeda et al., 1999). Experimentos de Real Time RT-PCR foram realizados para quantificar a expressão de ceramidase ácida (CA) em amostras de tecido tireoideano normal, bócio, adenoma e carcinomas papilífero, folicular e anaplásico. Embora haja grandes variações entre amostras de um mesmo tipo e grande sobreposição entre amostras de tipos distintos, de modo geral o gene *Ca* é mais expresso em tecidos normais, seguido dos com bócio (Maeda et al., 1999).

1.3.5. Moléculas com outras funções

Um processo de subtração foi usado para identificar RNAs mais abundantes em uma linhagem de carcinoma anaplásico do que em uma de carcinoma papilífero. Um dos genes isolados foi o da *Thymosin β -10*, uma proteína ácida pequena que inibe a polimerização e formação dos filamentos de actina e que está relacionada com crescimento e proliferação celulares (Califano et al., 1998). O mRNA de *Thymosin β -10* está aumentado tanto em tecidos quanto em linhagens celulares de carcinomas em comparação com de adenomas e normais. Nos carcinomas anaplásicos os níveis do transcrito são pelo menos duas vezes maiores do que nos papilíferos e foliculares, sugerindo que a expressão do gene está correlacionada com o grau de malignidade dos tumores (Califano et al., 1998).

As proteínas HMGI (alta mobilidade, grupo I) são proteínas nucleares envolvidas na regulação da estrutura e função da cromatina, e portanto da expressão gênica. O subtipo HMGI(Y) está altamente expresso durante a embriogênese e em tumores malignos, mas apresenta baixos níveis nos tecidos adultos normais (Chiappetta et al., 1998). A positividade de tecidos tireoideanos para a proteína (por imunohistoquímica) e para o RNA (por RT-PCR) HMGI(Y) são, respectivamente: 22 e 40% em adenomas e 96 e 100% em carcinomas (foliculares, papilíferos e anaplásicos). Das 12 amostras de PAAF analisadas por imunohistoquímica e RT-PCR, apenas as de carcinoma são positivas, ilustrando o potencial uso desse marcador para o diagnóstico de tumores tireoideanos (Chiappetta et al., 1998).

A proteína ligante de β -galactosídeo (GBP), é uma das lecitinas melhor caracterizadas (Chiariotti et al., 1992). Os níveis de mRNA de GBP foram analisados por Northern blot em linhagens de células tireoideanas de rato transfectadas com diferentes oncogenes. Os autores descrevem que a expressão de GBP foi 5 a 10 vezes maior nas células pouco tumorigênicas e aproximadamente 100 vezes maior nas altamente tumorigênicas, em relação às linhagens celulares não-transfectadas (Chiariotti et al., 1992). Um experimento semelhante mostra aumento do RNA de GBP em carcinoma papilífero em comparação com bócio e tecidos normais. Infelizmente, estes experimentos de Northern blot não podem ser considerados quantitativos pela ausência de um gene controle para normalização dos dados. Ainda assim, a diferença observada foi tão grande que pode refletir diferenças biológicas.

A imunoreatividade para o antígeno epitelial de membrana (EMA) é significativamente maior em carcinomas papilíferos com metástase à distância do que nos demais carcinomas papilíferos: positividade em 50% das células é observada em 47% dos tumores do subtipo mais agressivo em comparação com 0% e 3% dos tumores sem metástase à distância e ocultos, respectivamente. A reatividade a EMA pode ser um marcador útil para estimar o risco de metástases a distância ou letalidade para indivíduos com carcinomas papilíferos (Yamamoto et al., 1992).

A expressão dos genes de calcitonina, antígeno carcinoembrionário (*Cea*) e *Ret*, foi usada para distinção entre carcinomas medulares e carcinomas foliculares, papilíferos, adenomas, bóciolos, e tecidos normais (Takano et al., 1999a). Os transcritos dos três genes apresentaram amplificação por RT-PCR em amostras de PAAF de todos os carcinomas medulares, mas em nenhuma das amostras dos demais tecidos (Takano et al., 1999a).

Comparando o perfil dos RNAs presentes em carcinomas papilíferos, carcinomas foliculares, adenomas e tecidos normais por uma técnica semelhante ao DDRT-PCR e por RT-PCR observou-se superexpressão do gene da fibronectina oncofetal em carcinomas papilíferos (Takano et al., 1997). Analisando um número maior de amostras por Real-Time RT-PCR, utilizando tireoglobulina para normalização dos dados, observou-se que carcinomas papilíferos e anaplásicos apresentaram um aumento significativo na expressão da fibronectina, apesar da grande variação entre as amostras (Takano et al., 1999b).

O gene *Rassf1* é um supressor de tumor, cuja isoforma *Rassf1A* está inativada em diversos tumores (Schagdarsurengin et al., 2002). O estado de metilação do promotor de *Rassf1A* foi analisado em linhagens celulares de carcinoma tireoideano e em amostras de carcinomas anaplásicos, papilíferos, foliculares, medulares e tecidos normais. Setenta e

um % dos tumores e 100% das linhagens de carcinoma apresentam hipermetilação, em comparação com 25% dos tecidos normais, sugerindo que esta alteração seja um evento precoce na patogênese do câncer de tireóide (Schagdarsurengin et al., 2002).

1.3.6. Busca de marcadores em câncer de tireóide

Algumas técnicas de busca de genes diferencialmente expressos tem sido usadas para identificar marcadores para o câncer de tireóide, dentre elas DDRT-PCR, Microarray e SAGE. Um trabalho utilizando SAGE comparou o perfil de expressão de mais de 600 genes em carcinoma folicular, papilífero, anaplásico, adenoma e tecido normal (Takano et al., 2000). Os genes foram classificados como diferencialmente expressos quando seus transcritos foram pelo menos 10 vezes mais abundantes em um dos tecidos. Menos de 12 genes diferenciais foram identificados em cada uma das comparações: tecido normal e carcinomas diferenciados, adenoma e carcinomas diferenciados, carcinoma papilífero e folicular. Por outro lado, 45 e 34 genes apresentaram expressão diferencial entre carcinoma anaplásico e papilífero, e anaplásico e folicular, respectivamente, entre eles *Gapdh*, *Tireoglobulina* e α -*Tubulina* (Takano et al., 2000).

Um trabalho de nosso grupo utilizou DDRT-PCR para isolar genes diferencialmente expressos nas comparações entre tecido normal, bócio e carcinoma papilífero e entre tecido normal, adenoma e carcinoma folicular (Stolf et al., 2003). Duzentos e treze genes foram isolados e fixados em arrays para hibridizações comparativas com material de pacientes distintos. Três destes genes foram selecionados por sua expressão diferencial nas hibridizações dos arrays, e sua expressão foi avaliada em aproximadamente 10 amostras de cada tipo de tecido por RT-PCR seguido de Southern blot. A expressão de dois desses genes foi capaz de separar amostras de tecido normal e bócio (não-malignos) de carcinoma papilífero (maligno) com precisões de 80 e 89% para as classificações não-maligno e maligno, respectivamente. Dois destes genes foram capazes de discriminar amostras de tecido normal de adenoma e carcinoma folicular (neoplásicos) com precisões de 78 e 88% para classificações de normal e neoplásico, respectivamente (Stolf et al., 2003).

Muito poucos trabalhos utilizando arrays foram publicados em câncer de tireóide. Um deles utilizou um microarray comercial com oligonucleotídeos correspondentes a 12.000 genes humanos para comparar o perfil de expressão de 8 pares de carcinomas papilíferos e tecidos normais (Huang et al., 2001). 143 genes tiveram expressão pelo menos duas vezes

aumentada em 5 ou mais carcinomas papilíferos, enquanto que 83 genes foram pelo menos duas vezes menos expressos em pelo menos 5 dos tumores. Dentre os genes diferencialmente expressos encontram-se genes relacionados com funções tireoideanas especializadas (tireoperoxidase e transportador de sódio-iodo, por exemplo, menos expressos nos carcinomas), genes previamente associados com carcinoma papilífero (*Fibronectina-1* e *Met*, mais expressos neste tumor), genes de moléculas de adesão celular (*Laminina- β 3* e *Mucina-1*, mais expressos no carcinoma) e outros, alguns dos quais relacionados com tumores de outros órgãos. Alguns destes genes tiveram sua expressão analisada por RT-PCR semi-quantitativo em amostras de RNA provenientes de carcinomas papilíferos e tecidos normais, e os dados obtidos confirmaram os do microarray. A presença das proteínas dos genes *Cited1* e *Sfibp*, identificados pelo microarray como superexpressos em carcinomas papilíferos, foi analisada em um array de tecidos contendo amostras de carcinomas papilíferos, foliculares e anaplásicos e tecidos normais. CITED1 e SFTBP apresentaram marcação positiva em 93 e 78%, respectivamente, das amostras de carcinoma papilífero e em nenhum outro tipo de tecido (Huang et al., 2001).

Um outro estudo avaliou a abundância de 6 proteínas em tecidos normais, adenomas e carcinomas de Hurthle por array de tecidos (Hoos et al., 2002). As proteínas analisadas estão relacionadas com ciclo celular, proliferação e apoptose, e dois dos genes correspondentes já foram descritos como diferencialmente expressos em adenomas e carcinomas de Hurthle. Embora poucas diferenças tenham sido observadas entre os adenomas e os carcinomas, duas proteínas mostraram aumento significativo nestes tumores e uma apresentou redução gradual na evolução de tecido normal para adenoma, carcinoma minimamente invasivo e altamente invasivo (Hoos et al., 2002).

Várias dificuldades são encontradas na busca de marcadores moleculares para o câncer de tireóide. Uma delas é a grande heterogeneidade entre os pacientes com um mesmo tipo de tumor. Diferenças genéticas individuais fazem com que muitos dos genes diferencialmente expressos em uma amostra não o sejam em outras amostras de mesmo tipo. A necessidade de se aumentar a amostragem para a obtenção de resultados conclusivos aumenta o custo do estudo e dificulta a análise de tipos tumorais mais raros. Outra dificuldade é que a maior parte dos estudos para a identificação de marcadores requerem uma quantidade considerável de RNA de cada amostra, seja pelas necessidades dos métodos utilizados, seja pela realização de análises repetidas da mesma amostra.

Muitas vezes a amostra cirúrgica recebida não tem o tamanho necessário para a obtenção dessa massa de RNA, e não pode ser incluída no estudo.

O aprimoramento das técnicas de Biologia Molecular, de modo a reduzir a massa de RNA necessária e o custo dos experimentos têm contribuído para os estudos na área. A colaboração entre médicos e pesquisadores têm aumentado a qualidade e quantidade das amostras. A associação entre técnicas eficazes para a busca de marcadores e dados clínicos e patológicos permitirá um avanço no diagnóstico e melhorias no manejo dos tumores da glândula.

1.4. Diagnóstico

Ao contrário dos tumores malignos, alterações tireoideanas são bastantes comuns. A incidência de nódulos clinicamente aparentes na população geral é de 4-5%, e este valor aumenta para 37% com as análises de autópsia (Shaha, 2000). O manejo desses nódulos é um assunto extremamente controverso, principalmente no que se refere ao diagnóstico e extensão da tireoidectomia (Shaha, 2000).

A maioria destes nódulos é benigna, menos de 5% são malignos (Weiss et al., 1996). A tireoidectomia é o método de escolha para o tratamento dos carcinomas e as complicações cirúrgicas variam de 2 a 23%, dependendo da extensão da ressecção e da experiência do cirurgião (Weiss et al., 1996).

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).

Diversos métodos são atualmente empregados nas análises pré-operatórias de nódulos tireoideanos, incluindo ultra-sonografia, cintilografia e biópsia de aspiração por agulha fina (PAAF) (Inohara et al., 1999). O uso da PAAF, um dos mais importantes deles, trouxe melhoria no manejo clínico dos nódulos, mas a caracterização pré-operatória de lesões foliculares continua pouco precisa (Bartolazzi et al., 2001). O método apresenta excelente precisão no diagnóstico de carcinomas anaplásicos e medulares, mas gera falsos-negativos em carcinomas papilíferos e não é capaz de distinguir entre adenoma e carcinoma folicular, já que o diagnóstico do último baseia-se na invasão capsular (Inohara et al., 1999).

Apenas aproximadamente 30% dos pacientes submetidos à cirurgia apresenta tumor maligno, como atestado pelas análises pós-cirúrgicas de fragmentos tumorais incluídos em parafina. Parte desses indivíduos são submetidos a tireoidectomia pela presença de nódulos benignos que causam compressão, ocasionando problemas respiratórios e de deglutição. Parte, porém, apresenta nódulos classificados na análise pré-operatória como adenomas ou carcinomas foliculares, e diante da dúvida, são geralmente submetidos a tireoidectomia total. A distinção entre esses dois tipos de tumor é, de fato, um dos maiores desafios da cirurgia tireoideana.

Nesse sentido, estudos utilizando expressão gênica podem contribuir para a identificação de genes diferencialmente expressos entre adenomas e carcinomas foliculares, permitindo o diagnóstico correto destes dois tumores. O uso destas técnicas permite também comparar o perfil de expressão gênica entre tecido normal, bócio e carcinoma papilífero, o que contribuirá para a melhor compreensão da relação entre os mesmos.

2. SOBRE A ABORDAGEM METODOLÓGICA

2.1. Expressão diferencial

Visto que o câncer resulta do acúmulo de mutações que ocasionam alterações qualitativas e quantitativas no padrão de expressão gênica, a identificação de genes mutados ou de expressão alterada em câncer de tireóide pode contribuir para a compreensão da relação entre as diversas doenças tireoideanas e para o aprimoramento dos testes diagnósticos, especialmente se associada a PAAF.

Alguns anos atrás, os cientistas estudavam a patogênese de maneira linear, identificando e examinando a expressão de um gene por vez. Recentemente, a implementação de tecnologias de maior escala tornou possível o estudo global da expressão gênica de tecidos ou células (Lakhani and Ashworth, 2001). De fato, a análise das diferenças na expressão gênica entre dois ou mais tipos de tecidos/células tem sido tema de diversos estudos.

Após o sequenciamento do genoma humano, duas grandes questões surgiram: a) como a expressão diferencial da informação contida no genoma está associada com saúde e doença e b) como mutações ou pequenas variações naturais nas sequências produzem desordens genéticas e/ou aumentam o risco de se desenvolver doenças. A primeira dessas questões é o fundamento da genômica funcional, área que estuda os genes expressos por um tipo celular específico e as mudanças nesse padrão de expressão durante o desenvolvimento, doença ou alterações ambientais (Freeman et al., 2000).

Grande parte da pesquisa em câncer dos últimos 50 anos dedicou-se à análise de genes diferencialmente expressos em tecidos tumorais em relação aos normais (Zhang et al., 1997). Essa abordagem levou ao conhecimento dos oncogenes e dos genes supressores de tumor (Zhang et al., 1997).

Diversos métodos são utilizados para analisar a expressão diferencial de genes: "differential display" (DDRT-PCR) (Liang and Pardee, 1992); impressões digitais por PCR arbitrário de RNA (AP-PCR fingerprinting) (Welsh et al., 1992); hibridação subtrativa (SH) (Lee et al., 1991); análise de representação diferencial (RDA) (Lisitsyn et al., 1993), subtração eletrônica (ES); análise seriada da expressão gênica (SAGE) (Velculescu et al., 1995) e "DNA microarray".

O "AP-PCR RNA fingerprinting" consiste da transcrição reversa em baixa temperatura de RNA total utilizando iniciadores arbitrários, seguida da síntese da segunda

fita de DNA pela enzima Taq, e amplificação por PCR radioativo de alta estringência. As reações correspondentes a RNAs de populações celulares distintas são analisadas em gel de poliacrilamida e o DNA das bandas de interesse é recuperado, amplificado e usado como sonda em Northern blot (Welsh et al., 1992).

A hibridação subtrativa (SH) baseia-se na remoção de sequências comuns a duas populações celulares de modo que, ao final do processo, restem somente sequências presentes apenas nas células de interesse. As moléculas da população de interesse são denominadas “tester” enquanto que as moléculas da população usada para subtração são chamadas “driver”. Vários métodos foram descritos nessa categoria, para comparação entre genomas ou entre sequências expressas em uma e outra situação, utilizando estratégias diversas para o isolamento de moléculas “tester” não pareadas com “driver” (Lamar and Palmer, 1984); (Wieland et al., 1990); (Lee et al., 1991).

O método de RDA associa subtração e amplificação. cDNA total de duas populações celulares é digerido e amplificado separadamente. Aos fragmentos da população de interesse (denominados “tester”) são ligados novos adaptadores. Fragmentos “tester” e “driver” (os últimos provenientes da população utilizada para subtração, sem adaptadores) são desnaturados e reunidos num tubo para pareamento. O produto dessa incubação é submetido a amplificação por PCR utilizando iniciadores complementares aos adaptadores dos fragmentos “tester”. Deste modo, o produto do PCR é enriquecido em sequências presentes exclusivamente nos fragmentos “tester”, já que apenas os “testers” que não parearam com os “drivers” sofrem amplificação exponencial (Lisitsyn et al., 1993). Estes produtos podem ser clonados para sequenciamento e uso como sondas (Hubank and Schatz, 1994).

A subtração eletrônica (SE) corresponde à comparação do número de moléculas de mRNA em duas populações de células. Na prática, a comparação é feita pelo sequenciamento de cDNAs selecionados aleatoriamente a partir de bibliotecas de interesse (Wan et al., 1996).

A técnica SAGE permite uma análise quantitativa e simultânea de grande número de mRNAs, partindo do princípio de que um “tag” de 9-10 pares de bases identifica unicamente um transcrito. Essas pequenas sequências “tag”, situadas próximo à cauda poliA, são concatenadas e clonadas. O sequenciamento dos clones de duas populações de células assim obtidos fornece informações quantitativas sobre a expressão gênica (Velculescu et al., 1995). A técnica de SAGE é um sistema aberto (não limitado a genes isolados ou selecionados), bastante trabalhoso e que apresenta um viés na direção dos

genes mais expressos. De fato, em torno de 10-20% dos tags correspondem a genes únicos, e é necessário sequenciar um grande número de genes para encontrar os de baixa expressão (Zhang et al., 2001).

Um trabalho comparando os métodos de SH, ES e DDRT-PCR avaliou as vantagens de cada técnica (Wan et al., 1996). Quanto à quantidade de RNA necessária, ES e SH são desvantajosos, pois requerem 1-5µg de RNA poliA, em comparação com 5µg de RNA total (50 a 100ng de poliA) exigidos pelo DDRT-PCR. Enquanto que a ES permite identificar apenas RNAs abundantes, SH e DDRT-PCR levam ao isolamento tanto de transcritos raros quanto de abundantes. O isolamento dos RNAs por DDRT-PCR depende do par de iniciadores utilizado e não da prevalência do mRNA, como ocorre para ES e SH. O processo de SH leva ao isolamento de RNAs exclusivos de um tipo celular (expressão tudo ou nada); já a SE e o DDRT-PCR mostram diferenças a partir de 2 vezes na expressão gênica (Wan et al., 1996).

Embora o DDRT-PCR seja um método bastante rápido, a verificação e obtenção da região codificadora dos verdadeiros positivos pode ser trabalhosa. Por utilizar como um dos iniciadores da amplificação o oligonucleotídeo T₁₁VN, as sequências obtidas são geralmente próximas da extremidade 3' do RNA e muitas vezes correspondem a trechos não traduzidos (Wan et al., 1996).

O uso de pouco material pelos métodos DDRT-PCR e RDA (Hubank and Schatz, 1994) dispensa a cultura das células, o que pode ser problemático caso a expressão gênica *in situ* seja distinta daquela em cultura (Lee et al., 1991). A comparação de dados de SAGE de células de tumor colo-retal de tecido e de cultura mostrou considerável semelhança no padrão de expressão, mas alertou em relação ao uso de linhagens em cultivo como modelos para tumores presentes no organismo (Zhang et al., 1997).

Uma vantagem do RDA em comparação ao DDRT-PCR é a facilidade de análise dos resultados, já que no primeiro caso o cDNA apresenta-se enriquecido em sequências diferencialmente expressas, por conta da subtração. Além disso, o uso de adaptadores garante a amplificação da maioria dos cDNAs. No caso do DDRT-PCR, aproximadamente 1 em 12 cDNAs é amplificado por uma combinação particular de iniciadores, de modo que 20-25 iniciadores 5', associados aos 4 iniciadores ancorados 3' são necessários para a análise de toda a população de cDNAs (Hubank and Schatz, 1994).

Uma das vantagens da técnica de SAGE é a confiabilidade dos dados referentes à abundância relativa de transcritos, já que a determinação desta não é afetada por diferenças na clonagem ou na eficiência de amplificação por PCR (Zhang et al., 1997).

Uma das desvantagens é a necessidade de clonagem e sequenciamento exaustivos para obtenção dos dados. Embora o SAGE seja menos desgastante do que a subtração eletrônica em termos de sequenciamento (cada reação permite analisar aproximadamente 40 cDNAs ao invés de apenas um), esta etapa ainda é extensa (Wan et al., 1996).

2.2. DNA microarray

Embora o primeiro trabalho de microarray denomine o material marcado (flutuante) de sonda (“probe”) (Schena et al., 1995), a maior parte dos trabalhos subsequentes consideram sonda os fragmentos fixados e alvo o material marcado (Southern et al., 1999). Essa nomenclatura baseia-se fundamentalmente no fato de o material fixado ter sequência conhecida, como as sondas utilizadas em experimentos de Southern e Northern blot. Neste trabalho, mantemos a nomenclatura de sonda como o material fixado e alvo como o flutuante.

Alguns termos específicos são utilizados nos experimentos de microarray:

- Spot corresponde à área ocupada por cada clone na lâmina
- Amostra teste: amostra de interesse biológico
- Foreground ou região de sinal: região ocupada pelo spot
- Ruído: falta de contribuição de algumas moléculas fixadas para o valor do sinal
- Background: imagem de fundo da lâmina, regiões sem DNA fixado
- Artefato: sinais intensos por sujeira ou hibridização inespecífica no background

A escolha da técnica de DNA microarray

A heterogeneidade dos pacientes tem sido descrita como uma das grandes dificuldades para o estabelecimento de perfis gênicos para várias doenças. Por essa razão, a identificação de genes diferencialmente expressos que possam ser usados como marcadores específicos exige a comparação de um grande número de genes em amostras de múltiplos indivíduos.

A técnica de DNA arrays (Schena et al., 1995) permite a análise da expressão de centenas a milhares de genes em um só ensaio, e o resultado combinado de experimentos distintos permite o estabelecimento de perfis gênicos para cada condição patológica, identificando genes super ou subexpressos (Quackenbush, 2001). Além disso, o

agrupamento de genes em classes de expressão fornece pistas sobre a função biológica e a relevância de genes desconhecidos (Quackenbush, 2001).

Os principais tipos de array de DNA são macroarrays, microarrays, arrays de oligonucleotídeos de alta densidade e arrays microeletrônicos (Freeman et al., 2000). Os macroarrays ou simplesmente arrays são obtidos pela fixação robótica de fragmentos de DNA (clones de cDNA, ESTs ("expressed sequence tags"), produtos de PCR ou oligonucleotídeos) em membranas de nylon. As membranas geralmente contêm de 200 a 5000 sequências de DNA e são hibridizadas com alvos radioativos, preparados por transcrição reversa de RNA total ou RNA poliA⁺ (Freeman et al., 2000). A comparação entre o padrão de marcação das membranas hibridizadas com sondas de duas populações permite identificar genes diferencialmente expressos. Os microarrays são obtidos pela fixação robótica de clones de DNA ou cDNA, produtos de PCR ou oligonucleotídeos em grande número em lâminas de vidro. Os alvos são marcados com análogos de nucleotídeos fluorescentes e geralmente duas amostras marcadas com análogos diferentes são hibridizadas com um mesmo microarray (Schena et al., 1995); (Freeman et al., 2000).

Para os arrays de oligonucleotídeos de alta densidade, os fragmentos podem ser gerados *in situ* na superfície da lâmina por fotolitografia ou sintetizados e posteriormente fixados (Schena, 2003). Um dos maiores fabricantes de arrays gerados *in situ* é a Affymetrix, que produz lâminas contendo 40.000 a 60.000 sequências, incluindo controles negativos para todos os genes (Freeman et al., 2000). Os arrays microeletrônicos consistem de um conjunto de eletrodos cobertos com uma fina camada de agarose. A deposição de cada fragmento e a hibridização da sonda são controlados individualmente pelos campos elétricos dos eletrodos correspondentes (Freeman et al., 2000). Outra forma de microarray foi elaborada utilizando fibras ópticas nas quais foram fixados oligonucleotídeos. O sistema foi testado por hibridação com oligonucleotídeos marcados com fluoresceína, e o sinal captado pelos sensores é analisado por um software disponível comercialmente (Ferguson et al., 1996).

A identificação dos genes de interesse pode ser feita pela marcação direta do cDNA alvo ou por marcação indireta, pela ligação não-covalente de uma molécula intermediária, à qual se liga um composto fluorescente (Schena, 2003).

A utilização de cDNA microarrays na análise comparativa da expressão gênica tem crescido muito nos últimos anos. Sua capacidade de estudar milhares de genes simultaneamente tem justificado o abandono de técnicas de menor escala, especialmente em situações nas quais as células ou tecidos comparados são bastante semelhantes.

Ensaios utilizando Microarray têm levado a progressos na compreensão do câncer. Os estudos realizados até agora podem ser divididos em três grupos: os que buscam identificar perfis de tumores, a maioria visando melhoras no diagnóstico de subtipos em tumores morfológicamente homogêneos (Lakhani and Ashworth, 2001), os que identificam subtipos correlacionados com diferentes prognósticos e os que buscam alvos terapêuticos ou novos marcadores moleculares (Macoska, 2002).

Um dos primeiros estudos utilizando microarray permitiu a distinção precisa entre leucemias agudas mielóides (AML) e linfoblásticas (ALL), o que é essencial para o sucesso do tratamento a ser adotado (Golub et al., 1999). A partir da hibridização de RNAs de 38 amostras de aspirados de medula óssea com microarrays de 6800 oligonucleotídeos Affymetrix foram selecionados 50 genes capazes de distinguir os dois tipos de leucemia. Estes foram utilizados para a construção de um preditor, que mostrou alta precisão na distinção de novas amostras de AML e ALL (Golub et al., 1999).

Outro estudo utilizou arrays preparados a partir de bibliotecas de cDNA de câncer pancreático para selecionar genes de expressão alterada neste tipo de tumor (Gress et al., 1996). Estes perfizeram 4% dos clones das bibliotecas, compreendendo um total de 369 ESTs distintos sendo 32,5% deles sequências novas ou ESTs já sequenciados mas de função desconhecida. Os dados obtidos foram a primeira evidência das complexas alterações primárias e secundárias responsáveis pelo desenvolvimento do câncer pancreático (Gress et al., 1996).

A comparação de linfomas utilizando microarray permitiu distinguir facilmente entre linfomas foliculares, leucemias linfocíticas crônicas e linfomas difusos de células B (Alizadeh et al., 2000). Esse último tipo foi ainda subdividido em linfomas de células B de centro germinal e periféricas ativadas. Esses dois tumores apresentam prognósticos bastante distintos, sendo a sobrevida em 5 anos 76 e 16% para os de célula germinal e periférico, respectivamente (Alizadeh et al., 2000). Um trabalho publicado dois anos depois analisou amostras de linfomas difusos de células B de pacientes submetidos a quimioterapia, e subdividiu o grupo nas categorias curada e refratária ao tratamento, marcadas por taxas de sobrevida em cinco anos de 70 e 12%, respectivamente (Shipp et al., 2002). Os dados deste trabalho e os obtidos pelo grupo de Alizadeh apresentam diferenças que foram atribuídas às coleções de genes fixadas nos dois arrays, ao tipo de array utilizado (oligonucleotídeo e cDNA), ao grupo de pacientes e ao tipo de abordagem computacional (Shipp et al., 2002).

Um modelo de câncer de próstata transplantado em camundongo foi usado para comparar genes diferencialmente expressos entre tumores responsivos a andrógenos (de melhor prognóstico) e refratários (de estágio terminal, não curável) (Bubendorf et al., 1999). Dos 5000 genes analisados por cDNA array, 10 foram superexpressos e 14 subexpressos em pelo menos dois dos quatro tumores refratários a hormônio. Dois dos genes superexpressos nesses tumores, IGFBP2 e HSP27, tiveram seus produtos analisados em arrays de tecidos contendo tumores de próstata malignos responsivos e refratários, tumores benignos e tecidos normais (Bubendorf et al., 1999). IGFBP-2 estava aumentada em 100% dos tumores refratários e 36% dos responsivos, enquanto que HSP27 apresentou aumento em 31% dos tumores refratários e 5% dos responsivos. Nenhuma das proteínas estava aumentada nos tumores benignos (Bubendorf et al., 1999).

O perfil de transcrição de amostras de adenomas e adenocarcinomas de cólon e dos tecidos normais pareados foi analisado em um array de 6000 oligonucleotídeos (Notterman et al., 2001). Dezenove transcritos apresentaram expressão aumentada e 88 diminuída nas amostras tumorais em comparação com as normais. Onze das sequências diferencialmente expressas foram validadas por RT-PCR e apresentaram resultados coerentes com os de microarray. A distribuição das amostras segundo a expressão gênica mostrou três grupos principais, correspondendo a tecidos normais, adenomas e adenocarcinomas, sendo os tecidos neoplásicos mais semelhantes entre si do que com os normais (Notterman et al., 2001).

Carcinomas renais e tecidos normais também foram comparados segundo a expressão de 20.000 genes fixados em um cDNA microarray (Takahashi et al., 2001). Mais de 100 genes apresentaram expressão 3 vezes aumentada ou diminuída em pelo menos 75% dos tumores. Com base na expressão de uma fração dos genes do array foi possível separar os tumores em dois grupos, caracterizados por sobrevidas em cinco anos de 0 e 100% (Takahashi et al., 2001).

Um trabalho de nosso grupo identificou genes diferencialmente expressos entre tecido gástrico normal e tumoral utilizando um array de 4500 clones ORESTES (Meireles et al., 2003). Avaliando as razões de expressão entre os dois tipos de tecido e a variabilidade dessa expressão entre as diversas amostras de um mesmo tecido foram selecionados 80 genes capazes de separar perfeitamente tecidos normais e malignos. Diversos trios de genes dentre os 80 foram também capazes de realizar essa discriminação (Meireles et al., 2003).

Microarrays foram também utilizados nos estudos de tumores de mama. Um deles diferenciou entre tumores clinicamente classificados como ER positivos, associados a células luminais, e tumores ER-negativos, subdivididos molecularmente em três grupos (Perou et al., 2000). Um trabalho posterior identificou ainda três subgrupos nos tumores de tipo luminal, sendo um de melhor prognóstico que os demais (Sorlie et al., 2001). Outro trabalho utilizou DDRT-PCR para identificar genes diferencialmente expressos em células normais e tumorais de mama (Martin et al., 2001). Os genes isolados foram fixados em cDNA arrays, os quais foram hibridizados com RNA extraído do sangue de pacientes normais e com tumor de mama. Doze genes apresentaram expressão aumentada no sangue de 77% dos pacientes com tumor (Martin et al., 2001). Um trabalho recente baseou-se no perfil de expressão de 70 genes identificados por microarray para prever a ocorrência de metástases em 295 pacientes nos cinco anos seguintes ao tratamento (van de Vijver et al., 2002). Os resultados mostraram a existência de dois padrões de expressão dos 70 genes analisados, fortemente correlacionados a bom e mau prognóstico em termos de sobrevida e presença de metástase. Esse método molecular mostrou-se mais eficiente na predição da ocorrência de metástases do que os métodos classicamente utilizados na Europa e Estados Unidos (van de Vijver et al., 2002).

As amostras utilizadas como alvos para microarray

Várias fontes de variabilidade estão envolvidas na aquisição e análise dos dados de microarray (Zhang et al., 2001), entre elas as relacionadas às amostras. Primeiro, a heterogeneidade do material biológico. Um mesmo tipo de tumor de diferentes pacientes apresenta perfis genéticos e epigenéticos distintos. Essas diferenças são maiores entre tumores do que entre tecidos normais de indivíduos distintos. Segundo, tumores removidos cirurgicamente podem diferir em tamanho e estágio, na composição de células (células tumorais, estromais, endoteliais e inflamatórias), na presença de linfócitos e de necrose. Terceiro, as amostras coletadas podem corresponder a regiões diferentes do tumor, o que acrescenta variabilidade (Zhang et al., 2001).

Um dos cuidados utilizados para reduzir a heterogeneidade das amostras de tecido das quais se obtém o RNA para os experimentos de microarray é a micro (ou semi-micro) dissecação dos tecidos provenientes da cirurgia. Esse procedimento permite retirar áreas de tecido normal das amostras tumorais e áreas de tecido tumoral das amostras normais, bem como descartar regiões ou amostras com extensos infiltrados linfocitários ou necrose.

Outros métodos também são usados, com menor frequência, para esse fim, entre eles seleção por “cell sorting” e por marcadores de superfície (Zhang et al., 2001).

Uma técnica implementada mais recentemente é a microdissecção com captura a laser, ou LCM (Emmert-Buck et al., 1996). Esse método permite a obtenção de células isoladas com precisão e pouca manipulação (Bonner et al., 1997). Uma das críticas contra a LCM é a criação de uma amostra de células excessivamente purificada, distinta da encontrada no indivíduo e que não forneceria uma visão real do tumor. De fato, as células tumorais não existem isoladamente, e a patologia do câncer trata de uma mistura de células tumorais, estroma, células inflamatórias e endoteliais (Zhang et al., 2001). Todos esses tipos celulares podem ter papéis importantes na progressão tumoral e apresentar perfis de expressão informativos (Ramaswamy and Golub, 2002). Além disso, a diferença entre a composição celular de uma amostra pura obtida por LCM e uma amostra cirúrgica, de endoscopia ou de punção aspirativa dificulta o uso de marcadores identificados nas amostras purificadas para diagnóstico de amostras de pacientes. Outra crítica ao uso de LCM é o pequeno número de células obtido após a captura, geralmente insuficiente para os experimentos de microarray. Essa limitação torna necessário o uso de métodos de amplificação de RNA (Zhang et al., 2001).

A elaboração dos experimentos de microarray

Uma das maiores dificuldades dos experimentos de array é analisar o grande número de dados resultantes, identificando os genes significativamente diferencialmente expressos entre dois ou mais tipos de tecidos/condições.

De fato, a enorme quantidade de dados gerados deve ser compreendida e interpretada em termos de significado biológico. Protocolos biológicos para coleta dos dados devem ser integrados a protocolos matemáticos e estatísticos para analisá-los precisamente e a algoritmos computacionais para implementá-los (Zhang et al., 2001). Para tal, biólogos devem se tornar mais quantitativos e informáticos devem aprender a apreciar a natureza às vezes idiossincrática da biologia (Zhang et al., 2001).

Até o momento, há pouca padronização para elaboração e implementação de bancos de dados de expressão gênica, o que dificulta a comparação entre resultados de laboratórios distintos (Ramaswamy and Golub, 2002). Em 2002 foi fundado o banco público internacional ArrayExpress (www.ebi.ac.uk/arrayexpress), cujos principais objetivos são armazenar dados de publicações, possibilitar à comunidade o fácil acesso a

informações de expressão gênica de alta qualidade em um formato padronizado e facilitar a troca de desenhos experimentais e protocolos. Os dados podem ser submetidos nos formatos MIAME (minimum information about a microarray experiment-www.mged.org) ou MAGE-ML (microarray gene expression markup language).

Diversos cuidados devem ser considerados na elaboração, análise e interpretação dos resultados de experimentos de microarray, dentre eles a escolha do método para normalização da fluorescência dos dois canais, o número de replicatas, a forma de expressar os resultados (por exemplo pela razão da expressão ou pelo log da mesma) (Quackenbush, 2001) e o uso ou não de uma amostra referência em todas as hibridizações (Yang and Speed, 2002).

A normalização, feita após o processamento da imagem, visa igualar a intensidade relativa dos dois fluorocromos, principalmente visto que a intensidade observada para Cy5 é geralmente mais baixa do que a observada para Cy3. Ela ajusta diferenças na marcação e na eficiência de detecção dos dois corantes, bem como diferenças nas quantidades iniciais das amostras (Quackenbush, 2001). Diversos métodos são usados para normalização, baseados em ajustes globais ou locais. Todas as abordagens assumem que os genes como um todo ou um subgrupo de genes-controle representados no array devem ter razão de expressão igual a um entre os dois canais. Os métodos mais usados são a normalização por energia total e por regressão linear ou não-linear e por genes controle (Quackenbush, 2001). A normalização por energia total compara a soma das intensidades nos dois canais e calcula um fator de normalização que será multiplicado aos valores de cada spot. O método de regressão linear global utiliza um gráfico de dispersão (“scatter plot”) das intensidades dos spots dos canais Cy3 x Cy5, e ajusta a reta obtida para inclinação um. Os métodos de regressão locais, como lowess (Cleveland, 1979) e loess (programa R, bioconductor) corrigem de forma linear ou não-linear (dependendo dos ajustes) as regressões de subregiões do array, visando corrigir variações das razões entre os canais devido a intensidade e a outras variações sistemáticas desconhecidas. As subregiões podem corresponder aos subarrays (setores ou grupos de spots fixos com um mesmo pino ou “print-tip”), caso os pinos apresentem diferenças na eficiência de fixação dos spots que devam ser corrigidas (Dudoit et al., 2003). A normalização baseada em genes controle considera que as amostras em estudo possuem a mesma quantidade de transcritos de genes essenciais, como os genes “housekeeping”. A razão da marcação destes genes nos dois canais é ajustada para um, e são calculados limites de confiança para identificar genes diferencialmente expressos (Quackenbush, 2001). Duas críticas importantes a esse método

são o uso de um número reduzido de genes para corrigir todos os genes fixados no array, o que reduz a robustez da análise, e a dificuldade de se obter genes controle de expressão semelhante em dois tipos de amostra (Lee et al., 2002).

Após a normalização, os dados são geralmente apresentados como razões de expressão ou os logaritmos das mesmas. Genes de potencial interesse são frequentemente selecionados com base em um aumento ou redução de 2x em sua razão de expressão (Quackenbush, 2001). Esse método apresenta desvantagens, entre elas a falta de base teórica para determinação da razão para seleção e a falta análises como intervalos de confiança e variância (Quackenbush, 2001); (Jin et al., 2001).

Um único experimento de cDNA microarray está sujeito a grande variabilidade, e resultados confiáveis requerem a execução de réplicas. As maiores variabilidades referem-se à localização dos spots no array, a preparação das sondas, o uso de lâminas diferentes e de arrays produzidos em momentos diferentes (Lee et al., 2000). O número ideal de réplicas depende de diversos fatores como o tipo de equipamento utilizado para a produção dos arrays, as técnicas de laboratório, as condições de preparo das amostras e sobretudo a disponibilidade de material biológico (Lee et al., 2000). Experimentos controlados mostraram que aproximadamente 5% dos genes expressos em um tecido geraram sinais falso-negativos porque o mRNA presente no tecido não estava representado na amostra usada para sonda ou não hibridizou com o cDNA correspondente, fixo na lâmina. Por outro lado, 10% dos spots com sinal corresponderam a falso-positivos em algumas replicatas (Lee et al., 2000).

Existem três tipos principais de fontes de variações em experimentos de array: erros de medida, associados às leituras dos valores de fluorescência, afetados por artefatos como poeira; variações técnicas, introduzidas durante extração do RNA, marcação e hibridização; e variações biológicas, intrínsecas às amostras (organismos) (Churchill, 2002). Três tipos de replicatas técnicas são utilizadas para corrigir as variações associadas aos métodos utilizados. Um primeiro tipo são as replicatas internas (mais de um spot de um mesmo fragmento de DNA), que fornecem maior confiabilidade para a medida da marcação do gene no array. Essas replicatas partilham das mesmas condições de fixação do DNA no array, hibridização, etc. O segundo tipo de replicata, baseado em hibridizações de um mesmo RNA, permite corrigir diferenças na fabricação dos arrays e condições de hibridização, e pode ser útil nas situações em que esses métodos não estão totalmente estabelecidos. O terceiro tipo utiliza RNAs extraídos independentemente de trechos de um mesmo tecido, e corrige simultaneamente diferenças na produção de array, extração de

RNA e hibridização (Yang and Speed, 2002). O uso de um ou outro tipo de replicata depende da fonte de variação que se busque corrigir (Churchill, 2002).

Uma abordagem frequentemente utilizada é a replicata com inversão de corantes (“dye swap”). Cada hibridização é feita duas vezes, sendo as amostras marcadas com corantes invertidos nos dois experimentos (Yang and Speed, 2002).

A amostra de referência é uma amostra de RNA cujos dados não interessam ao estudo, marcada com um fluorocromo e hibridizada juntamente com as amostras de interesse, marcadas com um fluorocromo distinto. Os valores de intensidade dos spots nas amostras de interesse são divididos pelos valores obtidos na amostra referência que funciona como normalização indireta para as demais amostras, permitindo a comparação entre elas (Yang and Speed, 2002). Ela é especialmente útil em experimentos que envolvem diversos tipos de RNA, mas seu uso exige maior quantidade de recursos, uma vez que o número de experimentos é duplicado (Yang and Speed, 2002). Diversos trabalhos adotam a amostra de referência nos experimentos de microarray, dentre eles dois que identificam subclasses de tumores de mama com diferentes prognósticos (Perou et al., 2000); (Sorlie et al., 2001) e um que subdivide linfomas de células B grandes em dois tipos associados a taxas de sobrevivência bastante distintas (Alizadeh et al., 2000). Outros, porém, condenam seu uso, alegando que ela não acrescenta dados de interesse, que pode introduzir viés e afetar o poder de discriminação de genes diferencialmente expressos, aumentando a variância associada à comparação. (Jin et al., 2001). O uso ou não da amostra de referência dependerá do desenho experimental (número de classes de amostras, número de amostras por classe, grau de precisão desejado e disponibilidade de recursos), que deve refletir a pergunta principal do estudo (Churchill, 2002).

Determinação da intensidade dos spots de microarrays

Após a hibridização, uma imagem digital é adquirida para cada comprimento de onda (fluorocromo). Cada gene é então identificado por sua posição no array e a intensidade de sua marcação (sinal) para cada um dos corantes é calculada. Para isso, é necessário segmentar a imagem em três classes: subarrays, caixa do spot (“spot box”, região retangular que contém o spot) e spot (Hirata et al., 2003). Cada pixel da caixa do spot é caracterizado como pertencente ao foreground (região de sinal) ou background.

A maior parte dos programas disponíveis baseia-se no ajuste manual das caixas dos spots, o que leva a uma baixa reprodutibilidade e grande gasto de tempo no caso de experimentos com muitas imagens (Hirata et al., 2003).

Um dos programas usados com frequência em experimentos de microarray, utilizado neste trabalho, é o Quantarray (Packard BioScience, BioChip Technologies LLC). Este programa quantifica spots e background por três métodos: histograma, círculo fixo e adaptativo (Yang et al., 2002).

O primeiro constrói um histograma dos valores dos pixels de cada spot, e permite fixar os percentis que serão usados para os cálculos das intensidades de sinal e do background. Estes cálculos podem se basear na soma, média, moda ou mediana das intensidades dos pixels das duas regiões (spot e background). O método histograma é simples e estável, e exclui pixels de características extremas do cálculo das intensidades do spot e do background. Um defeito desse método é o caráter subjetivo da escolha dos percentis para os cálculos. O uso de intervalos pequenos pode levar à superestimação do valor do spot, uma vez que pixels “fracos” são excluídos do cálculo. Artefatos como ciscos na região externa do spot também podem afetar o cálculo da intensidade do mesmo, já que não há filtros para a localização dos pixels.

O método de círculo fixo utiliza regiões distintas, concêntricas, para o cálculo das intensidades de spot e background. Estas são determinadas de acordo com o diâmetro do spot, e dois diâmetros externos ao do spot que delimitam um “anel” onde é calculado o background. Os valores são calculados com base na soma, média, moda ou mediana das intensidades dos pixels do spot e background, nos percentis escolhidos. Uma das vantagens deste método é separar espacialmente pixels que serão usados para cálculo de spot e background, evitando a influência de artefatos de um parâmetro sobre o outro. Por outro lado, o uso de círculo fixo não é recomendável para arrays com spots irregulares (como os obtidos por detector CCD) pois esses não se enquadram nas regiões circulares de spot e background traçadas pelo método.

O método adaptativo utiliza regiões semelhantes às traçadas pelo círculo fixo para cálculo de spot e background. Um teste estatístico compara 8 pixels do spot com os 8 pixels medianos do background, ajustando essas duas regiões com base no nível de significância do teste, repetido até englobar toda a região previamente definida pelo diâmetro do spot. Experiências do nosso grupo mostraram que esse método apresenta maiores erros para spots de baixa intensidade em relação aos outros dois (Gustavo Esteves, comunicação pessoal), e sua utilização foi portanto descartada neste trabalho.

Recentemente, um programa para localização automática de subarrays e spots, denominado Bioinfo, foi elaborado no IME-USP baseado em morfologia matemática (ferramentas MMORPH, do MATLAB) (Hirata et al., 2003). A assistência do usuário é necessária apenas para selecionar os parâmetros de segmentação e checar o alinhamento final. A utilização deste programa reduzirá o tempo necessário para a segmentação e quantificação dos dados e aumentará a reprodutibilidade dos mesmos (Hirata et al., 2003).

2.3. Métodos de análise

As análises de cluster têm como objetivo encontrar grupos (“clusters”) de genes ou de amostras que se comportam de modo semelhante nas condições experimentais estudadas (Freeman et al., 2000); (Zhang et al., 2001). A clusterização de amostras complexas permite identificar subgrupos que, em estudos de doenças, por exemplo, podem apresentar diferentes prognósticos ou respostas a tratamentos.

Os métodos de análise de padrão ou agrupamento podem ser não-supervisionados (clusters) ou supervisionados (Triche et al., 2001); (Quackenbush, 2001). No primeiro caso, a busca da relação entre os genes e entre as amostras é feita sem informação adicional, ou seja, é baseada apenas na expressão gênica, nos dados obtidos no experimento (Ramaswamy and Golub, 2002); (Triche et al., 2001). No segundo caso, informações sobre a classificação patológica das amostras, “outcome” do paciente ou dados biológicos sobre genes funcionalmente relacionados são usados para guiar ou direcionar o algoritmo de agrupamento (Triche et al., 2001).

Dentre os métodos de clusterização não-supervisionados mais utilizados destacam-se o hierárquico, o “K-means”, o “self-organizing-maps” (SOM) e a análise de componentes principais (Quackenbush, 2001). O clustering hierárquico baseia-se numa matriz de distância par a par, a partir da qual os genes ou amostras mais semelhantes são reunidos dois a dois, sucessivamente, formando uma única árvore hierárquica. O método de “K-means” exige um conhecimento prévio do número de clusters. Os objetos (genes ou amostras) são distribuídos aleatoriamente nos grupos, um vetor de expressão média é calculado para cada cluster e são determinadas as distâncias entre os clusters. Os clusters são então redefinidos de modo a incluir as amostras mais próximas ao vetor médio, as distâncias intra e inter-clusters são novamente medidas e os vetores de expressão calculados. O processo é repetido até atingir a menor variabilidade intra-cluster (Quackenbush, 2001). Para o clustering por “self-organizing maps” é definida uma

configuração geométrica aleatória para os nós, cujo número é pré-escolhido. São gerados vetores para cada nó, treinados por um processo iterativo. Cada interação envolve a seleção de um gene ao acaso e o ajuste dos vetores dos nós, sendo essa correção maior quanto maior a proximidade em relação ao mapeamento do gene. O processo é repetido por 20.000 a 50.000 iterações (Tamayo et al., 1999). A análise de componentes principais gera uma projeção de conjuntos de dados complexos em um espaço reduzido, de fácil visualização. Ela busca visualizações que permitem a melhor separação dos dados, dando pesos diferentes para cada variável e descartando as que não contribuem para o agrupamento dos dados. Esse método pode ser utilizado para agrupamento de genes ou de amostras (Quackenbush, 2001).

As ferramentas de análise supervisionadas buscam construir classificadores que associem classes pré-definidas a determinados padrões de expressão gênica. Tais classificadores são tipicamente construídos com base nos dados de parte das amostras classificadas a priori, utilizados para a seleção dos melhores caracteres (no caso, genes) capazes de separar os grupos conhecidos (Brazma and Vilo, 2000); (Ramaswamy and Golub, 2002). Os classificadores são então testados em outras amostras de classificação também conhecida para validar sua capacidade de distinção (Brazma and Vilo, 2000); (Ramaswamy and Golub, 2002). Um dos métodos supervisionados mais usados é a máquina de suporte vetorial (SVM). Ela é um classificador binário cujo treinamento utiliza genes relacionados por função e genes não relacionados para ensinar a distinguir membros e não membros da classe com base nos dados de expressão (Quackenbush, 2001). As classes são separadas por uma superfície linear ou não linear que dá maior peso para a distinção entre pontos próximos a ela.

2.4. Amplificação de RNA

A maior parte dos métodos para análise global da expressão gênica requer grandes quantidades de material, nem sempre disponíveis para alguns tecidos ou tumores (Mahadevappa and Warrington, 1999). Para a técnica de microarray, resultados semelhantes são obtidos utilizando-se RNA total ou poliA como sonda (Mahadevappa and Warrington, 1999), mas nem sempre há quantidades suficientes de RNA para realização de replicatas de uma mesma amostra.

Nesses casos, uma opção que tem sido utilizada é a amplificação da população de mRNAs. Dois métodos principais podem ser utilizados para esse fim. Um deles baseia-se

em amplificação logarítmica por PCR, o outro em amplificação linear utilizando transcrição *in vitro*. O primeiro método parte de quantidades mínimas de RNA (até 1ng) para produção de cDNAs, aos quais são ligados adaptadores complementares aos iniciadores usados na amplificação por PCR (Puskas et al., 2002b). O segundo utiliza um oligonucleotídeo oligo dT ligado à sequência do promotor T7 para transcrição reversa, e o cDNA resultante é usado como molde para transcrição *in vitro* T7- dependente (Puskas et al., 2002b).

O primeiro método para amplificação linear de RNA baseava-se na transcrição reversa na presença de oligodT-T7 (oligodT contendo a sequência do promotor da T7 RNA polimerase), síntese da segunda fita após clivagem do RNA com RNaseH e transcrição *in vitro* tendo como molde o DNA dupla fita obtido (Van Gelder et al., 1990). Este processo levou ao aumento de 80 vezes na quantidade de RNA partindo de 3ng de cDNA, mantendo a heterogeneidade de tamanho dos transcritos da amostra inicial (Van Gelder et al., 1990). Um protocolo derivado deste incluiu um ciclo adicional de amplificação iniciado pela transcrição reversa do RNA amplificado na presença de oligonucleotídeos aleatórios e resultou em amplificações de 5-20 vezes e 200-400 vezes para o primeiro e o segundo ciclo, respectivamente (Baugh et al., 2001). Um outro método de amplificação linear foi elaborado com base nas propriedades da enzima transcriptase reversa adicionar nucleotídeos extras (especialmente dCTPs) na extremidade 3' do cDNA e trocar de molde ("switch template") (Matz et al., 1999). Este protocolo inclui dois ciclos de amplificação, iniciados por reações de transcrição reversa que utilizam oligodT-T7 e oligonucleotídeos aleatórios, e finalizados pela etapa de transcrição *in vitro* T7-dirigida (Wang et al., 2000). Comparações entre microarrays utilizando sondas de RNA total e RNA amplificado mostraram confiabilidade partindo de até 250ng de RNA total para um ciclo de amplificação e de até 30ng para dois (Wang et al., 2000).

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

Um dos trabalhos que utilizaram amplificação exponencial comparou por microarray a expressão de duas amostras de RNA total, os mesmos RNAs submetidos a amplificação linear e submetidos a amplificação exponencial pelo método SMART (Clontech) (Puskas et al., 2002b). Algumas alterações no perfil de expressão gênica foram observadas

comparando dados obtidos a partir de RNA total e amplificado, especialmente pelo método exponencial (Puskas et al., 2002b). Um método de amplificação baseado em PCR e transcrição *in vitro* foi utilizado para obtenção de grandes quantidades de RNA de referência para experimentos de microarray, e sua hibridização apresentou resultados positivos para a grande maioria dos spots, mostrando a manutenção da diversidade de transcritos (Puskas et al., 2002a).

3. OBJETIVOS

- Padronizar a metodologia de microarray utilizando lâminas de vidro e RNA proveniente de tecidos tireoidianos
- Comparar o perfil de expressão gênica de tecido normal, bócio, adenoma, carcinoma papilífero e carcinoma folicular utilizando arrays contendo clones pré-selecionados por DDRT-PCR ou ORESTES
- Identificar genes diferencialmente expressos entre tecido normal, hiperplasia (bócio), neoplasia benigna (adenoma) e maligna (carcinomas papilíferos e foliculares)

4. MATERIAIS E MÉTODOS

4.1. Nucleotídeos radioativos e fluorescentes

$\alpha^{32}\text{P}$ dCTP (Amersham Biosciences)

$\gamma^{32}\text{P}$ ATP (Amersham Biosciences)

Cy3-dCTP (Amersham Biosciences)

Cy5-dCTP (Amersham Biosciences)

4.2. Enzimas

Superscript II RT (Invitrogen)

DNase (Promega)

Advantage cDNA polymerase (Clontech)

T7 Enzyme mix (Ribomax- Promega)

T4 kinase (New England Biolabs)

Ribonuclease H (Invitrogen)

Taq DNA polymerase (Invitrogen)

Sequencing grade Taq DNA polymerase (Promega)

Taq Man PCR mix (Perkin Elmer Life Sciences)

Enzimas de restrição (New England Biolabs)

4.3. Vetores

Sureclone (Amersham Biosciences)

PGEM-T Easy (Promega)

4.4. Oligonucleotídeos ou iniciadores

Comerciais

“Random” hexanucleotídeos- pd(N)₆ (Amersham Biosciences)

oligo(dT) 12-18 (Invitrogen)

Sintetizados

Tabela I. Oligonucleotídios utilizados nesse trabalho

Oligont.	Sequência
aleatório 13	5' CTGATCCATG 3'
aleatório 14	5' CTGCTCTCAA 3'
aleatório 15	5' CTTGATTGCC 3'
Opa 17	5' GACCGCTTGT 3'
Opa 18	5' AGGTGACCGT 3'
Bauer 1	5' TACAACGAGG 3'
Bauer 2	5' TGGATTGGTC 3'
pUC forward	5' CGCCAGGGTTTCCAGTCACGAC 3'
pUC reverse	5' TTTCACACAGGAAACAGCTATGAC 3'
18 for	5' CAGCAGTATGGACAGCTGGA 3'
18 rev	5' TTAAATGAGGGCTGAAACGG 3'
44 for	5' CCTGTTGATGGGATGTTTCC 3'
44 rev	5' CTGCTGTTGCAGGAAAATGA 3'
199 for	5' GGAGGTGAGACAGGGTCTA 3'
199 rev	5' GAACTCAGGTGATACTCCCA 3'
234 for	5' AGTGCAAAGAGTGCAAATG 3'
234 rev	5' CAAAGGGGTCAAGATTGTAG 3'
TBP for	5' CCACAGCTCTTCCACTCACA 3'
TBP rev	5' ATCCTCATGATTACCGCAGC 3'
GAPDH for	5' CTGCACCACCAACTGCTTA 3'
GAPDH rev	5' CTAGACGGCAGGTCAGGTC 3'
18 f (Real T)	5' AACTGAGGACCTCGGAATCTCTAG 3'
18r (Real T)	5' GGTGTCTTTTGTAGCTTTTGTGCATCT 3'
44f (Real T)	5' CATTTATGAAACGGCAATATTTGG 3'
44r (Real T)	5' GCTGTTGCAGGAAAATGATTA 3'
18S f (Real T)	5' CGGCTACCACATCCAAGGAA 3'
18S r (Real T)	5' GCTGGAATTACCGCGGCT 3'
Sonda 18S-JOE	5' TGCTGGCACCAGACTTGCCCTC 3'

Sonda 18-FAM	5' AAGGGCTCAACTTCGAAAATGGCAACAAC 3'
Sonda 44-FAM	5' TTCCTTGTGTCTTCTGGTAAGTCTCTGCACTATCA 3'
oligodT-T7	5' AAACGACGGCCAGTGAATTGTAATACGACTCACTATAGGCGCT 3'
TS	5' AAGCAGTGGTAACAACGCAGAGTACGCGGG 3'

4.5. Coleta das amostras

Para a obtenção de um número de amostras de carcinoma folicular suficiente para as análises estatísticas foi fundado o “Grupo sul-americano de estudo de tireóide”, do qual fazem parte o Hospital do Câncer A. C. Camargo, o Instituto de Enfermedades Neoplasicas- INEN de Lima, Peru, o Hospital Araújo Jorge de Goiânia e o Hospital Heliópolis de São Paulo. Em todas essas instituições houve aprovação do projeto pelo Comitê de Ética local e foi exigido um consentimento assinado pelo paciente, além do relatório com dados clínicos e cirúrgicos. As amostras pós-cirúrgicas foram coletadas pelo patologista de plantão no centro cirúrgico e congeladas em nitrogênio líquido. O transporte das amostras de outros hospitais foi feito em gelo seco.

A tabela II mostra a distribuição das amostras utilizadas para extração de RNA e as disponíveis no banco, de acordo com a instituição de origem e o tipo de tecido. As amostras de carcinoma e adenoma foliculares são do tipo clássico, as de Hurthle foram excluídas deste estudo.

Tabela II. Amostras de tireóide utilizadas para extração de RNA ou armazenadas no banco de tumores, segundo diagnóstico e instituição.

DIAGNÓSTICO	RNA extraído		No banco de tumores		
	H.CÂNCER	INEN	H.CÂNCER	H.ARAÚJO JORGE	H.HELIOÓPOLIS
Normal	109	50	293	164	61
Bócio	33	16	164	85	46
Adenoma	18	18	10	29	-
C. Folicular	2	5	-	3	-
C. Papilífero	70	18	58	31	5
Total	235	110	562	315	122

Nesta etapa do trabalho utilizamos apenas amostras provenientes de nosso hospital e do INEN. A seleção das amostras para o estudo foi feita com base no laudo definitivo. A análise das lâminas de cortes de parafina para determinação do laudo definitivo e a semi-microdissecção de todas as amostras foram realizadas pelo Dr Fernando Soares, chefe da equipe de anatomia patológica do Hospital do Câncer, com o intuito de manter a uniformidade da classificação histológica. As lâminas contendo os cortes de parafina provenientes do INEN foram enviadas juntamente com as amostras e os relatórios correspondentes. As lâminas para semi-microdissecção foram preparadas no Hospital do Câncer, a partir do material mantido em gelo seco, e a semi-microdissecção também foi realizada com as amostras congeladas. Uma vez descongeladas, estas foram utilizadas imediatamente para extração de RNA total.

4.6. Extração de RNA total, tratamento com DNase I

Para extração de RNA total fragmentos de tecido foram transferidos para tubos contendo 1ml de TRIZOL (Life Technologies, USA) e triturados com auxílio de um homogenizador (Polytron).

Quando desejado, o RNA total foi tratado com DNase I a fim de eliminar possíveis contaminações com DNA genômico. Para cada 50µg de RNA adicionou-se 30U de DNaseI, 120U de RNasin, DTT para 1mM, tampão DNase e água DEPC para 300µl. A mistura foi incubada a 37°C por 1h. Acrescentou-se 75µl de stop solution (Promega) e 225µl de água e procedeu-se extração com fenol-clorofórmio e clorofórmio. A precipitação foi feita com 10µl NaCl e 525µl de etanol 80%, a -70°C durante a noite. Centrifugou-se a 14000rpm por 30 minutos e o precipitado foi lavado com etanol 75%, centrifugado, seco e ressuspensão em um volume de água para uma concentração aproximada de 1,5-2µg/µl de RNA. A quantificação foi feita em espectrofotômetro.

O RNA de referência, utilizado nas hibridizações de microarray, consistiu de um pool de RNAs de quatro amostras grandes de tecidos tireoideanos normais. Cada RNA foi marcado e hibridizado em microarray para assegurar qualidade satisfatória. Após a reunião das amostras e quantificação em espectrofotômetro, a qualidade deste pool de RNAs também foi avaliada por marcação e hibridização em microarray.

4.7. Avaliação da qualidade do RNA: Northern blot e gel de agarose

Northern Blot

Northern blots foram realizados tanto para avaliar a qualidade dos RNAs, através de hibridização com GAPDH, quanto para analisar a expressão dos transcritos correspondentes às bandas de DDRT-PCR. Para analisar a qualidade dos RNAs foram preparadas membranas com 5µg de cada preparação, enquanto que para estudo dos níveis de expressão utilizou-se 15-20µg de RNA de cada pool (geralmente 3 amostras de cada tecido).

As amostras de RNAs foram preparadas em 10% glicerol, 10mM ácido bórico pH 8,3, 0,4mM EDTA, 25% formamida e 7,5% formaldeído, aquecidas a 65°C por 5 min e aplicadas em um gel de agarose 1% contendo 0,02M ácido bórico pH 8,3, 0,75mM EDTA e 16,6% formaldeído. A eletroforese foi realizada a 70V em tampão 0,02M ácido bórico pH 8,3, 0,5mM EDTA e 5% formaldeído. Após a eletroforese o RNA foi transferido para membrana de nylon por capilaridade (em 10x SSC, por aproximadamente 12hs) e fixado a esta utilizando-se o aparelho UV Crosslinker (FisherBiotech).

As sondas de GAPDH e dos fragmentos de DDRT-PCR foram preparadas a partir do fragmento de digestão dos clones contendo os insertos de interesse. A marcação foi feita com auxílio do kit Ready-to-Go (Amersham Biosciences), utilizando $\alpha^{32}\text{PdCTP}$, seguindo as recomendações do fabricante. Após a marcação, a sonda radioativa foi purificada por centrifugação através de coluna Sephadex G50 equilibrada em TE.

As hibridizações basearam-se no protocolo descrito por Church & Gilbert (Church & Gilbert, 1984). A pré-hibridização, a hibridização e as lavagens foram feitas em forno de hibridização (Robbins Scientific). Para a pré-hibridização as membranas foram incubadas em 50% solução fosfato (0,5M Na, 0,25M PO_4), contendo 7% SDS, 1% BSA e 1mM EDTA a 65°C por 15 min. A sonda foi então aquecida a 95°C por 5 min e acrescentada à solução de pré-hibridização. A hibridização foi feita a 65°C por pelo menos 12 horas. Foram realizadas duas lavagens de 30 min a 65°C em 50% solução fosfato, 1% SDS, 1mM EDTA. As membranas ainda úmidas foram expostas contra filme auto-radiográfico e Phosphor Screen.

A figura 1 mostra um Northern blot de 11 amostras de RNA de tireóide, hibridizado com sonda do gene GAPDH.

Figura 1. Northern blot para avaliar a qualidade das amostras de RNA total. Cada amostra corresponde a 1 μ g de RNA de tecido normal (N), carcinoma papilífero (CP) ou bócio (B), de um único paciente. Os RNAs foram fracionados em gel de agarose 1% desnaturante, transferidos para membrana de nylon e hibridizados com sonda do gene *GAPDH*.

Gel de agarose com brometo de etídio

Outro método utilizado para avaliação da qualidade do RNA foi a eletroforese em gel de agarose com brometo de etídio. Amostras de 1 μ g de RNA em tampão com 3,5M de uréia, 15% glicerol e azul de bromofenol foram analisadas em gel de agarose 1% em TAE contendo 0,5 μ g/ml de brometo de etídio. A qualidade foi estimada pela intensidade das bandas dos rRNAs 28S e 18S. Um gel representativo é mostrado na figura 2.

Figura 2. Gel de agarose 1% para avaliar a qualidade do RNA. Cada amostra contém 1 μ g de RNA total de tireóide de um único paciente, em tampão desnaturante. As duas bandas mais intensas correspondem aos RNAs ribossômicos 18 e 28S.

4.8. Differential Display RT-PCR (DDRT-PCR)

Diversos tipos de sequências podem ser fixados nas membranas ou lâminas de DNA microarray. Clones provenientes de projetos genoma são bastante utilizados por sua abundância e pela facilidade de sua obtenção. O uso de clones pré-selecionados como diferencialmente expressos através de métodos como DDRT-PCR aumenta a possibilidade de se encontrar marcadores.

A técnica de DD-RT-PCR (Liang and Pardee, 1992) baseia-se na transcrição reversa de mRNAs utilizando oligonucleotídeos (iniciadores ou “primers”) dT₁₁VN (onde V= nucleotídeos A, C ou G e N= A, T, C ou G) e na amplificação do cDNA assim obtido por PCR (Saiki et al., 1988) radioativo de baixa estringência, utilizando como iniciadores o mesmo oligonucleotídeo da transcrição reversa e um oligonucleotídeo arbitrário. Os produtos de reações feitas a partir de RNA de diferentes populações celulares (células normais e tumorais, por exemplo) são analisados em gel de poliacrilamida, e o padrão de bandas é comparado. Essa técnica já foi usada para busca de mRNAs de meia-vida curta, superexpressos em cânceres de tireóide (Gonsky et al., 1997).

Neste trabalho utilizamos o método desenvolvido por Liang e Pardee (Liang and Pardee, 1992), adaptado por Silva (Silva et al., 1999). A transcrição reversa foi feita com 0,2µg do RNA tratado com DNase, utilizando 200U de enzima transcriptase reversa Superscript II (Invitrogen) e 25 pmoles do iniciador ancorado (dT₁₁VN, onde V corresponde a A, C, ou G e N é um nucleotídeo fixo). O RNA e o oligonucleotídeo (em um volume total de 10µl) foram aquecidos a 70°C e transferidos para gelo. A reação de transcrição foi feita a 42°C por 1 hora após adição de enzima, 10mM DTT, 20µM dNTPs e tampão 1st strand. A enzima foi inativada por aquecimento a 70°C por 15 minutos. Dois microlitros da reação de síntese da primeira fita de cDNA foram utilizados na etapa de PCR. A reação foi preparada com 2,5U de enzima Taq DNA-polimerase seq Grade (Promega), 20µM dNTPs, 2,5mM MgCl₂, 0,2µCi/µl de dCTP α³²P, 2,5µM do iniciador T₁₁VN e 0,5µM de um iniciador aleatório (decâmero contendo aproximadamente 50%GC) e tampão-PCR (Invitrogen- Tris-HCl 10mM pH 8,4, KCl 50mM), em um volume final de 20 µL. A mistura foi submetida a aquecimento a 94°C por 4min, 40 ciclos de 94°C/30seg., 40°C/2min, 72°C/30seg., e um último ciclo de extensão a 72°C/5min. Os cDNAs amplificados foram fracionados em Gel de Sequenciamento 6% e o perfil de bandas foi analisado após exposição em filme auto-radiográfico e Phosphor Screen.

4.9. Recuperação das bandas do gel de DDRT-PCR, re-amplificação, clonagem

Os fragmentos do gel de DDRT-PCR contendo as bandas de interesse foram excisados e transferidos para tubos eppendorf. Adicionou-se 100µl de ddH₂O e incubou-se a 95°C por 15 minutos. Centrifugou-se por 2 minutos e recuperou-se o sobrenadante. O DNA foi precipitado com acetato de sódio 3M (0,1 volume) e etanol (2,5 volumes) a -70°C por 30min, seguido de centrifugação a 12000 rpm a 4°C por 30 minutos. Após lavagem com etanol 80%, o precipitado foi dissolvido em 40µl de ddH₂O. Para a re-amplificação foram utilizadas as mesmas condições de PCR descritas no item 4.8, com algumas alterações: o volume final da reação foi aumentado para 40µl, foram utilizados 2µl da preparação de cDNA, utilizou-se Taq polimerase (Invitrogen) e não Taq seqGrade e não foi utilizado dNTP marcado com $\alpha^{32}\text{PdCTP}$. Os produtos foram fracionados em gel de agarose 1,0% contendo 0,5µg/ml de brometo de etídio.

Os produtos da re-amplificação foram purificados a partir das bandas do gel de agarose por centrifugação através de coluna montada com lã de vidro, seguida de uma extração com fenol-clorofórmio e duas extrações com clorofórmio. O DNA presente no eluato foi precipitado com 0,1volumes de acetato de sódio 3M e 2,5volumes de etanol a -70°C por 1h ou a -20°C durante a noite. Alternativamente, nos casos em que o produto de amplificação se mostrou livre de dímeros dos iniciadores ("primer-dimers"), procedemos sua precipitação com acetato e etanol, sem purificação prévia. Nos dois casos o precipitado foi lavado com etanol 80% e ressuspenso em 20 ou 40µl de água, de acordo com a quantidade de DNA visualizada no gel.

A clonagem foi feita com auxílio dos kits SureClone (Amersham Biosciences) -que utiliza o vetor pUC18- e PGEM-T Easy (Promega). De modo geral utilizou-se 10µl de DNA para o kit Sureclone e 4µl para o PGEM-T Easy.

Preparação de bactérias competentes e transformação

Foram inoculadas 10 colônias de *Escherichia coli* DH5 ou JM109 (recém plaqueadas a partir de estoque congelado) em 260ml de meio LB (5g/l NaCl, 5g/l extrato de levedura, 10g/l tripton, pH 7,5), em frasco de 2l. O frasco foi mantido sob agitação a 22°C até a cultura atingir A₆₀₀ 0,6 (aproximadamente 24-30hs) e transferido para gelo por 10min. A cultura foi distribuída em 6 tubos de 50ml, com 42ml cada. Centrifugou-se a 3500rpm

(Sorval RT6000, rotor H1000B) por 15 min a 4°C e ressuspendeu-se cada pellet em 12ml de TB (10mM Pipes, 55mM MnCl₂, 15mM CaCl₂, 250mM KCl pH6,7) gelado. Os conteúdos de 3 tubos foram reunidos em um único frasco, incubados em gelo por 10min e centrifugados como descrito. Cada precipitado foi ressuspensão em 9,3ml em TB gelado, o conteúdo de 2 tubos foi reunido e adicionou-se lentamente 1,4ml de DMSO, agitando suavemente. Incubou-se em gelo por 10 min e distribuiu-se em alíquotas de 0,5ml, prontamente transferidas para nitrogênio líquido.

Para transformação descongelou-se bactéria a temperatura ambiente e adicionou-se 100µl a cada eppendorf contendo 5-6µl de reação de ligação. Os tubos foram mantidos em gelo por 30min, a 42°C por 2min e em gelo novamente por 1min. Adicionou-se então 1ml de LB e manteve-se a 37°C sob agitação por 1h. Os tubos foram centrifugados e os precipitados ressuspensos em 100µl de LB. O plaqueamento foi feito em LB amp (100µg/ml), na presença de Xgal (22µg/ml) e IPTG (55µg/ml).

Seleção de clones positivos

Foram selecionadas quatro colônias brancas de cada placa e cada uma foi inoculada em 20µl de LB e mantida sob agitação a 37°C por 1h. Para cada colônia foi preparada uma reação de PCR de 10µl contendo 1µl da cultura de bactéria, 1µM dos iniciadores pUC forward e pUC reverse (tab. I), 2,5mM MgCl₂, 20µM dNTPs, 1U de Taq polimerase (Invitrogen) e tampão PCR. O programa utilizado consistiu de uma etapa de 94°C por 4min, 35 ciclos de (94°C por 30s, 58°C por 30s, 72°C por 1min45s) e uma incubação a 72°C por 7min. Os produtos das amplificações foram analisados em gel de agarose 1%.

Dois clones de cada preparação, contendo insertos de tamanho compatível com as bandas originais (gel de DDRT-PCR), foram selecionados para preparação de DNA plasmidial.

Preparação de DNA plasmidial recombinante em pequena escala (Miniprep)

O DNA plasmidial foi isolado a partir de 2,5ml de cultura, com auxílio do kit Wizard Plus Minipreps (Promega). A confirmação da identidade dos clones foi feita pela análise de digestões de 4µl das preparações com *Eco*RI (para clonagens em PGEM-T) ou *Bam*HI e *Kpn*I (no caso de clonagens em pUC18- Sureclone).

4.10. Sequenciamento

Para o sequenciamento em maior escala, em placa, utilizamos o MEGAbace. O DNA foi preparado por miniprep em placa, utilizando culturas de bactéria incubadas a 37°C por 22h. A miniprep foi feita pelo método de lise alcalina (SDS-NaOH), o RNA foi digerido com RNase e o DNA foi precipitado com acetato de potássio e isopropanol e ressuspensão em 40µl de água.

As reações de sequenciamento foram feitas no Laboratório de Genética do Câncer, no Instituto Ludwig, utilizando geralmente 2µl do DNA proveniente da miniprep.

4.11. Real-Time RT-PCR

O método de Real-Time PCR consiste de uma cinética de amplificação que detecta o acúmulo do produto de PCR ao longo das reações de amplificação. Baseado na quantificação da clivagem de uma sonda fluorescente que pareia com um trecho central do amplicon pela atividade 5' nuclease da Taq polimerase, este método mede com precisão e reprodutibilidade a quantidade inicial do gene ou transcrito de interesse no espécime original (Heid et al., 1996). Experimentos de Real-Time foram usados para análises da expressão de genes específicos em leucemia (Preudhomme et al., 1999), tumores de pulmão (Calvo et al., 2000), de mama (Sgroi et al., 1999); (Bieche et al., 1999); (Martin et al., 2001) e de tireóide (Takano et al., 1999b).

Dois µg de RNA foram submetidos a transcrição reversa utilizando 2µM de “random” hexanucleotídeos (Amersam Biosciences). Como controles para amplificação de DNA genômico foram preparados tubos com todos os reagentes exceto Superscript. As reações foram incubadas a 42°C por 60min e inativadas a 70°C por 15min. Adicionou-se a cada tubo 30µl de água.

As reações de amplificação foram feitas em placas de PCR, utilizando 2,5µl de mix Primer/Probe do gene alvo, 2,5µl de mix Primer/Probe do gene controle, 12,5µl de master mix (Perkin Elmer Life Sciences) e 5µl de água. O programa consistiu de incubações de 50°C por 2min, 94°C por 10min e 55 ciclos de 94°C por 30s, 55°C por 30s e 72°C por 30s.

A reação foi feita no aparelho ABI 7700 (Applied Biosystems), e os resultados foram importados para o programa MS-Office Excel para normalização e demais análises.

O valor utilizado para cálculo da massa dos RNAs de interesse foi o Ct, que corresponde ao número de ciclos necessários para a fluorescência atingir um limiar fixado para cada experimento.

Além das amostras de interesse foram preparadas diluições seriadas de uma reação de transcrição reversa de RNA de tireóide normal para elaboração das curvas padrão para os genes de interesse e gene controle. A partir da curva padrão ($C_t \times \text{ngRNA}$), calculou-se a quantidade do RNA dos genes de interesse e do controle interno em cada amostra e o DNA correspondente dos controles negativos. A quantidade de DNA dos controles negativos foi subtraída do RNA correspondente e a massa dos transcritos dos genes 18 e 44 em cada amostra foi normalizada com base na quantidade de RNA do gene controle presente na mesma amostra.

4.12. RT-PCR semi-quantitativo

Foram realizadas amplificações dos transcritos de interesse, bem como dos genes TBP e GAPDH, utilizados como controles internos, em reações independentes.

Incubou-se 1µg de RNA e 2µM de “random” hexanucleotídeos (Amersham Biosciences) em 15µl totais a 70°C por 10min e transferiu-se para gelo. Foram acrescentados dNTPs para 0,25mM, DTT para 0,5mM e tampão 1st strand e incubou-se a 42°C por 2 min. Adicionou-se 200U de Superscript II (Invitrogen) e manteve-se a 42°C por 50min e a 70°C por 15min.

A reação de amplificação foi feita em 25µl, na presença de 0,4µM de cada oligonucleotídeo, 1,5mM MgCl₂, 20µM dNTPs, 0,625U Taq polimerase (Invitrogen) e tampão PCR. O programa consistiu de uma etapa inicial de desnaturação a 94°C por 5min, 18, 28 ou 35 ciclos de (94°C por 30s, temperatura de annealing por 30s e 72°C por 1min) e uma etapa adicional de extensão a 72°C por 7min. A temperatura de annealing utilizada em cada caso coincidiu com a T_m dos oligonucleotídeos empregados na reação.

Aliquotas de 10-15µl de cada reação foram analisadas em gel de agarose 1% com 0,5µg/ml de brometo de etídio.

4.13. Southern blot

Após eletroforese, o gel foi incubado em solução desnaturante (1,5M NaCl; 0,5M NaOH) por 20 min sob agitação. Após breve lavagem com água, o gel foi mantido em solução neutralizante (0,5M Tris-HCl pH8; 1,5M NaCl) por 30 min, sob agitação. Transferiu-se então para SSC 2x e preparou-se o aparato de transferência, realizada em 10x SSC por pelo menos 3h.

A membrana foi lavada em 2x SSC e submetida (ainda úmida) a cross-link em UV de 1220J.

Hibridização de Southern blot com sonda de oligonucleotídeo

Os oligonucleotídeos reversos utilizados para RT-PCR dos genes 18N, 44C, 199, 234, TBP e GAPDH foram marcados radioativamente com $\gamma^{32}\text{PdATP}$. Para tal, foram preparadas reações contendo 20pmol de oligonucleotídeo, 2,5 μl de tampão quinase 10x (NEB), 1,0 μl de T4 quinase (NEB), 2,5 μl de $\gamma^{32}\text{PdATP}$ (3000mCi/mmol) e água qsp 25 μl . Incubou-se a 37°C por 30min, acrescentou-se 5 μl de EDTA 0,5M e purificou-se em coluna Sephadex G25, equilibrada com TE.

O bloqueio da membrana para ligações inespecíficas da sonda, a pré-hibridização, a hibridização e a segunda lavagem foram feitas a 5°C abaixo da T_m do oligonucleotídeo usado como sonda, em forno de hibridização.

Para o bloqueio, a membrana foi incubada em 0,1x SSPE, 0,5% SDS por 20-30min. A pré-hibridização foi feita por incubação em 2x SSPE, 0,1% SDS por 20-30 min. Após esse período adicionou-se a sonda e incubou-se durante a noite.

A remoção da sonda não ligada especificamente foi feita através de uma lavagem de 10 min a temperatura ambiente em 2x SSPE, 0,1% SDS e duas lavagens de 15 min na temperatura calculada em 2xSSPE, 0,1% SDS.

Hibridização de Southern blot com sonda de fragmento de DNA

Em alguns casos, especialmente aqueles em que o produto da amplificação por RT-PCR era de pequeno tamanho, a utilização do oligonucleotídeo iniciador como sonda gerou imagens nas quais o rastro de marcação do iniciador livre foi próximo às bandas de interesse. Nessas situações optamos por fazer uma reação de RT-PCR utilizando os iniciadores correspondentes e RNA de tecido normal, purificar o fragmento amplificado e

utilizá-lo como sonda. A marcação da sonda, pré-hibridização, hibridização e lavagem das membranas foram feitas conforme descrito no item 4.7.

Quantificação das bandas do Southern blot

Após as lavagens, as membranas foram expostas em Phosphor screen por tempos variáveis, de acordo com o nível de expressão do gene e com o tipo de sonda (oligonucleotídeo ou fragmento de DNA). As sondas de oligonucleotídeo resultavam em marcações mais fracas, e portanto exigiam exposições mais prolongadas.

A quantificação foi feita com auxílio do programa ImageQuant (Molecular Dynamics), com base na determinação da área da banda radioativa em retângulos de mesmo tamanho para todas as amostras. A normalização dos dados de cada amostra foi feita dividindo-se a intensidade da banda do gene de interesse pela intensidade da banda referente ao gene *TBP* ou ao gene *GAPDH*, amplificado a partir da mesma reação de transcrição reversa.

Os valores obtidos após a normalização foram utilizados para a construção de gráficos e Box plots. Os dados da normalização por *TBP* foram também utilizados nas análises de cluster e árvore de decisão.

4.14. Métodos de classificação

Análise de cluster

As análises foram realizadas com auxílio do Matlab (Math Works ®) e suas ferramentas. Foram tomados os logaritmos dos dados normalizados por *TBP* (ver acima, 4.13) e os resultados foram analisados por cluster hierárquico, distância Euclidiana e ligação completa. Estratégias de reamostragem foram adotadas para gerar 100.000 datasets simulados de expressão gênica. Aplicando o método de cluster a cada um desses datasets, calculou-se a proporção dos que apresentaram agrupamento semelhante ao do dataset real. Essa proporção fornece uma estimativa da possibilidade de se obter, ao acaso, o agrupamento observado no experimento.

A capacidade de cada gene discriminar tipos distintos de amostras também foi estimada pelos valores P de dois testes não paramétricos- Wilcoxon Rank Sum e Kolmogorov-Smimov, Pw e Pks, respectivamente.

Árvore de decisão

O método de árvore de decisão baseia-se na separação gradativa das amostras em duas classes, utilizando dados de um gene por vez. A eficiência do método foi avaliada pela técnica de validação cruzada (“cross-validation”) denominada “leave-one-out”, na qual uma das amostras era retirada da análise e as restantes eram utilizadas para construir um classificador, testado pela amostra excluída. O processo foi repetido para todas as combinações possíveis e a porcentagem de classificações corretas foi calculada.

4.15. Construção de Arrays

4.15.1. Membranas de nylon

Os arrays em membranas de nylon utilizados nesse trabalho foram elaborados com clones de DDRT-PCR. Duas reações de PCR de 100µl foram preparadas para cada clone, conforme discutido no item anterior. As duas reações foram reunidas em um poço único da placa e o produto de PCR foi purificado e ressuspenso em 50µl de água, com auxílio do QIAquick 96 PCR purification kit (QIAGEN). Cada clone foi fixado em triplicata, e os controles (positivos e negativos) foram fixados em diluições seriadas de um fator de 2x, desde o concentrado até a diluído 1:2048 ($1:2^{11}$).

As amostras foram fixadas pelo robô de microarray Flexys (Genomic Solutions). As membranas foram mantidas por 5 minutos em NaOH, SSPE 10x e SSC 2x e submetidas a cross-link de 1200J em UV Crosslinker (Fisher Scientific).

Hibridização com cDNA radioativo

cDNAs radioativos foram preparados a partir de pools de RNA (3 amostras/pool) de tecido normal, bócio e carcinoma papilífero, em dois experimentos independentes.

Aproximadamente 40µg de RNA de cada pool foram secos e ressuspenso em água e submetidos a transcrição reversa na presença de 2µg OligodT, Superscript e 5µl $\alpha^{32}\text{PdCTP}$, a 42°C por 2 horas.

As etapas de pré-hibridização, hibridização e lavagens foram feitas conforme descrito em 4.7.

Digitalização da imagem e quantificação do sinal dos spots

As membranas foram expostas em Phosphorscreen por 2 dias. Para digitalização da imagem foi utilizado o scanner STORM 840 (Amersham Biosciences) e o programa Scanner Control.

A quantificação do sinal de cada hibridização feita pelo programa ArrayVision (Amersham Biosciences).

Análise dos dados

Para os arrays em membranas de nylon utilizamos o método de normalização por energia total. A soma das intensidades de marcação de todos os spots foi comparada entre as membranas de um mesmo experimento, e igualadas utilizando fatores de correção.

Foram calculadas as razões entre as intensidades de cada spot nas três comparações possíveis (tecido normal x bócio, bócio x carcinoma e tecido normal x carcinoma), e os genes cuja razão foi igual ou maior do que dois em alguma comparação foram considerados diferencialmente expressos.

Dois experimentos distintos foram realizados (com alvos provenientes de pacientes diferentes) e os genes diferencialmente expressos nas duas análises foram selecionados.

4.15.2. Lâminas de vidro

Lâminas de vidro silanizadas (Corning) foram utilizadas para fixação do produto de PCR dos clones de ORESTES. Sequências ORESTES contidas em plasmídios pUC foram amplificadas em placas de 96 poços e o produto foi purificado em placa com resina G50. Aliquotas destes produtos foram depositadas em lâminas de vidro silanizadas (Corning) com auxílio do robô Flexys (Genomic Solutions). As lâminas foram tratadas e submetidas a crosslink em luz ultravioleta.

As lâminas utilizadas neste trabalho, denominadas 3.8K, contêm em torno de 3800 clones de cDNA (ORESTES), de aproximadamente 400-500pb, provenientes de bibliotecas de estômago, laringe e mama, bem como alguns dos clones de DDRT-PCR também fixados na membrana de nylon. A maior parte dos clones está representada em uma única cópia. Como controles positivos da marcação e hibridização, cDNAs dos genes

Q do fago λ , *Trp*, *DapB*, *ThrB*, *PheB* e *LysA* foram fixados na lâmina. Esses controles são utilizados apenas para análise da qualidade dos dados, e não para normalização dos mesmos.

Hibridização com cDNA fluorescente

Alíquotas de 25µg de RNA de amostras individuais de tecido normal, adenoma, bócio ou carcinoma papilífero foram precipitadas em acetato de sódio e etanol juntamente com uma alíquota de uma mistura de RNAs dos controles positivos transcritos *in vitro* (45ng para o gene *Q* do fago λ , 30ng para o gene *Trp*, 15ng para o gene *DapB*, 3ng para o gene *ThrB*, 1,5ng para o gene *PheB* e 0,3ng para o gene *LysA*). O RNA precipitado foi ressuspenso em 10µl de água, acrescentou-se 2µg de oligodT, aqueceu-se a 70°C por 10 min e resfriou-se em gelo por 1 min. A reação foi feita em um volume de 30µl, acrescentando-se 1st strand buffer, DTT, dNTPs, 20U RNasin, 500U Superscript II (Invitrogen) e 3nmoles de nucleotídeo Cy3-dCTP ou Cy5-dCTP (Amersham Biosciences). A mistura foi incubada a 42°C por duas horas e o RNA foi degradado pela adição de NaOH e EDTA e incubação a 70°C por 20 min. O pH foi neutralizado por adição de HCl e o cDNA marcado foi purificado em coluna Autoseq G50 (Amersham Biosciences). Para cada hibridização preparou-se uma sonda a partir de 25µg do RNA de referência, marcada com o nucleotídeo distinto do da amostra. Após a purificação com Autoseq G50, as duas sondas (amostra e referência) foram reunidas em um só tubo, adicionou-se 4µg de DNA polyA (Amersham Biosciences), 4µg de DNA Cot1 e 20µg de BSA e reduziu-se o volume para 12,25µl. Foram adicionados 47,5µl de tampão de hibridização 2x (10x SSC, 0,2% SDS), 23,75µl de formamida, 9,5µl de solução Denhardt's 50x e 10µg de esperma de salmão.

Antes do uso, as lâminas foram incubadas em tampão de pré-hibridização (5x SSC, 0,2% SDS, 1% BSA, 5x Denhardt's) por pelo menos 6h a 42°C. Foram então lavadas em água, centrifugadas para secagem e montadas na estação de hibridização.

As sondas foram aquecidas a 95°C por 5 minutos e mantidas a 70° até aplicação sobre a lâmina. Após a hibridização, as lâminas foram submetidas às seguintes lavagens: 5 min em 2x SSC, duas incubações de 10 min em 0,1x SSC-0,1%SDS e duas incubações de 2,5 min em 0,1x SSC. Seguiu-se uma centrifugação breve para secagem da lâmina.

OBS: Para preparação de cDNA marcado a partir de RNA amplificado, utiliza-se 3µg de aRNA e 3µg de oligonucleotídeos aleatórios (“random” hexanucleotídeos). Os demais procedimentos foram feitos conforme descrito para marcação a partir de RNA total.

Digitalização das imagens e análise da qualidade da hibridização

A imagem foi digitalizada pelo detector de câmera CCD (GeneTac2000, Genomic Solutions) e avaliada com auxílio do programa GeneTac Analyser (Genomic Solutions). Após a realização de mais da metade das hibridizações, nosso laboratório adquiriu um scanner a laser (ScanArray Express, Packard BioScience), o qual não foi utilizado pela dificuldade de se comparar dados gerados por scanners distintos.

Gel de agarose para análise da qualidade do cDNA fluorescente

Este protocolo foi implementado em nosso laboratório na fase final deste trabalho, logo apenas algumas amostras de cDNA marcado foram analisadas. Ele permitiu analisar, indiretamente, a qualidade do RNA amplificado a partir do qual foi preparado o cDNA marcado. O gel de agarose 1% em TAE filtrado foi preparado em uma cuba apropriada. Amostras de 1µl das reações de marcação com cada corante foram misturadas a 3µl de tampão e 1µl de água. O marcador foi preparado a partir de uma alíquota de 1µl da reação de digestão de Pel com *HindIII*, à qual foram adicionados 3µl de tampão e 1µl de Vista Green.

A eletroforese foi realizada a 60V por aproximadamente 40min. O gel foi lavado em água e a imagem adquirida por detector de câmera CCD.

Quantificação da intensidade dos spots

Parte das hibridizações foi inicialmente analisada com auxílio do programa ScanAlyze (Stanford University). Após uma análise inicial, optamos pelo programa Quantarray (version 3.0, Packard BioScience, BioChip Technologies LLC), o qual foi adotado para captura dos valores de intensidade dos spots de todos os arrays.

As quantificações das marcações de spot e background de todos os arrays foram feitas pelos métodos histograma e círculo fixo. Após uma avaliação, optamos pelos dados

obtidos 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.

Análise dos dados de microarray

As tabelas obtidas a partir do programa Quantarray foram encaminhadas aos nossos colaboradores do IME-USP para as análises estatísticas.

A filtragem dos dados foi feita da seguinte maneira: seleção dos spots com sinal maior que background, seleção (dentre esses) dos spots com sinal/background $> 1,05$.

Calculou-se então o background médio de cada canal de um array.

A correção dos valores dos spots foi feita subtraindo do sinal de cada spot o valor médio do background da respectiva lâmina, e a normalização dos dados foi feita usando o método de loess do projeto Bioconductor (<http://www.bioconductor.org>), com $\text{span}=0,4$ e grau 2.

Para a identificação dos genes diferencialmente expressos em cada uma das comparações entre os quatro tipos de tecidos analisados determinamos, para cada gene, os valores de t e $-\log_2 p$ (ver página 75), utilizando o t -test e assumindo variâncias diferentes para cada gene em cada tipo de amostra. Realizamos ainda o teste de Wilcoxon não pareado para cada gene, para efeito de comparação dos métodos t -test e Wilcoxon. Identificamos os 100 genes com maior valor de $-\log_2 p$, e selecionamos quais deles permaneceram entre os 100 de maior valor em quatro diferentes métodos de normalização e correção de background (sinal - média do background, sinal - background local, normalização por energia total, normalização por loess).

Comparamos ainda os genes diferenciais entre todas as comparações A x B que envolvem um mesmo tecido (no nosso caso, 3 comparações por tecido). Os genes que mantiveram um mesmo perfil de expressão nas três comparações podem ser considerados marcadores daquele tipo de tecido.

Identificamos ainda genes cuja expressão estava alterada em dois tipos de comparação distintas, AB x CD (sendo A, B, C e D tipos de tecido tireoideano). Para tal, nos baseamos nos valores de S ($-\log_2 p \times \text{sinal do fold change}$) para cada comparação (AB e CD). Foram selecionados genes que apresentaram valores de S maiores do que 10 ou menores do que -10 (p valor do gene inferior a 0,00098) em AxB e/ou CxD.

4.16. Amplificação do RNA

O protocolo de amplificação de mRNA baseou-se em Wang e colaboradores (2000) com modificações, e foi padronizado em nosso laboratório por Luciana Gomes (FAPESP 01/03780-0) e Alex Fiorini de Carvalho. Um esquema do processo é mostrado na figura 3, gentilmente cedida por Luciana Gomes.

A transcrição reversa de 3µg de RNAtotal foi feita na presença de oligodT-T7 e a síntese de cDNA dupla fita baseou-se na propriedade de troca de molde ou “template switching”. A transcrição *in vitro* foi feita pela T7 polimerase, a partir do promotor T7. Um segundo ciclo de amplificação, iniciado pela transcrição reversa do RNA amplificado na presença de oligonucleotídeos aleatórios foi utilizado para aumentar o rendimento da reação (figura 3).

Figura 3. Esquemas do método de amplificação de RNA, realizado em dois ciclos, adaptado a partir de Wang *et al.*, 2000 (elaborado e gentilmente cedido por Luciana Gomes)

5. RESULTADOS

Um aspecto curioso deste trabalho é que ele reflete a sucessão cronológica de técnicas utilizadas na área. A sequência dos métodos utilizados segue a aquisição de equipamentos e programas relacionados com microarray em nosso laboratório. Como todo projeto utilizando técnicas em desenvolvimento, um número considerável de experimentos foi realizado para padronização e os dados obtidos refletem as condições experimentais disponíveis nas diversas etapas do trabalho.

5.1. DDRT-PCR e cDNA array em membrana de nylon

Em um primeiro momento, comparamos o perfil molecular de amostras de tecido normal (N), bócio (B), carcinoma papilífero (CP), adenoma (A), carcinoma folicular invasivo (CFI) e carcinoma folicular minimamente invasivo (CFMI) através de differential display-RT-PCR (DDRT-PCR). Cada tecido foi representado por um pool de três amostras de RNA (três pacientes distintos) tratadas com DNase, com exceção de carcinomas foliculares, dos quais só havia uma amostra de cada tipo. Os tecidos citados foram comparados do seguinte modo: normal, carcinoma papilífero e bócio, e normal, adenoma, carcinoma folicular invasivo e minimamente invasivo. Foram utilizados quatro iniciadores ancorados em combinação com 7 iniciadores aleatórios, gerando um total de 24 combinações distintas, apresentadas na tabela III. A figura 4 mostra um experimento representativo de DDRT-PCR.

Figura 4. Autorradiograma mostrando um gel de DDRT-PCR representativo. As reações de RT-PCR foram elaboradas com pools de RNA de tecido normal (N), carcinoma papilífero (C) e bócio (B), e os iniciadores ancorado T11VC e aleatórios 13, 14 e 15. Alguns exemplos de bandas diferenciais são apontados pelas setas

Tabela III. Combinações de iniciadores utilizadas nos experimentos de DDRT-PCR

Iniciador ancorado	Iniciador aleatório
T11VA	13, 14, 15, Opal7, Opal8, Bauer1, Bauer2
T11VC	13, 14, 15, Opal7, Opal8, Bauer1, Bauer2
T11VG	13, 14, 15, Opal7, Opal8
T11VT	13, 14, 15, Opal7, Opal8

Foram seleccionadas 213 bandas diferenciais entre pelo menos dois tecidos. O número de bandas mais intensas em cada tecido é mostrado na tabela IV

Tabela IV. Caracterização dos clones isolados por DDRT-PCR

	Tecido no qual a banda de DDRT-PCR é mais intensa					
	Normal	Bócio	CP	Adenoma	CF I	CF MI
nº. bandas	39	54	65	21	18	16

As bandas purificadas e reamplificadas foram clonadas. Seleccionamos um total de 429 clones (aproximadamente 2 clones por banda), que foram fixados em triplicata em uma membrana de nylon. Estas membranas foram hibridizadas, em dois experimentos, com alvos radioativos preparados a partir de pools de 3 amostras de RNA de tecido normal, bócio e carcinoma papilífero. Os pools dos dois experimentos foram preparados a partir de RNAs de pacientes distintos. Em ambos os experimentos foram usadas amostras de RNA diferentes das utilizadas no experimento de DDRT-PCR nos quais os clones foram identificados. A figura 5 mostra uma membrana representativa do experimento 1, hibridizada com sonda proveniente de tecido normal. Foram identificados 41 e 36 clones com expressão aumentada pelo menos 2x em um dos tecidos, nos experimentos 1 e 2, respectivamente (tabela V).

Figura 5. Membrana de nylon contendo clones de DDRT-PCR, fixados em triplicata. Este array foi hibridizado com alvo radioativo preparado a partir de pool de 3 amostras de RNAs de tecido normal, totalizando 30µg.

Tabela V. Número de clones com expressão aumentada pelo menos duas vezes nos dois experimentos de cDNA array

	Clones com razão de expressão maior do que 2						
	B/CP	CP/B	N/B	B/N	N/CP	CP/N	Total
Exp 1	5	0	5	23	5	3	41
Exp 2	0	17	0	1	1	17	36

Quando comparamos os dados referentes aos dois experimentos encontramos apenas 2 clones com o mesmo padrão de expressão diferencial: clone 199, com razão de expressão CP/N 2,09 e 2,10 nos experimentos 1 e 2, respectivamente; e clone 234, com razão de expressão B/N 2,41 e 2,09 nos experimentos 1 e 2, respectivamente. Com base nestes dados, considerando significativas apenas razões de expressão superiores a 2, conclui-se que o nível de expressão do clone 199 permite distinguir carcinoma papilífero de tecido normal em ambos experimentos. A expressão do clone 234 distingue bócio e tecido normal em ambos os experimentos.

Dois clones, denominados 18 e 44, apresentaram bandas muito mais intensas em carcinoma do que em tecido normal nos géis de DDRT-PCR (ver figura 6), mas não tiveram razões superiores a dois nos experimentos de Microarray. O clone 18 teve razões CP/N 1,54 e 1,06 e o clone 44 teve razões CP/N 1,22 e 1,50. Ainda assim, devido a grande diferença entre as intensidades das bandas destes clones no DDRT-PCR, decidimos avaliar a expressão de 18 e 44 nas mesmas amostras de RNA utilizadas no DDRT-PCR por um método bastante quantitativo: Real-Time RT-PCR.

Figura 6. Detalhe das bandas 18 e 44 (setas) no autorradiograma do gel de DDRT-PCR. N, CP e B correspondem a pools de RNA de tecido normal, carcinoma papilífero e bócio, respectivamente

5.2. Identidade dos clones 18, 44, 199 e 234

Antes de “desenhar” os iniciadores para Real-Time, as sequências dos insertos dos quatro clones foram determinadas e analisadas utilizando o programa BLAST do NCBI (Altschul et al., 1990). O clone 18 corresponde a um fragmento de 175pb com 100% de identidade com a proteína 5 ligante do fator de crescimento do tipo insulina (*IGFBP-5*, posições 3480-3654, número de acesso Hs107169). O clone 44 (depositado no GenBank com número de acesso AF486818) corresponde a um fragmento de cDNA localizado em um clone BAC (acc AC002543), apresentando 99% de identidade com dois cDNAs previamente sequenciados (FLJ22131, número de acesso AK025784, e clone IMAGE 4452921, acc BC016592). A busca de homologia contra o genoma humano mostrou que o clone 44 mapeia no cromossomo 7. O clone 199 (número de acesso do GenBank AF486819) mapeia em uma sequência genômica humana localizada no cromossomo 4

(AC099397, human chromosome 4 working draft). No genoma, a sequência do clone 199 está dividida em duas metades, separadas por aproximadamente 16kb. A segunda metade (nucleotídeos 130 a 248) é homóloga a uma única EST depositada no GenBank (número de acesso T60530). O clone 234 corresponde a um trecho do gene de metalotioneína IG (MTIG, número de acesso T03910).

5.3. Real-Time RT-PCR

Devido à disponibilidade do aparelho ABI PRISM 7700 em nosso instituto e ao aspecto quantitativo e preciso da técnica de Real-Time RT-PCR, decidimos utilizá-la para comparar a expressão dos genes 18 e 44 em amostras de tecido normal, bócio e carcinoma papilífero.

Muitos genes são utilizados como controles internos em análises de expressão gênica. Cuidados especiais devem ser tomados na escolha de controles internos para as comparações entre tecidos tumorais e não-tumorais, já que diversas vias metabólicas estão alteradas nos primeiros (Brasileiro Fo et al., 1993), invalidando o uso de muitos dos genes tradicionalmente considerados constitutivos. Para os experimentos de Real-Time RT-PCR utilizamos como controle interno o gene ribossômico *18S*, já adotado por vários grupos. Visto que as sondas para os genes alvo e a sonda para o controle interno são marcadas com fluorocromos distintos (FAM e JOE, respectivamente), as duas amplificações foram feitas em um mesmo poço da placa de PCR, como descrito em material e métodos.

Dois métodos principais são utilizados para a quantificação da expressão em experimentos de Real-Time RT-PCR. Um deles é o da curva padrão, utilizado por nós e mencionado em material e métodos. O outro é o denominado Ct comparativo, que usa os dados do controle interno para corrigir os valores do gene alvo sem construir uma curva padrão. Um pré-requisito para o uso deste método é que as eficiências de amplificação dos genes alvo e controle sejam semelhantes (ABI PRISM 7700, user bulletin). Construímos um gráfico Ct gene alvo -Ct gene controle para diluições seriadas de RNA total de tireóide normal e verificamos que a amplificação do gene controle é muito mais eficiente do que as dos genes 18 e 44, o que invalida o uso do método de Ct comparativo neste trabalho.

Após a padronização das concentrações de iniciadores e sondas para ambos os genes e para o controle interno realizamos dois experimentos utilizando os mesmos RNAs nas reações de transcrição reversa distintas. Na figura 7 são apresentadas as curvas referentes a emissão pelo repórter JOE (utilizado a sonda do controle interno) e pelo repórter FAM

(utilizado as sondas dos genes 18 e 44). Visto que os iniciadores podem amplificar tanto DNA genômico quanto cDNA gerando produtos idênticos, preparamos controles negativos para todas as amostras, os quais continham todos os componentes para transcrição reversa e PCR exceto a transcriptase reversa (Superscript). Como observado na figura 7A, referente à amplificação do gene *18S*, os controles negativos começaram a emitir fluorescência acima do limiar (“threshold”, linha preta) a partir do ciclo 30, enquanto que as amostras tiveram amplificação significativa a partir de 12 ciclos. Na figura 7B é mostrado o gráfico referente à amplificação dos genes 18 e 44. As amostras emitiram fluorescência acima do limiar a partir de 22 ciclos enquanto que os controles a partir de 36 ciclos. Os valores de Ct dos controles negativos foram usados para o cálculo da quantidade de DNA genômico contaminante em cada amostra, a qual foi subtraída da quantidade de RNA da mesma amostra. O gene *18S* é um gene de múltiplas cópias e alta expressão, o que justifica sua expressão com menor número de ciclos do que 18 e 44, genes pouco expressos. Após normalização com base no controle interno (item 4.11) obtivemos as razões de expressão mostradas na tabela VI.

Figura 7. Gráficos de Real-Time RT-PCR nos quais a emissão (ΔRn) pelos corantes JOE (A) e FAM (B) está plotada em função do número de ciclos (cycle). Cada linha colorida corresponde a uma amostra contendo iniciadores e sonda para o controle interno, o gene ribossômico *18 S* (A), bem como para os genes *18* e *44* (B). As linhas de menor inclinação (amplificação tardia, ver seta em B) nos dois gráficos correspondem aos controles negativos de cada amostra.

Tabela VI. Razões de expressão para os genes 18 e 44 em dois experimentos de Real-Time PCR

Experimento	Razões de expressão entre CP, N e B			
	clone	CP/N	CP/B	B/N
1	18	2,37	26,80	0,08
	44	5,37	8,44	0,64
2	18	2,69	32,29	0,08
	44	5,56	9,39	0,59

Realizamos outro experimento de Real-Time RT-PCR com amostras distintas de CP, B e N, e não observamos expressão diferencial dos genes 18 e 44. Isso mostra que pacientes distintos apresentam perfis diferentes de expressão desses genes, indicando a necessidade de analisar um número maior de amostras para determinar o verdadeiro potencial discriminatório de 18 e 44.

Visto que os experimentos de Real-Time, com a quantidade de controles realizados, não permitem a análise de grande número de amostras por placa, optamos por outra metodologia: RT-PCR semi-quantitativo seguido de Southern blot.

5.4. RT-PCR semi-quantitativo

Em um primeiro momento, verificamos se a técnica de RT-PCR era capaz de refletir qualitativamente as diferenças entre as expressões dos genes 18 e 44 nas mesmas amostras avaliadas por Real-Time RT-PCR e cDNA array. Os dados referentes à amplificação dos dois genes por 14, 28 e 35 ciclos em duas amostras de carcinoma papilífero, tecido normal e bócio são mostrados na figura 8. A amplificação do transcrito de *GAPDH* é apresentada como comparação.

Figura 8. Análise da expressão dos genes 18 e 44 por RT-PCR. Os transcritos dos genes 18 e 44 e do gene controle *GAPDH* foram amplificados por 14, 28 e 35 em duas amostras (1 e 2) de carcinoma papilífero (CP), tecido normal (N) e bócio (B). As reações controle sem RNA são indicadas pelos 0.

Visto que as observações qualitativas da RT-PCR foram coerentes com os dados de Real-Time RT-PCR e cDNA array, decidimos comparar a expressão dos 4 genes selecionados pelo array em um maior número de amostras de cada tipo de tecido tireoideano por esse método.

Dez amostras de RNA de tecido normal (N), 10 de bócio (B), 10 de adenoma (A), 9 de carcinoma papilífero (CP) e 7 de carcinoma folicular (CF) foram submetidas a RT-PCR para amplificação dos transcritos dos genes 18, 44, 199 e 234 por 20 e 28 ciclos. As comparações foram feitas entre três tecidos: N, B, CP e N, A, CF. Após eletroforese e Southern blot selecionamos a condição de amplificação não saturante (20 ciclos) e quantificamos a intensidade de marcação de cada amostra para os 4 genes. Como controles internos para normalização amplificamos tanto o transcrito de *GAPDH* quanto do gene *TBP* (componente do complexo ligante de DNA TFIID) (Bieche et al., 1999).

Após a quantificação, os valores de intensidade das bandas referentes aos genes 18, 44, 199 e 234 foram divididos pelos os valores das bandas de *TBP* ou *GAPDH* amplificadas a partir das mesmas amostras. A dispersão dos dados para amostras de um mesmo tecido foi maior para a normalização com *GAPDH* (dados não mostrados) do que com *TBP*. Isso sugere que o gene *GAPDH*, considerado constitutivo em inúmeros estudos, não é eficaz como controle para a comparação normal-tumor de tireóide. De fato, ele parece ser superexpresso, juntamente com outros genes da via glicolítica, em tecidos neoplásicos, metabolicamente mais ativos do que os tecidos normais (van Erp et al., 1991); (Gress et al., 1996);(Lee et al., 2002).

Observamos grandes variações na expressão dos quatro genes, em especial dos genes 199 e 234, nas amostras de carcinoma papilífero (dados não mostrados). Ainda assim, a mediana dos valores dos genes 18, 199 e 44 para cada tecido foi condizente com o padrão de expressão dos genes no array, DDRT-PCR e Real-Time RT-PCR: superexpressão em carcinoma papilífero. O gene 234, identificado por array como mais expresso em bócio do que em tecido normal, teve maior expressão em bócio e tecido normal do que em a carcinoma papilífero. A expressão dos genes 18 e 44, mais eficientes para a diferenciação de N, B e CP, é apresentada em forma de “Box plot” na figura 9.

Figura 9. Box plot representando a expressão (normalizada pela expressão do gene *TBP*) dos genes 18 e 44 em amostras de bócio (G), tecido normal (N) e carcinoma papilífero (CP). A linha vermelha representa a mediana dos valores em cada tecido; cada retângulo corresponde a metade dos dados e as linhas representam os valores extremos na distância de 1,5 interquartil.

Quando realizamos a mesma análise para a comparação entre 9 amostras de N, 10 de A e 7 de CF, observamos que os 4 genes apresentam expressões bastante semelhantes entre A e CF, e que as maiores diferenças são observadas entre tecido normal e neoplásico (adenoma e carcinoma folicular). Os genes mais diferenciais entre os três tecidos foram o 44 e o 199, cujos dados são representados na figura 10.

Figura 10. Box plot representando a expressão (normalizada pela expressão do gene *TBP*) dos genes 44 e 199 em amostras de tecido normal (N), adenoma (A) e carcinoma folicular (CF). A linha vermelha representa a mediana dos valores para cada tecido; cada retângulo corresponde a metade dos dados e as linhas representam os valores extremos na distância de 1,5 interquartil. Os outliers são indicados por uma cruz.

A distribuição das amostras de acordo com a expressão dos dois genes mais diferenciais para cada comparação também pode ser visualizada nos scatter plots da figura 11. Observa-se que há um plano capaz de separar 7 das 9 amostras de carcinoma papilífero das amostras de tecido normal e de bócio (fig. 11A). Para a comparação N, A e CF (fig 11B) o melhor plano separa tecidos normais dos adenomas e 6 dos 7 carcinomas foliculares, mas há amostras muito próximas deste plano, mostrando que essa divisão é pouco precisa.

Figura 11. Scatter plot mostrando a distribuição das amostras de acordo com os logs das expressões (normalizadas pela expressão do gene *TBP*) dos genes 18 e 44 (A) e 44 e 199 (B). Em (A) são representadas amostras de tecido normal (■), bócio (●) e carcinoma papilífero (▲); em (B), amostras de tecido normal (■), adenoma (●) e carcinoma folicular (▲).

5.5. Análise de cluster

As amostras de N, B e CP foram submetidas à análise de cluster com base nos valores de expressão dos quatro genes, normalizados pelos de *TBP*. O cluster resultante do agrupamento utilizando os dados referentes aos genes 18 e 44 é mostrado na figura 12. Um teste de bootstrap mostrou que dentre 100.000 datasets simulados de forma aleatória, apenas 3 apresentaram melhor separação do que o cluster obtido, indicando a confiabilidade do mesmo. Clusters elaborados com base nas expressões dos quatro genes ou de outras combinações dos mesmos apresentaram pior capacidade classificatória, e por essa razão não são mostrados.

Os valores de P (P_w e P_{ks}) dos testes não-paramétricos para cada gene mostraram que o gene 18 é capaz de discriminar bócio de não-bócio: para B x N, $P_w=0.0019$ e $P_{ks}=0.0012$, para B x PC, $P_w=0.00097$ e $P_{ks}=0.00033$. O gene 44 permite a separação entre carcinoma papilífero e não-carcinoma: para B x CP, $P_w=0.0018$ e $P_{ks}=0.0026$, para N x CP, $P_w=0.0018$ e $P_{ks}=0.0026$.

A análise de cluster foi também aplicada aos dados de tecido normal, adenoma e carcinoma folicular, mas nenhum cluster de significado biológico foi obtido. Os testes não

paramétricos indicaram que o gene 44 distingue amostras normais de neoplásicas: para CF x N, $P_w=0.0015$ e $P_{ks}=0.011$, para A x N, $P_w=0.0015$ e $P_{ks}=0.0017$. O gene 199 separa com precisão amostras de adenoma das de tecido normal ($P_w=0.000087$ e $P_{ks}=0.00026$).

Figura 12. Cluster hierárquico das amostras de tecido normal (N), carcinoma papilífero (C) e bócio (G). A escala de cores de verde a vermelho representa a gradação da expressão dos genes 18 (coluna 1) e 44 (coluna 2). Amostras pertencentes aos cluster contendo principalmente N e G são mostradas em vermelho, e amostras do cluster contendo predominantemente C são mostradas em azul (3). Os dados não clusterizados são apresentados em (A), os dados clusterizados em (B) e o dendrograma obtido após a clusterização em (C).

5.6. Árvore de decisão

Utilizamos os mesmos dados das duas comparações (N, B, CP e N, A e CF) para a construção de dois classificadores baseados no método árvore de decisão. A precisão do classificador N+B x CP, testada pelo método “leave-one-out”, foi de 80 e 89% para classificações de amostras não-malignas e malignas, respectivamente. A precisão do classificador N x A+CF foi de 78 e 88% para classificação de amostras normais e neoplásicas, respectivamente.

5.7. Os experimentos de microarray

5.7.1. O uso de amostra de referência e a inversão de corantes

A inclusão de um RNA de referência nos experimentos de array é uma questão ainda controversa, conforme apresentado na introdução (3.2). Em nossos experimentos, comparamos quatro tipos de tecido tireoideano (tecido normal, bócio, adenoma e carcinoma papilífero), e optamos pelo uso do RNA de referência para normalização interna de cada microarray e comparação entre os dados das diversas amostras. O RNA referência utilizado foi exatamente o mesmo para todas as hibridizações, e consistiu de um pool de quatro amostras grandes de tecido tireoideano normal. Cada amostra foi testada individualmente por hibridização em array, bem como o pool final.

Cada lâmina foi hibridizada simultaneamente com cDNA obtido a partir do RNA de referência, marcado com um dos corantes (Cy3 ou Cy5), e com cDNA obtido a partir do RNA de uma das amostras, marcado com o outro corante. As amostras foram hibridizadas em duplicata, com inversão de corantes, e as imagens foram adquiridas em detector CCD. A figura 13 mostra duas imagens de trechos de lâminas hibridizadas com uma mesma amostra, segundo essa estratégia.

Figura 13. Trechos de lâminas 3.8K hibridizadas com RNA de uma amostra de tecido normal (N) e RNA de referência (ref), mostrando o efeito da inversão dos corantes.

(A) RNA N marcado com Cy3, RNA ref com Cy5. (B) N Cy5, ref Cy3.

Nosso objetivo inicial era ter hibridizações de 10 amostras de cada tipo de tecido (N, B, A, CP, CF) em duplicata com inversão de corantes. Infelizmente, as amostras de carcinoma folicular eram poucas e pequenas, e obtivemos apenas 3 hibridizações em duplicata. Por essa razão, estas não foram incluídas nessa etapa do estudo.

Diversas hibridizações apresentaram marcação fraca e/ ou background alto, além de artefatos devido a partículas de poeira, e foram descartadas. Ao final, após cerca de 150 hibridizações, capturamos os resultados de 86 lâminas, correspondentes a:

Tecido normal (N): 8 amostras

Bócio (B): 10 amostras

Adenoma (A): 10 amostras

Carcinoma Papilífero (CP): 10 amostras

5.7.2. Normalização das intensidades dos dois canais

A fim de escolher um método adequado para a normalização das marcações nos dois canais (Cy3 e Cy5) de uma mesma lâmina, analisamos as variações destas intensidades de acordo com alguns parâmetros. Tomamos como exemplo uma hibridização de uma

amostra de b6cio, na qual o RNA de interesse foi marcado com Cy3 e o refer6ncia com Cy5, cujos dados s6o analisados nas figuras 14, 15 e 16.

A figura 14 mostra a distribui66o da marca666o do background e dos spots no array, dividido em subarrays. Na figura 15 s6o apresentadas as distribui666es dos valores de M (log das raz66es entre as intensidades nos dois canais) por subarray antes e ap6s a normaliza666o. A figura 16 mostra o chamado “MA plot”, um gr6fico que relaciona, para cada spot, o valor de M com o aumento de intensidade A (log da raiz quadrada do produto da intensidade dos dois canais).

Figura 14. Esquemas mostrando a marca666o de Cy3 (verde) e Cy5 (vermelho) no background e nos spots nos 32 blocos (subarrays) que comp6em o array de 3.8K. (A) Intensidade absoluta de marca666o de Cy3 e Cy5 no background. (B) Valores de M (log Cy3/Cy5 nos spots) antes (esq) e ap6s (dir) normaliza666o por subarray.

Figura 15. Distribuição dos spots de cada subarray (barras coloridas, eixo x) segundo o valor de M (log Cy3/Cy5 nos spots, eixo y). (A) Valores não normalizados. (B) Valores normalizados por regressão não linear por subarray.

Figura 16. MA plot: distribuição dos spots segundo os valores de M (log Cy3/Cy5, eixo y) e A (log da raiz quadrada do produto Cy3 x Cy5, eixo x). (A) Valores não normalizados. (B) Valores normalizados por regressão não linear por subarray (curvas coloridas). As curvas coloridas representam as regressões por subarray.

Analisando as figuras apresentadas observa-se que o background, embora forte, é relativamente homogêneo, com exceção de artefatos (fig.14). O sinal, embora fraco, também é praticamente homogêneo ao longo da lâmina (fig.15), e parece não sofrer interferência da localização em subarrays. De fato, a correção das regressões por subarray não alterou significativamente o padrão dos dados, e não foi adotada (fig.16). Neste trabalho, optamos por usar o menor número possível de correções para não afetar a robustez das análises.

5.7.3. “Validação” da quantificação pelo programa Quantarray

Uma das críticas ao uso de métodos de quantificação de ajuste manual, como o Quantarray, é a baixa reprodutibilidade entre duas quantificações de um mesmo experimento devido à subjetividade do ajuste. Para testar a variabilidade entre ajustes de um mesmo usuário, quantificamos os dados de uma mesma lâmina duas vezes, em dias distintos. Os resultados são mostrados na figura 17.

Figura 17. Gráficos mostrando a reprodutibilidade entre duas quantificações independentes de uma mesma hibridização. (A) Razão entre os dois canais para cada spot, nas duas quantificações (original e requantificado). (B) Razão entre as razões obtidas nas duas quantificações. O= original, R= requantificado

Comparamos ainda os dados da quantificação de quatro lâminas (uma hibridização para cada tipo de amostra- A9, B10, CP10, N5) digitalizadas por scanner a laser, pelos programa Quantarray e Bioinfo, de localização automática (ver item 3.2). Os dados são apresentados na figura 18.

Figura 18. Gráficos comparando as quantificações das hibridizações de quatro amostras (A9, B10, CP10, N5) pelos programas Quantarray e Bioinfo. (A) Razão entre os dois canais para cada spot, pelos dois programas. (B) Razão das razões dos canais entre os dois programas. Os valores para cada spot são médias das duplicatas (inversão de corantes).

5.7.4. Comparações A x B, ranking dos valores de p

A normalização das marcações nos dois canais (Cy3 e Cy5) foi feita utilizando o método loess de regressão não linear local e subtraiu-se a intensidade do background médio do array dos valores de cada spot. Para cada dois tipos de amostras foram calculados os valores de p e t de cada spot. Os valores de p derivam do teste paramétrico t-test e do teste não paramétrico de Wilcoxon. O parâmetro t corresponde à diferença entre o logaritmo da intensidade de um gene em uma amostra de interesse e na amostra referência, normalizada pelos desvios-padrão deste gene nos dois tipos de amostra. Este valor mostra quantos desvios-padrão separam as médias das intensidades nos dois tipos de amostra, incluindo informações não apenas do valor de expressão de um gene, mas também de sua variabilidade dentro do conjunto de amostras de um mesmo tipo. A fórmula para o cálculo de t é:

$$t^k = \frac{\overline{A1^k} - \overline{A2^k}}{\sqrt{\frac{S_{A1^k}^2}{n} + \frac{S_{A2^k}^2}{n}}}$$

onde $\overline{A1^k}$ e $\overline{A2^k}$ indicam a expressão média do gene k na amostra tipo A1 e no tipo A2, respectivamente, calculados como $\overline{A^k} = \frac{1}{n} \sum_{j=1}^n N_j^k$ para um conjunto de n amostras de cada tipo. S_{A1^k} e S_{A2^k} indicam os desvios-padrão dos dois tipos de amostra.

Calculou-se também o parâmetro $\log_2 p$. Conforme mostrado na figura 19A, que compara amostras de tecido normal e bócio, os valores de p calculados pelos métodos t-test e Wilcoxon tem grande correlação, devido à concordância entre os valores de média e mediana de cada spot (figura 19B). Optamos pelo uso dos valores de p dados pelo t-test.

Figura 19. Comparação entre os testes t e Wilcoxon. (A) Correlação entre os valores de $-\log_2 p$ dos testes t (eixo x) e Wilcoxon (eixo y) para cada spot, na comparação entre médias (eixo x) e log da razão das medianas (eixo y) de cada spot na comparação entre tecido normal e bócio.

Observamos grande concordância entre os valores de t e p de cada gene, e utilizamos os valores de p para seleção dos genes de maior interesse.

Para cada uma das seis comparações A x B possíveis entre os quatro tipos de tecido estudados (N, B, A, CP), selecionamos os 100 genes de maior valor S ($-\log_2 p$ x sinal do fold change), e portanto menor p (dados não mostrados).

Fizemos ainda normalização por energia total e subtração de background local, o que nos levou a um total de 4 combinações de análise: normalização loess-background médio, loess-background local, energia total- background médio, energia total- background local. Selecionamos os 100 clones de maior valor S para cada uma destas metodologias. Identificamos então os clones presentes nas quatro listas de cada comparação, ou seja, aqueles que permaneceram entre os 100 mais diferenciais independentemente do método de normalização e correção de background. O número desses clones para cada comparação é mostrado na tabela VII.

Tabela VII. Número de clones diferenciais nas quatro análises (duas normalizações e correções de background), para cada comparação A x B.

Comparação A x B	Número de clones diferenciais
N x A	47
N x B	43
N x CP	47
A x B	28
A x CP	57
B x CP	35

Tabela VIII. Número de clones apresentando expressão aumentada em cada comparação AxB. Os genes mostrados são mais expressos nos tecidos em vermelho em relação aos em verde.

tecido	N	A	B	C
N	-	19	0	12
A	28	-	4	30
B	43	24	-	26
C	35	27	9	-

As tabelas IX a XIV mostram os clones identificados em cada comparação, juntamente com os valores de $\log_2 p$. Os valores de t seguem o mesmo padrão (ranking) dos valores de $\log_2 p$, e por essa razão não são mostrados. Observamos 160 ORESTES distintos, já que muitas das comparações identificaram os mesmos clones. A maior parte desses clones foi sequenciada e a sequência resultante foi comparada com as armazenadas nos bancos de dados Hs e Nr.

Tabela IX. Comparação NxA, mostrando a identidade dos clones diferenciais (tabs VII e VIII), juntamente com o valor de $\log_2 p$. Os clones mais expressos em N são mostrados em vermelho, os mais expressos em A em verde. Clones no match não apresentaram identidade significativa com nenhuma sequência dos bancos Hs e Nr. Clones ns não foram sequenciados.

Genes ou ORESTES		log ₂ p
RPS25 ribosomal protein S25	Hs.429729	22.042
RPL15 ribosomal protein L15	Hs.74267	19.945
LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357	18.894
IL3-CT0220-051199-027-F11- ns		16.969
ANXA2 annexin A2	Hs.217493	15.856
RPS4X ribosomal protein S4, X-linked	Hs.430084	15.573
RC5-HT0582-110400-013-G04- ns		13.732
MR2-ST0129-111199-014-A10- ns		12.716
ESTs	Hs.224732	12.623
RPL10A	Hs.425293	12.884
MR2-ST0129-191199-003-G12- no match		12.311
DHCR24 24-dehydrocholesterol reductase	Hs.75616	12.570
translation elongation factor eEF-1 gamma chain	Hs.381229	12.248
RC5-HT0582-110400-013-E01- ns		11.775
IL0-CT0031-221099-113-E08- ns		11.679
IL3-CT0219-281099-024-B11- ns		10.943
RPL36AL ribosomal protein L36a-like	Hs.419465	10.597
RNF13 ring finger protein 13	Hs.6900	10.169
CALM2 calmodulin 2 (phosphorylase kinase, delta)	Hs.425808	9.822
EFHU1 translation elongation factor eEF-1 alpha-1 chain	Hs.369597	9.488
CDKN1B cyclin-dependent kinase inhibitor 1B (p27, Kip1)	Hs.238990	9.052
CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867	8.880
RPL10 ribosomal protein L10	Hs.412900	8.023
Homo sapiens, clone IMAGE:4074138, mRNA	Hs.30089	7.881
RPL36 ribosomal protein L36	Hs.433411	7.655
XBP1 X-box binding protein 1	Hs.149923	7.654
RL26 60S ribosomal protein L26	Hs.356175	6.382
UBA52 ubiquitin A-52 residue ribosomal prot fusion prod. 1	Hs.5308	6.371
RC0-BT0561-210100-012-B12- ns		12.639
RC0-BT0561-210100-032-A06- ns		12.256
NADH-ubiquinone oxidoreductase chain 5	Hs.414669	10.496
RC5-ST0171-201099-012-C02- ns		9.787
CTSB cathepsin B	Hs.297939	9.709
NADH-ubiquinone oxidoreductase chain 5	Hs.400458	9.549
IL3-ST0023-300799-002-F02- ns		9.589
Human HepG2 partial cDNA, clone hmd3c03m5.	Hs.434261	9.439
NADH-ubiquinone oxidoreductase chain 5	Hs.414669	9.032
Homo sapiens chromosome 16 clone RP11-420O6	AC040165	8.390
EIF2B2 eukaryotic translation initiation factor 2B, subunit 2β	Hs.170001	8.173
Homo sapiens haplotype E13K mitochondrion	AY195765	8.393
G3P2 Glyceraldehyde 3-phosphate dehydrogenase	Hs.356772	8.379
NADH-ubiquinone oxidoreductase chain 5	Hs.414669	8.311
Homo sapiens cDNA clone IMAGE:3510006 5'	Hs.279327	7.770
RC3-BT0568-301299-011-H07- ns		7.665
NADH-ubiquinone oxidoreductase chain 5	Hs.414669	7.463
Homo sapiens cDNA clone IMAGE:3885707 3'	Hs.316269	6.952
DNAJB12 DnaJ (Hsp40) homolog, subfamily B, member 12	Hs.7960	6.457

Tabela X. Comparação NxB, mostrando a identidade dos clones diferenciais (tabs VII e VIII), juntamente com o valor de $\log_2 p$. Os clones mais expressos em N são mostrados em vermelho. Clones no match não apresentaram identidade significativa com nenhuma sequência dos bancos Hs e Nr. Clones ns não foram sequenciados.

Gene ou ORESTES		$\log_2 p$
EEF1A1 eukaryotic translation elongation factor 1 alpha 1	Hs.422118	20.485
Homo sapiens cDNA FLJ34095 fis	Hs.290475	17.629
RPS25 ribosomal protein S25	Hs.429729	16.190
RPS4X ribosomal protein S4, X-linked	Hs.430084	15.218
RPL15 ribosomal protein L15	Hs.74267	14.538
LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357	13.986
CALM2 calmodulin 2 (phosphorylase kinase, delta)	Hs.425808	13.785
Homo sapiens prothymosin α (gene sequence 28) (PTMA)	NM_002823	12.832
ZNF9 zinc finger protein 9	Hs.2110	12.726
DHCR24 24-dehydrocholesterol reductase	Hs.75616	12.024
IL3-CT0220-051199-027-F11- ns		12.136
CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867	11.899
APP amyloid beta (A4) precursor protein (nexin-II)	Hs.177486	12.069
MGC4836 similar to hypothetical protein L.I.H 3 region	Hs.343206	11.782
RPL4 ribosomal protein L4	Hs.286	11.225
PDCD4 programmed cell death 4 (neoplastic transformation inhibitor)	Hs.326248	11.371
CALM1 calmodulin 1 (phosphorylase kinase, delta)	Hs.282410	11.011
RPL36AL ribosomal protein L36a-like	Hs.419465	9.612
IL3-CT0219-281099-024-B11- ns		10.173
A Human DNA sequence from clone CTA-215D11	HS215D11	10.033
IPW imprinted in Prader-Willi syndrome	Hs.5022	9.816
ubiquitin C	Hs.183704	9.615
SFRS5 splicing factor, arginine/serine-rich 5	Hs.166975	9.628
CDKN1B cyclin-dependent kinase inhibitor 1B (p27, Kip1)	Hs.238990	9.526
NSEP1 nuclease sensitive element binding protein 1	Hs.74497	9.211
MR2-ST0131-111199-016-G07- no match		9.243
QV4-BT0536-271299-059-E05- ns		8.956
NTRK2 neurotrophic tyrosine kinase, receptor, type 2	Hs.47860	8.830
Homo sapiens, clone IMAGE:4696119, mRNA	Hs.109390	8.897
PPIB peptidylprolyl isomerase B (cyclophilin B)	Hs.394389	8.918
IL0-BT0168-170899-100-F01- ns		8.757
NIT1 nitrilase 1	Hs.146406	8.699
ATP1A1 ATPase, Na ⁺ /K ⁺ transporting, α 1	Hs.76549	8.241
MR2-ST0129-111199-014-A10- ns		8.258
Homo sapiens, clone IMAGE:4074138	Hs.30089	7.984
SFRS5 splicing factor, arginine/serine-rich 5	Hs.166975	7.766
EIF3S6IP eukaryotic translation initiation factor 3, subunit 6 interacting prot	Hs.119503	7.618
EFHU1 translation elongation factor eEF-1 alpha-1 chain	Hs.369597	7.129
ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549	6.701
RC3-BT0568-301299-011-B09- no match		6.399

ESTs, Moderately similar to ribosomal protein L4	Hs.381104	6.286
ZNF255 zinc finger protein 255	Hs.181696	5.898
PABPC1 poly(A) binding protein, cytoplasmic 1	Hs.172182	5.549

Tabela XI. Comparação NxC, mostrando a identidade dos clones diferenciais (tabs VII e VIII), juntamente com o valor de $\log_2 p$. Os clones mais expressos em N são mostrados em vermelho, os mais expressos em C em verde. Clones no match não apresentaram identidade significativa com nenhuma sequência dos bancos Hs e Nr. Clones ns não foram sequenciados.

Gene ou ORESTES		$\log_2 p$
ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549	28.332
CM0-BT0304-111199-093-C11- ns		24.254
CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867	23.413
RC5-HT0582-110400-013-E01- ns		21.688
ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549	18.049
ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549	15.422
ubiquitin C	Hs.183704	16.839
DHCR24 24-dehydrocholesterol reductase	Hs.75616	16.694
CLTC clathrin, heavy polypeptide (Hc)	Hs.178710	14.059
IPW imprinted in Prader-Willi syndrome	Hs.5022	13.113
RNF13 ring finger protein 13	Hs.6900	12.301
Homo sapiens prothymosin α (gene sequence 28) (PTMA)	NM_002823	11.928
GRP58 glucose regulated protein, 58kDa	Hs.13751	11.596
MR2-ST0131-111199-016-G07- no match		10.984
GRP58 glucose regulated protein, 58kDa	Hs.13751	10.523
A Human DNA sequence from clone CTA-215D11	HS215D11	10.637
NSEPI nuclease sensitive element binding protein 1	Hs.74497	10.474
XBP1 X-box binding protein 1	Hs.149923	10.540
RC4-ST0185-191099-012-A07	AC096546	10.257
LOC51260 hypothetical protein LOC51260	Hs.128764	10.007
RC5-HT0582-110400-013-A03- ns		9.943
CM0-BT0615-050200-174-E10- ns		9.635
GDI2 GDP dissociation inhibitor 2	Hs.56845	9.403
RPL15 ribosomal protein L15	Hs.74267	9.228
RC5-HT0580-050400-021-C09- ns		9.170
EIF3S6IP eukaryotic translation initiation factor 3, subunit 6 interacting prot	Hs.119503	8.745
MAGED1 melanoma antigen, family D, 1	Hs.398474	8.803
CDKN1B cyclin-dependent kinase inhibitor 1B (p27, Kip1)	Hs.238990	8.142
hypothetical protein FLJ13213	Hs.331328	8.128
RPL7 ribosomal protein L7	Hs.421257	8.034
Homo sapiens, clone IMAGE:4074138, mRNA	Hs.30089	7.744
T14766 hypothetical protein DKFZp564K247.1	Hs.355732	7.502
QV3-BT0537-221299-047-E07- ns		7.448
SDC1 syndecan 1	Hs.171595	6.887
MAGED2 melanoma antigen, family D, 2	Hs.4943	6.534

S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004	20.132
CTSB cathepsin B	Hs.297939	13.502
HLA-DQB1 MHC class II, DQ beta 1	Hs.73931	12.523
SERPINA1-alpha-1 antiproteinase, antitrypsin	Hs.297681	11.264
S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004	10.718
HLA-C MHC class I, C	Hs.277477	9.852
MAD1L1 MAD1 mitotic arrest d	Hs.7345	7.721
CM4-ST0227-111199-041-D12- ns		7.227
SDC1 syndecan 1	Hs.82109	6.989
SDC4 syndecan 4 (amphiglycan, ryudocan)	Hs.252189	6.793
S100A2 S100 calcium binding protein A2	Hs.38991	6.167
EST	Hs.401918	6.018

Tabela XII. Comparação AxB, mostrando a identidade dos clones diferenciais (tabs VII e VIII), juntamente com o valor de $\log_2 p$. Os clones mais expressos em A são mostrados em vermelho, os mais expressos em B em verde. Clones ns não foram sequenciados.

Gene ou ORESTES		$\log_2 p$
PDCD4 programmed cell death 4 (neoplastic transformation inhibitor)	Hs.326248	15.318
SFRS5 splicing factor, arginine/serine-rich 5	Hs.166975	8.649
Homo sapiens haplotype E13K mitochondrion	AY195765	8.546
PKD2L2 polycystic kidney disease 2-like 2	Hs.272418	8.130
GOLGIN-67 golgin-67	Hs.182982	7.318
RC3-BT0046-310899-002-B06- ns		7.200
IL0-BT0168-170899-100-F01- ns		7.110
CTSB cathepsin B	Hs.297939	6.363
A2M alpha-2-macroglobulin	Hs.74561	6.053
PPIA peptidylprolyl isomerase A (cyclophilin A)	Hs.401787	6.006
PUM1 pumilio homolog 1	Hs.153834	5.961
IL0-ST0162-280999-123-B08- ns		5.686
ESTs	Hs.253677	5.269
CLTC clathrin, heavy polypeptide (Hc)	Hs.178710	4.942
CALM1 calmodulin 1 (phosphorylase kinase, delta)	Hs.282410	4.693
HSD17B4 hydroxysteroid (17-beta) dehydrogenase 4	Hs.75441	4.605
RC3-BT0046-310899-002-G12- ns		4.237
CM0-ST0078-140999-055-A09- ns		4.205
GOLGA4 golgi autoantigen, golgin subfamily a, 4	Hs.183773	4.113
RC3-BT0046-310899-002-B04- ns		4.140
EIF3S6IP eukaryotic translation initiation factor 3, subunit 6 interacting prot	Hs.119503	4.059
RC3-BT0568-301299-011-H07- ns		3.621
RC0-BT0561-210100-032-A06- ns		3.402
SBF1 SET binding factor 1	Hs.112049	3.339
Homo sapiens chromosome 16 clone CTC-527H23	AC023813	7.970
IL3-ST0142-131099-018-E09- ns		5.853
QV3-BT0511-081299-035-A06- ns		4.716
QV0-CT0224-091299-070-G10- ns		4.662

Tabela XIII. Comparação Ax-C, mostrando a identidade dos clones diferenciais (tabs VII e VIII), juntamente com o valor de $\log_2 p$. Os clones mais expressos em A são mostrados em vermelho, os mais expressos em C em verde. Clones no match não apresentaram identidade significativa com nenhuma sequência dos bancos Hs e Nr. Clones ns não foram sequenciados.

Gene ou ORESTES		$\log_2 p$
GRP58 glucose regulated protein, 58kDa	Hs.13751	9.523
ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549	8.945
NQO1 NAD(P)H dehydrogenase, quinone 1	Hs.406515	8.750
IPW imprinted in Prader-Willi syndrome	Hs.5022	8.171
CM0-BT0304-111199-093-C11- ns		7.950
HDAC1 histone deacetylase 1	Hs.88556	7.547
CM0-BT0615-050200-174-E10- ns		7.470
RC5-HT0582-110400-013-E01- ns		7.121
PDCD4 programmed cell death 4 (neoplastic transf. inhibitor)	Hs.326248	6.715
GDI2 GDP dissociation inhibitor 2	Hs.56845	6.515
PKD2L2 polycystic kidney disease 2-like 2	Hs.272418	6.382
CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867	6.300
RC3-BT0568-301299-011-H07- ns		6.218
IL0-BT0168-170899-100-F01- ns		5.873
HSPC194 hypothetical protein HSPC194	Hs.30376	5.781
GRP58 glucose regulated protein, 58kDa	Hs.13751	5.707
ATP1A1 ATPase, Na ⁺ /K ⁺ transporting, alpha 1 polypeptide	Hs.76549	5.571
IL3-ST0018-110899-003-A07- ns		5.570
SNTA1 syntrophin, alpha 1	Hs.31121	5.573
ATP1A1 ATPase, Na ⁺ /K ⁺ transporting, α 1	Hs.76549	5.403
LOC51260 hypothetical protein LOC51260	Hs.128764	5.296
ESTs, cytochrome c oxidase VIII H.sapiens	Hs.374980	5.273
HSPC182 HSPC182 protein	Hs.30026	4.826
HDAC1 histone deacetylase 1	Hs.88556	4.595
MR1-ST0206-280100-009-A11- ns		4.403
UQCRH ubiquinol-cytochrome c reductase hinge protein	Hs.73818	4.320
XBPI X-box binding protein 1	Hs.149923	4.214
RPS4X ribosomal protein S4, X-linked	Hs.430084	13.336
RPL36 ribosomal protein L36	Hs.433411	12.758
HLA-C MHC class I, C	Hs.277477	12.556
HLA-DQB1MHC class II, DQ beta 1	Hs.73931	12.137
SDC4 syndecan 4 (amphiglycan, ryudocan)	Hs.252189	11.259
RPS8 ribosomal protein S8	Hs.399720	10.822
SERPINA1 - alpha-1 antiproteinase, antitrypsin	Hs.297681	9.507
IGHG3 immunoglobulin heavy constant gamma 3	Hs.413826	9.048
S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004	8.238
MR2-ST0129-191199-003-G12- no match		8.026
CM0-ST0285-061299-118-B02- ns		8.111

HLA-A major histocompatibility complex, class I, A	Hs.181244	8.043
MAD1L1 MAD1 mitotic arrest d	Hs.7345	8.029
MR3-ST0290-290100-026-B08- ns		7.676
DKFZP761L0424 hypothetical protein DKFZp761L0424	Hs.149377	7.617
RPS2 ribosomal protein S2	Hs.356360	6.815
SAT spermidine/spermine N1-acetyltransferase	Hs.28491	6.511
MR3-ST0191-140200-013-B12- no match		6.424
CM4-ST0227-111199-041-D12- ns		6.123
ARCN1 archain 1	Hs.33642	5.921
UQCRC1 ubiquinol-cytochrome c reductase	Hs.3712	5.803
DIM1 similar to S. pombe dim1+	Hs.433683	5.748
ATIC 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase	Hs.90280	5.419
RAB27A member RAS oncogene family	Hs.50477	5.366
CLDN7 claudin 7	Hs.278562	5.232
MR2-ST0129-111199-014-A10- ns		5.196
RPL10 ribosomal protein L10	Hs.412900	4.677
MR2-ST0129-191099-011-G08- ns		4.379
IL0-ST0150-051099-121-F11- ns		4.164
hypothetical protein FLJ12649	Hs.432506	4.055

Tabela XIV. Comparação BxC, mostrando a identidade dos clones diferenciais (tabs VII e VIII), juntamente com o valor de $\log_2 p$. Os clones mais expressos em B são mostrados em vermelho, os mais expressos em C em verde. Clones ns não foram sequenciados.

Gene ou ORESTES		$\log_2 p$
ATP1A1 ATPase, Na ⁺ /K ⁺ transporting, α 1	Hs.76549	18.538
CM0-BT0304-111199-093-C11-ns		16.789
CM0-BT0615-050200-174-E10- ns		9.772
RC5-HT0582-110400-013-E01- ns		9.559
RC5-HT0582-110400-013-A03- ns		6.108
ACADM acyl-Coenzyme A dehydrogenase	Hs.79158	6.068
GDI2 GDP dissociation inhibitor 2	Hs.56845	5.788
UQCRH ubiquinol-cytochrome c reductase hinge protein	Hs.73818	5.719
MR1-ST0206-280100-009-A11- ns		3.974
S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004	14.597
HLA-DQB1 MHC class II, DQ beta 1	Hs.73931	12.258
RPS4X ribosomal protein S4, X-linked	Hs.430084	11.810
HLA-C MHC class I, C	Hs.277477	11.656
RC3-BT0046-310899-002-G12- ns		10.270
CTSB cathepsin B	Hs.297939	10.198
S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004	9.645
SDC4 syndecan 4	Hs.252189	8.573
RPS8 ribosomal protein S8	Hs.399720	7.703
SERPINA1 alpha-1 antiproteinase, antitrypsin	Hs.297681	7.435
RPL36 ribosomal protein L36	Hs.433411	6.884
GOLGIN-67 golgin-67	Hs.182982	6.703
TSC22 TGF beta-stimulated protein TSC-22	Hs.114360	6.584
MR0-ST0032-280999-002-B10- ns		6.564
BETA2 Human MEN1 region clone epsilon/beta	Hs.240443	6.336
APLP2 amyloid beta (A4) precursor-like protein 2	Hs.279518	6.174
HLA-A MHC class I, A	Hs.181244	6.145
CM0-ST0285-061299-118-B02- ns		5.887
EGFR-RS epidermal growth factor receptor	Hs.57988	5.529
RC3-BT0046-310899-002-B06- ns		5.461
EEF1A1 eukaryotic translation elongation factor 1 α 1	Hs.422118	5.125
MR3-ST0290-290100-026-B08- ns		5.035
Homo sapiens cDNA FLJ34095 fis	Hs.290475	4.479
CTSB cathepsin B	Hs.297939	4.532
NIT1 nitrilase 1	Hs.146406	4.462
RC3-BT0046-310899-002-B04- ns		3.743

Nota-se que alguns clones correspondem a um mesmo gene, com entradas iguais ou distintas no banco Hs. Isso será discutido posteriormente.

Identificamos então os genes comuns entre as comparações que envolviam um mesmo tipo de tecido, por exemplo A x N, A x B, A x CP. Os resultados são apresentados em forma de diagramas de Venn na figura 20.

Figura 20. Diagramas de Venn mostrando os genes diferenciais (tabela VI) entre todas as comparações envolvendo um mesmo tipo de tecido.

5.7.5. Comparações AB x CD

Realizamos também um outro tipo de análise, em que os genes diferenciais não foram selecionados por meio de ranking de p valores mas sim pelos valores absolutos de S ($-\log_{10} p \times \text{sinal do fold change}$). Adotamos a normalização por loess e correção por subtração do background médio. Estabelecemos valores de corte para S: S maior do que 10 ou menor do que -10, o que corresponde a valores de p inferiores a 0,00098 para cada gene. Todos os genes foram plotados em um gráfico segundo os valores de S nas comparações entre A e B e entre C e D. A figura 21 mostra como exemplo o gráfico BN x BC.

Figura 21. Comparação entre B x N (eixo x) e B x C (eixo y). Os genes são plotados de acordo com os valores de S. As cores distinguem genes de regiões diferentes (numeradas de 1 a 8).

Dentre as 15 comparações possíveis, analisamos em detalhe as 12 que envolviam um tecido em comum: AB x AC, AN x AC, AN x AB, BC x AC, BN x BC, NA x NC, NC x AC, NC x BC, NB x AB, NB x NA, NB x NC e BA x BC. Identificamos genes alterados em uma ou ambas comparações. A tabela XV resume os dados.

Tabela XV. Número de genes em cada região (ver fig. 21), por comparação

COMPARAÇÃO	REGIÃO					
	1	2	4	5	6	8
AB x AC	-	1	6	-	-	-
AN x AC	-	4	4	2	21	-
AN x AB	-	4	-	-	23	1
BC x AC	-	3	3	3	4	-
BN x BC	-	1	6	1	19	3
NA x NC	4	19	6	-	4	18
NC x AC	-	22	5	1	5	-
NC x BC	3	19	4	3	3	-
NB x AB	1	19	-	-	1	-
NB x NA	11	9	4	-	1	-
NB x NC	6	14	6	-	1	16
BA x BC	-	-	7	-	1	3

Praticamente todos os genes mencionados na tabela acima correspondem aos identificados pelo ranking dos valores de $-\log_2 p$ nas comparações A x B. Apenas quatro deles não foram identificados pelo ranking, e foram sequenciados.

As tabelas XVI a XXVII mostram a identidade dos genes encontrados em cada comparação, por região.

Tabela XVI. Comparação ABxAC, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV)

Região	Gene ou ORESTES	
2	PDCD4 programmed cell death 4 (neoplastic transformation inhibitor)	Hs.326248
4	HLA-DQB1 MHC class II, DQ beta 1	Hs.73931
	RPL36 ribosomal protein L36	Hs.433411
	RPS4X ribosomal protein S4, X-linked	Hs.430084
	HLA-C MHC class I, C	Hs.277477
	RPS8 ribosomal protein S8	Hs.399720
	SDC4 syndecan 4 (amphiglycan, ryudocan)	Hs.252189

Tabela XVII. Comparação ANxAC, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV). Clones no match não apresentaram identidade significativa com nenhuma sequência dos bancos Hs e Nr. Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
2	NU5MNADH-ubiquinone oxidoreductase chain 5 D37	Hs.414669
	RC0-BT0561-210100-032-A06- ns	
	RC0-BT0561-210100-012-B12- ns	
	EIF3S6IP eukaryotic translation initiation factor 3, subunit 6 interacting prot	Hs.119503
4	HLA-DQB1 MHC class II, DQ beta 1	Hs.73931
	RPL36 ribosomal protein L36	Hs.433411
	HLA-C major histocompatibility complex, class I, C	Hs.277477
	SDC4 syndecan 4 (amphiglycan, ryudocan)	Hs.252189
5	RPS4X ribosomal protein S4, X-linked	Hs.430084
	RPS8 ribosomal protein S8	Hs.399720
6	ANXA2 annexin A2	Hs.217493
	RPL4 ribosomal protein L4	Hs.286
	ESTs, Moderately similar to 60S ribosomal protein L4	Hs.381104
	RPL10A	Hs.425293
	MR2-ST0129-191199-003-G12- no match	
	MR2-ST0129-111199-014-A10- ns	
	RPS25 ribosomal protein S25	Hs.429729
	IL3-CT0220-051199-027-F11- ns	
	IL3-CT0219-281099-024-B11- ns	
	EEF1A1eukaryotic translation elongation factor 1 alpha 1	Hs.422118
	RC5-HT0582-110400-013-E01- ns	
	ESTs	Hs.224732
	weakly similar to Homo sapiens SARDH mRNA.	Hs.290475
	ESTs, translation elongation factor eEF-1 gamma chain	Hs.381229
	RPL36AL ribosomal protein L36a-like	Hs.419465
	IL0-CT0031-221099-113-E08- ns	
	DHCR24 24-dehydrocholesterol reductase	Hs.75616
	LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357
	RC5-HT0582-110400-013-G04- ns	
	RPL15 ribosomal protein L15	Hs.74267
	RNF13 ring finger protein 13	Hs.6900

Tabela XVIII. Comparação ANxAB, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV). Clones no match não apresentaram identidade significativa com nenhuma sequência dos bancos Hs e Nr. Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
2	NU5MNADH-ubiquinone oxidoreductase chain 5 D37	Hs.414669
	RC0-BT0561-210100-032-A06- ns	
	RC0-BT0561-210100-012-B12- ns	
	EIF3S6IP eukaryotic translation initiation factor 3, subunit 6 interacting prot	Hs.119503
6	ANXA2 annexin A2	Hs.217493
	RPL4 ribosomal protein L4	Hs.286
	ESTs, Moderately similar to 60S ribosomal protein L4	Hs.381104
	RPL10A	Hs.425293
	MR2-ST0129-191199-003-G12- no match	
	MR2-ST0129-111199-014-A10- ns	
	RPS25 ribosomal protein S25	Hs.429729
	IL3-CT0220-051199-027-F11- ns	
	RPS4X ribosomal protein S4, X-linked	Hs.430084
	IL3-CT0219-281099-024-B11- ns	
	EEF1A1eukaryotic translation elongation factor 1 alpha 1	Hs.422118
	RC5-HT0582-110400-013-E01- ns	
	ESTs	Hs.224732
	weakly similar to Homo sapiens SARDH mRNA.	Hs.290475
	S22655 translation elongation factor eEF-1 gamma chain	Hs.381229
	RPL36AL ribosomal protein L36a-like	Hs.419465
	IL0-CT0031-221099-113-E08- ns	
	DHCR24 24-dehydrocholesterol reductase	Hs.75616
	LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357
	RPS8 ribosomal protein S8	Hs.399720
	RC5-HT0582-110400-013-G04- ns	
	RPL15 ribosomal protein L15	Hs.74267
	RNF13 ring finger protein 13	Hs.6900
8	PDCD4 programmed cell death 4 (neoplastic transformation inhibitor)	Hs.326248

Tabela XIX. Comparação BCxAC, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV). Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
2	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting, alpha 1 polypeptide	Hs.76549
	GRP58 glucose regulated protein, 58kDa	Hs.13751
	CM0-BT0304-111199-093-C11- ns	
4	RPL36 ribosomal protein L36	Hs.433411
	RPS8 ribosomal protein S8	Hs.399720
	SDC4 syndecan 4 (amphiglycan, ryudocan)	Hs.252189
5	HLA-DQB1 MHC class II, DQ beta 1	Hs.73931
	RPS4X ribosomal protein S4, X-linked	Hs.430084
	HLA-C major histocompatibility complex, class I, C	Hs.277477
6	CLTA clathrin, light polypeptide (Lca)	Hs.104143
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
	CTSB cathepsin B	Hs.297939
	RC3-BT0046-310899-002-G12- ns	

Tabela XX. Comparação BNxBC, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV). Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
2	RC0-BT0505-221199-031-A06- ns	
4	HLA-DQB1 MHC class II, DQ beta 1	Hs.73931
	CLTA clathrin, light polypeptide (Lca)	Hs.104143
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
	CTSB cathepsin B	Hs.297939
	RC3-BT0046-310899-002-G12- ns	
	HLA-C major histocompatibility complex, class I, C	Hs.277477
5	RPS4X ribosomal protein S4, X-linked	Hs.430084
6	ZNF9 zinc finger protein 9	Hs.2110
	RPL4 ribosomal protein L4	Hs.286
	RPS25 ribosomal protein S25	Hs.429729
	IL3-CT0220-051199-027-F11- ns	
	IL3-CT0219-281099-024-B11- ns	
	CALM2 calmodulin 2 (phosphorylase kinase, delta)	Hs.425808
	EEF1A1 eukaryotic translation elongation factor 1 alpha 1	Hs.422118
	CALM1 calmodulin 1 (phosphorylase kinase, delta)	Hs.282410
	A Human DNA sequence from clone CTA-215D11	HS215D11
	Homo sapiens prothymosin α (gene sequence 28) (PTMA)	NM_002823
	CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867
	Homo sapiens cDNA FLJ34095 fis	Hs.290475
	DHCR24 24-dehydrocholesterol reductase	Hs.75616
	MGC4836 similar to hypothetical protein (L1H 3 region)	Hs.343206
	LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357
	APP amyloid beta (A4) precursor protein (protease nexin-II)	Hs.177486
	RPL15 ribosomal protein L15	Hs.74267
	RNF13 ring finger protein 13	Hs.6900
	PDCD4 programmed cell death 4 (neoplastic transformation inhibitor)	Hs.326248
8	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	GRP58 glucose regulated protein, 58kDa	Hs.13751
	CM0-BT0304-111199-093-C11- ns	

Tabela XXI. Comparação NAXNC, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV). Clones no match não apresentaram identidade significativa com nenhuma sequência dos bancos Hs e Nr. Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
1	RC5-HT0582-110400-013-E01- ns	
	DHCR24 24-dehydrocholesterol reductase	Hs.75616
	LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357
	RNF13 ring finger protein 13	Hs.6900
2	ANXA2 annexin A2	Hs.217493
	RPL4 ribosomal protein L4	Hs.286
	ESTs, Moderately similar to 60S ribosomal protein L4	Hs.381104
	RPL10A	Hs.425293
	MR2-ST0129-191199-003-G12- no match	
	MR2-ST0129-111199-014-A10- ns	
	RPS25 ribosomal protein S25	Hs.429729
	IL3-CT0220-051199-027-F11- ns	
	RPS4X ribosomal protein S4, X-linked	Hs.430084
	IL3-CT0219-281099-024-B11- ns	
	EEF1A1 eukaryotic translation elongation factor 1 alpha 1	Hs.422118
	ESTs	Hs.224732
	Homo sapiens cDNA FLJ34095 fis	Hs.290475
	translation elongation factor eEF-1 gamma chain	Hs.381229
	RPL36AL ribosomal protein L36a-like	Hs.419465
	IL0-CT0031-221099-113-E08- ns	
	RPS8 ribosomal protein S8	Hs.399720
	RC5-HT0582-110400-013-G04- ns	
	RPL15 ribosomal protein L15	Hs.74267
4	SERPINA1- alpha-1 antiproteinase, antitrypsin	Hs.297681
	HLA-DQB1 MHC class II, DQ beta 1	Hs.73931
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
	CTSB cathepsin B	Hs.297939
	hypothetical protein dJ465N24.2.1	Hs.8084
6	NU5MNADH-ubiquinone oxidoreductase chain 5 D37	Hs.414669
	RC0-BT0561-210100-032-A06- ns	
	RC0-BT0561-210100-012-B12- ns	
	EIF3S6IP eukaryotic translation initiation factor 3, subunit 6 interacting prot	Hs.119503
8	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	Homo sapiens BAC clone CTD-2005P16 from 2	AC096546
	XBP1 X-box binding protein 1	Hs.149923
	MR2-ST0131-111199-016-G07- no match	
	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	A Human DNA sequence from clone CTA-215D11	HS215D11
	Homo sapiens prothymosin α (gene sequence 28) (PTMA)	NM_002823

LOC51260 hypothetical protein LOC51260	Hs.128764
GRP58 glucose regulated protein, 58kDa	Hs.13751
CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867
UBC-ubiquitin C	Hs.183704
CLTC clathrin, heavy polypeptide (Hc)	Hs.178710
GRP58 glucose regulated protein, 58kDa	Hs.13751
ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
IPW imprinted in Prader-Willi syndrome	Hs.5022
hypothetical protein MGC29784	Hs.33719
CM0-BT0304-111199-093-C11- ns	
NSEP1 nuclease sensitive element binding protein 1	Hs.74497

Tabela XXII. Comparação NCxAC, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV). Clones no match não apresentaram identidade significativa com nenhuma sequência dos bancos Hs e Nr. Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
2	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	Homo sapiens BAC clone CTD-2005P16 from 2	AC096546
	XBP1 X-box binding protein 1	Hs.149923
	MR2-ST0131-111199-016-G07- no match	
	RC5-HT0582-110400-013-E01- ns	
	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	A Human DNA sequence from clone CTA-215D11	HS215D11
	Homo sapiens prothymosin α (gene sequence 28) (PTMA)	NM_002823
	LOC51260 hypothetical protein LOC51260	Hs.128764
	GRP58 glucose regulated protein, 58kDa	Hs.13751
	CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867
	UBC-ubiquitin C	Hs.183704
	CLTC clathrin, heavy polypeptide (Hc)	Hs.178710
	GRP58 glucose regulated protein, 58kDa	Hs.13751
	DHCR24 24-dehydrocholesterol reductase	Hs.75616
	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357
	IPW imprinted in Prader-Willi syndrome	Hs.5022
	hypothetical protein MGC29784	Hs.33719
	CM0-BT0304-111199-093-C11- ns	
	NSEP1 nuclease sensitive element binding protein 1	Hs.74497
	RNF13 ring finger protein 13	Hs.6900
4	RPL36 ribosomal protein L36	Hs.433411
	RPS4X ribosomal protein S4, X-linked	Hs.430084
	HLA-C MHC class I, C	Hs.277477
	RPS8 ribosomal protein S8	Hs.399720
	SDC4 syndecan 4 (amphiglycan, ryudocan)	Hs.252189
5	HLA-DQB1 MHC class II, DQ beta 1	Hs.73931
6	SERPINA1- alpha-1 antiproteinase, antitrypsin	Hs.297681
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
	CTSB cathepsin B	Hs.297939
	hypothetical protein dJ465N24.2.1	Hs.8084

Tabela XXIII. Comparação NCxBC, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV). Clones no match não apresentaram identidade significativa com nenhuma sequência dos bancos Hs e Nr. Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
1	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	GRP58 glucose regulated protein, 58kDa	Hs.13751
	CM0-BT0304-111199-093-C11- ns	
2	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	Homo sapiens BAC clone CTD-2005P16 from 2	AC096546
	XBP1 X-box binding protein 1	Hs.149923
	MR2-ST0131-111199-016-G07- no match	
	RC5-HT0582-110400-013-E01- ns	
	A Human DNA sequence from clone CTA-215D11	HS215D11
	Homo sapiens prothymosin α (gene sequence 28) (PTMA)	NM_002823
	LOC51260 hypothetical protein LOC51260	Hs.128764
	CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867
	UBC-ubiquitin C	Hs.183704
	CLTC clathrin, heavy polypeptide (Hc)	Hs.178710
	GRP58 glucose regulated protein, 58kDa	Hs.13751
	DHCR24 24-dehydrocholesterol reductase	Hs.75616
	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting, α 1	Hs.76549
	LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357
	IPW imprinted in Prader-Willi syndrome	Hs.5022
	hypothetical protein MGC29784	Hs.33719
4	NSEP1 nuclease sensitive element binding protein 1	Hs.74497
	RNF13 ring finger protein 13	Hs.6900
	RPS4X ribosomal protein S4, X-linked	Hs.430084
	CLTA clathrin, light polypeptide (Lca)	Hs.104143
5	RC3-BT0046-310899-002-G12- ns	
	HLA-C MHC class I, C	Hs.277477
	HLA-DQB1 MHC class II, DQ beta 1	Hs.73931
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
6	CTSB cathepsin B	Hs.297939
	SERPINA1- alpha-1 antiproteinase, antitrypsin	Hs.297681
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
	hypothetical protein dJ465N24.2.1	Hs.8084

Tabela XXIV. Comparação NBxAB, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV). Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
1	PDCD4 programmed cell death 4 (neoplastic transformation inhibitor)	Hs.326248
2	ZNF9 zinc finger protein 9	Hs.2110
	RPL4 ribosomal protein L4	Hs.286
	RPS25 ribosomal protein S25	Hs.429729
	IL3-CT0220-051199-027-F11	NS
	RPS4X ribosomal protein S4, X-linked	Hs.430084
	IL3-CT0219-281099-024-B11- ns	
	CALM2 calmodulin 2 (phosphorylase kinase, delta)	Hs.425808
	EEF1A1 eukaryotic translation elongation factor 1 alpha 1	Hs.422118
	CALM1 calmodulin 1 (phosphorylase kinase, delta)	Hs.282410
	A Human DNA sequence from clone CTA-215D11	HS215D11
	Homo sapiens prothymosin α (gene sequence 28) (PTMA)	NM_002823
	CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867
	Homo sapiens cDNA FLJ34095 fis	Hs.290475
	DHCR24 24-dehydrocholesterol reductase	Hs.75616
	MGC4836 similar to hypothetical protein (L1H 3 region)	Hs.343206
	LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357
	APP amyloid beta (A4) precursor protein (protease nexin-II)	Hs.177486
	RPL15 ribosomal protein L15	Hs.74267
	RNF13 ring finger protein 13	Hs.6900
6	RC0-BT0505-221199-031-A06- ns	

Tabela XXV. Comparação NBxNA, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV). Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
1	RPL4 ribosomal protein L4	Hs.286
	RPS25 ribosomal protein S25	Hs.429729
	IL3-CT0220-051199-027-F11- ns	
	CM0-BT0301-121199-094-E12	Hs.430084
	IL3-CT0219-281099-024-B11- ns	
	EEF1A1 eukaryotic translation elongation factor 1 alpha 1	Hs.422118
	Homo sapiens cDNA FLJ34095 fis	Hs.290475
	DHCR24 24-dehydrocholesterol reductase	Hs.75616
	LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357
	RPL15 ribosomal protein L15	Hs.74267
	RNF13 ring finger protein 13	Hs.6900
2	ZNF9 zinc finger protein 9	Hs.2110
	CALM2 calmodulin 2 (phosphorylase kinase, delta)	Hs.425808
	CALM1 calmodulin 1 (phosphorylase kinase, delta)	Hs.282410
	A Human DNA sequence from clone CTA-215D11	HS215D11
	Homo sapiens prothymosin α (gene sequence 28) (PTMA)	NM_002823
	CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867
	MGC4836 similar to hypothetical protein (L1H 3 region)	Hs.343206
	APP amyloid beta (A4) precursor protein (protease nexin-II)	Hs.177486
	PDCD4 programmed cell death 4 (neoplastic transformation inhibitor)	Hs.326248
4	NU5MNADH-ubiquinone oxidoreductase chain 5 D37	Hs.414669
	RC0-BT0561-210100-032-A06- ns	
	RC0-BT0561-210100-012-B12- ns	
	EIF3S6IP eukaryotic translation initiation factor 3, subunit 6 interacting prot	Hs.119503
6	RC0-BT0505-221199-031-A06- ns	

Tabela XXVI. Comparação NBxNC, mostrando a identidade dos clones diferenciais de cada região (ver figura 2 1 e tabela XV). Clones no match não apresentaram identidade significativa com nenhuma sequência dos bancos Hs e Nr. Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
1	A Human DNA sequence from clone CTA-215D11	HS215D11
	Homo sapiens prothymosin α (gene sequence 28) (PTMA)	NM_002823
	CYR61 cysteine-rich, angiogenic inducer, 61	Hs.8867
	DHCR24 24-dehydrocholesterol reductase	Hs.75616
	LAMR1 laminin receptor 1 (ribosomal protein SA, 67kDa)	Hs.181357
	RNF13 ring finger protein 13	Hs.6900
2	ZNF9 zinc finger protein 9	Hs.2110
	RPL4 ribosomal protein L4	Hs.286
	RPS25 ribosomal protein S25	Hs.429729
	IL3-CT0220-051199-027-F11- ns	
	RPS4X ribosomal protein S4, X-linked	Hs.430084
	IL3-CT0219-281099-024-B11- ns	
	CALM2 calmodulin 2 (phosphorylase kinase, delta)	Hs.425808
	EEF1A1 eukaryotic translation elongation factor 1 alpha 1	Hs.422118
	CALM1 calmodulin 1 (phosphorylase kinase, delta)	Hs.282410
	Homo sapiens cDNA FLJ34095 fis	Hs.290475
	MGC4836 similar to hypothetical protein (L1H 3 region)	Hs.343206
	APP amyloid beta (A4) precursor protein (protease nexin-II)	Hs.177486
	RPL15 ribosomal protein L15	Hs.74267
	PDCD4 programmed cell death 4 (neoplastic transformation inhibitor)	Hs.326248
4	SERPINA1- alpha-1 antiproteinase, antitrypsin	Hs.297681
	HLA-DQB1 MHC class II, DQ beta 1	Hs.73931
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
	CTSB cathepsin B	Hs.297939
	hypothetical protein dJ465N24.2.1	Hs.8084
6	RC0-BT0505-221199-031-A06- ns	
8	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	Homo sapiens BAC clone CTD-2005P16 from 2	AC096546
	XBP1 X-box binding protein 1	Hs.149923
	MR2-ST0131-111199-016-G07- no match	
	RC5-HT0582-110400-013-E01- ns	
	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	LOC51260 hypothetical protein LOC51260	Hs.128764
	GRP58 glucose regulated protein, 58kDa	Hs.13751
	UBC-ubiquitin C	Hs.183704
	CLTC clathrin, heavy polypeptide (Hc)	Hs.178710
	GRP58 glucose regulated protein, 58kDa	Hs.13751
	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting α 1	Hs.76549
	IPW imprinted in Prader-Willi syndrome	Hs.5022

hypothetical protein MGC29784	Hs.33719
CM0-BT0304-111199-093-C11- ns	
NSEP1 nuclease sensitive element binding protein 1	Hs.74497

Tabela XXVII. Comparação BxBC, mostrando a identidade dos clones diferenciais de cada região (ver figura 21 e tabela XV). Clones ns não foram sequenciados.

Região	Gene ou ORESTES	
4	HLA-DQB1 MHC class II, DQ beta 1	Hs.73931
	RPS4X ribosomal protein S4, X-linked	Hs.430084
	CLTA clathrin, light polypeptide (Lca)	Hs.104143
	S100A11 S100 calcium binding protein A11 (calgizzarin)	Hs.417004
	CTSB cathepsin B	Hs.297939
	RC3-BT0046-310899-002-G12- ns	
	HLA-C MHC class I, C	Hs.277477
6	PDCD4 programmed cell death 4 (neoplastic transformation inhibitor)	Hs.326248
8	ATP1A1 ATPase, Na ⁺ /K ⁺ transporting, alpha 1 polypeptide	Hs.76549
	GRP58 glucose regulated protein, 58kDa	Hs.13751
	CM0-BT0304-111199-093-C11- ns	

5.8. Amplificação de RNA

A padronização do protocolo de amplificação de RNA em nosso laboratório tornou possível o estudo das amostras de carcinoma folicular por microarrays. Partindo de 3µg de RNA total, amplificamos o mRNA das 10 amostras disponíveis de carcinoma folicular, o mesmo número de amostras de adenoma e quantidade suficiente de RNA de referência para as hibridizações. Os rendimentos obtidos foram bastante variáveis, possivelmente por diferenças na qualidade (pureza) do RNA total das amostras ou ainda na eficiência dos kits utilizados, em especial o Ribomax, cuja eficiência decai sensivelmente com o uso. Os dados são mostrados na tabela XVIII.

Tabela XXVIII. Rendimento das reações de amplificação dos RNAs de amostras de adenoma, carcinoma folicular e RNA referência (RNA ref).

TECIDO	AMOSTRA	massa (μg)
ADENOMA	aA1	77
	aA2	32
	aA3	72
	aA4	94
	aA5	66
	aA6	60
	aA7	58
	aA8	26
	aA9	12
	aA10	52
CARCINOMA FOLICULAR	aCF1	35
	aCF2	65
	aCF3	77
	aCF4	30
	aCF5	17
	aCF6	79
	aCF7	64
	aCF8	56
	aCF9	19
	aCF10	16
RNA REF.	média	52

A qualidade dos RNAs amplificados foi avaliada pelo perfil eletroforético do cDNA obtido por transcrição reversa na presença de nucleotídeos fluorescentes. A imagem de um gel representativo é mostrada na figura 22.

Figura 22. Gel de agarose contendo alíquotas de marcações de RNA amplificado de amostras de RNA de tireóide e do RNA referência, misturadas em proporções iguais. (A) Migração do cDNA marcado em relação ao marcador Pel *Hind* III, marcado com Vistra Green, nos comprimentos de emissão de Cy3 (esq) e Cy5 (dir). (B) Visualização dos cDNAs marcados nos comprimentos de onda correspondentes às emissões de Cy3 (esq), Cy5 (centro) e ambos (dir, imagem composta). As cores de Cy3 e Cy5 estão invertidas devido ao processamento da imagem pelo programa GeneTac.

As amostras de RNA amplificado serão hibridizadas com uma lâmina contendo 4800 clones ORESTES, selecionados segundo os critérios: tamanho superior a 300pb, localização mais próxima à região 3' do gene, sequência única no genoma. Esta lâmina será adotada ao invés da 3.8K pela maior quantidade de clones, seleção mais cuidadosa dos mesmos e uso da nova metodologia de fixação de DNA (ver discussão adiante).

6. DISCUSSÃO

6.1. Discussão sobre os resultados obtidos

6.1.1. DDRT-PCR e cDNA array em membrana de nylon

O uso de DDRT-PCR para seleção de genes para fixação em arrays é um procedimento muito trabalhoso e que requer muito tempo. Além disso, os genes selecionados em um experimento usando amostra de um paciente nem sempre são diferenciais para outros pacientes, analisados no array. Um problema adicional é que como as intensidades das bandas de DDRT-PCR não refletem a abundância dos transcritos correspondentes, parte dos genes selecionados mostram uma expressão muito reduzida no array. Todos esses fatores nos levaram a abandonar a pré-seleção por DDRT-PCR.

6.1.2. IGFBP-5

O clone 18, um dos quatro genes selecionados pelo array de clones de DDRT-PCR e analisados por RT-PCR, corresponde ao gene da IGFBP-5 ou proteína 5 ligante do fator de crescimento do tipo insulina. Este é o primeiro relato da superexpressão desse gene em câncer de tireóide, embora ele já tenha sido descrito como aumentado em tecido prostático maligno (Tennant et al., 1996) e como estimulador do crescimento para osteoblastos (Miyakoshi et al., 2001); (Richman et al., 1999); (Slootweg et al., 1996) e células de tumor de próstata (Miyake et al., 2000) e de mama (Perks et al., 1999; Perks et al., 2000). Os níveis de outras proteínas da família das IGFBPs, bem como dos IGFs (fatores de crescimento do tipo insulina) estão aumentados ou diminuídos em diversos tipos de câncer (Holly, 2000).

6.1.3. Real-Time RT-PCR

Embora forneça medidas quantitativas da expressão gênica, o método de Real-Time RT-PCR foi utilizado apenas uma vez nesse trabalho, devido a seu alto custo e pouca processividade (poucas amostras e genes analisados por placa, devido ao grande número de controles e diluições para curva padrão).

Quando realizado com as mesmas amostras nas quais os genes 18 e 44 foram identificados, esse método mostrou resultados coerentes com os de DDRT-PCR. As mesmas amostras, analisadas por microarray, tiveram resultados coerentes com os demais métodos, embora as razões entre as expressões tenham sido maiores por Real-Time RT-PCR do que por microarray, observação já descrita na literatura (Feldman et al., 2002).

Quando analisamos a expressão desses dois genes em amostras distintas, porém, observamos um perfil bastante diferente. Esse resultado reflete a heterogeneidade da expressão gênica entre indivíduos, especialmente em tecidos tumorais, e reforça a necessidade de se analisar o maior número possível de amostras de cada tecido, o que não é viável por nosso protocolo de Real-Time RT-PCR.

6.1.4. RT-PCR, cluster, árvore de decisão

Embora o método de RT-PCR com Southern blot envolva diversas etapas, o que aumenta o erro e reduz a reprodutibilidade, ele permite a comparação entre amostras quando executado com cuidado e com os controles adequados.

Os resultados obtidos por esse método mostraram que a seleção de genes diferenciais por microarray é eficiente, mas que são necessários mais de dois genes para diferenciar precisamente tecido normal, bócio e carcinoma papilífero, e provavelmente um número bem maior para distinguir adenoma e carcinoma foliculares.

Os métodos de cluster hierárquico e árvore de decisão foram usados para classificação das amostras nessa etapa devido a sua fácil aplicação e à pouca complexidade dos dados. Para datasets grandes e complexos o cluster hierárquico pode não refletir precisamente as relações entre os padrões de expressão gênica, pois a natureza determinística do método pode levar dados a serem agrupados com base em decisões locais, sem oportunidade de checagem (Tamayo et al., 1999).

6.1.5. Microarray

6.1.5.1. Os genes encontrados

Como mencionado no item 5.7.4, foram identificados 160 clones diferencialmente expressos entre quaisquer dois tecidos, cujas sequências foram determinadas e comparadas com as dos bancos de dados. Alguns desses clones correspondiam a um mesmo gene, com

identificações Hs iguais ou distintas, de acordo com a localização das nossas sequências nos clones do banco de dados. Mantivemos todos os clones nas tabelas e observamos que seus valores de $\log_2 p$ em uma dada comparação foram bastante semelhantes, o que atesta a favor da confiabilidade das análises. Como exemplo pode-se citar o gene *NADH ubiquinone-oxidoreductase*, representado por 5 clones mais expressos em A na comparação Nx A (tabela IX), cujos valores de $\log_2 p$ variaram de 10,4 a 7,4.

Diversos genes interessantes foram identificados como diferencialmente expressos. Parte deles estão menos expressos em tumores do que nas amostras de tecido normal. Dentre eles encontra-se o correspondente à ATPase transportadora de Na/K, cuja expressão está reduzida nos carcinomas em relação a tecidos normais, bócios e adenomas. Essa enzima está também diminuída em carcinomas de bexiga em relação aos tecidos benignos, segundo comparações utilizando microarray tecidual (Espineda et al., 2003). O gene da proteína PDCD4, envolvida na morte celular programada, está mais expresso em adenomas e tecidos normais em comparação com carcinomas e bócios, podendo estar relacionado com redução da apoptose e aumento da longevidade celular na hiperplasia e no tumor maligno. Dados de SAGE (SAGE Genie, do CGAP) mostram maior expressão do gene em tecido normal do que nos tumores de cólon. O gene do inibidor de quinase ciclina-dependente P27 ou KIP1, que inibe a transição da quiescência para a proliferação, é menos expresso em adenomas, bócios e carcinomas do que nos tecidos normais. A redução dos níveis dessa proteína também foi associada a mau prognóstico em tumores renais (Hedberg et al., 2003). Diversos trabalhos avaliaram a expressão de P27 em tumores tireoideanos (Wang et al., 1998); (Resnick et al., 1998). Um deles mostrou 89% de marcação fraca de P27 por imunohistoquímica em carcinomas papilíferos da variante folicular e 100% de marcação intensa em adenomas, sendo a diferença entre os dois tumores estatisticamente significativa (Wang et al., 1998). Outro estudo obteve resultados semelhantes, observando maior abundância da proteína P27 em tecidos normais, seguidos de adenomas e por fim de carcinomas (Resnick et al., 1998). Um outro gene de expressão reduzida em bócio, adenoma e carcinoma papilífero em relação ao tecido normal é o da proteína CYR 61 (cystein-rich angiogenic inducer 61), um regulador de angiogênese.

Alguns genes estão mais expressos nos carcinomas do que nos demais tecidos. Um deles foi o gene da $\alpha 1$ -anti-tripsina, um inibidor de serina protease que foi identificado como superexpresso em adenocarcinoma ductal do pâncreas por cDNA array e imunohistoquímica (Trachte et al., 2002) e em carcinoma papilífero da tireóide por Western blot, como mencionado em 1.3.3 (Poblete et al., 1996). Outro gene identificado

foi o da calgizzarin (S100A11), uma proteína ligante de cálcio que transmite sinais regulatórios dependentes de cálcio modulando a atividade de enzimas alvo, também superexpresso em carcinomas coloretais em comparação com a mucosa normal, por Northern blot (Tanaka et al., 1995a; Tanaka et al., 1995a). O gene *Sdc4* ou *syndecan 4* também foi identificado como mais expresso nos carcinomas papilíferos. Ele regula ligação e sinalização de fosfolípidios de inositol, e participa de mecanismos de adesão e quimiotaxia. A expressão de *cathepsina B* foi maior em carcinoma papilífero do que em bócio e tecido normal, conforme observado em dados de SAGE e proteína da literatura (Shuja and Mumane, 1996). Os genes referentes ao MHC classe II DQ β 1 (HLA-DQB1) e MHC classe I C (HLA-C) também estão mais expressos em carcinomas do que nos outros três tecidos, mas não temos explicação biológica clara para essa observação.

6.1.5.2. Quantarray

Duas críticas usualmente feitas a programas de quantificação manual de spots, como o Quantarray, foram abordadas neste trabalho. Uma delas é a falta de reprodutibilidade entre os dados de duas quantificações de uma mesma lâmina. A outra é a subjetividade do método e seu erro em relação a métodos automatizados.

Como apresentado na fig.17, as duas quantificações da mesma hibridização pelo método histograma do programa Quantarray foram muito semelhantes, sugerindo que ajustes feitos por um mesmo usuário são reprodutíveis.

A comparação das quantificações de um experimento por esse mesmo método e pelo programa automático Bioinfo (fig.18) foi também animadora, sugerindo que o uso do Quantarray não introduz erros significativos em nossas análises.

6.1.5.3. Normalização

Dado o conhecimento limitado das fontes de variação sistemática, a normalização é um aspecto importante e desafiador, que não pode ser abordado de maneira simples e genérica, sem verificação (Dudoit et al., 2003). A primeira recomendação antes de se escolher um método de normalização é examinar, para cada array, a variação das razões entre os sinais dos dois corantes de acordo com as intensidades, usando M-A plots (5.7.2 e fig.16). Essa representação ajuda a identificar artefatos dos spots e características específicas da lâmina, como marcações desiguais relacionadas com localização em

subarrays (print-tips). Isso auxilia na escolha do método de normalização mais adequado para cada situação (Dudoit et al., 2003). Conforme mencionado nos resultados (5.7.2), optamos por não utilizar normalização por subarrays.

Métodos de regressão local como lowess ou loess são ferramentas promissoras para gerar normalizações robustas e flexíveis, por corrigirem variações sistemáticas locais ao invés de adotar normalizações globais (Dudoit et al., 2003). Por essa razão adotamos o método de loess para normalização de nossos dados.

6.2. Discussão sobre os problemas dos resultados obtidos

Embora tenhamos identificado genes capazes de distinguir tecido normal, bócio, adenoma e carcinoma papilífero, o erro associado à identificação destes genes foi bastante alto. A amplitude desse erro foi estimada pelo grande efeito que a retirada de uma amostra teve sobre os dados e pela baixa correlação entre as duplicatas com inversão de corantes (dados não mostrados). Uma análise que reforça a baixa correlação entre duplicatas com inversão dos corantes é mostrada na figura 23. Neste gráfico os spots são representados de acordo com os valores de $-\log_2 p$ (altos valores de $-\log_2 p$ = baixos valores de p e alta significância) da comparação entre tecido normal e bócio, tomando-se a média das duplicatas e sem tomar a média. Observa-se maior número de spots com altos valores de $-\log_2 p$ quando se faz a média das duplicatas e baixos valores quando não se faz a média (região azul) do que o contrário (região verde). Isso sugere que as duplicatas apresentam grande variabilidade, a qual é reduzida ou mascarada ao se fazer a média. Por essa razão utilizamos as duplicatas independentemente, sem fazer média.

Figura 23. Scatter plot mostrando os valores de $-\log_2 p$ para cada spot na comparação entre amostras de tecido normal e bócio, utilizando a média das duplicatas (swaps) de cada amostra (eixo x) e sem utilizar a média dos swaps (eixo y). A região azul delimita spots com valores de $-\log_2 p$ maiores do que 5 quando se faz a média dos swaps e menores do que 5 quando não se faz média. A região verde mostra spots com $-\log_2 p$ superior a 5 sem média dos swaps e inferior a 5 com a média.

Muitas das dificuldades encontradas ao longo desse trabalho levaram a implementações de protocolos ou equipamentos que culminaram com a grande melhora na qualidade dos dados observada hoje. Nesta seção serão apresentados alguns dos problemas encontrados e as providências tomadas para resolvê-los.

6.2.1. Câmera CCD x laser

Aspectos importantes da hibridização em microarrays são background baixo e uniforme e altos valores de sinal/ ruído (Hegde et al., 2000). De fato, se a intensidade de um spot for baixa, mesmo que a razão entre os valores dos dois corantes for alta, a medida pode não ser confiável (Brazma and Vilo, 2000).

Em nossos experimentos obtivemos valores de sinal baixos e de background altos devido ao uso de detector CCD. De fato, enquanto que a maioria dos spots apresenta sinal

visível em uma digitalização utilizando scanner a laser, muitos spots não são visíveis em imagens obtidas por CCD. Em nossas lâminas, mais da metade dos spots não apresentou valor positivo após a subtração do background médio. A figura 24 mostra imagens de duas lâminas iguais hibridizadas com a mesma amostra porém digitalizadas com scanners distintos.

Figura 24. Imagens de lâminas hibridizadas com uma mesma amostra, obtidas com detector CCD (A) ou scanner a laser (B).

6.2.2. Fixação do DNA na lâmina

Em algumas hibridizações com a lâmina 3.8k observamos um arraste partindo da região do spot. Isso levou nosso grupo a testar uma nova metodologia de fixação do DNA

à lâmina, que consistiu de uma incubação extra em estufa a 80°C por 2 horas, após a exposição à luz UV. A figura 25 mostra uma mesma amostra hibridizada com a lâmina 3.8k (método de fixação antigo) e com outra lâmina contendo 3800 clones fixados com a metodologia nova, ambas digitalizadas com scanner a laser.

Figura 25. Lâminas contendo 3800 genes fixados com métodos diferentes. (A) Lâmina 3.8K usada neste trabalho. (B) Lâmina QTG, usada em outro projeto.

6.2.3. Qualidade do RNA de tireóide

Muitas das hibridizações com RNA de tireóide apresentaram sinais fracos, background alto e artefatos. Suspeitando de problemas na qualidade do RNA de tireóide, realizamos marcações em paralelo com RNA de esôfago. De fato, os resultados foram bem melhores com esse segundo tecido, sugerindo que as amostras de tireóide contêm

algum tipo de impureza que atrapalha a marcação (dados não mostrados). Aproximadamente metade do volume da tireóide corresponde ao colóide, que se acumula na luz do folículo e é composto principalmente por tireoglobulina, uma glicoproteína que participa da produção dos hormônios tireoideanos. É possível que parte dessa massa glicoproteica resista ao método de purificação de RNA utilizado ou retenha impurezas que não são removidas no processo, e seja responsável pelos resultados de qualidade inferior observados para esse tecido.

A presença de impurezas no RNA de tireóide parece também afetar o rendimento da amplificação de RNA, que é menor para amostras deste tecido do que para amostras de esôfago e sobretudo de mama, seguindo o mesmo protocolo (dados não mostrados).

O processo de amplificação de RNA aumenta não só a quantidade mas também a pureza do RNA. A figura 26 mostra o resultado da hibridização utilizando RNA total e amplificado de uma mesma amostra, sendo a primeira imagem obtida por detector CCD e a segunda por scanner a laser. O uso de scanners distintos é certamente responsável por parte da melhora na imagem obtida, mas a amostra amplificada também tem um efeito importante. De fato, um trabalho de nosso grupo mostrou que o uso de RNA amplificado levou a um aumento na intensidade de marcação nos dois canais (Gomes et al., 2003), aumentando o número de spots com valores acima do background. Além disso, outro trabalho mostrou que quando se utiliza RNA amplificado a correlação entre duas marcações com inversão de corantes é mais alta do que quando se utiliza RNA total (Feldman et al., 2002).

Figura 26. Lâminas hibridizadas com RNA total (A) e amplificado (B) de uma mesma amostra de carcinoma folicular. Imagens obtidas em CCD (A) e laser(B).

7. CONCLUSÕES E PERSPECTIVAS

Este trabalho foi o primeiro utilizando arrays em lâminas de vidro de nosso laboratório, logo foi necessário um intenso investimento em padronizações.

Identificamos 160 genes diferencialmente expressos entre tecido tireoideano normal, bócio, adenoma e carcinoma papilífero. Alguns desses genes terão sua expressão avaliada por RT-PCR-Southern blot ou Real-Time RT-PCR e tissue array.

Os dados quantitativos têm baixa precisão devido a inúmeras dificuldades, entre elas o uso de detector CCD, a metodologia de fixação do DNA à lâmina e a má qualidade do RNA de tireóide. Assim, buscamos tornar nossas observações qualitativamente confiáveis utilizando métodos de seleção bastante restritivos. A confiabilidade dos dados foi sustentada pela identificação de vários genes já descritos como diferencialmente expressos na literatura de tireóide ou de outros tumores e/ou nas análises de SAGE de tumores. Como exemplos, podemos citar os genes da ATPase transportadora de Na/K, menos expresso nos carcinomas em relação a tecidos normais, bócios e adenomas e cuja enzima está também diminuída em carcinomas de bexiga em relação aos tecidos benignos (Espineda et al., 2003) e o gene *P27*, menos expresso em adenomas, bócios e carcinomas do que nos tecidos normais, cuja proteína está também reduzida em tumores tireoideanos (Wang et al., 1998); (Resnick et al., 1998) e renais (Hedberg et al., 2003). Genes identificados como mais expressos nos carcinomas do que nos demais tecidos também são sustentados por observações de outros grupos. O RNA e a proteína α 1-anti-tripsina são descritos como aumentados em adenocarcinoma ductal do pâncreas (Trachte et al., 2002) e a proteína em carcinoma papilífero da tireóide (Poblete et al., 1996), o gene da calgizzarin (*S100A11*) é superexpresso em carcinomas coloretais em comparação com a mucosa normal (Tanaka et al., 1995a; Tanaka et al., 1995a) e o da *cathepsina B* é mais expresso em carcinoma papilífero do que em bócio e tecido normal segundo dados de SAGE e proteína (Shuja and Murnane, 1996).

Na próxima etapa hibridizaremos os RNAs de carcinoma folicular e adenoma, juntamente com RNA de referência, já amplificados, em lâminas de 4800 ORESTES. Identificaremos os genes diferencialmente expressos entre esses dois tipos de neoplasia e buscaremos elaborar um classificador molecular baseado na expressão do menor número possível de genes para distinguir os dois tipos de tumor.

8. RESUMO

Doenças tireoideanas são bastante comuns, sendo sua maioria benigna. A relação entre os diversos tipos de doenças tireoideanas, bem como seus aspectos moleculares, são pouco conhecidos. O bócio (hiperplasia), por exemplo, é descrito por alguns como relacionado com carcinoma (tumor maligno) papilífero, enquanto que outros afirmam não haver relação causal entre as duas doenças. A questão mais desafiante, porém, refere-se à distinção entre adenoma (tumor benigno) e carcinoma folicular, que atualmente é feita apenas após à cirurgia, não permitindo tratamento diferenciado para os dois tipos de tumor.

Este trabalho buscou identificar genes diferencialmente expressos entre tecido tireoideano normal, bócio, adenoma e carcinoma papilífero utilizando microarrays. Carcinomas foliculares não foram incluídos devido ao número e tamanho reduzidos das amostras.

Dois tipos de array foram utilizados: arrays em membranas de nylon, contendo 213 clones obtidos por DDRT-PCR de amostras de tireóide, e arrays em vidro, contendo 3800 clones ORESTES. Experimentos utilizando o primeiro tipo de array identificaram três genes diferencialmente expressos, cuja expressão foi analisada por RT-PCR em 10 amostras de cada tipo de tecido. Dois deles foram capazes de diferenciar carcinomas papilíferos de tecido normal e bócio com 89% de precisão para o tumor maligno e 80% para os tecidos não malignos.

Os arrays em vidro foram utilizados para avaliar o perfil de expressão de aproximadamente 10 amostras de cada tipo de tecido tireoideano. Foram identificados 160 clones diferencialmente expressos entre quaisquer dois tipos de tecido, cujas sequências foram determinadas e comparadas com as dos bancos de dados. Dentre os genes mais interessantes destacam-se o correspondente à ATPase Na/K, cuja expressão está reduzida nos carcinomas em relação a tecidos normais e adenomas, o da proteína PDCD4, envolvida em morte celular programada, mais expresso em adenomas e tecidos normais em comparação com carcinomas e bócios, e os da calgizzarin (S100A11) e da α 1-anti-tripsina, ambos mais ativos nos carcinomas do que nos demais tecidos. Todos esses genes já foram descritos como diferencialmente expressos em algum tipo de tumor.

Este trabalho levou à padronização da metodologia de microarray em lâminas de vidro em nosso laboratório, bem como à identificação de genes que podem elucidar as alterações envolvidas na formação do bócio, adenoma e carcinoma papilífero. A

implantação da técnica de amplificação de mRNA em nosso laboratório viabilizou a utilização de 10 amostras de carcinoma folicular, cuja massa de RNA total era insuficiente para as hibridizações. Essas amostras serão hibridizadas, juntamente com 10 amostras de adenoma, com microarrays contendo 4800 genes humanos conhecidos para a busca de genes diferencialmente expressos, de grande interesse diagnóstico.

9. ABSTRACT

Thyroid diseases are very common and are usually benign. The causal relationships among the different types of disease, as well as their molecular aspects, are not well understood. The goiter (hyperplasia), for instance, is described by some as related to papillary carcinoma (a malignant tumor), while others say there is no causal relationship between the two diseases. The most defying question, however, concerns the distinction between adenoma (benign tumor) and follicular carcinoma, which is currently made only after surgery, not allowing distinct treatments for the two kinds of tumor.

This work aimed to identify differentially expressed genes among normal thyroid tissue, goiter, adenoma and papillary carcinoma using microarrays. Follicular carcinomas were not included due to the reduced number and size of the samples.

Two kinds of array were used: arrays in nylon membranes, with 213 clones isolated from thyroid samples by differential display (DDRT-PCR); and glass slide arrays containing 3800 ORESTES clones. Experiments using the first type of array identified three differentially expressed genes, whose expression was analyzed by RT-PCR in 10 samples of each kind of tissue. Two of these genes were able to differentiate papillary carcinomas from goiters and normal tissues with precisions of 89% for the malignant tumor and 80% for the non-malignant tissues.

Glass slide arrays were used to evaluate gene expression profile of approximately 10 samples of each type of thyroid tissue. 160 clones differentially expressed between any two tissues were identified, and their sequences were determined and compared with databases. Among the most interesting genes are Na/K ATPase gene, whose expression is reduced in carcinomas compared to normal tissues and adenomas, the gene corresponding to PDCD4 protein, involved in program cell death, with elevated expression in adenomas and normal tissues than carcinomas and goiters, and the genes of calgizzarin (S100A11) and α 1-antitrypsin, both more active in carcinomas than the other tissues. All these genes have already been described as differentially expressed in at least one type of human cancer.

This work led to the standardization of glass slide microarray technology in our laboratory, and to the identification of genes that may clarify the alterations involved in the formation of goiter, adenoma and follicular carcinoma. The implementation of mRNA amplification technique in our laboratory allowed the utilization of 10 samples of follicular carcinoma, whose mass was insufficient for microarray hybridizations. These samples will

be hybridized along with 10 samples of adenomas, with microarrays containing 4800 known human genes to search for differentially expressed genes, of great diagnostic interest.

10. BIBLIOGRAFIA

- Alberts,B., Bray,D., Lewis,J., Raff,M., Roberts,J.A., and Watson,J.D. (1994). Cancer. In *Molecular biology of the cell*, B.Alberts, D.Bray, J.Lewis, M.Raff, J.A.Roberts, and J.D.Watson, eds. (New York: Garland Publishing Inc), pp. 1255-1294.
- Alizadeh,A.A., Eisen,M.B., Davis,R.E., Ma,C., Lossos,I.S., Rosenwald,A., Boldrick,J.C., Sabet,H., Tran,T., Yu,X., Powell,J.I., Yang,L., Marti,G.E., Moore,T., Hudson,J., Jr., Lu,L., Lewis,D.B., Tibshirani,R., Sherlock,G., Chan,W.C., Greiner,T.C., Weisenburger,D.D., Armitage,J.O., Warnke,R., Staudt,L.M., and . (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* *403*, 503-511.
- Altschul,S.F., Gish,W., Miller,W., Myers,E.W., and Lipman,D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* *215*, 403-410.
- Augustin-voss,M.G. and Pauli,B.U. (1994). Pathology of neoplasia:morphological and functional attributes . In *Developmental Biology and Cancer*, G.M.Hodges and C.Rowlat, eds. (Florida: CRC Press), pp. 3-28.
- Bartolazzi,A., Gasbarri,A., Papotti,M., Bussolati,G., Lucante,T., Khan,A., Inohara,H., Marandino,F., Orlandi,F., Nardi,F., Vecchione,A., Tecce,R., and Larsson,O. (2001). Application of an immunodiagnostic method for improving preoperative diagnosis of nodular thyroid lesions. *Lancet* *357*, 1644-1650.
- Battista,S., de Nigris,F., Fedele,M., Chiappetta,G., Scala,S., Vallone,D., Pierantoni,G.M., Mega,T., Santoro,M., Viglietto,G., Verde,P., Fusco,A., and Megar,T. (1998). Increase in AP-1 activity is a general event in thyroid cell transformation in vitro and in vivo. *Oncogene* *17*, 377-385.
- Baugh,L.R., Hill,A.A., Brown,E.L., and Hunter,C.P. (2001). Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.* *29*, E29.
- Bieche,I., Laurendeau,I., Tozlu,S., Olivi,M., Vidaud,D., Lidereau,R., and Vidaud,M. (1999). Quantitation of MYC gene expression in sporadic breast tumors with a real-time reverse transcription-PCR assay. *Cancer Res.* *59*, 2759-2765.
- Blume-Jensen,P. and Hunter,T. (2001). Oncogenic kinase signalling. *Nature* *411*, 355-365.
- Bonner,R.F., Emmert-Buck,M., Cole,K., Pohida,T., Chuaqui,R., Goldstein,S., and Liotta,L.A. (1997). Laser capture microdissection: molecular analysis of tissue. *Science* *278*, 1481,1483.
- Brabant,G., Hoang-Vu,C., Cetin,Y., Dralle,H., Scheumann,G., Molne,J., Hansson,G., Jansson,S., Ericson,L.E., and Nilsson,M. (1993). E-cadherin: a differentiation marker in thyroid malignancies. *Cancer Res.* *53*, 4987-4993.
- Brasileiro Fo,G., Guimarães,R.C., and Bogliolo,L. (1993). Distúrbios do crescimento e da diferenciação celular. In *Patologia Geral*, Brasileiro Fo., F.E.L.Pereira, J.E.H.Pitella, E.A.Bambirra, and A.J.A.Barbosa, eds. (Rio de Janeiro: Guanabara Koogan SA), pp. 144-185.

Brazma,A. and Vilo,J. (2000). Gene expression data analysis. *FEBS Lett.* 480, 17-24.

Bubendorf,L., Kolmer,M., Kononen,J., Koivisto,P., Mousses,S., Chen,Y., Mahlamaki,E., Schraml,P., Moch,H., Willi,N., Elkahloun,A.G., Pretlow,T.G., Gasser,T.C., Mihatsch,M.J., Sauter,G., and Kallioniemi,O.P. (1999). Hormone therapy failure in human prostate cancer: analysis by complementary DNA and tissue microarrays. *J. Natl. Cancer Inst.* 91, 1758-1764.

Califano,D., Monaco,C., Santelli,G., Giuliano,A., Veronese,M.L., Berlingieri,M.T., de,F., V, Berger,N., Trapasso,F., Santoro,M., Viglietto,G., and Fusco,A. (1998). Thymosin beta-10 gene overexpression correlated with the highly malignant neoplastic phenotype of transformed thyroid cells in vivo and in vitro. *Cancer Res.* 58, 823-828.

Calvo,R., West,J., Franklin,W., Erickson,P., Bemis,L., Li,E., Helfrich,B., Bunn,P., Roche,J., Brambilla,E., Rosell,R., Gemmill,R.M., and Drabkin,H.A. (2000). Altered HOX and WNT7A expression in human lung cancer. *Proc. Natl. Acad. Sci. U. S. A* 97, 12776-12781.

Cheung,C.C., Carydis,B., Ezzat,S., Bedard,Y.C., and Asa,S.L. (2001a). Analysis of ret/PTC gene rearrangements refines the fine needle aspiration diagnosis of thyroid cancer. *J. Clin. Endocrinol. Metab* 86, 2187-2190.

Cheung,C.C., Ezzat,S., Freeman,J.L., Rosen,I.B., and Asa,S.L. (2001b). Immunohistochemical diagnosis of papillary thyroid carcinoma. *Mod. Pathol.* 14, 338-342.

Chiappetta,G., Tallini,G., De Biasio,M.C., Manfioletti,G., Martinez-Tello,F.J., Pentimalli,F., de Nigris,F., Mastro,A., Botti,G., Fedele,M., Berger,N., Santoro,M., Giancotti,V., and Fusco,A. (1998). Detection of high mobility group I HMGI(Y) protein in the diagnosis of thyroid tumors: HMGI(Y) expression represents a potential diagnostic indicator of carcinoma. *Cancer Res.* 58, 4193-4198.

Chiappetta,G., Tallini,G., De Biasio,M.C., Pentimalli,F., de Nigris,F., Losito,S., Fedele,M., Battista,S., Verde,P., Santoro,M., and Fusco,A. (2000). FRA-1 expression in hyperplastic and neoplastic thyroid diseases. *Clin. Cancer Res* 6, 4300-4306.

Chiariotti,L., Berlingieri,M.T., De Rosa,P., Battaglia,C., Berger,N., Bruni,C.B., and Fusco,A. (1992). Increased expression of the negative growth factor, galactoside-binding protein, gene in transformed thyroid cells and in human thyroid carcinomas. *Oncogene* 7, 2507-2511.

Churchill,G.A. (2002). Fundamentals of experimental design for cDNA microarrays. *Nat. Genet.* 32 *Suppl*, 490-495.

Cleveland,W.S. (1979). Robust Locally Weighted Regional and smotthing scatterplots. *Journal of the American Statistical Association* 74, 829-836.

Cotran,R.S., Kumar,V., and Robbins,S.L. (1994). Neoplasia. In *Pathologic basis of disease*, F.J.Schoen, ed. (Philadelphia: W.B.Saunders Company), pp. 241-303.

Curran,T. and Franza,B.R., Jr. (1988). Fos and Jun: the AP-1 connection. *Cell* 55, 395-397.

de Nigris,F., Mega,T., Berger,N., Barone,M.V., Santoro,M., Viglietto,G., Verde,P., and Fusco,A. (2001). Induction of ETS-1 and ETS-2 transcription factors is required for thyroid cell transformation. *Cancer Res.* 61, 2267-2275.

de Nigris,F., Visconti,R., Cerutti,J., Califano,D., Mineo,A., Santoro,M., Santelli,G., and Fusco,A. (1998). Overexpression of the HIP gene coding for a heparin/heparan sulfate-binding protein in human thyroid carcinomas. *Cancer Res.* 58, 4745-4751.

Di Renzo,M.F., Olivero,M., Ferro,S., Prat,M., Bongarzone,I., Pilotti,S., Belfiore,A., Costantino,A., Vigneri,R., Pierotti,M.A., and . (1992). Overexpression of the c-MET/HGF receptor gene in human thyroid carcinomas. *Oncogene* 7, 2549-2553.

Dudoit,S., Yang,Y.H., Callow,M.J., and Speed,T.P. (2003). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. *Statistica Sinica* 12, 111-139.

Emmert-Buck,M.R., Bonner,R.F., Smith,P.D., Chuaqui,R.F., Zhuang,Z., Goldstein,S.R., Weiss,R.A., and Liotta,L.A. (1996). Laser capture microdissection. *Science* 274, 998-1001.

Espineda,C., Seligson,D.B., James,B.W., Jr., Rao,J., Palotie,A., Horvath,S., Huang,Y., Shi,T., and Rajasekaran,A.K. (2003). Analysis of the Na,K-ATPase alpha- and beta-subunit expression profiles of bladder cancer using tissue microarrays. *Cancer* 97, 1859-1868.

Evan,G.I. and Vousden,K.H. (2001). Proliferation, cell cycle and apoptosis in cancer. *Nature* 411, 342-348.

Fagin,J.A. (1994). Molecular genetics of human thyroid neoplasms. *Annu. Rev. Med.* 45, 45-52.

Farid,N.R., Zou,M., and Shi,Y. (1995). Genetics of follicular thyroid cancer. *Endocrinol. Metab Clin. North Am.* 24, 865-883.

Feldman,A.L., Costouros,N.G., Wang,E., Qian,M., Marincola,F.M., Alexander,H.R., and Libutti,S.K. (2002). Advantages of mRNA amplification for microarray analysis. *Biotechniques* 33, 906-12, 914.

Ferguson,J.A., Boles,T.C., Adams,C.P., and Walt,D.R. (1996). A fiber-optic DNA biosensor microarray for the analysis of gene expression. *Nat. Biotechnol.* 14, 1681-1684.

Freeman,W.M., Robertson,D.J., and Vrana,K.E. (2000). Fundamentals of DNA hybridization arrays for gene expression analysis. *Biotechniques* 29, 1042-1055.

Golub,T.R., Slonim,D.K., Tamayo,P., Huard,C., Gaasenbeek,M., Mesirov,J.P., Coller,H., Loh,M.L., Downing,J.R., Caligiuri,M.A., Bloomfield,C.D., and Lander,E.S. (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286, 531-537.

Gomes,L.I., Silva,R.L.A., Stolf,B.S., Cristo,E.B., Hirata,R., Soares,F.A., Reis,L.F.L., Neves,E.J., and Carvalho,A.F. (2003). Comparative analysis of amplified and

- nonamplified RNA for hybridization in cDNA microarray. *Analytical Biochemistry* 321, 244-251.
- Gonsky,R., Knauf,J.A., Elisei,R., Wang,J.W., Su,S., and Fagin,J.A. (1997). Identification of rapid turnover transcripts overexpressed in thyroid tumors and thyroid cancer cell lines: use of a targeted differential RNA display method to select for mRNA subsets. *Nucleic Acids Res.* 25, 3823-3831.
- Gress,T.M., Muller-Pillasch,F., Geng,M., Zimmerhackl,F., Zehetner,G., Friess,H., Buchler,M., Adler,G., and Lehrach,H. (1996). A pancreatic cancer-specific expression profile. *Oncogene* 13, 1819-1830.
- Hamburger,J.I. and Kaplan,M.M. (1996). Evaluation of thyroid nodules by needle biopsy. In Werner and Ingbar's- The thyroid. A fundamental and clinical text, L.E.Braverman and R.D.Utger, eds. (Philadelphia: Lippincott-Raven Publishers), pp. 447-460.
- Harach,H.R. and Williams,E.D.W. (1994). Pathology of thyroid cancer. In Diseases of the thyroid, M.H.Wheeler and J.H.Lazarus, eds. (London: Chapman Hall).
- Haugen,D.R., Akslen,L.A., Varhaug,J.E., and Lillehaug,J.R. (1996). Expression of c-erbB-3 and c-erbB-4 proteins in papillary thyroid carcinomas. *Cancer Res.* 56, 1184-1188.
- Hedberg,Y., Ljungberg,B., Roos,G., and Landberg,G. (2003). Expression of cyclin D1, D3, E, and p27 in human renal cell carcinoma analysed by tissue microarray. *Br. J. Cancer* 88, 1417-1423.
- Hegde,P., Qi,R., Abernathy,K., Gay,C., Dharap,S., Gaspard,R., Hughes,J.E., Snesrud,E., Lee,N., and Quackenbush,J. (2000). A concise guide to cDNA microarray analysis. *Biotechniques* 29, 548-4, 556.
- Heid,C.A., Stevens,J., Livak,K.J., and Williams,P.M. (1996). Real time quantitative PCR. *Genome Res.* 6, 986-994.
- Henderson,B.E., Ross,R.K., and Pike,M.C. (1991). Toward the primary prevention of cancer. *Science* 254, 1131-1138.
- Hermann,S., Sturm,I., Mrozek,A., Klosterhalfen,B., Hauptmann,S., Dorken,B., and Daniel,P.T. (2001). Bax expression in benign and malignant thyroid tumours: dysregulation of wild-type P53 is associated with a high Bax and P21 expression in thyroid carcinoma. *Int. J. Cancer* 92, 805-811.
- Hirata,R.J., Barrera,J., Hashimoto,R.F., Dantas,D., and Esteves,G.H. (2003). segmentation of microarray images by mathematical morphology. *Real-Time Imaging* 8, 491-505.
- Holly,J.M. (2000). Still more questions than answers: report on the 5th International Symposium on Insulin-like growth factors. *Trends Endocrinol. Metab.* 11, 151-153.
- Hoos,A., Stojadinovic,A., Singh,B., Dudas,M.E., Leung,D.H., Shaha,A.R., Shah,J.P., Brennan,M.F., Cordon-Cardo,C., and Ghossein,R. (2002). Clinical significance of molecular expression profiles of Hurthle cell tumors of the thyroid gland analyzed via tissue microarrays. *Am. J. Pathol.* 160, 175-183.

Huang,Y., Prasad,M., Lemon,W.J., Hampel,H., Wright,F.A., Kornacker,K., LiVolsi,V., Frankel,W., Kloos,R.T., Eng,C., Pellegata,N.S., and de la,C.A. (2001). Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. *Proc. Natl. Acad. Sci. U. S. A* 98, 15044-15049.

Hubank,M. and Schatz,D.G. (1994). Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res.* 22, 5640-5648.

Inohara,H., Honjo,Y., Yoshii,T., Akahani,S., Yoshida,J., Hattori,K., Okamoto,S., Sawada,T., Raz,A., and Kubo,T. (1999). Expression of galectin-3 in fine-needle aspirates as a diagnostic marker differentiating benign from malignant thyroid neoplasms. *Cancer* 85, 2475-2484.

Ivan,M., Bond,J.A., Prat,M., Comoglio,P.M., and Wynford-Thomas,D. (1997). Activated ras and ret oncogenes induce over-expression of c-met (hepatocyte growth factor receptor) in human thyroid epithelial cells. *Oncogene* 14, 2417-2423.

Jin,W., Riley,R.M., Wolfinger,R.D., White,K.P., Passador-Gurgel,G., and Gibson,G. (2001). The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet.* 29, 389-395.

Karayan,L., Hadjadj,S., Larsen,C.J., Marechaud,R., and Kraimps,J.L. (2002). [Update on differentiated thyroid carcinoma]. *Bull. Cancer* 89, 113-123.

Kato,M., Maeta,H., Kato,S., Shinozawa,T., and Terada,T. (2000). Immunohistochemical and in situ hybridization analyses of midkine expression in thyroid papillary carcinoma. *Mod. Pathol.* 13, 1060-1065.

Lakhani,S.R. and Ashworth,A. (2001). Microarray and histopathological analysis of tumours: the future and the past? *Nat. Rev. Cancer* 1, 151-157.

Lamar,E.E. and Palmer,E. (1984). Y-encoded, species-specific DNA in mice: evidence that the Y chromosome exists in two polymorphic forms in inbred strains. *Cell* 37, 171-177.

Learoyd,D.L., Messina,M., Zedenius,J., and Robinson,B.G. (2000). Molecular genetics of thyroid tumors and surgical decision-making. *World J. Surg.* 24, 923-933.

Lee,M.L., Kuo,F.C., Whitmore,G.A., and Sklar,J. (2000). Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc. Natl. Acad. Sci. U. S. A* 97, 9834-9839.

Lee,P.D., Sladek,R., Greenwood,C.M., and Hudson,T.J. (2002). Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome Res.* 12, 292-297.

Lee,S.W., Tomasetto,C., and Sager,R. (1991). Positive selection of candidate tumor-suppressor genes by subtractive hybridization. *Proc. Natl. Acad. Sci. U. S. A* 88, 2825-2829.

- Lemoine,N.R., Mayall,E.S., Wyllie,F.S., Williams,E.D., Goyns,M., Stringer,B., and Wynford-Thomas,D. (1989). High frequency of ras oncogene activation in all stages of human thyroid tumorigenesis. *Oncogene* 4, 159-164.
- Liang,P. and Pardee,A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257, 967-971.
- Lisitsyn,N., Lisitsyn,N., and Wigler,M. (1993). Cloning the differences between two complex genomes. *Science* 259, 946-951.
- Loeb,L.A. (2001). A mutator phenotype in cancer. *Cancer Res.* 61, 3230-3239.
- Macoska,J.A. (2002). The progressing clinical utility of DNA microarrays. *CA Cancer J. Clin.* 52, 50-59.
- Maeda,I., Takano,T., Matsuzuka,F., Maruyama,T., Higashiyama,T., Liu,G., Kuma,K., and Amino,N. (1999). Rapid screening of specific changes in mRNA in thyroid carcinomas by sequence specific-differential display: decreased expression of acid ceramidase mRNA in malignant and benign thyroid tumors. *Int. J. Cancer* 81, 700-704.
- Mahadevappa,M. and Warrington,J.A. (1999). A high-density probe array sample preparation method using 10- to 100-fold fewer cells. *Nat. Biotechnol.* 17, 1134-1136.
- Martin,K.J., Graner,E., Li,Y., Price,L.M., Kritzman,B.M., Fournier,M.V., Rhei,E., and Pardee,A.B. (2001). High-sensitivity array analysis of gene expression for the early detection of disseminated breast tumor cells in peripheral blood. *Proc. Natl. Acad. Sci. U. S. A* 98, 2646-2651.
- Matz,M., Shagin,D., Bogdanova,E., Britanova,O., Lukyanov,S., Diatchenko,L., and Chenchik,A. (1999). Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res.* 27, 1558-1560.
- Meireles,S.I., Carvalho,A.F., Hirata,R., Montagnini,A.L., Martins,W.K., Runza,F.B., Stolf,B.S., Termini,L., Neto,C.E., Silva,R.L., Soares,F.A., Neves,E.J., and Reis,L.F. (2003). Differentially expressed genes in gastric tumors identified by cDNA array. *Cancer Lett.* 190, 199-211.
- Miyake,H., Nelson,C., Rennie,P.S., and Gleave,M.E. (2000). Overexpression of insulin-like growth factor binding protein-5 helps accelerate progression to androgen-independence in the human prostate LNCaP tumor model through activation of phosphatidylinositol 3'-kinase pathway. *Endocrinology* 141, 2257-2265.
- Miyakoshi,N., Richman,C., Kasukawa,Y., Linkhart,T.A., Baylink,D.J., and Mohan,S. (2001). Evidence that IGF-binding protein-5 functions as a growth factor. *J. Clin. Invest* 107, 73-81.
- Nakayama,T., Ito,M., Ohtsuru,A., Naito,S., Nakashima,M., and Sekine,I. (1999). Expression of the ets-1 proto-oncogene in human thyroid tumor. *Mod. Pathol.* 12, 61-68.
- Notterman,D.A., Alon,U., Sierk,A.J., and Levine,A.J. (2001). Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res.* 61, 3124-3130.

Perks,C.M., Bowen,S., Gill,Z.P., Newcomb,P.V., and Holly,J.M. (1999). Differential IGF-independent effects of insulin-like growth factor binding proteins (1-6) on apoptosis of breast epithelial cells. *J. Cell Biochem.* 75, 652-664.

Perks,C.M., McCaig,C., and Holly,J.M. (2000). Differential insulin-like growth factor (IGF)-independent interactions of IGF binding protein-3 and IGF binding protein-5 on apoptosis in human breast cancer cells. Involvement of the mitochondria. *J. Cell Biochem.* 80, 248-258.

Perou,C.M., Sorlie,T., Eisen,M.B., van de,R.M., Jeffrey,S.S., Rees,C.A., Pollack,J.R., Ross,D.T., Johnsen,H., Akslen,L.A., Fluge,O., Pergamenschikov,A., Williams,C., Zhu,S.X., Lonning,P.E., Borresen-Dale,A.L., Brown,P.O., and Botstein,D. (2000). Molecular portraits of human breast tumours. *Nature* 406, 747-752.

Poblete,M.T., Nualart,F., del Pozo,M., Perez,J.A., and Figueroa,C.D. (1996). Alpha 1-antitrypsin expression in human thyroid papillary carcinoma. *Am. J. Surg. Pathol.* 20, 956-963.

Ponder,B.A. (2001). Cancer genetics. *Nature* 411, 336-341.

Preudhomme,C., Revillion,F., Merlat,A., Hornez,L., Roumier,C., Duflos-Grardel,N., Jouet,J.P., Cosson,A., Peyrat,J.P., and Fenaux,P. (1999). Detection of BCR-ABL transcripts in chronic myeloid leukemia (CML) using a 'real time' quantitative RT-PCR assay. *Leukemia* 13, 957-964.

Puskas,L.G., Zvara,A., Hackler,L., Jr., Micsik,T., and van Hummelen,P. (2002a). Production of bulk amounts of universal RNA for DNA microarrays. *Biotechniques* 33, 898-900, 902, 904.

Puskas,L.G., Zvara,A., Hackler,L., Jr., and van Hummelen,P. (2002b). RNA amplification results in reproducible microarray data with slight ratio bias. *Biotechniques* 32, 1330-4, 1336, 1338, 1340.

Quackenbush,J. (2001). Computational analysis of microarray data. *Nat. Rev. Genet.* 2, 418-427.

Ramaswamy,S. and Golub,T.R. (2002). DNA microarrays in clinical oncology. *J. Clin. Oncol.* 20, 1932-1941.

Resnick,M.B., Schacter,P., Finkelstein,Y., Kellner,Y., and Cohen,O. (1998). Immunohistochemical analysis of p27/kip1 expression in thyroid carcinoma. *Mod. Pathol.* 11, 735-739.

Richman,C., Baylink,D.J., Lang,K., Dony,C., and Mohan,S. (1999). Recombinant human insulin-like growth factor-binding protein-5 stimulates bone formation parameters in vitro and in vivo. *Endocrinology* 140, 4699-4705.

Said,S., Schlumberger,M., and Suarez,H.G. (1994). Oncogenes and anti-oncogenes in human epithelial thyroid tumors. *J. Endocrinol. Invest* 17, 371-379.

Saiki,R.K., Gelfand,D.H., Stoffel,S., Scharf,S.J., Higuchi,R., Horn,G.T., Mullis,K.B., and Erlich,H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487-491.

Schagdarsurengin,U., Gimm,O., Hoang-Vu,C., Dralle,H., Pfeifer,G.P., and Dammann,R. (2002). Frequent epigenetic silencing of the CpG island promoter of RASSF1A in thyroid carcinoma. *Cancer Res.* 62, 3698-3701.

Schena,M. (2003). *Microarray analysis.*, M.Schena, ed. (New Jersey: John Wiley & Sons).

Schena,M., Shalon,D., Davis,R.W., and Brown,P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467-470.

Serini,G., Trusolino,L., Saggiorato,E., Cremona,O., De Rossi,M., Angeli,A., Orlandi,F., and Marchisio,P.C. (1996). Changes in integrin and E-cadherin expression in neoplastic versus normal thyroid tissue. *J. Natl. Cancer Inst.* 88, 442-449.

Sgroi,D.C., Teng,S., Robinson,G., LeVangie,R., Hudson,J.R., Jr., and Elkahoul,A.G. (1999). In vivo gene expression profile analysis of human breast cancer progression. *Cancer Res.* 59, 5656-5661.

Shaha,A.R. (2000). Controversies in the management of thyroid nodule. *Laryngoscope* 110, 183-193.

Shipp,M.A., Ross,K.N., Tamayo,P., Weng,A.P., Kutok,J.L., Aguiar,R.C., Gaasenbeek,M., Angelo,M., Reich,M., Pinkus,G.S., Ray,T.S., Koval,M.A., Last,K.W., Norton,A., Lister,T.A., Mesirov,J., Neubergh,D.S., Lander,E.S., Aster,J.C., and Golub,T.R. (2002). Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med.* 8, 68-74.

Shuja,S. and Murnane,M.J. (1996). Marked increases in cathepsin B and L activities distinguish papillary carcinoma of the thyroid from normal thyroid or thyroid with non-neoplastic disease. *Int. J. Cancer* 66, 420-426.

Silva,A.M., Pires,E.G., Abrantes,E.F., Ferreira,L.R., Gazzinelli,R.T., and Reis,L.F. (1999). Application of the differential display RT-PCR strategy for the identification of inflammation-related mouse genes. *Braz. J. Med. Biol. Res.* 32, 845-852.

Slootweg,M.C., Ohlsson,C., van Elk,E.J., Netelenbos,J.C., and Andress,D.L. (1996). Growth hormone receptor activity is stimulated by insulin-like growth factor binding protein 5 in rat osteosarcoma cells. *Growth Regul.* 6, 238-246.

Sorlie,T., Perou,C.M., Tibshirani,R., Aas,T., Geisler,S., Johnsen,H., Hastie,T., Eisen,M.B., van de,R.M., Jeffrey,S.S., Thorsen,T., Quist,H., Matese,J.C., Brown,P.O., Botstein,D., Eystein,L.P., and Borresen-Dale,A.L. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U. S. A* 98, 10869-10874.

Southern,E., Mir,K., and Shchepinov,M. (1999). Molecular interactions on microarrays. *Nat. Genet.* 21, 5-9.

Stolf,B.S., Carvalho,A.F., Martins,W.K., Runza,F.B., Brun,M., Hirata,R., Jr., Jordao,N.E., Soares,F.A., Postigo-Dias,I., Kowalski,L.P., and Reis,L.F. (2003). Differential expression of IGFBP-5 and two human ESTs in thyroid glands with goiter, adenoma and papillary or follicular carcinomas. *Cancer Lett.* 191, 193-202.

Takahashi,M., Rhodes,D.R., Furge,K.A., Kanayama,H., Kagawa,S., Haab,B.B., and Teh,B.T. (2001). Gene expression profiling of clear cell renal cell carcinoma: gene identification and prognostic classification. *Proc. Natl. Acad. Sci. U. S. A* 98, 9754-9759.

Takano,T., Hasegawa,Y., Matsuzuka,F., Miyauchi,A., Yoshida,H., Higashiyama,T., Kuma,K., and Amino,N. (2000). Gene expression profiles in thyroid carcinomas. *Br. J. Cancer* 83, 1495-1502.

Takano,T., Matsuzuka,F., Sumizaki,H., Kuma,K., and Amino,N. (1997). Rapid detection of specific messenger RNAs in thyroid carcinomas by reverse transcription-PCR with degenerate primers: specific expression of oncofetal fibronectin messenger RNA in papillary carcinoma. *Cancer Res.* 57, 3792-3797.

Takano,T., Miyauchi,A., Matsuzuka,F., Liu,G., Higashiyama,T., Yokozawa,T., Kuma,K., and Amino,N. (1999a). Preoperative diagnosis of medullary thyroid carcinoma by RT-PCR using RNA extracted from leftover cells within a needle used for fine needle aspiration biopsy. *J. Clin. Endocrinol. Metab* 84, 951-955.

Takano,T., Miyauchi,A., Yokozawa,T., Matsuzuka,F., Maeda,I., Kuma,K., and Amino,N. (1999b). Preoperative diagnosis of thyroid papillary and anaplastic carcinomas by real-time quantitative reverse transcription-polymerase chain reaction of oncofetal fibronectin messenger RNA. *Cancer Res.* 59, 4542-4545.

Tamayo,P., Slonim,D., Mesirov,J., Zhu,Q., Kitareewan,S., Dmitrovsky,E., Lander,E.S., and Golub,T.R. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. U. S. A* 96, 2907-2912.

Tanaka,M., Adzuma,K., Iwami,M., Yoshimoto,K., Monden,Y., and Itakura,M. (1995a). Human calgizzarin; one colorectal cancer-related gene selected by a large scale random cDNA sequencing and northern blot analysis. *Cancer Lett.* 89, 195-200.

Tanaka,T., Umeki,K., Yamamoto,I., Sakamoto,F., Noguchi,S., and Ohtaki,S. (1995b). CD26 (dipeptidyl peptidase IV/DPP IV) as a novel molecular marker for differentiated thyroid carcinoma. *Int. J. Cancer* 64, 326-331.

Tennant,M.K., Thrasher,J.B., Twomey,P.A., Birnbaum,R.S., and Plymate,S.R. (1996). Insulin-like growth factor-binding proteins (IGFBP)-4, -5, and -6 in the benign and malignant human prostate: IGFBP-5 messenger ribonucleic acid localization differs from IGFBP-5 protein localization. *J. Clin. Endocrinol. Metab* 81, 3783-3792.

Terrier,P., Sheng,Z.M., Schlumberger,M., Tubiana,M., Caillou,B., Travagli,J.P., Fragu,P., Parmentier,C., and Riou,G. (1988). Structure and expression of c-myc and c-fos proto-oncogenes in thyroid carcinomas. *Br. J. Cancer* 57, 43-47.

Trachte,A.L., Suthers,S.E., Lerner,M.R., Hanas,J.S., Jupe,E.R., Sienko,A.E., Adesina,A.M., Lightfoot,S.A., Brackett,D.J., and Postier,R.G. (2002). Increased

- expression of alpha-1-antitrypsin, glutathione S-transferase pi and vascular endothelial growth factor in human pancreatic adenocarcinoma. *Am. J. Surg.* 184, 642-647.
- Triche,T.J., Schofield,D., and Buckley,J. (2001). DNA microarrays in pediatric cancer. *Cancer J.* 7, 2-15.
- van de Vijver,M.J., He,Y.D., van't Veer,L.J., Dai,H., Hart,A.A., Voskuil,D.W., Schreiber,G.J., Peterse,J.L., Roberts,C., Marton,M.J., Parrish,M., Atsma,D., Witteveen,A., Glas,A., Delahaye,L., van,d., V, Bartelink,H., Rodenhuis,S., Rutgers,E.T., Friend,S.H., and Bernards,R. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347, 1999-2009.
- van Erp,H.E., van Unnik,J.A., Rijksen,G., Smits,J.G., and Staal,G.E. (1991). Cellular expression of K-type pyruvate kinase in normal and neoplastic human tissues. *Cancer* 68, 2595-2601.
- Van Gelder,R.N., von Zastrow,M.E., Yool,A., Dement,W.C., Barchas,J.D., and Eberwine,J.H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. U. S. A* 87, 1663-1667.
- Velculescu,V.E., Zhang,L., Vogelstein,B., and Kinzler,K.W. (1995). Serial analysis of gene expression. *Science* 270, 484-487.
- Wan,J.S., Sharp,S.J., Poirier,G.M., Wagaman,P.C., Chambers,J., Pyati,J., Hom,Y.L., Galindo,J.E., Huvar,A., Peterson,P.A., Jackson,M.R., and Erlander,M.G. (1996). Cloning differentially expressed mRNAs. *Nat. Biotechnol.* 14, 1685-1691.
- Wang,E., Miller,L.D., Ohnmacht,G.A., Liu,E.T., and Marincola,F.M. (2000). High-fidelity mRNA amplification for gene profiling. *Nat. Biotechnol.* 18, 457-459.
- Wang,S., Wu,J., Savas,L., Patwardhan,N., and Khan,A. (1998). The role of cell cycle regulatory proteins, cyclin D1, cyclin E, and p27 in thyroid carcinogenesis. *Hum. Pathol.* 29, 1304-1309.
- Weiss,M., Baruch,A., Keydar,I., and Wreschner,D.H. (1996). Preoperative diagnosis of thyroid papillary carcinoma by reverse transcriptase polymerase chain reaction of the MUC1 gene. *Int. J. Cancer* 66, 55-59.
- Welsh,J., Chada,K., Dalal,S.S., Cheng,R., Ralph,D., and McClelland,M. (1992). Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res.* 20, 4965-4970.
- Wieland,I., Bolger,G., Asouline,G., and Wigler,M. (1990). A method for difference cloning: gene amplification following subtractive hybridization. *Proc. Natl. Acad. Sci. U. S. A* 87, 2720-2724.
- Yamamoto,Y., Izumi,K., and Otsuka,H. (1992). An immunohistochemical study of epithelial membrane antigen, cytokeratin, and vimentin in papillary thyroid carcinoma. Recognition of lethal and favorable prognostic types. *Cancer* 70, 2326-2333.
- Yang,Y.H., Buckley,M.J., Dudoit,S., and Speed,T.P. (2002). Comparison of methods for image analysis on cDNA microarray data. *Journal of Computational and Graphical Statistics* 11, 108-136.

Yang,Y.H. and Speed,T. (2002). Design issues for cDNA microarray experiments. *Nat. Rev. Genet.* 3, 579-588.

Zhang,L., Zhou,W., Velculescu,V.E., Kern,S.E., Hruban,R.H., Hamilton,S.R., Vogelstein,B., and Kinzler,K.W. (1997). Gene expression profiles in normal and cancer cells. *Science* 276, 1268-1272.

Zhang,W., Laborde,P.M., Coombes,K.R., Berry,D.A., and Hamilton,S.R. (2001). Cancer genomics: promises and complexities. *Clin. Cancer Res.* 7, 2159-2167.

CURRICULUM VITAE

NOME: Beatriz Simonsen Stolf

LOCAL E DATA DE NASCIMENTO: São Paulo, 17 de novembro de 1973

EDUCAÇÃO

Segundo Grau: Colégio Santa Cruz (1989 a 1991)

Superior: Universidade de São Paulo

Bacharelado em Ciências Biológicas (1992-1995)

Licenciatura em Ciências Biológicas (1992-1996)

Mestrado em Bioquímica, Instituto de Química, Universidade de São Paulo (1996-1999)

OCUPAÇÃO

Doutorado em Bioquímica, Instituto de Química, Universidade de São Paulo

Parte experimental realizada no Instituto Ludwig de Pesquisa sobre o câncer

Desde 07/1999

Bolsista FAPESP

PUBLICAÇÕES

- Floeter-Winter, LM, Souto, RP, Stolf, BS, Zingales, B & Buck, GA. Can activity of the rRNA promoter be used as a marker of speciation? *Exp. Parasitol.* 1997; 86, 232-234
- Zingales, B, Stolf, BS, Souto, RP, Fernandes, O & Briones, M. Epidemiology, biochemistry and evolution of *Trypanosoma cruzi* lineages based on ribosomal RNA sequences. *Mem Inst Oswaldo Cruz.* 1999; 94 Suppl 1:159-64
- Briones, MRS, Souto, RP, Stolf, BS & Zingales, B. Evolution of *Trypanosoma cruzi* lineages and its possible correlation with American faunal exchanges in the Cenozoic. *Mol. Biochem Parasitol.* 1999; 104(2):219-32
- Stolf, BS, Souto, RP, Pedroso, A & Zingales, B Two types of ribosomal RNA genes in hybrid *Trypanosoma cruzi* strains- *Mol. Biochem. Parasitol.* 2003; 126, 73-80

No período do doutorado:

- Stolf, BS, Carvalho, AF, Martins, WK, Runza, FB, Brun, M, Hirata Jr., R, Neves, EJ, Soares, FA, Postigo-Dias, J, Kowalski, LP & Reis, LFL. (2003) Differential expression of IGFBP-5 and two human ESTs in thyroid glands with goiter, adenoma and papillary or follicular carcinomas. *Cancer Letters* 191, 193-202

- Meireles, SI, Carvalho, AF, Hirata, R, Montagnini, AL, Martins, WK, Runza, FB, Stolf, BS, Termini, L, Neto, CE, Silva, RL, Soares, FA, Neves, EJ & Reis, LF. (2003). Differentially expressed genes in gastric tumors identified by cDNA array. *Cancer Letters* 190, 199-211
- Gomes, L.I., Silva, R.L., Stolf, B.S., Cristo, E.B. *et al.* Comparative analysis of amplified and non-amplified RNA for hybridization in cDNA microarray. *Analytical Biochemistry* 321, 244-51
- Stolf, B. S. Marcadores tumorais nas neoplasias da glândula tireóide, do livro Glândulas tireóide e paratireóides- abordagem clínico-cirúrgica. Editora Atheneu, no prelo

Differential expression of IGFBP-5 and two human ESTs in thyroid glands with goiter, adenoma and papillary or follicular carcinomas

Beatriz S. Stolf^{a,b}, Alex F. Carvalho^a, Waleska K. Martins^a, Franco B. Runza^a,
Marcel Brun^c, Roberto Hirata Jr.^{c,d}, Eduardo Jordão Neves^e, Fernando A. Soares^e,
Juan Postigo-Dias^f, Luis P. Kowalski^e, Luiz F.L. Reis^{a,e,*}

^aLudwig Institute for Cancer Research, Rua Professor Antonio Prudente 109, São Paulo 01509-010, Brazil

^bInstituto de Química, Universidade de São Paulo, São Paulo, Brazil

^cInstituto de Matemática e Estatística, Universidade de São Paulo, São Paulo, Brazil

^dSENAC College of Computer Science and Technology, São Paulo, São Paulo, Brazil

^eHospital do Câncer A. C. Camargo, São Paulo, São Paulo, Brazil

^fInstituto de Enfermedades Neoplásicas-INEN, Lima, Peru

Received 15 March 2002; received in revised form 28 April 2002; accepted 10 May 2002

Abstract

Here, we describe the identification of three human genes with altered expression in thyroid diseases. One of them corresponds to insulin-like growth factor binding protein 5 (IGFBP5), which has already been described as over expressed in other cancers and, for the first time, is identified as overexpressed in thyroid tumors. The other genes, named 44 and 199, are ESTs with yet unknown function and were mapped on human chromosomes seven and four, respectively. We determined by RT-PCR the expression level of these genes in ten samples of disease-free thyroid, ten of goiter, nine of papillary carcinoma, ten of adenoma and seven of follicular carcinoma and the significance of observed differences was statistically determined. IGFBP-5 and gene 44 were significantly overexpressed in papillary carcinoma when compared to normal and goiter. Genes 44 and 199 were differentially expressed in follicular carcinoma and adenoma when compared to normal thyroid tissue.

© 2002 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Thyroid tumors; Goiter; cDNA array; Differential-display RT-PCR; Tumor markers

1. Introduction

Approximately 4–5% of the general population has palpable thyroid nodules [1]. The most common thyroid disorder responsible for nodule formation is multinodular goiter. The molecular events associated

with goiters appear to be related to genetic heterogeneity of follicular cells regarding growth and function and the acquisition of inheritable genetic changes during replication [2]. Further hyperplasia occurs in response to increased thyroid stimulating hormone (due to iodine deficiency, goitrogens or defects in thyroid hormone production) or other thyroid-stimulating factors (EGF and insulin-like growth factors 5 (IGFs)), which aggravate cell

* Corresponding author. Tel.: +55-11-2704922; fax: +55-11-2707001.

E-mail address: lreis@ludwig.org.fr (L.F.L. Reis).

variability, resulting in growth and focal autonomic function [2].

According to world statistics (OncoLink- <http://www.oncolink.com>), thyroid carcinomas correspond to 1.1% of the malignant tumors, with an incidence of 1–10/100 000 people. The initial detection of these tumors is often made after clinical examination and identification of thyroid nodules. Four to 17% of multinodular goiters removed surgically contain carcinoma, usually of papillary type [3–5]. Some authors believe that the incidence of thyroid cancer does not change according to iodine deficiency [6] whereas others have shown that these tumors are more prevalent in areas of endemic goiter [7,8]. Nevertheless, causal relationship between carcinoma and goiter has not been formally determined [7]. Patients submitted to thyroidectomy may have their glands more carefully analyzed than the general population, which could introduce a bias. It is also questionable whether carcinomas arise from adenomas. A strong argument against this hypothesis is that most carcinomas are papillary, while the majority of the adenomas are of follicular type [9]. In fact, histological features of adenomas are very similar to follicular carcinomas and the distinction between these two pathological conditions is based solely on detection of capsule invasion [9].

The molecular differences among pathological conditions of the thyroid gland are not well determined and several groups have reported studies aiming to identify genes that are differentially expressed in thyroid diseases. It was shown by Maeda and co-workers that the mRNA level for the acid ceramidase is lower in follicular adenomas and in follicular, papillary and anaplastic carcinomas than in normal thyroid tissue [10]. Higher levels of oncofetal fibronectin were observed in papillary and anaplastic carcinomas when compared to follicular carcinomas and benign diseases [11]. Also, it has been well documented that RET oncogene is frequently rearranged in papillary carcinomas leading to its activation [12,13]. Activation of RET is also a frequent finding in medullary carcinomas where a point mutation appears to be responsible for augmented RET mRNA levels [14,15]. Interestingly, activation of RET led to transformation of NTH3T3 cells [16,17] and has been implicated in the oncogenic process of these thyroid tumors [18]. Finally,

expression of the genes encoding for calcitonin and carcinoembryonic antigen appears to be increased exclusively in medullary carcinomas when compared to follicular and papillary carcinomas, goiters and adenomas [15].

In this study, we describe the identification, by differential-display RT-PCR (DDRT-PCR), of three genes with altered expression in goiters, adenomas, carcinomas (follicular and papillary), or normal thyroid tissues. One gene, IGFBP-5, has already been identified as over expressed in other malignant diseases whereas the other two genes, named 44 and 199, are human Expressed Sequence tag (EST)s with yet unknown function. The expression levels of these three genes were determined in 46 samples of normal, goiter, and benign and malignant thyroid tumors by semi-quantitative RT-PCR. Statistical analysis confirmed their differential expression.

2. Materials and methods

2.1. Tissue samples

Post-surgical tissues were obtained from the Head and Neck Department, Hospital do Câncer, São Paulo and Instituto de Enfermedades Neoplásicas-INEN, Lima, Perú. All patients signed an informed consent and the procedure was approved by the Internal Ethical Committee from both institutions. Excess of tissue not used for histopathology was immediately transferred to liquid nitrogen and kept frozen until RNA extraction. Before RNA extraction histopathological diagnosis was re-confirmed, all samples were micro-dissected and only tissues with at least 70% of tumor cells and no visible infiltrating inflammatory cells were used as tumor. Morphologically disease-free tissue obtained from surgical margins was considered as normal. Detailed description of samples and patients can be obtained at our web page as complementary information (<http://array.ludwig.org.br>, public area, Stolf et al., 2002). Five types of thyroid tissues were used in this work: normal tissue (N), goiter (G), papillary carcinoma (PC), adenoma (A) and follicular carcinoma (FC).

2.2. RNA extraction and DNaseI treatment

RNA extraction was performed using TRIZOL (Life Technologies) and a mechanical tissue homogenizer (Polytron). Each 50 µg of total RNA was submitted to DNaseI treatment at 37 °C for 1 h using 30 U of DNaseI, 120 U of RNasin, 1 mM DTT and DNase buffer in a total volume of 300 µl. The reaction was terminated by the addition of 75 µl stop solution (Promega) and RNA was purified by phenol-chloroform extractions. Samples with OD_{260/280} ratio > 1.7 were used in the analysis.

2.3. DDRT-PCR and selection of differentially expressed genes

Differential-display experiments, isolation of bands of interest and re-amplification of recovered fragments were performed as previously described [19]. cDNAs were reamplified by PCR and cloned into pUC18 using the Sureclone ligation kit (Amersham Pharmacia Biotech) or into PGEM using PGEM-T Easy system (Promega). Differentially represented bands were excised from gels and, at least two independent clones were kept from each band. The insert from each clone was amplified by PCR with M13 forward and reverse primers and immobilized on a nylon membrane with the aid of a robot (Flexys, Genome Solutions). For screening of differentially expressed genes, a pool of RNA from three patients with a given pathology was hybridized with three identical arrays. For each tissue type (N, G, or PC), two independent pools were used. Clones showing a consistent expression ratio higher than two when any two kinds of thyroid tissue were compared were considered as differentially expressed. For these clones, plasmid DNA was isolated using Wizard Plus Minipreps (Promega) and inserts were sequenced using pUCIM13 forward and reverse primers in an ABI 377 (PE Applied Biosystems). Sequence homology were analyzed using the BLAST program available at the National Center of Biotechnology Information home page (<http://www.ncbi.nlm.nih.gov>).

2.4. Semi-quantitative RT-PCR and Southern blot

For RT-PCR, aliquots of 2 µg of DNA-free total

RNA were reverse transcribed using random hexanucleotides (Amersham Pharmacia) and Superscript II reverse transcriptase (GIBCO BRL).

For the establishment of non-saturating RT-PCR conditions, an experiment with genes 18 and 44 was carried out, using RNA samples from two normal tissues, two samples of goiter and two papillary carcinomas. As reference, we amplified the cDNA corresponding to the glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA from the same samples. Amplification was performed for 14, 28 and 35 cycles. Amplified products were fractionated on agarose gel and stained with ethidium bromide. Based on the parameters obtained, comparative analysis of RNA levels for genes 18, 44 and 199 was performed with 20 cycles. As internal reference, we amplified (28 cycles) a fragment of the TATA-binding protein (TBP) cDNA [20]. The levels of mRNA for these four genes were determined in ten samples of normal tissue, ten goiters, nine papillary carcinomas, ten adenomas and seven follicular carcinomas. RT-PCR products were fractionated through agarose gel, blotted onto nylon membranes and hybridized with ³²P-czdCTP-labeled probes. Filters were exposed to a Phosphorimager (Storm 840, Molecular Dynamics) and the volume (sum of all the pixel intensity values in a given band minus the background) for each band was determined. Normalization was achieved by dividing the volume of the genes in each sample by the volume of TBP gene in the corresponding sample. The ratio was considered as arbitrary units.

The primers used for PCR amplification of 18, 44, 199, TBP and GAPDH genes were as follows: for gene 18, 5' CAGCAGTATGGACAGCTGGA 3' and 5' TTAAATGAGGGCTGAAACGG 3'; for gene 44, 5' CCTGTTGATGGGATGTTTCC 3' and 5' CTGCTGTTGCAGGAAAATGA 3'; for gene 199, 5' GGAGGTGAGACAGGGTCTA 3' and 5' GAACTCAGGTGATACTCCCA 3'; for TBP gene, 5' CCACAGCTCTTCCACTCACA 3' and 5' ATCCTCATGATTACCGCAGC 3'; and for GAPDH, 5' CTGCACCACCAACTGCTTA 3' and 5' CTAGACGGCAGGTCAGGTC 3'.

We used boxplot (implemented in Matlab, MathWorks®) to represent graphically the expression levels in each kind of tissue. Each box represents half of the data points and the line in the middle of the box

Fig. 1. Identification of expressed sequences by differential display. DDRT-PCR experiments using pools of DNA-free total RNA from samples of normal thyroid tissue (N), Papillary Carcinoma (PC) and Goiter (G), or Follicular Carcinoma (FC), Adenoma (A) and N. The arrowhead indicates the bands corresponding to genes 18, 44 and 199. The amplification reactions used primers T11VA/random 14 (5'CTGCTCTCAA 3'), T11VC/random 15 (5'CTTGATTGCC 3') and T11VC/random 14 for the identification of 18, 44 and 199 genes, respectively.

represents the median of the sample distribution. The lines extending from the end of the box represent the extension of the extreme values of the data within 1.5 interquartile distances from the median and the cross indicates the outliers.

2.5. Unsupervised classification: clustering analysis

Data analysis was done using Matlab (Math Works®) and its statistical toolbox. Primary data was log-transformed and analyzed by hierarchical clustering, Euclidean distance, and complete linkage. To evaluate the statistical significance of the resulting clusters we applied bootstrap strategies. Using this resampling technique, several simulated gene expression datasets were generated under the null hypothesis that the gene expression probability distributions are the same for the conditions under comparison. The classification method was applied to each simulated dataset and the proportion with an equivalent or better classification performance provides an estimate for the *P*-value.

2.6. Supervised classification: decision tree classifier

To determine whether the differentially expressed genes could eventually be used in a classifier we used two simple approaches. First we searched for the best linear separation for the gene expression datapoints corresponding to the conditions of interest. Next we used the Decision Tree method [21,22] because it is

easy to implement and may provide structure with meaningful interpretation. The method we implemented was based on splitting the classified samples (based on gene expression levels) recursively into two classes until the remaining classes are completely homogeneous with respect to the classification criteria (Twoing). The efficiency of the Decision Tree Method was evaluated by a cross validation procedure. To do that, we trained a decision tree with all samples but one. The sample that was excluded from the training procedure was then used to test the classifier. Doing that for all possible combinations, we had the known leave one out cross validation. The evaluation of the method was done by counting the number of times it correctly classified the class of a sample in relation to the number of samples in that class.

3. Results

3.1. Identification of genes 18, 44 and 199

The mRNA profiles of normal thyroid tissue, goiter, adenoma and carcinomas were compared by DDRT-PCR. We identified and isolated a total of 213 bands corresponding to transcripts showing different intensity in either of two types of comparison: among N, G, PC or among normal tissue, FC and A (Fig. 1). For each band excised and sub-cloned we selected at least two colonies, leading to a total of 429 clones.

Fig. 2. RT-PCR of genes 18 and 44 in two samples of PC, N and G. Two micrograms of DNA-free total RNA from two samples of N, G or PC were reversed transcribed with SuperScript (Amersham) and oligodT as primer. Amplification of the six samples and the negative control (0-RT reaction without Superscript) was performed for 14, 28 and 35 cycles. GAPDH was amplified from the same RT reactions as an internal control. Amplified samples were fractionated through an agarose gel and stained with ethidium bromide.

The 429 clones arrayed in nylon membranes were hybridized to pools of RNA (three patients) of each pathological condition, in two independent experiments. Clones 18, 44, and 199 were selected based on their differential expression in all replicas of at least one independent hybridization (data not shown).

3.2. Identity of clones 18, 44 and 199

Inserts from the four clones were sequenced and analyzed for sequence homology using the BLAST tool available at the NCBI homepage [23]. Clone 18 corresponds to a fragment of 175 bp showing 100% identity to insulin-like growth factor binding protein 5 (IGFBP-5; positions 3480–3654, acc. number Hs 107169). Clone 44 (deposited on genebank with acc. number 486818) corresponds to a cDNA fragment located on a sequenced BAC clone (acc. AC002543) and showed 99% identity to two previously sequenced cDNAs, FLJ22131 (accession number AK025784) and IMAGE clone 4452921, acc. BC016592). Homology search against the human genome sequence revealed that clone 44 mapped on chromosome 7. Clone 199 (deposited on genebank with acc. number 486819) mapped into a human genome sequence

Fig. 3. Box plots showing the normalized expression values of genes 18 (IGFBP5), 44, and 199 in thyroid tissues as determined by semi-quantitative RT-PCR. In Panel (A), we represent expression levels of genes 18 and 44 in ten samples of G, ten N and nine PC. In Panel (B) we represent expression levels of genes 44 and 199 in ten samples of A, nine samples of N and seven samples of FC. Arbitrary unit values determined by dividing the volume of bands from genes 18, 44, or 199 in each sample by the volume for the TBP band in the corresponding sample are represented in box plots. The horizontal line inside the boxes shows the median of the samples (values are shown) in each tissue type; each box represents half of the data points and the lines extending from the end of the box represent the extension of the extreme values of the data within 1.5 interquartile distance from the median. Outliers are indicated by a cross.

located on chromosome 4 (AC099397, human chromosome 4 working draft). In the genome, sequence from clone 199 is spilled into two halves, separated by about 16 kb. The second half (nucleotides 130–248) showed homology to a single human EST deposited in GenBank (acc. number T60530).

Fig. 4. Cluster analysis of the ten samples of N, nine PC (C) and ten G using normalized expression values for genes 18 and 44. Color scale from green to red represent increasing expression levels for genes 18 (column one) and 44 (column two) in all 30 samples (rows). Samples belonging to the cluster containing predominantly N and G are shown in red, and samples of the cluster containing predominantly C are represented in blue. Non-elusterized (original) data are shown in the left color table. Clusterized data are shown in the right color table (column three). Dendrogram obtained after clustering is shown in the right side of the figure.

3.3. Semi-quantitative RT-PCR

In order to analyze the expression profile of genes 18, 44 and 199 in individual samples, we performed RT-PCR under non-saturating conditions using individual samples of DNaseI- treated RNAs from each tissue. In Fig. 2 we show RNA levels of genes 18 and 44 in two representative samples of N, G and PC. The same procedure followed by Southern blot was adopted for genes 18, 44 and 199 in all 46 RNA samples. Comparison was performed among N, G and PC samples and among N, A and FC samples. The corresponding gene expression boxplots are shown in Figs. 3A and 3B. This analysis revealed a pattern of

Fig. 5. Scatter plot of log-transformed expression levels of genes 18, 44, and 199 in 46 thyroid samples. The log-transformed value of expression levels of genes 18, 44, and 199 was determined as described in Fig. 3 and plotted pair wise. In panel (A), we represent the mRNA levels for genes 18 and 44 in ten samples of N (squares), nine samples of PC (triangles) and ten samples of G (circles). In panel (B), we represent the mRNA levels for genes 44 and 199 in nine samples of N (squares), ten samples of A (circles) and seven samples of FC (triangles).

expression compatible with that expected from the DDRT-PCR: genes 18 and 44 showed higher expression in PC than in G and N (compare Figs. 1 and 3) and gene 199 is more expressed in A than in N and, to a lesser extent, in FC (compare Figs. 1 and 3). Moreover, gene 44 showed higher expression in N than A and FC. Gene 199 was expressed at comparable levels in N, G and PC whereas gene 18 showed similar expression in N, A, FC (data not shown).

3.4. Unsupervised and supervised classification

Based upon the expression values for genes 18 and 44 in samples of G, PC and N depicted in Fig. 3, we performed hierarchical cluster analysis with Euclidean distance and complete linkage. Three and two clusters grouping were evaluated. For the three-cluster grouping, after the early dichotomy, the branch containing predominantly goiters and normal tissues was further subdivided in a small group containing one normal and one carcinoma sample (data not shown). We therefore chose two-cluster analysis as the most meaningful. Results are presented as a color dendrogram (left) and a color diagram for the expression values (right), in Fig. 4. We conclude that for G, PC and N, malignant and benign diseases are discriminated with considerable sensitivity and specificity, since one of the clusters (red) contains all G, nine N and two PC, while the other (blue) contains seven PC and one N. To evaluate whether such automatic discrimination is not simply due to randomness we performed the bootstrap test described in Section 2. For 100 000 simulated datasets, only three resulted in an equivalent or better classification performance.

To assess the discrimination performance of each gene we used two non-parametric hypothesis tests: the Wilcoxon Rank Sum test and the Kolmogorov–Smirnov test [24]. Let us denote by P_w and P_{KS} respectively, the P -values for the Wilcoxon and Kolmogorov–Smirnov tests. We determined that IGFBP5 (gene 18) expression allows distinction between goiter and non-goiter samples ($P_w = 0.0019$ and $P_{KS} = 0.0012$ for $G \times N$; $P_w = 0.00097$ and $P_{KS} = 0.00033$ for $G \times PC$), whereas gene 44 allows discrimination of cancer from non-cancer samples ($P_w = 0.028$ and $P_{KS} = 0.013$ for $G \times PC$; $P_w = 0.0018$ and $P_{KS} = 0.0026$ for $N \times PC$). This discrimination capability can also be visualized by examining the expression levels of both genes in Figs. 3A and 5A.

The best linear separator that classifies malignant (papillary carcinoma) and non-malignant (goiter plus normal tissues) conditions yields two discrepant results, both corresponding to papillary carcinoma that were grouped with normal and goiter samples. A binary tree supervised classifier for distinction between papillary carcinoma and non-papillary car-

cinoma has been developed based on the expression levels of genes 18 and 44. Using cross-validation, the accuracy of this classifier was 80% and 89% for the non-malignant and malignant classifications, respectively.

When analyzing the expression of genes 18, 44 and 199 in N, A and FC samples no reliable cluster was observed. Using again the non-parametric tests and notations as above, we determined that gene 44 allows distinction between normal and neoplastic samples ($P_w = 0.0079$ and $P_{KS} = 0.011$ for $FC \times N$; $P_w = 0.0015$ and $P_{KS} = 0.0017$ for $A \times N$), while gene 199 allows good discrimination between adenomas and normal tissues ($P_w = 0.000087$ and $P_{KS} = 0.00026$). The best linear separator allows perfect separation between normal and neoplastic samples. The same method also separates follicular carcinomas from the remaining samples with three misclassifications, corresponding to adenomas incorrectly classified as FCs. With the cross validation, a binary tree classifier to distinguish normal from neoplastic samples built using the expression values of those samples (Fig. 5B) had accuracies of 78% and 88% for the percentages of normal or neoplastic correct classifications, respectively.

4. Discussion

The aim of this work was to identify genes differentially expressed in thyroid diseases that could help the understanding of the molecular processes involved in these disorders and, eventually, serve as molecular markers.

DDRT-PCR technique enabled us to identify 213 bands that showed different intensities when we compared the expression profile of several thyroid diseases and could potentially represent differentially expressed genes. Two of these genes, 18 and 44, were identified as over expressed in papillary carcinomas, and one, 199, was identified as over expressed in adenomas.

Gene 18 corresponds to IGFBP-5, which we identify, for the first time, as overexpressed in thyroid cancer. IGFBPs comprise a family of proteins that interact with IGFs and play a major role in regulating cell proliferation, differentiation, apoptosis and trans-formation (review in Ref. [25]). The levels of IGFs

and IGFBPs have been determined in various cancers and they can be found either augmented or diminished. (Reviewed in Ref. [26]).

The precise functions and patterns of expression of IGFBP-5 have not been determined (reviewed in Ref. [27]). IGFBP-5 has been implicated in growth stimulation of osteoblast proliferation [28–30] as well as of prostate cancer cells [31] and its mRNA is augmented in malignant prostate epithelium compared to benign tissue [32]. IGFBP-5 also improves survival of breast cancer cells [33,34] but, in the presence of some specific signals such as Vitamin D, IGFBP-5 inhibits proliferation of breast carcinoma-derived cell lines [35]. The involvement of IGFBP-5 in promoting cell growth or in improving survival would be very much in agreement with our findings of higher expression in thyroid papillary carcinoma.

Clone 44 is probably derived from a rare transcript since it showed homology to only two previously described ESTs. One of these ESTs (AK025784) was cloned from a cDNA library derived from the human hepatoma cell line HepG2 by the NEDO human cDNA sequencing project. The second EST (BCO 16592) was cloned from a kidney hypernephroma.

Clone 199 appears to have two exons and both contain repetitive sequences of the Alu family, what accounts for their homology to several genomic clones and ESTs. A single previously identified human EST showed significant similarity to the entire second exon, including non-repetitive sequences. Thus, it appears that clones 44 and 199 are both derived from rare transcripts corroborating the idea that, besides a bias for detecting differentially expressed genes, DDRT-PCR is a powerful methodology that can amplify both rare as well as more abundant transcripts [36,37].

The expression of genes 18, 44, and 199 in 46 different thyroid samples was determined by RT-PCR and Southern Blot. As observed in other studies [10, 11], there were variations in the expression levels of each gene among samples, even between those of the same pathological condition. It has been extensively documented, especially after the introduction of microarray technology, that tumors from a given organ differ significantly regarding to the expressed genes, what is apparently responsible for the differences in their biology [38,39]. In order to interpret the data from our analysis, a variety of statistical tools

were applied. Despite individual variations, unsupervised and supervised procedures using the expression values of genes 18 and 44 confirmed that the differences in their expression among normal, goiter, and papillary carcinomas were statistically significant. Either bootstrap analysis of the cluster or non-parametric tests have supported the distinction between benign (normal tissue and goiter) and malignant disease. This distinction was made with 83% certainty by a classifier built based upon the same expression values. Similar analysis was performed to confirm the differential expression of these genes in normal thyroid tissue and in those with adenomas and with follicular carcinomas.

Adenomas and follicular carcinomas showed overlapping expression profiles, as expected from the high structural and cytological similarities between the two diseases. In fact, the major difference between follicular carcinomas and adenomas is invasion of the capsule, which might be related to or accompanied by changes in gene expression [9]. The identification of genes differentially expressed in these two pathologies would be of great importance for improving specificity of diagnosis, especially in samples from fine needle aspiration biopsy [40]. An important question that needs to be addressed is whether all the cells from follicular carcinomas are capable of invading the capsule or whether the tumor is composed of different cell populations, derived from an adenoma. If so, it is possible that some follicular carcinomas are more similar to adenomas than others. Again, the statistical significance of the differential expression of genes 44 and 199 in normal tissues, adenomas, and follicular carcinomas was confirmed by the two non-parametric hypothesis tests with significantly low *P* values.

Our results show that DDRT-PCR and cDNA array are efficient methods to identify differentially expressed genes in thyroid diseases. Further functional characterization of these genes, especially genes 44 and 199 that are being described for the first time in thyroid tissues, could contribute to the better understanding of thyroid pathologies. Eventually, these genes could also shade some light into the causal relationship among them. The identification of new genes using these technologies may allow the development of a classifier of great accuracy, which will greatly improve diagnosis of thyroid tumors.

Acknowledgements

We would like to thank Lara Termini and Chamberlein E. Mendonça for helping with RNA extraction from surgical specimens and all members of the Head and Neck surgery and Pathology groups from Hospital do Câncer for tissue collection. We also thank Carlos Ferreira and Miyuki F. da Silva for helping with dissection procedures, Anna Christina de Mattos Salin and Elizangela Monteiro for sequencing data, Dr Junior Barrera for discussions, and Drs Luisa L. Villa and Luiz DeMarco for critically reading the manuscript. This work was supported in part by a grant from FAPESP/CEPID (98/1435-2). B.S.S. is a pre-doctoral fellow from FAPESP

References

- [1] A.R. Shaha, Controversies in the management of thyroid nodule, *Laryngoscope* 110 (2000) 183–193.
- [2] H.J. Peter, H. Gerber, H. Studer, S. Smeds, Pathogenesis of heterogeneity in human multinodular goiter: a study on growth and function of thyroid tissue transplanted onto nude mice, *J. Clin. Invest.* 76 (1985) 1992–2002.
- [3] A. Sugeno, H. Masuda, M. Komatsu, S. Yokoyama, T. Shimizu, M. Fujimori, S. Kobayashi, F. Iida, Adenomatous goitre: therapeutic strategy, postoperative outcome, and study of epidermal growth factor receptor, *Br. J. Surg.* 79 (1992) 404–406.
- [4] M.R. Pelizzo, A. Piotto, D. Rubello, D. Casara, A. Fassina, B. Busnardo, High prevalence of occult papillary thyroid carcinoma in a surgical series for benign thyroid disease, *Tumori* 76 (1990) 255–257.
- [5] K.B. Koh, K.W. Chang, Carcinoma in multinodular goitre, *Br. J. Surg.* 79 (1992) 266–267.
- [6] A. Belfiore, G.L. La Rosa, G.A. La Porta, D. Giuffrida, G. Milazzo, L. Lupo, C. Regalbuto, R. Vigneri, Cancer risk in patients with cold thyroid nodules: relevance of iodine intake, sex, age, and multinodularity, *Am. J. Med.* 93 (1992) 363–369.
- [7] H.W. Wahner, C. Cuello, P. Correa, L.F. Uribe, E. Galtan, Thyroid carcinoma in an endemic goiter area Cali, Colombia, *Am. J. Med.* 40 (1966) 58–66.
- [8] F. Balçiri, F.K. Djemli, L.A. Mokrane, F.K. Djidel, The relative roles of endemic goiter and socioeconomic development status in the prognosis of thyroid carcinoma, *Cancer* 82 (1998) 1146–1153.
- [9] H.R. Haraeh, E.D. Williams, Pathology of Thyroid Cancer, in: M.H. Wheeler, J.H. Lazarus (Eds.), *Diseases of the Thyroid*, Chapman & Hall, London, 1994, pp. 341–363.
- [10] I. Maeda, T. Takano, F. Matsuzuka, T. Maruyama, T. Higashiyama, G. Liu, K. Kuma, N. Amino, Rapid screening of specific changes in mRNA in thyroid carcinomas by sequence specific-differential display: decreased expression of acid ceramidase mRNA in malignant and benign thyroid tumors, *Int. J. Cancer* 81 (1999) 700–704.
- [11] T. Takano, A. Miyauchi, T. Yokozawa, F. Matsuzuka, I. Maeda, K. Kuma, N. Amino, Preoperative diagnosis of thyroid papillary and anaplastic carcinomas by real-time quantitative reverse transcription-polymerase chain reaction of oneofetal fibronectin messenger RNA, *Cancer Res.* 59 (1999) 4542–4545.
- [12] M. Grieco, M. Santoro, M.T. Berlingieri, R.M. Melillo, R. Donghi, I. Bongarzone, M.A. Pierotti, P.G. Della, A. Fusco, G. Vecchio, PTC is a novel rearranged form of the ret proto-oncogene and is frequently detected in vivo in human thyroid papillary carcinomas, *Cell* 60 (1990) 557–563.
- [13] S. Said, M. Schlumberger, H.G. Suarez, Oncogenes and anti-oncogenes in human epithelial thyroid tumors, *J. Endocrinol. Invest.* 17 (1994) 371–379.
- [14] M. Santoro, R. Rosati, M. Grieco, M.T. Berlingieri, G.L. D'Amato, V. de F., A. Fusco, The ret proto-oncogene is consistently expressed in human pheochromocytomas and thyroid medullary carcinomas, *Oncogene* 5 (1990) 1595–1598.
- [15] T. Takano, A. Miyauchi, F. Matsuzuka, G. Liu, T. Higashiyama, T. Yokozawa, K. Kuma, N. Amino, Preoperative diagnosis of medullary thyroid carcinoma by RT-PCR using RNA extracted from leftover cells within a needle used for fine needle aspiration biopsy, *J. Clin. Endocrinol. Metab.* 84 (1999) 951–955.
- [16] A. Fusco, M. Grieco, M. Santoro, M.T. Berlingieri, S. Pilotti, M.A. Pierotti, P.G. Della, G. Vecchio, A new oncogene in human thyroid papillary carcinomas and their lymph- nodal metastases, *Nature* 328 (1987) 170–172.
- [17] M. Santoro, F. Carlomagno, A. Romano, D.P. Bottaro, N.A. Dathan, M. Grieco, A. Fusco, G. Vecchio, B. Matoskova, M.H. Kraus, Activation of RET as a dominant transforming gene by germline mutations of MEN2A and MEN2B, *Science* 267 (1995) 381–383.
- [18] H. van, V. Genetics: one gene, four syndromes, *Nature* 367 (1994) 319–320.
- [19] A.M. Silva, L.F. Reis, Sodium salicylate induces the expression of the immunophilin FKBP5 1 and biglycan genes and inhibits p34cdc2 mRNA both in vitro and in vivo, *J. Biol. Chem.* 275 (2000) 36388–36393.
- [20] I. Bieche, I. Laurendeau, S. Tozlu, M. Olivi, D. Vidaud, R. Lidereau, M. Vidaud, Quantitation of MYC gene expression in sporadic breast tumors with a real-time reverse transcription-PCR assay, *Cancer Res.* 59 (1999) 2759–2765.
- [21] J.R. Quinlan, *Programs for Machine Learning*, The Morgan Kaufmann Series in Machine Learning, Morgan Kaufmann, 1992.
- [22] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and Regression Trees*, The Wadsworth Statistics/Probability Series, Wadsworth International Group, 1984.
- [23] S.F. Abtschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman,

- Basic local alignment search tool, *J. Mol. Biol.* 215 (1990) 403–410.
- [24] J.H. Zar, *Biostatistical Analysis*, Prentice Hall, 2002.
- [25] H. Yu, T. Rohan, Role of the insulin-like growth factor family in cancer development and progression, *J. Natl. Cancer Inst.* 92 (2000) 1472–1489.
- [26] J.M. Holly, Still more questions than answers: report on the 5th International Symposium on Insulin-like Growth Factors, Brighton, UK, 31 October–4 November 1999, *Trends Endocrinol. Metab.* 11 (2000) 151–153.
- [27] M.R. Schneider, E. Wolf, A. Hoefflich, H. Lahm, IGF-binding protein-5: flexible player in the IGF system and effector on its own, *J. Endocrinol.* 172 (2002) 423–440.
- [28] N. Miyakoshi, C. Richman, Y. Kasukawa, T.A. Linkhart, D.J. Baylink, S. Mohan, Evidence that IGF-binding protein-5 functions as a growth factor, *J. Clin. Invest.* 107 (2001) 73–81.
- [29] C. Richman, D.J. Baylink, K. Lang, C. Dony, S. Mohan, Recombinant human insulin-like growth factor-binding protein-S stimulates bone formation parameters in vitro and in vivo, *Endocrinology* 140 (1999) 4699–4705.
- [30] M.C. Sootweg, C. Ohlsson, E.J. van Elk, J.C. Netelenbos, D.L. Andress, Growth hormone receptor activity is stimulated by insulin-like growth factor binding protein 5 in rat osteosarcoma cells, *Growth Regul.* 6 (1996) 238–246.
- [31] H. Miyake, C. Nebson, P.S. Rennie, M.E. Gleave, Overexpression of insulin-like growth factor binding protein-S helps accelerate progression to androgen-independence in the human prostate LNCaP tumor model through activation of phosphatidylinositol 3'-kinase pathway, *Endocrinology* 141 (2000) 2257–2265.
- [32] M.K. Tennant, J.B. Thrasher, P.A. Twomey, R.S. Birnbaum, S.R. Plymate, Insulin-like growth factor-binding proteins (IGFBP)-4, -5, and -6 in the benign and malignant human prostate: IGFBP-5 messenger ribonucleic acid localization differs from IGFBP-5 protein localization, *J. Clin. Endocrinol. Metab.* 81 (1996) 3783–3792.
- [33] C.M. Perks, S. Bowen, Z.P. Gibr, P.V. Newcomb, J.M. Holby, Differential IGF-independent effects of insulin-like growth factor binding proteins (1–6) on apoptosis of breast epithelial cells, *J. Cell Biochem.* 75 (1999) 652–664.
- [34] C.M. Perks, C. McCaig, J.M. Hobby, Differential insulin-like growth factor (IGF)-independent interactions of IGF binding protein-3 and IGF binding protein-S on apoptosis in human breast cancer cells; involvement of the mitochondria, *J. Cell Biochem.* 80 (2000) 248–258.
- [35] F. Rozen, M. Polbak, Inhibition of insulin-like growth factor I receptor signaling by the vitamin D analogue EB 1089 in MCF-7 breast cancer cells: a role for insulin-like growth factor binding proteins, *Int. J. Oncol.* 15 (1999) 589–594.
- [36] P. Liang, A.B. Pardee, Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction, *Science* 257 (1992) 967–971.
- [37] J.S. Wan, S.J. Sharp, G.M. Poirier, P.C. Wagaman, J. Chambers, J. Pyati, Y.L. Hom, J.E. Galindo, A. Huvar, P.A. Peterson, M.R. Jackson, M.G. Erbander, Cloning differentially expressed mRNAs, *Nat. Biotechnol.* 14 (1996) 1685–1691.
- [38] A.A. Alizadeh, D.T. Ross, C.M. Perou, R.M. van de, Towards a novel classification of human malignancies based on gene expression patterns, *J. Pathol.* 195 (2001) 41–52.
- [39] O.P. Kallioniemi, Biochip technologies in cancer research, *Ann. Med.* 33 (2001) 142–147.
- [40] A. Bartolazzi, Improving accuracy of cytology for nodular thyroid lesions, *Lancet* 355 (2000) 1661–1662.