

KARINA LUMI MINE

ANÁLISE TRANSCRIPTÔMICA EM CÂNCER DO COLO DO ÚTERO

Tese apresentada à Universidade
Federal de São Paulo para obtenção
do título de Doutor em Ciências

São Paulo

2009

KARINA LUMI MINE

ANÁLISE TRANSCRIPTÔMICA EM CÂNCER DO COLO DO ÚTERO

Tese apresentada à Universidade
Federal de São Paulo para obtenção
do título de Doutor em Ciências

Orientador: Prof^a Dr^a Maria Gerbase de Lima

Co- orientadores: Dr^a Natalia Shulzhenko

Dr Andrey Morgun

São Paulo

2009

Mine, Karina Lumi

Análise transcriptômica em câncer do colo do útero. / Karina Lumi

Mine. -- São Paulo, 2009.

x, 100f.

Tese (Doutorado) - Universidade Federal de São Paulo. Escola Paulista de Medicina. Programa de Pós-graduação em Ciências Aplicadas à Pediatria.

Título em inglês: Transcriptome analysis in cervical cancer

1. Neoplasias do colo do útero. 2. Expressão gênica. 3. Análises de microsséries. 4. Humanos.

Este trabalho foi desenvolvido no Setor de Imunogenética do Departamento de Pediatria da Universidade Federal de São Paulo.

Aos meus pais e meus irmãos,
pelo apoio e carinho constante mesmo
de longe em muitos momentos.

Ao Élcio,
pelo amor e companheirismo
sempre presentes.

Agradecimentos

À Profa Dra Maria Gerbase de Lima, pelos valiosos ensinamentos e por me acolher por todos esses anos no laboratório com muito carinho, incentivo e confiança.

Aos Drs Andrey Morgun e Natalia Shulzhenko, pela participação e interesse constante neste trabalho, pela participação valiosa na minha formação científica e, pela amizade, atitudes e momentos que para mim serão inesquecíveis.

À Dra Gerdine Sanson, pela amizade, pela importante participação neste trabalho e, pelos grandes momentos de apoio e incentivo.

Ao Prof Dr Anatoli Yambartsev, pela colaboração e pelos ensinamentos e discussões de matemática e estatística.

À Profa Dra Chloé Camba Musatti, pelos ensinamentos de imunologia.

Aos Prof Dr Ismael Dale Cotrim Guerreiro da Silva e Dra Carmen Carvalho, pelas valiosas discussões e colaboração neste trabalho.

Às Dra Sylvia Brenna e Dra Julisa Ribalta pela participação clínica e importante colaboração neste trabalho.

À todos meus amigos do Igen, pelo apoio, pela convivência e pelos vários momentos agradáveis que passamos juntos.

Aos meus amigos, pelo carinho e apoio, mesmo com nossos encontros pouco frequentes.

Aos meus tios e primos, que sempre me incentivaram e apoiaram.

APOIO FINANCEIRO:
Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)
Processo n° 05/51329-6

Sumário

Dedicatória	iv
Agradecimentos	v
Lista de abreviaturas	viii
Resumo	ix
1 INTRODUÇÃO	1
1.1. Análise de expressão gênica no estudo de resposta ao tratamento em câncer de colo do útero	3
1.2. Análise de expressão gênica no estudo de mecanismos moleculares envolvidos no câncer de colo do útero	5
2 OBJETIVOS	8
3 CASUÍSTICA E MÉTODO	10
3.1. Casuística	11
3.2. Extração de RNA total	12
3.3. Metodologia de <i>microarray</i>	13
3.4. Análise dos dados clínicos	15
3.5. Análise dos dados de <i>microarray</i>	15
4 RESULTADOS	27
4.1. Busca de um perfil de expressão gênica capaz de prever a resposta ao tratamento	28
4.2. Busca de genes com expressão alterada em amostras de tumores de colo do útero	34
4.3. Análise de grupos funcionais de genes	38
4.4. Comparação dos genes diferencialmente expressos entre tumor e normal com dados de outros trabalhos publicados	49
5 DISCUSSÃO	54
5.1. Busca de um perfil de expressão gênica capaz de prever a resposta ao tratamento	55
5.2. Identificação de genes e grupos funcionais de genes com expressão alterada em amostras de tumor de colo do útero	57
6 CONCLUSÕES	68
7 ANEXOS	70
8 REFERÊNCIAS	82
Abstract	
Bibliografia consultada	

Lista de abreviaturas

ADC	adenocarcinoma
aRNA	<i>antisense</i> RNA (RNA anti-sense)
cDNA	<i>complementary</i> DNA (DNA complementar)
DAVID	<i>Database for Annotation, Visualization and Integrated Discovery</i>
DWD	<i>Distance Weighted Discrimination</i>
ERK	<i>Extracellular Signal-Regulated Kinases</i>
FatiGO	<i>Fast Assignment and Transference of Information using Gene Ontology</i>
FDR	<i>False Discovery Rate</i> (taxa de descobertas falsas)
FIGO	Federação Internacional de Ginecologia e Obstetrícia
GEO	<i>Gene Expression Omnibus</i>
GSEA	<i>Gene Set Enrichment Analysis</i>
GO	<i>Gene Ontology</i>
HPV	<i>Human papillomavirus</i>
IPA	<i>Ingenuity Pathway Analysis</i>
JNK	<i>Jun N-terminal kinases</i>
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes Pathway</i>
KLK	<i>Kallikrein-related peptidases</i>
KNN	<i>K- Nearest Neighbors</i>
LOOCV	<i>Leave-one-out cross-validation</i>
MAPK	<i>Mitogen-activated protein kinase</i>
MCODE	<i>Molecular Complex Detection</i>
MILANO	<i>Microarray Literature-based Annotation</i>
mRNA	<i>messenger</i> RNA (RNA mensageiro)
NR	não respondedoras ao tratamento
OXPHOS	<i>Oxidative Phosphorylation</i> (fosforilação oxidativa)
R	respondedoras ao tratamento
RT	radioterapia
RT+QT	radioterapia com quimioterapia
SCC	<i>Squamous Cell Carcinoma</i> (carcinoma de células escamosas)
SDS	<i>Sodium Dodecyl Sulfate</i>
SSC	<i>Standard Saline Citrate</i>

Resumo

O câncer de colo do útero é o segundo tipo de câncer mais comum entre mulheres. Cerca de 80% dos casos ocorrem em países em desenvolvimento onde a maior parte são diagnosticados em estádios relativamente avançados. Os **objetivos** desse trabalho foram: (i) identificar um perfil de expressão gênica capaz de prever a resposta ao tratamento, visto que, cerca de 35% das pacientes com câncer localmente avançado não respondem ao tratamento; (ii) identificar genes e grupos funcionais de genes (vias e redes gênicas) com expressão alterada em amostras de tecido tumoral em relação ao normal adjacente, para identificar genes e mecanismos envolvidos na patogênese do câncer de colo uterino. **Método.** Realizamos análise de expressão gênica em larga escala, utilizando *microarrays* contendo oligonucleotídeos impressos correspondentes a aproximadamente 14.000 genes. No estudo de resposta ao tratamento, utilizamos 23 lâminas hibridizadas com amostras de tumor e, no estudo comparativo entre tecido tumoral e normal, utilizamos 34 lâminas de tumor e 20 de tecido normal. Para a análise dos dados foram utilizados os seguintes programas: *BRB Array Tools* para busca de um preditor molecular e geração da lista de genes; *DAVID*, *Babelomics* e *Ingenuity Pathway Analysis* para verificar as vias representadas na lista; *MILANO* para verificar a literatura em relação aos nossos achados; *DAVID*, *CFinder* e *Cytoscape* para análises de redes gênicas. **Resultados:** (i) Não identificamos nenhum perfil de expressão gênica capaz de prever de forma consistente a resposta ao tratamento; (ii) identificamos 810 genes diferencialmente expressos entre tecido tumoral e normal, sendo que 341 tinham expressão aumentada no tumor e 469 expressão diminuída. Na análise de vias, identificamos 13 vias, entre as quais havia vias já descritas em câncer de colo uterino, como a via do ciclo celular e da p53 e vias ainda não estudadas nesse câncer, como a da fosforilação oxidativa e a do ribossomo. Além disso, verificamos que vários genes pertencentes as vias, mesmo as já descritas como relacionadas com o câncer de colo uterino, ainda não haviam sido estudados nesse câncer. Na análise de redes gênicas encontramos 23 sub-redes. Dentre essas, destacamos a rede formada pelos genes da família das calicreínas. Nossos resultados também sugerem que o perfil de genes identificados também se aplica a casos de câncer de colo de útero estudados em outros trabalhos assim como a câncer de esôfago. **Conclusão:** O uso da análise de expressão gênica em larga escala possibilitou a identificação de genes e grupos funcionais de genes, vias e redes

gênicas, alterados no câncer de colo uterino. Esses resultados podem ajudar no entendimento dos mecanismos moleculares envolvidos na patogenia deste câncer e os genes encontrados podem ser candidatos promissores para estudos de marcadores a serem utilizados no diagnóstico precoce e estudos de susceptibilidade genética em câncer de colo do útero.

1 INTRODUÇÃO

O câncer do colo do útero é o segundo tipo de câncer mais comum entre mulheres no mundo, com aproximadamente 500 mil casos novos (Parkin et al., 2005; Castellsagué et al., 2007) e 260 mil mortes por ano (Castellsagué, 2008). Cerca de 80% dos casos ocorrem em países em desenvolvimento (Parkin et al., 2005) onde a maioria dos casos são diagnosticados em estádios relativamente avançados. Para o Brasil, em 2008, eram esperados 18.680 casos novos, com um risco estimado de 19 a cada 100 mil mulheres (Brasil. Ministério da Saúde. Instituto Nacional de Câncer. INCA, 2007). O carcinoma de células escamosas (SCC, *squamous cell carcinoma*) é o tipo histológico mais comum desse câncer abrangendo aproximadamente 80% dos casos (Waggoner, 2003; Schiffman et al., 2007).

O uso da tecnologia de *microarray* espalhou-se rapidamente por toda a comunidade científica desde quando foi introduzida nos anos 90 (Fodor et al., 1991; Nguyen et al., 1995; Schena et al., 1995). Essa ferramenta possui muitas aplicações na pesquisa biomédica. Em oncologia, o estudo de expressão gênica por meio de *microarrays* é utilizado para buscar marcadores diagnósticos, definir perfis moleculares capazes de prever a evolução clínica ou resposta ao tratamento, melhor entender mecanismos biológicos envolvidos na oncogênese e para possivelmente descobrir novos alvos ou drogas terapêuticas (Alon et al., 1999; Golub et al., 1999; van't Veer et al., 2002; Potti et al., 2006; Watanabe et al., 2006; Ishigami et al., 2007; Perez-Diez et al., 2007). Em nosso grupo o *microarray* já foi utilizado como ferramenta para o estudo de eventos relacionados a transplante de órgãos (Morgun et al., 2006) e de células tronco hematopoiéticas (Arantes, 2009).

1.1. Análise de expressão gênica no estudo de resposta ao tratamento em câncer de colo do útero

O estadiamento do câncer do colo do útero é realizado de acordo com os critérios da Federação Internacional de Ginecologia e Obstetrícia (FIGO) (Benedet et al., 2000). Resumidamente, o estadio I (Ia, Ia1, Ia2, Ib1, Ib2) refere-se ao carcinoma estritamente limitado ao colo uterino. No estadio II (IIa e IIb) o tumor se estende mais além do colo uterino, mas não invade a parede pélvica e nem o terço inferior da vagina. No estadio III (IIIa e IIIb) o tumor invade a parede pélvica e/ou o terço inferior da vagina e/ou causa hidronefrose ou não funcionamento do rim. No estadio IV o tumor se estende mais além da pelve e/ou invade clinicamente a mucosa da bexiga e/ou reto (IVa) ou se dissemina para órgãos distantes (IVb). Os estadios iniciais do câncer de colo uterino incluem pequenos tumores confinados ao colo do útero com até 4 cm de dimensão (estadios Ia a Ib1), enquanto que os casos de câncer localmente avançados geralmente incluem tumores com estadios Ib2 a IVa (Rojas-Espailat, Rose, 2005; Barbera, Thomas, 2009). Muitas mulheres com estadios iniciais desse câncer são curadas com tratamento de cirurgia ou radioterapia. A sobrevida em cinco anos para esses casos é de aproximadamente 90%. No entanto, essa taxa cai drasticamente em estadios mais avançados, que é o caso da maioria das mulheres diagnosticadas nos países em desenvolvimento. A sobrevida em cinco anos atinge cerca de 50% para os casos localmente avançados (Tambaro et al., 2004; INCA, 2007).

A radioterapia, durante muitas décadas, foi o tratamento padrão de câncer de colo do útero localmente avançado. Recentemente, foi estabelecido o uso concomitante da radioterapia e quimioterapia como o novo tratamento padrão para esses casos (Rose, 2002; Monk et al., 2007). Embora tenha ocorrido uma melhora na sobrevida global, ainda cerca de 35% das pacientes não respondem ao tratamento (Santin et al., 2005; Gaffney et al., 2007). A caracterização de tumores com diferentes respostas ao tratamento poderia ajudar na seleção de terapias mais adequadas de acordo com a gravidade da doença. Assim, a assinatura biológica de um câncer poderia ser utilizada para identificar tumores que se comportam clinicamente de forma distinta.

Atualmente, com a metodologia de *microarray*, é possível realizar a análise de expressão de milhares de genes simultaneamente. O advento dessa tecnologia possibilitou a elaboração de estudos para gerar perfis de expressão gênica capazes de

predizer a evolução clínica ou resposta ao tratamento em vários tipos de câncer (Kihara et al., 2001; Kallioniemi, 2004; Watanabe et al., 2006; Ishigami et al., 2007). Em câncer de mama já existem marcadores biológicos para orientar o tratamento (van't Veer et al., 2002; Slodkowska, Ross, 2009). Com o *MammaPrint® Breast Cancer Test* é possível avaliar o nível de expressão de 70 genes, em uma amostra de tumor de mama removido cirurgicamente, e, após a utilização de um algoritmo específico e obtenção de um escore, determinar se a paciente é considerada com baixo ou alto risco de metastatase do câncer. As pacientes com alto risco podem ser indicadas para uma terapia mais agressiva. Este teste pode ser aplicado em mulheres com câncer de mama com menos de 61 anos, estadio I-II e linfonodo negativo (Li et al., 2008).

Até o ano de 2005, quando este estudo começou, havia apenas três trabalhos que utilizavam *microarrays* para buscar genes relacionados com a resposta ou sensibilidade ao tratamento em tecidos de pacientes com câncer de colo do útero (Kitahara et al., 2002; Wong et al., 2003; Harima et al., 2004). O número de pacientes estudadas nestes três trabalhos era pequeno e em um deles foi utilizado a termo-radiação como tratamento. Embora cada estudo tenha encontrado genes relacionados com a resposta do tumor ao tratamento houve pouca sobreposição dos genes entre eles. Assim, havia ainda a necessidade de estudos em busca de um perfil mais consistente que pudesse predizer a resposta ao tratamento das pacientes. Recentemente novos estudos foram publicados (Wong et al., 2006b; Klopp et al., 2008; Wheidhaas et al., 2009). No entanto, apesar dos resultados serem promissores ainda nenhum perfil de expressão gênica que possa predizer a resposta ao tratamento ou evolução clínica das pacientes com câncer de colo do útero foi consolidado.

1.2. Análise de expressão gênica no estudo de mecanismos moleculares envolvidos no câncer de colo do útero

A infecção por papilomavírus humano (HPV, *human papillomavirus*) de alto-risco é o principal fator de risco para o desenvolvimento do carcinoma de colo do útero (Walboomers et al., 1999; Bosch et al., 2002; zur Hausen, 2002). Embora essa infecção seja uma causa necessária ela não é suficiente para o desenvolvimento da doença, pois poucas mulheres infectadas desenvolvem o câncer. Assim, outros co-fatores devem contribuir para o processo carcinogênico (Muñoz et al., 2006; Biewenga et al., 2008; Castellsagué, 2008). Além de co-fatores já estabelecidos, como o uso prolongado de contraceptivo oral, alto número de gestações, tabagismo e tipo de HPV (Muñoz et al., 2006; Schiffman et al., 2007), fatores ligados a mulher, genéticos e imunológicos, podem também ter um papel importante no desenvolvimento do câncer de colo do útero (Hildesheim, Wang, 2002; de Araujo Souza, Villa, 2003; Muñoz et al., 2006).

Uma vez que a resposta imune deficiente pode estar relacionada tanto à persistência da infecção por HPV como ao desenvolvimento da neoplasia, polimorfismos genéticos que possivelmente influenciam esta resposta vêm sendo objetos de estudo de muitos pesquisadores (Maciag, Villa, 1999; Wang, Hildesheim, 2003; Kordi Tamandani et al., 2008; Wang et al., 2009; Yu et al., 2009), inclusive de nosso laboratório (Guzman et al., 2008).

Todos os genes estudados pelo nosso grupo foram selecionados com base em informações sobre seu papel na resposta imune, ou seja, os genes candidatos foram selecionados a partir de hipóteses geradas pela análise da literatura. No presente trabalho mudamos a estratégia de estudo e, a partir da análise de expressão simultânea de milhares de genes utilizando *microarrays*, buscamos identificar genes e grupos funcionais de genes alterados no câncer de colo do útero.

O advento de tecnologias de análise em larga escala, como os *microarrays*, trouxe novas perspectivas nos estudos para a descoberta de marcadores moleculares e para melhor entendimento de mecanismos envolvidos na carcinogênese, o que poderá ajudar no rastreamento, detecção precoce e tratamento do câncer (Martin et al., 2007). Embora os programas de rastreamento venham sendo efetivos em reduzir a incidência do câncer de colo do útero, ainda há a necessidade de estudos para

melhorar esse rastreamento devido à alta taxa de falsos positivos e negativos (Baldwin et al., 2003; Nijhuis et al., 2006). Além disso, a identificação de marcadores que pudessem distinguir lesões precursoras de alto grau do colo do útero, com potencial para progredir para o câncer invasivo das que regredem espontaneamente, seria de grande importância para guiar o tratamento das pacientes com lesões precursoras de alto grau (Steinau et al., 2007; Zhai et al., 2007).

Em poucos anos de uso de *microarrays* na pesquisa em câncer, centenas de experimentos em larga escala foram realizados gerando diversos perfis de expressão gênica nessa doença. Um grande desafio gerado a partir dos dados obtidos é a interpretação destes para melhor entender os fenômenos biológicos existentes. Uma abordagem comum é gerar listas de genes com expressão diferencial entre as classes estudadas para priorizar candidatos para estudos posteriores. Embora esse tipo de abordagem identifique diversos genes importantes em câncer, muitas vezes podem passar despercebidos importantes efeitos em vias biológicas, pois processos celulares frequentemente afetam grupos de genes agindo em conjunto. Por exemplo, o aumento de genes codificando membros de uma via metabólica pode dramaticamente alterar essa via e ser mais importante do que um grande aumento de um único gene (Subramanian et al., 2005). Assim, atualmente, alguns grupos tem focado suas pesquisas na identificação de grupos de genes que compartilham uma mesma via ou anotação funcional (Subramanian et al., 2005; Chuang et al., 2007; Savli et al., 2008; Werner, 2008; Wong et al., 2008).

Diversas ferramentas utilizam listas de genes para identificar grupos que compartilham funções biológicas e que estão mais ou menos representados em uma determinada situação (Al-Shahrour et al., 2005; Maere et al., 2005, Khatri et al., 2007; Huang et al., 2009). Muitas abordagens utilizam bancos de dados externos como o *Gene Ontology* (GO, Ashburner et al, 2000), *Kyoto Encyclopedia of Genes and Genomes Pathway* (KEGG, Ogata et al., 1999) e *BioCarta* (<http://www.biocarta.com>), ou interno como nos programas *Ingenuity Pathway Analysis* (IPA, Ingenuity® Systems, www.ingenuity.com) e *Gene Set Enrichment Analysis* (GSEA, Subramanian, 2005). Além das abordagens de termos do GO e vias (KEGG, BioCarta e IPA) há programas que também realizam a construção de redes gênicas (*gene networks*) a partir de dados que indicam interações entre os genes (Palla et al., 2005; Cline et al., 2007; Savli et al., 2008).

Em câncer de colo do útero o *microarray* vem sendo aplicado em vários tipos de análises (Acevedo Rocha et al., 2007; Chao et al., 2007). Até 2005, poucos trabalhos utilizaram essa tecnologia para identificar genes envolvidos na carcinogênese utilizando amostras clínicas (Shim et al., 1998; Chen et al., 2003; Wong et al., 2003). Recentemente o número de trabalhos publicados cresceu consideravelmente na tentativa de melhor definir marcadores e entender mecanismos envolvidos nesse câncer (Wong et al., 2006a; Miyatake et al., 2007; Pérez-Plasencia et al., 2007; Song et al., 2008; Zhai et al., 2007 ; Biewenga et al., 2008; Scotto et al., 2008; Wilting et al., 2008).

Nosso estudo foi elaborado para abranger dois aspectos do câncer de colo do útero a partir de duas hipóteses: (i) diferentes perfis de expressão gênica estão associados a diferentes respostas ao tratamento; (ii) há genes e grupos funcionais de genes com expressão alterada no tumor. Assim, neste estudo utilizando o *microarray* como ferramenta, buscamos identificar perfis de expressão gênica capazes de prever a resposta ao tratamento e identificar genes e grupos funcionais de genes com expressão alterada no tumor, que poderão contribuir para o entendimento de mecanismos moleculares envolvidos no câncer de colo do útero.

2 OBJETIVOS

1. Identificar um perfil de expressão gênica capaz de prever a resposta ao tratamento do câncer do colo do útero.
2. Identificar genes e grupos funcionais de genes (vias e redes gênicas) com expressão alterada em câncer do colo do útero.

3 CASUÍSTICA E MÉTODO

3.1. Casuística

Quarenta amostras de tecido tumoral e vinte de tecido normal adjacente foram obtidas de pacientes portadoras de carcinoma de células escamosas do colo do útero. Dessas, treze amostras de tecido tumoral e normal eram da mesma paciente. As amostras de tecido normal foram retiradas da parede vaginal. Todos os fragmentos foram coletados na ocasião da realização de biópsia com finalidade diagnóstica. As pacientes eram do Setor de Oncologia do Departamento de Ginecologia da Universidade Federal de São Paulo (UNIFESP) ou do Hospital e Maternidade Leonor Mendes de Barros. O estudo foi aprovado pelo Comitê de Ética da Universidade Federal de São Paulo (Anexo 1) e o Termo de Consentimento Livre e Esclarecido (Anexo 2) foi assinado por todas as pacientes.

A classificação do estadió clínico do tumor foi realizada de acordo com os critérios da FIGO. Vinte pacientes apresentavam tumor com estadió IIb, uma, com estadió IIIa, e 19, com estadió IIIb.

Para a pesquisa de um possível perfil de expressão gênica capaz de prever a resposta ao tratamento foram utilizadas amostras de tecido tumoral de 24 das 40 pacientes com dados clínicos disponíveis para essa análise (Anexo 3). O tratamento de 11 pacientes consistiu apenas de radioterapia (RT) e, de 13, de radioterapia e quimioterapia (RT+QT). As pacientes foram classificadas em dois grupos: as pacientes com sobrevida livre da doença em um ano foram classificadas em respondedoras (R) e as com sobrevida em um ano com presença de tumor ou que faleceram durante esse período devido à doença foram classificadas como não respondedoras (NR) ao tratamento. No total 14 foram classificadas como R e 10, como NR. A média da idade e o desvio padrão (em anos) no grupo R foi $53,9 \pm 12,8$ (variando de 31 a 75 anos) e no grupo NR foi de $47,8 \pm 13,3$ (variando de 30 a 74 anos). As pacientes classificadas como R e NR não apresentaram diferença estatisticamente significativa em relação à idade ($p=0,27$), estadió clínico ($p=0,41$) e tipo de tratamento ($p=0,70$) (Tabela 1).

Após a coleta, as amostras foram imediatamente imersas em *RNA later* (Ambion, Austin, TX, EUA) e armazenadas a -80°C para posterior extração do RNA.

Tabela 1. Características clínicas das pacientes classificadas como respondedoras (R) ou não respondedoras (NR) ao tratamento e valor de p obtido nas análises.

Característica Clínica	Classificação das pacientes em relação a resposta ao tratamento		Valor de p obtido no teste
	R (n=14)	NR (n=10)	
Idade			
média e desvio padrão em anos	53,9±12,8	47,8 ±13,3	0,27*
Estadio clínico			
IIb	9	4	0,41**
IIIb	5	6	
Tipo de tratamento			
Radioterapia	7	4	0,70**
Radioterapia com quimioterapia	7	6	

*teste t de *Student*; **teste exato de *Fisher*

3.2. Extração de RNA total

O RNA total foi extraído das amostras utilizando o reagente *TRIzol* (Invitrogen, Carlsbad, CA, EUA). Inicialmente, os fragmentos dos tecidos foram homogeneizados em 0,80 mL de *TRIzol* e 15 µL de glicogênio (20 mg/mL) utilizando o homogeneizador *Handishear AC* (VirTis Company, Gardiner, NY, EUA). Todas as outras etapas foram realizadas de acordo com as instruções do fabricante. O RNA extraído foi ressuspensão em 50 µl de água bidestilada autoclavada, contendo 0,1 g% de dietilpirocarbonato (Sigma-Adrich, St.Louis, MO, EUA) e quantificado no espectrofotômetro *Gene Quant II* (GE Healthcare, Uppsalla, Suécia). A qualidade do RNA foi analisada pelo equipamento *Agilent 2100 Bioanalyzer* utilizando o *kit RNA 6000 Nano LabChip* (Agilent Technologies, Santa Clara, CA, USA).

3.3. Metodologia de *microarray*

3.3.1. Amostras e lâminas utilizadas

Das 40 amostras de tecido tumoral, 12 foram agrupadas em pares (*pools* de duas amostras) devido à quantidade limitada de RNA após a amplificação. Assim, no total obtivemos 34 lâminas com amostras de tecido tumoral e 20 com amostras de tecido normal. Em cada lâmina além da amostra teste (tumor ou normal) foi hibridizada uma amostra de referência. A amostra de referência consistia de um *pool* de RNA amplificado de 24 amostras de tecido tumoral. Nesse tipo de desenho de estudo de *microarray*, o *reference design*, utilizamos nas análises a razão entre os valores de expressão da amostra teste e da amostra referência. Nas análises de resposta ao tratamento foram utilizadas 23 lâminas das 34 hibridizadas com tecido tumoral. Dentre elas, uma lâmina tinha sido hibridizada com duas amostras, portanto essa análise englobou 24 amostras.

As lâminas de *microarray* utilizadas foram preparadas no *Microarray Facility of National Institute of Allergy and Infectious Disease (NIAID), National Institutes of Health, Estados Unidos*. Essas lâminas contêm impressos oligonucleotídeos de 70-mer correspondentes a aproximadamente 14.000 genes humanos desenhados e sintetizados por Qiagen/Operon (Alameda, CA, EUA). Foram utilizadas três diferentes lotes das lâminas, Hsaa, Hsab e Hsac. Esses três tipos de lâminas possuem o mesmo conteúdo de sondas de genes, variando somente no formato da lâmina e alguns *spots* controles.

3.3.2. Amplificação do RNA mensageiro (mRNA, *messenger RNA*)

Para os experimentos de *microarray*, os mRNAs foram amplificados em dois ciclos. Um micrograma de RNA total foi amplificado no primeiro ciclo, purificado e 1 µg do RNA anti-sense (aRNA, *antisense RNA*) obtido foi convertido em DNA complementar (cDNA, *complementary DNA*) dupla fita, todas etapas seguindo um protocolo previamente descrito (Wang et al., 2000; Morgun et al., 2006). Para preparar o RNA para os *microarrays*, um segundo ciclo de transcrição *in vitro* foi realizada. Nesta reação, além dos nucleotídeos necessários, foi adicionado um nucleotídeo modificado (aminoallyl-dUTP) que permite a marcação do RNA amplificado com os corantes Cy3 ou Cy5 numa etapa posterior. A reação de transcrição *in vitro* consistia de 16 µL de cDNA purificado, 3 µL de 50 mmol/L de 5-(3-aminoallyl)-UTP, 2 µL de

75mmol/L de UTP, 4µL de 75 mmol/L de ATP, GTP e CTP, 4 µL de tampão de reação 10x e 4 µL de uma mistura de enzimas para transcrição (todos da Ambion) que foi incubada por 13 horas à 37°C. Depois de uma incubação adicional de 15 minutos com 4 µL de Dnase I (Ambion), o aRNA foi purificado com o *RNAeasy mini kit* (Qiagen, Valencia, CA, EUA) de acordo com as instruções do fabricante. A qualidade do aRNA foi verificado utilizando o equipamento *Agilent 2100 Bioanalyzer* utilizando o *kit RNA 6000 Nano LabChip* (Agilent Technologies).

3.3.3. Pré-hibridização

Uma hora antes da hibridização, as lâminas foram tratadas a 42°C por 1 hora com uma solução bloqueadora contendo 1% de albumina de soro bovino a 10%, 25% de 20x *standard saline citrate* (SSC) (Invitrogen) e 0,1% de dodecil sulfato de sódio (SDS, *sodium dodecyl sulfate*) a 10% (todos da Invitrogen) em água sem nucleases. Em seguida as lâminas foram submetidas a duas lavagens com água sem nucleases e uma com isopropanol.

3.3.4. Hibridização e obtenção dos valores de expressão

Cinco microgramas de aRNA foram secos a vácuo e ressuspensos em 4,5 µL de 0,1 mol/L de tampão de acoplamento (bicarbonato de sódio, pH 8,5 a 9,0) (Sigma-Adrich). Depois da adição de 2,5 µL de água sem nucleases, as amostras de aRNA foram incubadas por 1 hora à temperatura ambiente com 3 µL de Cy3 (Amersham Biosciences, Piscataway, NJ) ou 4 µL de Cy5 (Amersham), cada um previamente dissolvido em 45 µL de dimetil sulfoxido (Sigma-Adrich). Em cada lâmina, as amostras teste (tumor ou normal) foram marcadas com Cy3 e as amostras referência com Cy5. Em seguida foram adicionados 6 µL de 4 mol/L de hidroxilamina à reação que foi incubada por mais 15 minutos. O aRNA marcado foi purificado usando o *RNAeasy mini kit* (Qiagen) e concentrado a um volume de 10 µL utilizando colunas *Microcon YM-30* (Milipore, Billerica, Mass), de acordo com as instruções do fabricante. As amostras marcadas com Cy3 e Cy5 foram misturadas e, 1 µg de poly(dA) (Amersham), 10 µg de DNA Cot I humano e 4 µg de tRNA de levedura (ambos da Invitrogen) foram adicionados. A mistura foi incubada a 98°C por três minutos e a seguir foi adicionado volume igual de solução de hibridização que consiste em 50% de formamida (Sigma-Adrich), 50% 20x SSC e 0,2% de SDS. A solução final foi aplicada na lâmina

previamente tratada e em seguida incubada *overnight* à 60°C. Depois da hibridização, as lâminas foram lavadas duas vezes em 1x SSC com 0,05% de SDS e três vezes em 0,1x SSC a 55°C.

As lâminas foram escaneadas utilizando o aparelho *GenePix 4000B microarray scanner* (Axon Instruments, Foster City, CA, EUA). Os valores de expressão foram obtidos utilizando o programa *GenePix Pro 6.0* (Axon Instruments).

3.4. Análise dos dados clínicos

As características clínicas das pacientes nos grupos estudados foram comparadas utilizando teste t de *Student* para variáveis quantitativas e teste exato de *Fisher* para variáveis qualitativas no programa *GraphPad InStat* versão 3.10 (GraphPad Software Inc., San Diego, CA, EUA). Valores de $p < 0,05$ foram considerados significantes.

3.5. Análise dos dados de *microarray*

3.5.1. Filtragem e Normalização dos dados

Antes de cada análise, os valores de expressão obtidos pelo programa *GenePix Pro* foram submetidos a filtragem e normalização utilizando o programa *BRB Array Tools* versão 3.6, desenvolvido pelos Drs Richard Simon e Amy Peng Lam (<http://linus.nci.nih.gov/BRB-ArrayTools.html>).

Na filtragem, para cada lâmina, foram selecionados os *spots* que não foram marcados como ruins ou vazios durante a etapa de geração dos dados, os com diâmetro maiores que 10 μm e os com valores de intensidade das amostras teste e referência maiores que o mínimo determinado, depois da subtração do ruído de fundo (*background*). Genes em que os *spots* correspondentes estavam presentes em pelo menos 70% das lâminas foram selecionados.

O valor de intensidade mínima foi determinado pela análise dos *spots* de controle negativo. Para cada tipo de lâmina e cada canal, verde (amostra teste marcada com Cy3) ou vermelho (amostra referência marcada com Cy5), consideramos como intensidade mínima o valor do terceiro quartil das intensidades dos *spots* de controle negativo. Assim, para as lâminas Hsaa as intensidades mínimas foram 95 e 34

para os canais verde e vermelho respectivamente, para as Hsab 240 e 125 e para as Hsac 500 e 70 (Quadro 1).

Quadro 1. Valor do terceiro quartil das intensidades dos spots de controle negativo utilizados como intensidade mínima para cada tipo de lâmina em cada canal, verde (amostra teste marcada com Cy3) ou vermelho (amostra referência marcada com Cy5).

Tipo de lâmina	Valor do terceiro quartil das intensidades dos spots de controle negativo	
	Canal verde (amostra teste)	Canal vermelho (amostra referência)
Hsaa	95	34
Hsab	240	125
Hsac	500	70

A normalização utilizada foi a *lowess* ou intensidade-dependente. Nessa normalização utiliza-se diferentes fatores de normalização para diferentes níveis de intensidade, uma vez que o viés causado pelos corantes pode ser diferente para os spots de baixa em relação aos de alta intensidade. Este método minimiza o efeito decorrente da intensidade, calculando um fator de ajuste por regressão linear ponderada local (Simon, Lam, 2007).

Nas análises para identificar um perfil de expressão gênica relacionado com a resposta ao tratamento, além da normalização, para eliminar o viés causado pelo uso de diferentes lotes de lâminas, utilizamos o método chamado *Distance Weighted Discrimination* (DWD) (Benito et al., 2004). O DWD é uma ferramenta capaz de corrigir vieses sistemáticos presentes em grupos de dados ao fazer um ajuste global para compensar esses vieses. Ele corrige vieses como: diferentes fontes de RNA, diferentes plataformas de *microarrays* e diferentes lotes de *microarrays*. Os parâmetros utilizados na análise foram: *non-standardized DWD* e *center at 0*. Para utilizar esse método os dados têm que possuir valores para todos os spots em todas as lâminas. No entanto, alguns genes poderiam não ter valores de expressão em algumas lâminas devido aos critérios de filtro de spots aplicados inicialmente. Assim, para estimar valores para esses spots utilizamos o método *K- Nearest Neighbors* (KNN), que seleciona genes com perfil de expressão similar ao gene de interesse para dar um valor ao spot

(Troyanskaya et al., 2001). Esse método foi utilizado na ferramenta de análise *Significance Analysis of Microarray* (SAM) (Tusher et al., 2001) com $K=10$.

3.5.2. Análise não-supervisionada

A análise não-supervisionada foi realizada para verificar se havia algum padrão de agrupamento nos experimentos não esperado, que poderia interferir nos resultados encontrados nas análises seguintes. Nesse tipo de análise apenas os dados de expressão dos genes que passam pelos filtros iniciais são utilizados, nenhuma característica da amostra é fornecida e nenhuma seleção de genes é realizada. Para realizar essa análise utilizamos o método de agrupamento hierárquico, usando correlação centralizada e ligação média no programa *BRB Array Tools*. Essa análise produz um agrupamento sequencial das amostras de acordo com a similaridade ou distância entre perfis de expressão gênica. O resultado é representado por um dendrograma. Nessa visualização as amostras podem ser identificadas de acordo com várias características, essa mudança não altera o agrupamento, uma vez que as amostras são agrupadas somente pelo seu perfil de expressão gênica.

3.5.3. Análise para identificar um preditor molecular

Para desenvolver um preditor (ou classificador molecular) capaz de prever a classe de uma nova amostra baseado no perfil de expressão gênica utilizamos a ferramenta de predição de classes no programa *BRB Array Tools*. Essa ferramenta inclui sete diferentes métodos multivariados de classificação com validação cruzada dos resultados (*cross-validation*). Os métodos de classificação são: *Compound Covariate Predictor*, *Diagonal Linear Discriminant Analysis*, *K-Nearest Neighbor Predictor* para $K=1$ e 3 , *Nearest Centroid Predictor*, *Support Vector Machine Predictor* e *Bayesian Compound Covariate Predictor*. Os métodos incorporam genes que foram diferencialmente expressos entre as classes (R e NR) a um nível de significância de $p=0,001$ utilizando o teste t de variância aleatória (*random-variance t test*) (Wright, Simon, 2003).

Para prevenir um preditor com sobre-ajuste ou *overfitting*, estimamos o erro de predição de cada método utilizando a validação cruzada do tipo *Leave-one-out cross-validation* (LOOCV) como descrito (Simon et al., 2003). O *overfitting* pode ocorrer pois o número de parâmetros de um modelo é tão grande em relação ao número de casos disponíveis que os parâmetros do modelo são otimizados para esses dados. Assim, o

modelo servirá para os dados originais mas pode não prever dados independentes (Simon et al., 2003). Com o LOOCV, o processo de validação cruzada omite uma amostra cada vez. Para cada amostra omitida, a análise inteira é repetida desde o início, incluindo a determinação de quais genes são significantes na análise univariada nas amostras que restaram, chamadas amostras treino. Dessa lista de genes, um preditor multivariado é construído e aplicado na amostra que foi omitida. O programa grava se a predição foi correta ou não. Esse processo é repetido, até que todas as amostras sejam omitidas uma vez. O resultado mostra quais amostras foram corretamente e incorretamente classificadas pelo preditor. Para avaliar se a taxa de erro da validação cruzada para um método foi significativamente menor do que a esperada para uma predição aleatória, as classes foram aleatoriamente permutadas 1000 vezes e o processo de LOOCV repetido. Entre os sete métodos foram considerados os resultados dos que possuíam uma permutação com $p < 0,001$.

3.5.4. Análise para identificar genes diferencialmente expressos

Para identificar genes diferencialmente expressos entre tecido tumoral e normal utilizamos a ferramenta de comparação de classes do programa *BRB Array Tools*.

Inicialmente, as 54 lâminas para essa análise foram divididas em dois grupos. Para cada grupo foi realizada uma análise de comparação de classes entre grupos de *arrays*, que compara duas ou mais classes pré-definidas. Neste caso foram comparadas as classes normal e tumor, utilizando o teste t de variância aleatória e $p < 0,05$. Os genes em comum entre os resultados das análises dos dois grupos, significantes em ambos, foram selecionados para análises posteriores.

Foi realizada também a análise de comparação de classes entre os canais verde (amostra teste) e vermelho (amostra referência) utilizando análise de variância (ANOVA, analysis of variance). Essa ferramenta pode ser usada para encontrar genes diferencialmente expressos entre os dois canais. Como nas lâminas que foram hibridizadas as amostras de tecido normal também foram hibridizadas a amostra de referência, que é um *pool* de amostras de tumor, ao compararmos o canal verde (amostra de tecido normal) com o canal vermelho (amostra de referência) nessas lâminas, estamos realizando também uma análise do tipo normal vs tumor. Essa análise foi utilizada apenas para minimizar o número de resultados falso positivos que poderiam existir na lista de genes determinada anteriormente, uma vez que essa análise não tem o mesmo poder da anterior pois a referência era a mesma para todas

amostras de tecido normal. Assim, realizamos essa análise utilizando a lista de genes obtida e excluímos genes que: (a) não possuíam variação coerente em relação aos resultados da análise anterior, ou seja, em uma análise o gene era mais expresso na amostra de tumor e em outra mais expresso na amostra de tecido normal (ou vice-versa); (b) apresentavam $p > 0,1$.

3.5.5. Análise de grupos funcionais de genes

A análise de grupos de genes com a lista de genes diferencialmente expressos entre tecido tumoral e normal foi realizada por meio da análise de vias (*pathways analysis*) e de redes gênicas (*gene networks*).

Análise de vias

Após a determinação da lista de genes diferencialmente expressos entre tumor e normal, verificamos quais vias estão mais representadas tanto na lista com baixa quanto na com alta expressão no tumor.

Utilizamos três ferramentas e três bancos de dados distintos com o objetivo de aumentar a consistência do resultado final. Os programas utilizados foram :

- *Database for Annotation, Visualization and Integrated Discovery 2007* (DAVID, Dennis et al., 2003). Ferramenta utilizada: *Functional Annotation Tool*.
- *Babelomics* versão 2.0 (Al-Shahrour et al., 2005). Ferramenta utilizada: *Fast Assignment and Transference of Information using Gene Ontology Plus* (FatiGOplus).
- *Ingenuity Pathways Analysis* versão 5.5.1 (IPA, Ingenuity® Systems, www.ingenuity.com). Ferramenta utilizada: *Core Analyse*.

Os bancos de dados utilizados pelos programas foram:

- *Kyoto Encyclopedia of Genes and Genomes Pathway* (KEGG Pathway, Ogata et al., 1999): é uma coleção de mapas de vias representando nosso conhecimento nas interações moleculares e redes de reação para o metabolismo, processamento da informação genética, processamento da informação do ambiente, processos celulares e doenças humanas.
- *BioCarta* (<http://www.biocarta.com>): fornece modelos que mostram como os genes interagem e também cataloga e sumariza importantes fontes fornecendo informações de aproximadamente 120 mil genes de múltiplas espécies. As vias

são divididas em diversas categorias como, por exemplo, regulação do ciclo celular, sinalização celular e metabolismo.

- *Canonical Pathways*: vias clássicas pertencentes ao banco de dados do *Ingenuity Pathways Analysis*, IPA. São vias metabólicas e de sinalização celular bem caracterizadas que têm sido organizadas por pesquisadores. As informações contidas nessas vias vêm de artigos de revistas específicas, artigos de revisão, livros texto e do KEGG *ligand*.

No quadro 2 estão esquematizadas as características das análises efetuadas. Nos programas DAVID e *Babelomics* utilizamos o KEGG e o *BioCarta* e no IPA as informações das vias clássicas (*Canonical Pathways*). No *Babelomics* e no IPA é utilizado o teste exato de Fisher monocaudal e no DAVID uma modificação mais conservadora desse teste, o EASE score. Nas análises é verificado se na lista de genes submetida há uma representação significativa de vias em relação a uma lista de comparação. Essa lista de comparação consistia em um conjunto de genes de *background*. No caso do DAVID foram usados todos os genes do genoma humano, no *Babelomics* os genes contidos na nossa lâmina sem os genes diferencialmente expressos e, no IPA, moléculas pertencentes à biblioteca de vias clássicas do programa.

Quadro 2. Características dos três programas utilizados nas análises de vias: Database for Annotation, Visualization and Integrated Discovery (DAVID), *Babelomics* e Ingenuity Pathways Analysis (IPA).

Parâmetros	Programa de análise		
	DAVID	<i>Babelomics</i>	IPA
Ferramenta utilizada	<i>Functional Annotation Tool</i>	<i>FatiGOplus</i>	<i>Core Analyse</i>
Teste estatístico	Teste exato de Fisher modificado	Teste exato de Fisher	Teste exato de Fisher
Lista de <i>background</i>	Genes do genoma humano	Genes contidos na lâmina sem os genes diferencialmente expressos	Moléculas pertencentes à biblioteca de vias clássicas do IPA
Banco de dados	KEGG e BioCarta	KEGG e BioCarta	Vias Clássicas

Nas análises, em cada programa obtivemos inicialmente as vias representadas com nível de significância (p) menor que 0,10 nas listas de genes com baixa e com alta expressão no tumor. No entanto, para que uma via fosse selecionada ela deveria aparecer nos resultados de pelo menos dois programas e com um valor de p menor que 0,05 em pelo menos um.

Para explorar as vias obtidas em relação à literatura existente utilizamos uma ferramenta, o *Microarray Literature-based Annotation* (MILANO) (Rubinstein, Simon, 2005), para verificar o número de estudos já realizados com os genes que compõem cada via. O MILANO realiza buscas automáticas de artigos no *PubMed* (<http://www.ncbi.nlm.nih.gov/sites/entrez/>) contendo co-ocorrências de termos com uma lista de genes. Inicialmente, definimos os termos que analisaríamos junto com as vias. Os termos definidos foram:

- (cervical OR cervix) AND (cancer OR tumor OR carcinoma OR neoplasma OR neoplasia): para abordar o câncer de colo do útero.
- (cancer OR tumor OR carcinoma OR neoplasma OR neoplasia): para abordar o câncer em geral.

- squamous AND (cancer OR tumor OR carcinoma OR neoplasma OR neoplasia): para abordar câncer de células escamosas que é o mesmo tipo celular dos tumores que utilizamos.

Em seguida, para realizarmos a busca, definimos a lista de genes de cada via. Como para cada programa tínhamos uma lista, não necessariamente idêntica, consideramos que os genes que compõem uma via são aqueles que aparecem ao menos uma vez em um dos programas. Em cada análise do MILANO submetemos uma lista de genes correspondente a uma via junto com cada termo definido anteriormente. Assim, obtivemos como resultado o número de artigos em que um gene é citado junto com um termo. Por exemplo, para o gene CDK2, que faz parte da via do ciclo celular, e o termo que aborda o câncer de colo do útero ((cervical OR cervix) AND (cancer OR tumor OR carcinoma OR neoplasma OR neoplasia)) foram encontrados 38 artigos que citam ambos juntos.

Análise de redes gênicas

Nas análises de redes gênicas as interações entre os genes são formadas a partir de informações que indicam a ligação entre eles. É diferente das análises de vias onde as interações entre os genes já são estabelecidas dentro de cada via. Nos nossos dados as interações entre os genes foram formadas baseada na similaridade do perfil de anotação funcional, ou seja, haverá uma interação entre um par de genes se eles compartilharem um perfil de anotação funcional, se forem funcionalmente relacionados. Para medir essa relação utilizamos uma ferramenta do programa DAVID. Essa ferramenta identifica pares de genes relacionados ao medir a similaridade do seu perfil de anotação global. Essa análise é baseada na hipótese de que se dois genes possuem um perfil de anotação similar, eles deveriam ser funcionalmente relacionados (Huang et al., 2007). Primeiramente, uma matriz de anotação gene-termo é compilada em uma forma binária usando vários termos de anotação em 14 categorias coletadas no banco de dados do programa, o *DAVID knowledgebase*. As categorias incluem: *Gene Ontology (GO) Biological Process, GO Molecular Function, GO Cellular Component, KEGG Pathways, BioCarta Pathways, Swiss-Prot Keywords, BBID Pathways, SMART Domains, NIH Genetic Association DB, UniProt Sequence Features, COG/KOG Ontology, NCBI OMIM, InterPro Domains, e PIR Super Family Names*. Em seguida, a estatística *kappa* é adotada para medir a co-ocorrência de anotação de um dado par de genes. O valor de *kappa* varia de zero a um. Huang e colaboradores,

2007, realizaram um estudo para testar a ferramenta e verificaram que essa medida de similaridade funcional é capaz de detectar especificamente as relações gene-gene e que isso ocorre particularmente para pares com um *kappa* de 0,35 ou acima. Além disso, ao avaliarmos a distribuição nos nossos dados dos valores de *kappa* em relação ao número de *clusters* (ou grupos) de genes verificamos que a região onde a curva começa a subir com uma inclinação maior, ou seja, onde o número de *clusters* de genes começa a aumentar é em torno de 0,35. Assim, decidimos usar esse valor como ponto de corte mínimo de *kappa*.

Após a determinação da interação entre os genes, seguimos com a construção das redes. Inicialmente, realizamos a visualização das redes e determinação dos *clusters* ou sub-redes de genes utilizando dois programas:

- *CFinder* (Palla et al., 2005). O algoritmo do programa localiza as comunidades *k-cliques* (*clusters*) das redes. Assim, um *cluster* pode ser interpretado como uma união de sub-gráficos completos que compartilham vértices (genes). Tais sub-gráficos em uma rede são chamados *k-cliques*, onde *k* refere-se ao número de vértices no sub-gráficos (Adamcsek et al., 2006). Como para uma rede de interação de proteínas é indicada a utilização de um valor de *k* igual a quatro (Palla et al., 2005) utilizamos esse valor nas nossas redes.
- *Cytoscape* (Shannon et al., 2003; Cline et al., 2007). Esse programa permite a visualização das redes e possui *plugins* para realizar diversas análises. Utilizamos um desses *plugins*, o *Molecular Complex Detection* (MCODE) (Bader, Hogue, 2003), para gerar os *clusters* de genes existentes. Esse algoritmo detecta regiões densamente conectadas em amplas redes que podem representar complexos moleculares. Os *clusters* são ordenados de acordo com uma pontuação. Um resultado é significativo se tiver uma pontuação alta (maior que um) e número de genes em interações razoáveis (http://cytoscape.org/cgi-bin/moin.cgi/Presentations/07_Complexes).

Em seguida, comparamos os resultados obtidos nos dois programas e selecionamos sub-redes presentes em ambos. Por último, realizamos a anotação funcional das sub-redes selecionadas utilizando a ferramenta *GeneCodis* (Carmona-Saez et al., 2007). Com essa ferramenta realizamos a anotação funcional dos genes

nas seguintes categorias: *GO Biological Process*, *GO Molecular Function*, *GO Cellular Component* e *KEGG Pathways*.

3.5.6. Comparação dos genes diferencialmente expressos entre tumor e normal com dados de outros trabalhos publicados

Para verificar como o perfil de expressão gênica identificado em nosso trabalho se relacionava com o perfil gerado em outros trabalhos da literatura com os mesmos genes, realizamos duas análises comparativas: (i) análise dos genes identificados em um trabalho sobre câncer de colo do útero publicado por outros autores; (ii) análise dos genes identificados em vários outros trabalhos sobre câncer.

i) Análise dos genes identificados em um trabalho sobre câncer de colo do útero publicado por outros autores

Analisamos nossos resultados em relação aos dados de outro trabalho realizado em câncer de colo do útero, publicado recentemente (Scotto et al., 2008), para verificar se havia similaridade no perfil dos nossos genes entre os dois trabalhos. Escolhemos o trabalho acima mencionado devido às seguintes características: os dados brutos estão disponibilizados no site do *Gene Expression Omnibus* (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) sob o número GSE9750, possui amostras de tecido normal e tumoral, as amostras de tumor são do mesmo tipo histológico que as nossas e a maioria com o mesmo estágio clínico.

Na análise utilizamos todas as amostras de tecido normal (n=20) e, entre as amostras tumorais, apenas as com estágio clínico IIb a IIIb (n=20). Inicialmente, realizamos uma análise de comparação entre as classes tumor e normal utilizando o teste t de variância aleatória apenas com os genes da nossa lista. O objetivo era verificar para cada gene a razão entre a expressão da amostra de tecido normal em relação à tumoral (razão normal/tumor) bem como se o gene era significativo ($p < 0,05$) ou não nesse trabalho. Em seguida analisamos a concordância entre esse estudo e o nosso, utilizando o teste exato de Fisher.

(ii) Análise dos genes identificados em outros trabalhos de câncer

Para essa análise realizamos uma busca de dados de expressão gênica de *microarrays* disponíveis. Para obter dados de outros tipos de tumor utilizamos a base de dados de expressão gênica em câncer *BRB-ArrayTools Data Archive for Human Cancer Gene Expression* (Zhao, Simon, 2008). Nessa fonte de dados era possível importar os dados já no formato para ser utilizado na ferramenta *BRB Array Tools*, que utilizamos nas análises. Selecionamos 14 estudos que possuíam amostras de tecido tumoral e normal. Desses selecionamos sete estudos (Hippo et al., 2002; Haslinger et al., 2004; Gordon et al., 2005; Ishikawa et al., 2005; Kimchi et al., 2005; Talantov et al., 2005; Wachi et al., 2005) que tinham no mínimo cinco amostras em cada grupo. Para buscar outros trabalhos de câncer de colo do útero utilizamos o GEO. Selecionamos mais três trabalhos (Wong et al., 2003; Pyeon et al., 2007; Biewenga et al., 2008) além do utilizado na análise anterior (Scotto et al., 2008), com no mínimo cinco amostras em cada grupo. Esses três trabalhos não foram utilizados anteriormente pois os estádios clínico das amostras eram na maioria iniciais. Assim, no total ficamos com 11 trabalhos (Quadro 3).

Após a obtenção dos dados de expressão, para cada trabalho, identificamos a razão entre os valores de expressão das amostras de normal e tumor (razão normal/tumor) para todos os genes da nossa lista. Em seguida, realizamos um agrupamento hierárquico dos estudos, usando correlação centralizada e ligação média no programa MultiExperiment Viewer versão 4.2 (MeV) (<http://www.tm4.org/mev.html>), utilizando os dados da razão normal/tumor dos genes selecionados. O objetivo era verificar como os trabalhos se agrupavam de acordo com o perfil dos nossos genes utilizando os dados de expressão de cada trabalho.

Quadro 3. Informações sobre os estudos utilizados na análise da nossa lista de genes em outros trabalhos.

Localização ou tipo do tumor	Referência do estudo	Plataforma de <i>microarray</i>	Número de amostras (normal / tumor)
Colo do útero	Wong et al., 2003	<i>Microarray facility at the Cancer Institute of New Jersey (nylon membrane)</i>	8/25
Colo do útero	Pyeon et al., 2007	<i>Affymetrix GeneChip (U133 Plus 2.0)</i>	8/20
Colo do útero	Biewenga et al., 2008	<i>Agilent Whole genome human oligonucleotide microarrays</i>	5/35
Colo do útero	Scotto et al., 2008	<i>Affymetrix GeneChip (U133A)</i>	21/32
Estômago	Hippo et al., 2002	<i>Affymetrix GeneChip (HuGeneFL)</i>	8/22
Leucemia	Haslinger et al., 2004	<i>Affymetrix GeneChip (U95A)</i>	11/100
Mesotelioma pleural	Gordon et al., 2005	<i>Affymetrix GeneChip (U133A)</i>	10/44
Pâncreas	Ishikawa et al., 2005	<i>Affymetrix GeneChip (U133A)</i>	25/24
Esôfago	Kimchi et al., 2005	<i>Affymetrix GeneChip (U133A)</i>	8/8
Melanoma	Talantov et al., 2005	<i>Affymetrix GeneChip (U133A)</i>	7/45
Pulmão	Wachi et al., 2005	<i>Affymetrix GeneChip (U133A)</i>	5/5

4 RESULTADOS

4.1. Busca de um perfil de expressão gênica capaz de prever a resposta ao tratamento

4.1.1. Análises iniciais

Nas análises foram utilizadas 23 lâminas hibridizadas com amostras de tecido tumoral, sendo que em uma delas foi hibridizada um *pool* com duas amostras. Neste caso as amostras possuíam características clínicas iguais (pacientes classificadas como R, com estadió clínico IIb e que receberam como tratamento RT+QT). Seis amostras foram hibridizadas em lâminas do tipo Hsab e 17 em Hsac (Quadro 4).

Quadro 4. Distribuição das lâminas utilizadas na análise de resposta ao tratamento em relação ao tipo de lâmina, Hsab ou Hsac, e a classe da amostra, respondedora (R) ou não respondedora (NR) ao tratamento.

Classe das amostras	Tipo de lâmina		Total
	Hsab	Hsac	
R	2	11	13
NR	4	6	10
Total	6	17	23

Após o escaneamento das lâminas hibridizadas e a obtenção dos valores de expressão, os dados foram submetidos ao processo de filtragem e normalização (Figura 1). Com isso sobraram 7.289 genes para as análises.

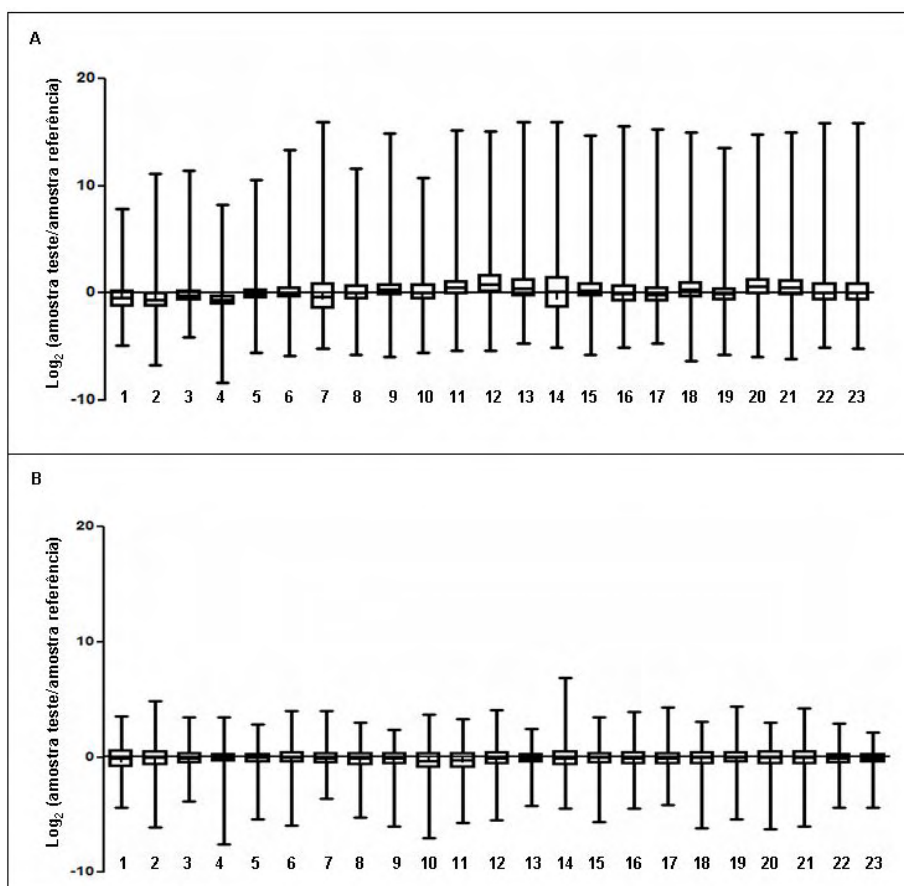


Figura 1. Gráficos do tipo *box-plot* do logaritmo da razão entre as intensidades do canal verde (amostra teste) e canal vermelho (amostra referência) na base 2, das 23 lâminas utilizadas nas análises antes (A) e após (B) o procedimento de normalização. As lâminas de 1 a 6 são do tipo Hsab e as lâminas de 7 a 23 do tipo Hsac.

Inicialmente, realizamos uma análise não supervisionada utilizando o agrupamento hierárquico. Visualizamos o agrupamento com as seguintes características: tipo de lâmina (Hsab ou Hsac), estadio clínico (IIb ou IIIb) e tipo de tratamento (RT ou RT+QT). Quando visualizamos o agrupamento pelo tipo de lâmina, observamos que as amostras formam dois grupos primários cujas amostras, com exceção de duas, distribuem-se de acordo com o tipo de lâmina (Figura 2).

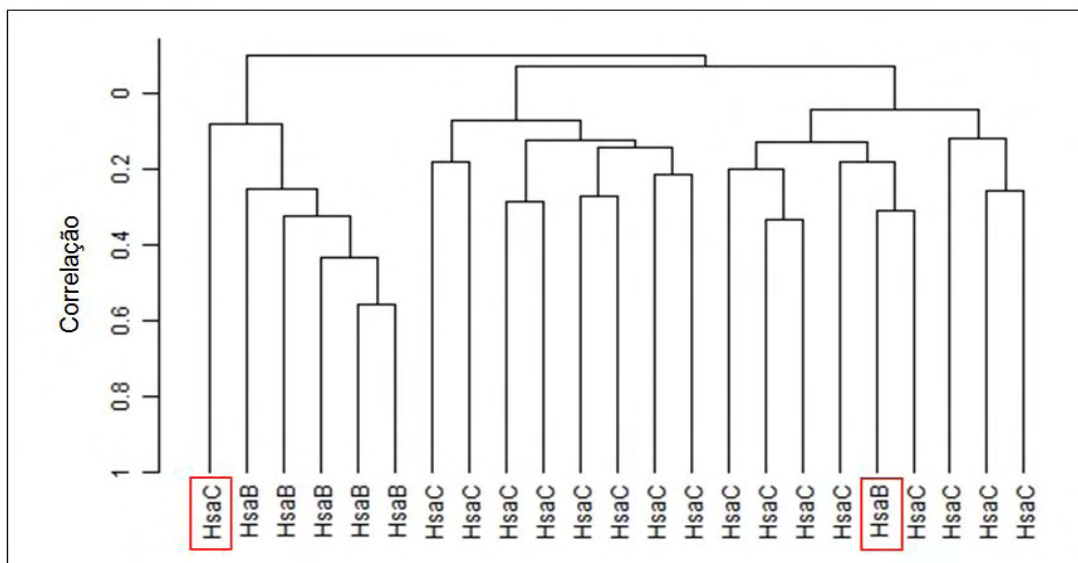


Figura 2. Dendrograma do agrupamento hierárquico das amostras. Verificamos que as amostras se distribuem em dois grupos de acordo com o tipo de lâmina em que foram hibridizadas, com exceção de duas (indicadas pelo quadrado vermelho).

Para corrigir esse viés causado pelo uso de diferentes impressões de lâminas utilizamos a ferramenta DWD. Após submetermos os dados ao DWD realizamos um novo agrupamento e não observamos mais o padrão de agrupamento relacionado com o tipo de lâmina (Figura 3).

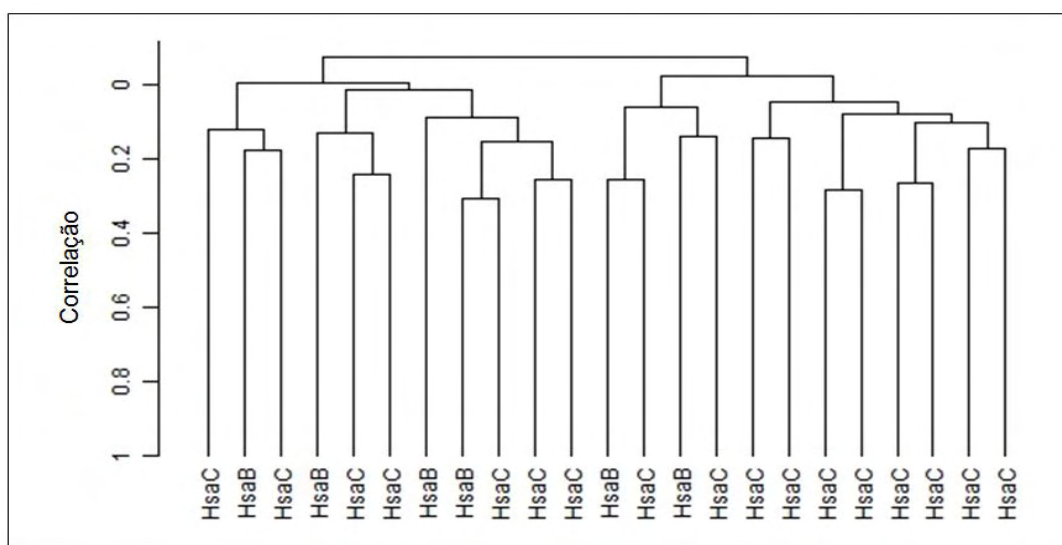


Figura 3. Dendrograma do agrupamento hierárquico das amostras após submeter os dados ao método *Distance Weighted Discrimination* (DWD).

4.1.2. Análise para identificação de um preditor molecular

Para tentar identificar um preditor, ou perfil de expressão gênica, capaz de prever a resposta ao tratamento utilizamos a ferramenta de predição de classes.

Como não há consenso sobre o melhor algoritmo matemático para predição de classes e considerando que o uso de múltiplos algoritmos aumenta a confiança e a validade dos resultados (Simon et al., 2003) utilizamos sete diferentes algoritmos de predição de classes assim como a validação cruzada do tipo *leave-one-out* (LOOCV). No nosso trabalho, nenhum método de predição teve um valor de p da permutação significativo. Apenas para se ter idéia da atuação dos classificadores ou preditores construídos em cada análise de LOOCV verificamos quantas amostras foram classificadas corretamente. Das 23 amostras analisadas, considerando que uma amostra é classificada corretamente quando seis dos sete métodos acertam a classificação, apenas 10 (43,5%) foram corretamente classificadas (Quadro 5). O número de genes que passaram pelo critério de $p < 0,001$ da análise univariada foi baixo, seis genes sendo que só três deles participaram de todos os preditores gerados durante a LOOCV. Assim, não verificamos nenhum preditor consistente capaz de prever a classe das amostras.

Quadro 5. Atuação dos preditores durante as validações cruzadas em cada um dos sete métodos utilizados nas 23 lâminas analisadas. O acerto da classe da amostra é indicado por “A”, o erro por “E” e quando a amostra não pôde ser classificada pelo método, por “NC”. Em amarelo na última coluna estão as amostras que foram classificadas corretamente (acerto em pelo menos seis dos sete métodos).

Amostras	Classe das amostras	Atuação dos preditores durante a validação cruzada em cada método Acerto (A), Erro (E), Não classificada (NC)							Número de acertos
		<i>Compound Covariate Predictor</i>	<i>Diagonal Linear Discriminant Analysis</i>	<i>1-Nearest Neighbor</i>	<i>3-Nearest Neighbors</i>	<i>Nearest Centroid</i>	<i>Support Vector Machines</i>	<i>Bayesian Compound Covariate Predictor</i>	
1	NR	A	A	A	A	A	A	A	7
2	NR	E	E	E	E	E	E	NC	0
3	NR	E	E	E	E	E	E	E	0
4	NR	E	E	E	E	E	E	E	0
5	NR	E	E	E	E	E	A	E	1
6	NR	A	A	A	A	A	A	A	7
7	NR	E	E	E	E	E	E	E	0
8	NR	A	A	A	A	A	A	A	7
9	NR	A	A	A	A	A	A	A	7
10	NR	E	E	A	E	E	E	E	1
11	R	E	E	E	E	E	E	NC	0
12	R	E	E	A	E	E	E	NC	1
13	R	E	E	E	A	A	E	NC	2
14	R	A	A	E	A	A	A	NC	5
15	R	E	E	E	E	E	E	E	0
16	R	A	A	E	A	A	E	A	5
17	R	A	A	A	A	A	A	A	7
18	R	A	A	A	A	A	E	A	6
19	R	A	A	E	A	A	A	A	6
20	R	E	E	E	E	A	E	NC	1
21	R	A	A	E	A	A	A	A	6
22	R	A	A	A	A	A	A	NC	6
23	R	A	A	E	A	A	A	A	6

Esse resultado nos levou a duas hipóteses: (i) a correção do viés pelo DWD pode ter removido o sinal de genes importantes relacionados com a resposta ao tratamento ou (ii) não há nos nossos dados genes que possam ser utilizados para elaborar um perfil de expressão gênica capaz de prever a resposta ao tratamento.

Para verificar a primeira hipótese analisamos separadamente as lâminas hibridizadas em apenas um tipo de lote, de maneira a não precisar utilizar o método DWD. Para isso, selecionamos as lâminas Hsac pois era o tipo que tinha mais amostras hibridizadas (n=17).

Realizamos as etapas de filtragem e normalização apenas com as lâminas Hsac e sobraram 8.170 genes para a análise. Na análise de predição de classes, novamente, nenhum método de predição foi significativo. Embora tenha aumentado a porcentagem de amostras classificadas corretamente, 64,7% (Quadro 6), a atuação dos preditores

4.2. Busca de genes com expressão alterada em amostras de tumores de colo do útero

4.2.1. Análises iniciais

Nas análises foram utilizadas 34 lâminas hibridizadas com amostras de tecido tumoral e 20 com amostras de tecido normal. Seis lâminas das amostras de tecido tumoral foram hibridizadas com um *pool* formado por duas amostras. As amostras de tumor foram hibridizadas em dois tipos de lâminas, Hsab e Hsac, enquanto que as de tecido normal foram hibridizadas nas lâminas Hsaa (Quadro 7).

Quadro 7. Distribuição das lâminas utilizadas na análise de comparação entre tecido tumoral e normal em relação ao tipo de lâmina, Hsaa, Hsab ou Hsac, e a classe da amostra, normal ou tumor.

Classe das amostras	Tipo de lâmina			Total
	Hsaa	Hsab	Hsac	
Normal	20	0	0	20
Tumor	0	13	21	34
Total	20	13	21	54

Após o escaneamento das lâminas hibridizadas e a obtenção dos valores de expressão, os dados foram submetidos ao processo de filtragem e normalização (Figura 4). Com isso sobraram 7.468 *spots* para as análises.

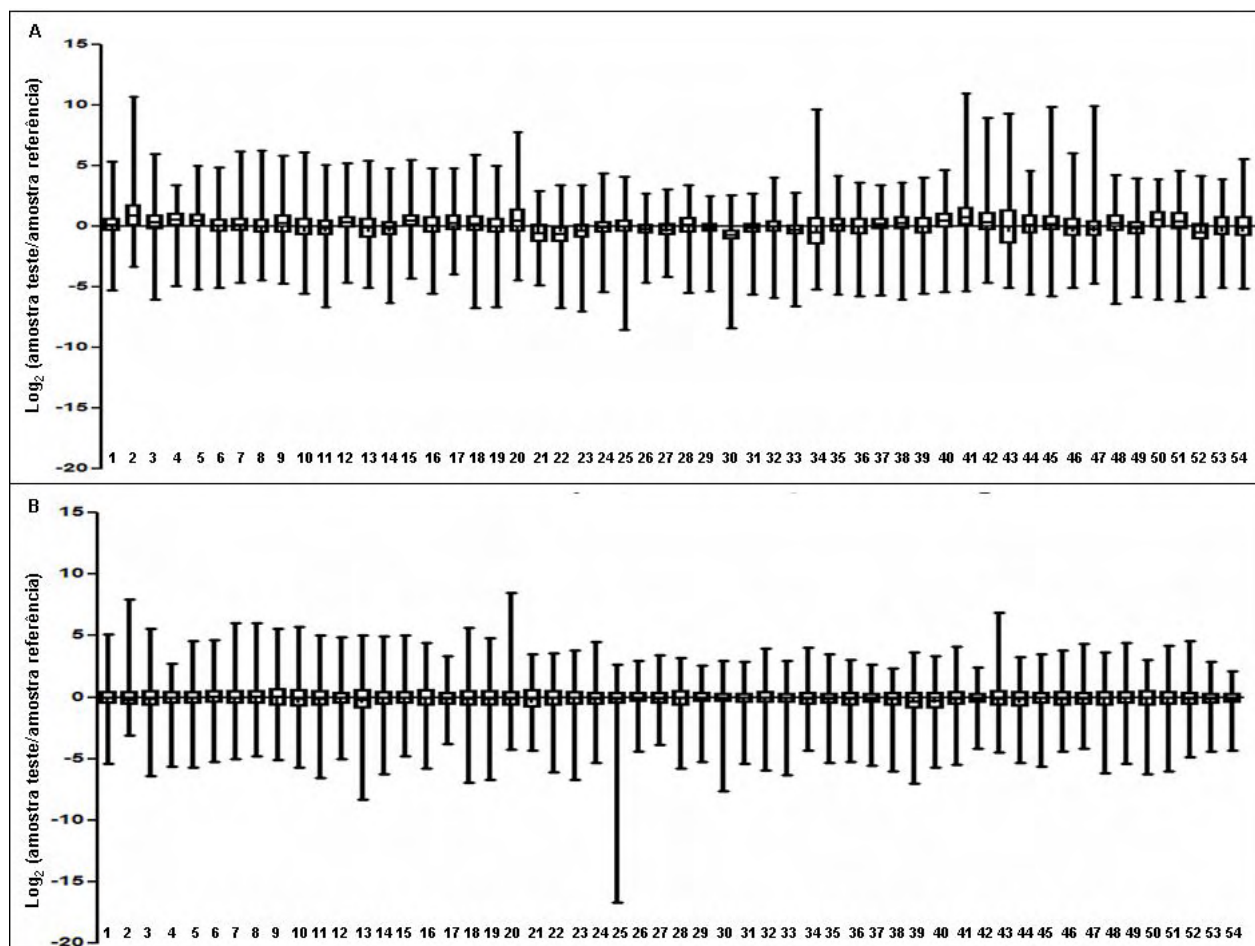


Figura 4. Gráficos do tipo *box-plot* do logaritmo da razão entre as intensidades do canal verde (amostra teste) e canal vermelho (amostra referência) das 54 lâminas utilizadas nas análises antes (A) e após (B) o procedimento de normalização. As lâminas de 1 a 20 são do tipo Hsaa hibridizadas com amostras de tecido normal, as de 21 a 33, Hsab com tecido tumoral e as de 34 a 54, Hsac com tecido tumoral.

4.2.2 Análises para geração de uma lista de genes diferencialmente expressos entre tecido tumoral e normal

Como iríamos utilizar as mesmas lâminas de tecido tumoral utilizadas na análise anterior (reposta ao tratamento) e mais um outro tipo de lâmina Hsaa, esperávamos observar novamente um agrupamento das amostras de acordo com o tipo de lâmina. No entanto, nesses dados o método DWD não se aplicava, pois como todas as amostras Hsaa são de tecido normal, ao retirar o viés entre Hsaa/Hsab e Hsaa/Hsac retiraríamos o sinal biológico entre tumor e normal. Devido a impossibilidade de uso do DWD, realizamos a seleção dos genes diferencialmente expressos entre tecido tumoral e normal em duas etapas para minimizar o número de resultados (genes) falso positivos na nossa lista.

i) Análise de comparação entre as classes tumor e normal em dois grupos de amostras

Dividimos as 54 lâminas em dois grupos, sendo que cada grupo possuía quantidades semelhantes dos três tipos de lâminas (Quadro 8).

Quadro 8. Distribuição das lâminas hibrizadas com tecido tumoral e normal em dois grupos.

Grupos	Hsaa	Hsab	Hsac	Total
I	10	6	11	27
II	10	7	10	27

Em seguida, realizamos uma análise de comparação entre as classes normal e tumor para os dois grupos independentemente, utilizando o teste t com variância aleatória. Obtivemos 2.565 *spots* significantes na análise do grupo I (GI) e 1.863 no grupo II (GII). Desses, 1.038 *spots* eram comuns aos dois grupos com taxa de descobertas falsas (FDR, *false discovery rate*) máxima de 15%. Analisamos esses *spots* para verificar se a variação deles nos dois grupos era coerente, ou seja, se o *spot* era mais expresso na mesma classe (ou no tumor ou no normal) nos dois grupos. Quatorze *spots* foram eliminados após essa análise, restando 1.024 *spots*.

ii) Análise de comparação entre os canais verde (amostra teste) e vermelho (amostra referência) nas lâminas do tipo Hsaa

Nas lâminas Hsaa foram hibridizadas amostras de tecido normal (canal verde) e amostra de referência (canal vermelho) que era um *pool* de amostras de tumor. Assim, também era possível investigar as diferenças entre normal e tumor comparando um canal contra o outro nessas lâminas. Apesar de usar apenas um lote de lâmina, essa análise não tem o mesmo poder que a anterior pois a referência era a mesma para todas as amostras de tecido normal. Dessa forma, essa análise foi usada apenas para minimizar o número de resultados falso positivos que ainda poderiam existir na lista determinada anteriormente.

Realizamos uma comparação de classes entre o canal verde (normal) e vermelho (tumor) usando a lista dos 1.024 *spots*. Com essa análise eliminamos 168 *spots* que, ou não possuíam variação coerente em relação a análise anterior (i) (73 *spots*) , ou apresentavam $p > 0,10$ (95 *spots*).

Dos 856 *spots* restantes, oito estavam presentes em duplicatas e 38 não tinham *gene symbol*, que é a abreviação oficial do nome do gene de acordo com o *HUGO Gene Nomenclature Committee* (<http://www.hugo-international.org/>) e é o identificador que seria utilizado nas análises de grupos de genes. Ficamos no total com 810 *spots* que representavam genes únicos.

Assim, a lista final de genes diferencialmente expressos entre tecido tumoral e normal é composta de 810 genes, dos quais 341 com alta (Anexo 4) e 469 com baixa (Anexo 5) expressão no tumor.

4.3. Análise de grupos funcionais de genes

Após a determinação da lista de genes diferencialmente expressos entre tumor e normal verificamos quais grupos de genes estão mais representados tanto na lista com baixa quanto na com alta expressão no tumor. Com essa análise pretendíamos entender alguns dos mecanismos moleculares envolvidos no câncer de colo do útero. O estudo foi realizado por meio da análise de vias (*pathways analysis*) e de redes gênicas (*gene networks*).

4.3.1. Análise de vias

Realizamos as análises para os 469 genes com baixa expressão no tumor e para os 341 com alta expressão. Após a submissão da lista de genes, os programas realizam uma identificação dos genes e alguns deles não foram reconhecidos. Como cada programa possui seu sistema de identificação, o número de genes final para as análises é diferente em cada um. No caso da lista de genes com baixa expressão no tumor, dos 469 genes restaram 451 no DAVID (96%), 424 no *Babelomics* (90%) e 448 no IPA (95%). Dos 341 genes com alta expressão, ficaram 320 no DAVID (93%), 309 no *Babelomics* (90%) e 329 no IPA (96%) (Quadro 9). A alta porcentagem de genes comuns a todas as análises e os testes estatísticos similares permitiram a comparação dos resultados obtidos com as diferentes ferramentas.

Quadro 9. Número e porcentagem de genes inicialmente submetidos aos três programas de análises, *Database for Annotation, Visualization and Integrated Discovery* (DAVID), *Babelomics* e *Ingenuity Pathways Analysis* (IPA), e o número identificado em cada programa, tanto na lista com baixa quanto com alta expressão no tumor.

Listas de genes segundo expressão no tumor	Número e porcentagem de genes			
	Submetidos aos programas	Identificados pelo DAVID	Identificados pelo <i>Babelomics</i>	Identificados pelo IPA
Baixa	469 (100%)	451 (96%)	424 (90%)	448 (95%)
Alta	341 (100%)	320 (93%)	309 (90%)	329 (96%)

Inicialmente, obtivemos todas as vias representadas com nível de significância (p) menor que 0,10 nos três programas. Para a lista com baixa expressão no tumor identificamos 6 vias no DAVID, 18, no *Babelomics* e 13, no IPA; para a lista com alta expressão, identificamos 13, no DAVID, 32, no *Babelomics* e 13, no IPA (Quadro 10).

Quadro 10. Número de vias obtidas nos programas de análise, *Database for Annotation, Visualization and Integrated Discovery* (DAVID), *Babelomics* e *Ingenuity Pathways Analysis* (IPA), para cada lista de genes, com baixa ou alta expressão no tumor.

Listas de genes segundo expressão no tumor	Número de vias		
	DAVID	<i>Babelomics</i>	IPA
Baixa	6	18	13
Alta	13	32	13
Total	19	50	26

Para que uma via fosse selecionada ela deveria aparecer em pelo menos dois programas e com um valor de p menor que 0,05 em pelo menos um. Após realizarmos o cruzamento dos resultados, ficamos com 17 vias. Dentre elas, seis foram agrupadas, resultando em duas vias. Para agruparmos as vias, verificamos se elas apresentavam similaridades nos seus genes. As duas vias finais eram compostas por genes que apareciam ao menos uma vez nas vias independentes e foram nomeadas com o nome da via mais geral. Duas das seis vias: *Ubiquitin Mediated Proteolysis* (no DAVID e *Babelomics*) / *Protein Ubiquitination Pathway* (no IPA) e *Proteasome*, foram agrupadas em uma única chamada *Protein Ubiquitination Pathway* (Quadro 11). As outras quatro vias, a saber, *Cell Cycle, Cyclins and Cell Cycle Regulation*, *Cell Cycle: G1/S Checkpoint Regulation* e *Cell Cycle: G2/M Checkpoint Regulation*, foram agrupadas em uma via chamada *Cell Cycle* (Quadro 12).

Assim, ficamos com 13 vias: 6 para a lista de genes com baixa expressão no tumor e 7 para a com alta expressão (Quadro 13).

Quadro 11. Lista de genes das vias *Ubiquitin Mediated Proteolysis / Protein Ubiquitination Pathway* e *Proteasome* e da via final (agrupada) denominada *Protein Ubiquitination Pathway*

<i>Ubiquitin Mediated Proteolysis (no DAVID e Babelomics) ou Protein Ubiquitination Pathway (no IPA)</i>			<i>Proteasome</i>		<i>Protein Ubiquitination Pathway</i>
DAVID	<i>Babelomics</i>	IPA	DAVID	<i>Babelomics</i>	LISTA FINAL
ANAPC5	ANAPC5	ANAPC5			ANAPC5
CDC20	CDC20	CDC20			CDC20
		HLA-A			HLA-A
		HLA-B			HLA-B
NEDD4L	NEDD4L	NEDD4L			NEDD4L
		PSMB2	PSMB2	PSMB2	PSMB2
		PSMB4	PSMB4	PSMB4	PSMB4
		PSMB9			PSMB9
		PSMC4	PSMC4	PSMC4	PSMC4
		PSMD14	PSMD14	PSMD14	PSMD14
		PSME1			PSME1
		TAP1			TAP1
		UBB			UBB
		UBD			UBD
UBE1	UBE1	UBE1			UBE1
UBE2C	UBE2C	UBE2C			UBE2C
		UBE2M			UBE2M
		UBE2S			UBE2S

Quadro 12. Lista de genes das vias *Cell Cycle, Cyclins and Cell Cycle Regulation, Cell Cycle: G1/S Checkpoint Regulation* e *Cell Cycle: G2/M Checkpoint Regulation* e da lista de genes da via final (agrupada), denominada *Cell Cycle*.

<i>Cell Cycle</i>		<i>Cyclins and Cell Cycle Regulation</i>		<i>Cell Cycle: G1/S CheckPoint</i>		<i>Cell Cycle: G2/M Checkpoint</i>		<i>Cell Cycle</i>
DAVID	<i>Babelomics</i>	DAVID	<i>Babelomics</i>	<i>Babelomics</i>	IPA	<i>Babelomics</i>	IPA	LISTA FINAL
ANAPC5	ANAPC5							ANAPC5
BUB3	BUB3							BUB3
CCNB1	CCNB1	CCNB1	CCNB1			CCNB1	CCNB1	CCNB1
CCNB2	CCNB2						CCNB2	CCNB2
CDC20	CDC20							CDC20
CDC25C	CDC25C					CDC25C	CDC25C	CDC25C
CDC45L	CDC45L							CDC45L
CDK2	CDK2	CDK2	CDK2	CDK2	CDK2			CDK2
CDK4	CDK4	CDK4	CDK4	CDK4	CDK4			CDK4
CDKN2A	CDKN2A	CDKN2A	CDKN2A	CDKN2A	CDKN2A		CDKN2A	CDKN2A
DBF4	DBF4							DBF4
	E2F3				E2F3			E2F3
					HDAC6			HDAC6
MCM2	MCM2							MCM2
MCM5	MCM5							MCM5
MCM7								MCM7
PCNA	PCNA							PCNA
PLK1	PLK1					PLK1	PLK1	PLK1
PTTG1	PTTG1							PTTG1
PTTG2								PTTG2
							UBB	UBB
							UBD	UBD

Quadro 13. Vias observadas nas listas de genes com alta e baixa expressão no tumor nos três programas analisados, DAVID, *Babelomics* e IPA. As cores azul escuro ($p < 0,05$) e azul claro ($0,05 < p < 0,1$) indicam a faixa do valor de p obtido nas análises e a cor cinza indica que a via ou não foi estatisticamente significativa ($p > 0,1$) ou não foi analisada por não fazer parte do banco de dados do programa.

Vias	Programas			Listas de genes segundo expressão no tumor
	DAVID	<i>Babelomics</i>	IPA	
<i>Antigen Processing and Presentation</i>	Dark Blue	Dark Blue	Dark Blue	Alta
<i>Cell Adhesion Molecules (CAMs)</i>	Dark Blue	Light Blue	Grey	Alta
<i>Cell Cycle</i>	Dark Blue	Dark Blue	Dark Blue	Alta
<i>Methionine Metabolism</i>	Light Blue	Dark Blue	Light Blue	Alta
<i>p53 Signaling</i>	Light Blue	Dark Blue	Dark Blue	Alta
<i>Pentose Phosphate Pathway</i>	Grey	Dark Blue	Dark Blue	Alta
<i>Protein Ubiquitination Pathway</i>	Dark Blue	Dark Blue	Dark Blue	Alta
<i>Fatty Acid Metabolism</i>	Light Blue	Dark Blue	Dark Blue	Baixa
<i>MAPKinase Signaling Pathway</i>	Dark Blue	Light Blue	Grey	Baixa
<i>Oxidative Phosphorylation</i>	Light Blue	Dark Blue	Dark Blue	Baixa
<i>Riboflavin Metabolism</i>	Light Blue	Dark Blue	Grey	Baixa
<i>Ribosome</i>	Dark Blue	Dark Blue	Grey	Baixa
<i>Urea cycle and metabolism of amino groups</i>	Grey	Light Blue	Dark Blue	Baixa

Para explorar as vias obtidas em relação à literatura utilizamos o MILANO para verificar o número de estudos já realizados com os genes que compõem cada via. Como descrito nos métodos (item 3.5.5.) definimos os termos que analisaríamos junto às vias bem como a lista de genes que a representaria. No Quadro 14 é mostrado o número e a porcentagem de genes com artigos citados no *PubMed* encontrados na busca do MILANO. Notamos que, quando consideramos os estudos de câncer em geral, a maioria dos genes em quase todas as vias já foram estudados. Já quando consideramos o câncer de colo do útero, o número de estudos é bem reduzido. O número de genes com citação em relação aos tumores de células escamosas também é reduzido em relação ao câncer em geral. Esses resultados indicam que muitos genes da nossa lista ainda são poucos estudados no câncer de colo uterino apesar de serem estudados em outros tipos de câncer.

Esses resultados da análise de vias foram apresentados no 54º Congresso Brasileiro de Genética, Salvador, BA (Mine et al., 2008) (Anexo 6).

Quadro 14. Número e porcentagem de genes com artigos citados no *PubMed* encontrados na busca do programa MILANO. A análise foi realizada entre cada lista de genes de uma via e os termos relacionados ao câncer de colo do útero, câncer de células escamosas e câncer em geral. É mostrado também o número de genes em cada via presentes nas nossas listas de baixa ou alta expressão no tumor.

Via	Expressão no tecido tumoral em relação ao normal	Número de genes da nossa lista na via	Número e Porcentagem de genes com artigos citados no <i>PubMed</i>		
			Câncer de colo do útero	Câncer de células escamosas	Câncer
Antigen Processing and Presentation	Alta	12	5 (42%)	6 (50%)	12 (100%)
Cell Adhesion Molecules (CAMs)	Alta	9	5 (56%)	6 (67%)	8 (89%)
Cell Cycle	Alta	22	12 (55%)	15 (68%)	22 (100%)
Methionine Metabolism	Alta	3	3 (100%)	2 (67%)	3 (100%)
p53 Signaling	Alta	7	5 (71%)	5 (71%)	7 (100%)
Pentose Phosphate Pathway	Alta	3	0	1 (33%)	1 (33%)
Protein Ubiquitination Pathway	Alta	18	6 (33%)	7 (39%)	17 (94%)
Fatty Acid Metabolism	Baixa	8	2 (25%)	2 (25%)	7 (88%)
MAPKinase Signaling Pathway	Baixa	7	1 (14%)	2 (29%)	7 (100%)
Oxidative Phosphorylation	Baixa	9	0	1 (11%)	5 (56%)
Riboflavin Metabolism	Baixa	3	0	0	3 (100%)
Ribosome	Baixa	14	0	4 (29%)	12 (86%)
Urea Cycle and metabolism of amino groups	Baixa	3	0	0	3 (100%)

4.3.2. Análise de redes gênicas

Na construção das redes gênicas, as interações entre os genes foram formadas quando eles compartilhavam um perfil de anotação funcional. Como descrito nos métodos (item 3.5.5) para medir essa relação utilizamos uma ferramenta do programa DAVID. A co-ocorrência de anotação de um dado par de genes foi medida pela estatística *kappa*.

Após submeter as nossas listas de genes ao programa, obtivemos os valores de *kappa* para todos os pares de genes. Selecionamos os pares com valor de *kappa* maior ou igual a 0,35 para cada lista. Em seguida, realizamos as análises de acordo com os seguintes passos, como descrito nos métodos (item 3.5.5): (a) visualização das redes e determinação dos *clusters* (sub-redes) de genes utilizando os programas CFinder e *Cytoscape*; (b) comparação dos resultados obtidos nos dois programas e seleção de sub-redes presentes em ambos; (c) realização da anotação funcional das sub-redes selecionadas utilizando a ferramenta GeneCodis.

Para a lista de alta expressão no tumor ficaram 906 pares para análise. Obtivemos 14 sub-redes no CFinder e 10 no *Cytoscape*. Após cruzar os resultados, selecionamos 9 sub-redes. Para a lista com baixa expressão no tumor ficaram 1.663 pares para análise. Obtivemos 17 sub-redes no CFinder e 13 no *Cytoscape*. Após cruzar os resultados, selecionamos 14 sub-redes. Em seguida realizamos a anotação funcional das sub-redes (Quadro 15, Figuras 5 e 6).

Quadro 15. Anotação funcional das sub-redes obtidas nas análises das listas de genes com alta e baixa expressão no tumor.

Lista de genes com alta expressão no tumor	Lista de genes com baixa expressão no tumor
1. Transcription Cell Cycle 2. RNA related mechanisms 3. Cell Cycle/ DNA replication/ p53 signaling pathway 4. Ubiquitin-dependent protein catabolic process Protein Ubiquitination Pathway 5. Membrane Cell adhesion molecules (CAMs) 6. Membrane N-Glycan biosynthesis 7. Phosphatase activity 8. Membrane 9. Ubiquitin-dependent protein catabolic process Ubiquitin mediated proteolysis	1. Transcription 2. Protein amino acid phosphorylation MAPK signaling pathway 3. Ribosome / Translation 4. Membrane 5. Proteolysis 6. GTP binding/ GTPase activity 7. RNA related mechanisms 8. Ion transport/ Membrane 9. Membrane 10. Phosphatase activity 11. Keratinization 12. Ubiquitin-dependent protein catabolic process 13. Oxidative phosphorylation 14. Membrane

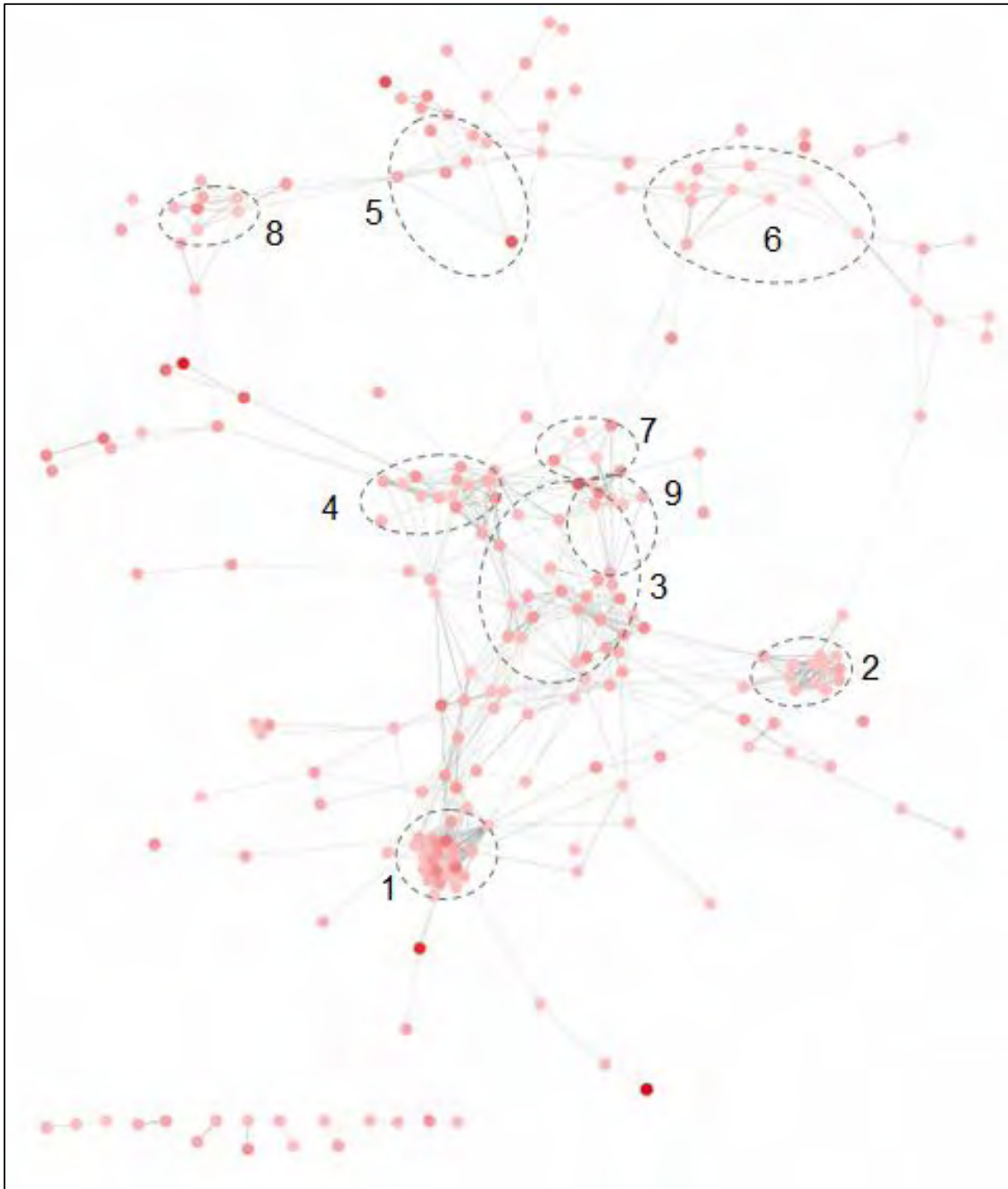


Figura 5. Redes dos genes com alta expressão no tumor usando a estatística *kappa*. Os números indicam as sub-redes listadas no Quadro 15 delimitadas pelos círculos tracejados. Os círculos representam os genes e as linhas denotam as ligações entre eles. A intensidade da cor dos círculos é proporcional à razão da expressão entre amostras de tumor e de tecido normal e a intensidade da cor das linhas é proporcional ao valor de *kappa* entre o par de genes.

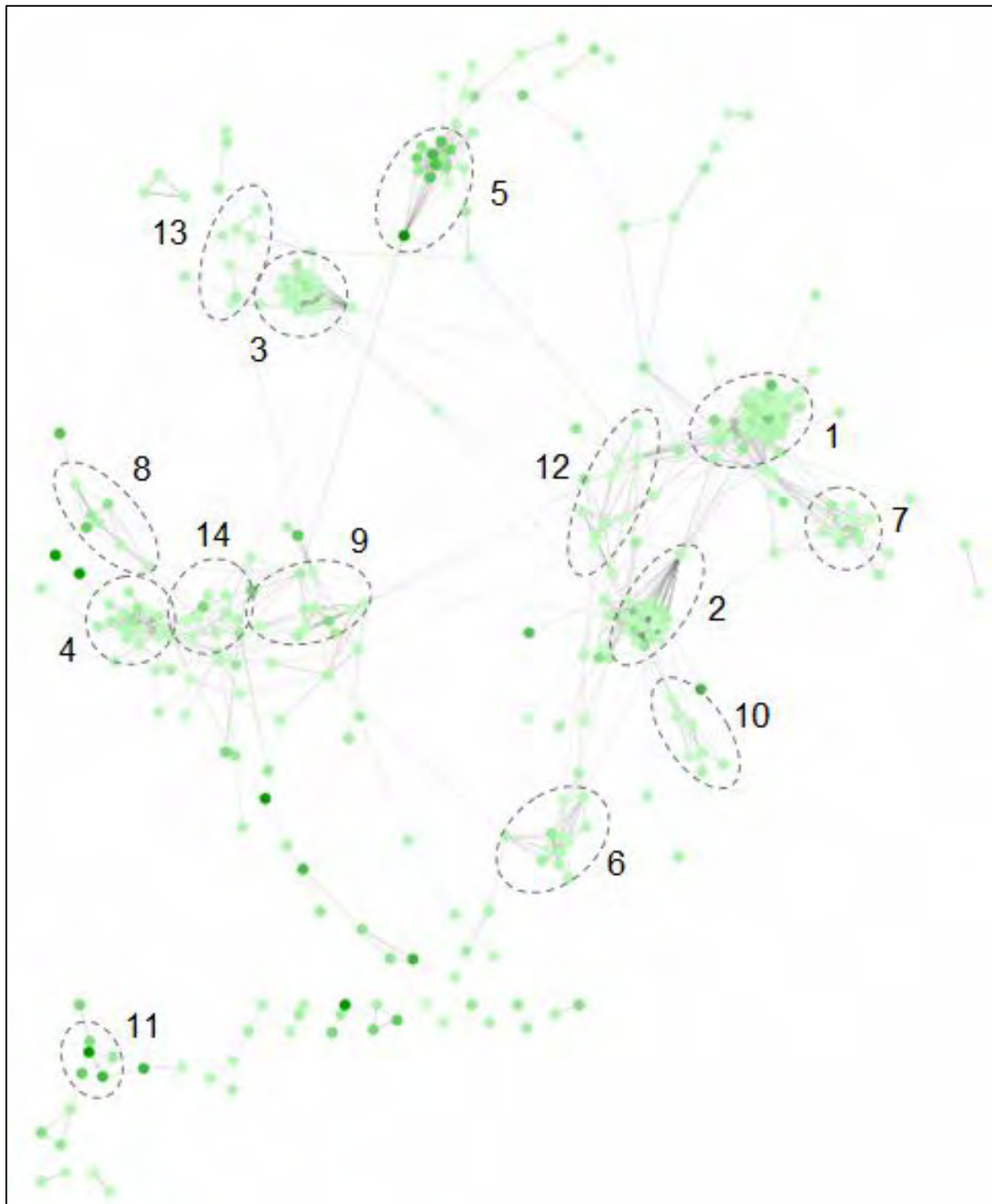


Figura 6. Rede dos genes com baixa expressão no tumor usando a estatística *kappa*. Os números indicam as sub-redes listadas no Quadro 15 delimitadas pelos círculos tracejados. Os círculos representam os genes e as linhas denotam as ligações entre eles. A intensidade da cor dos círculos é proporcional à razão da expressão entre amostras de tecido normal e tumor e a intensidade da cor das linhas é proporcional ao valor de *kappa* entre o par de genes.

Ao comparar esses resultados com os obtidos na análise de vias, verificamos que sete sub-redes apresentavam uma via correspondente: *Cell Adhesion Molecules*, *Cell Cycle*, *MAPK signaling pathway*, *Oxidative Phosphorylation*, *p53 signaling pathway*, *Protein Ubiquitination Pathway* e *Ribosome*. A Figura 7 representa a sub-rede 3 da lista de genes com alta expressão no tumor, relacionada com o ciclo celular, replicação do DNA e via de sinalização da p53. A Figura 8 representa a sub-rede 3 da lista de genes com baixa expressão no tumor, relacionada com o Ribosomo. Verificamos também que entre as sub-redes encontradas em cada lista, alta e baixa expressão no tumor, algumas possuíam a mesma anotação nas duas listas, *Transcription*, *RNA related mechanisms*, *Ubiquitin-dependent protein catabolic process*, *Membrane* e *Phosphatase activity*. Nas redes encontradas na lista de genes com baixa expressão no tumor, nos chamou a atenção a sub-rede 5 que é formada por vários membros de uma família, a das calicreínas (KLKs, *Kallikrein-related peptidase*) (Figura 9).

Esses resultados foram apresentados na *Molecular Medicine Tri-Conference 2009*, *San Francisco*, EUA (Mine et al., 2009) (Anexo 7).

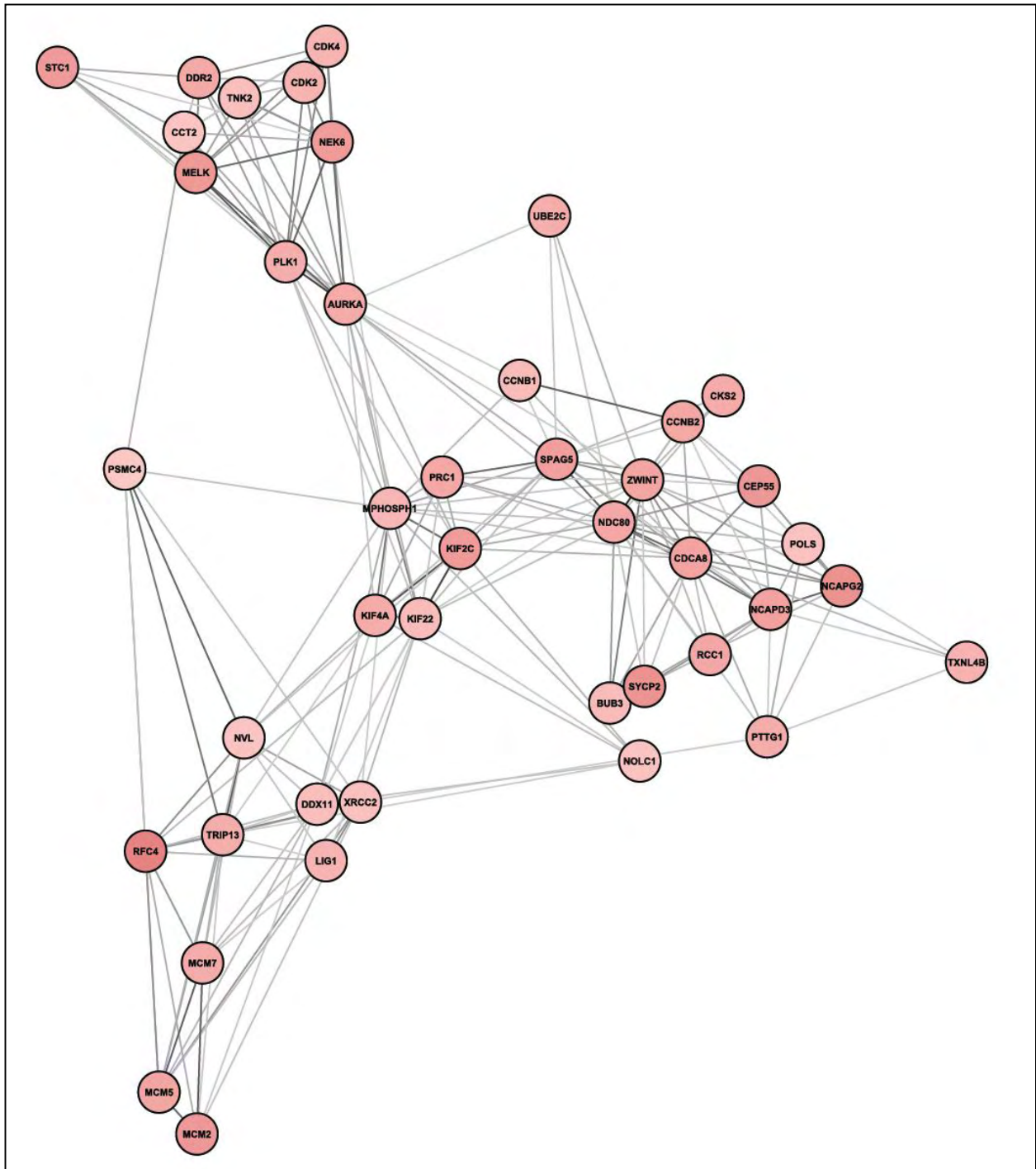


Figura 7. Sub-rede relacionada com o ciclo celular, replicação do DNA e via de sinalização da p53. A intensidade da cor dos círculos é proporcional à razão da expressão entre amostras de tecido tumoral e normal e a intensidade da cor das linhas é proporcional ao valor de *kappa* entre o par de genes.

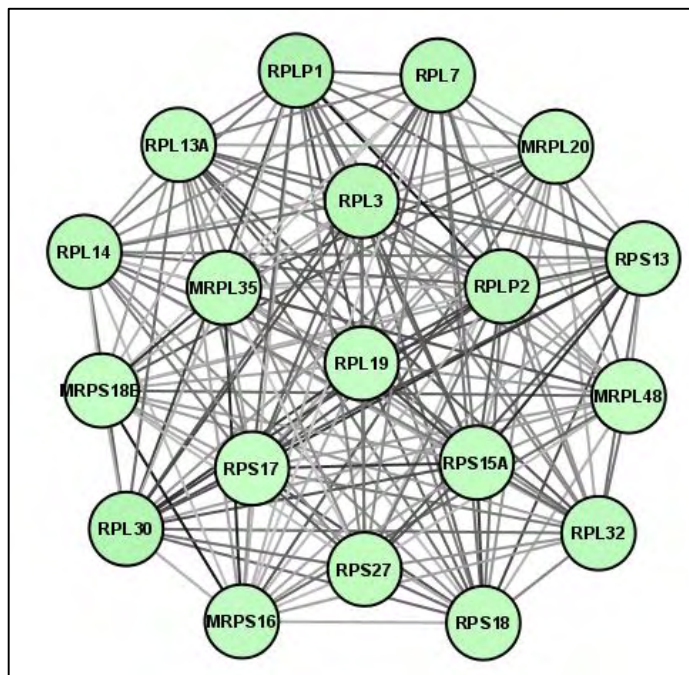


Figura 8. Sub-rede relacionada com o Ribossomo. A intensidade da cor dos círculos é proporcional à razão da expressão entre amostras de tecido normal e tumoral e a intensidade da cor das linhas é proporcional ao valor de *kappa* entre o par de genes.

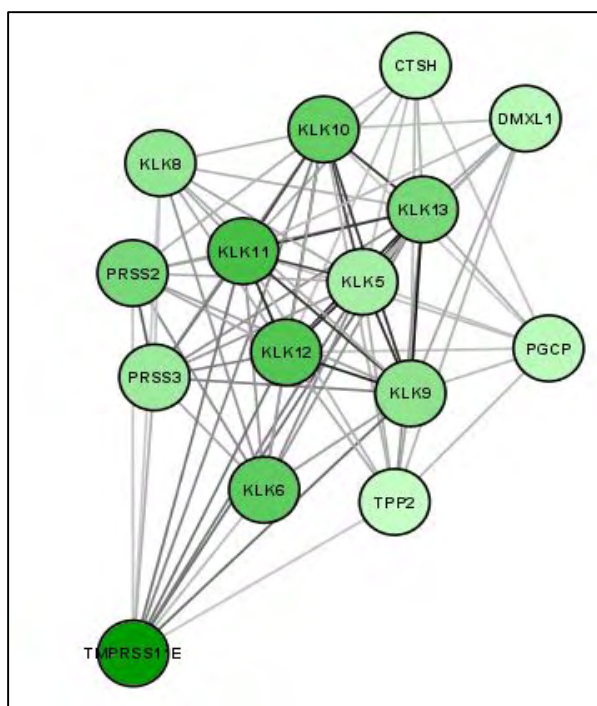


Figura 9. Sub-rede relacionada com os genes da família das caliceínas (KLKs, *Kallikrein-related peptidase*). A intensidade da cor dos círculos é proporcional à razão da expressão entre amostras de tecido normal e tumoral e a intensidade da cor das linhas é proporcional ao valor de *kappa* entre o par de genes.

4.4. Comparação dos genes diferencialmente expressos entre tumor e normal com dados de outros trabalhos publicados

Para verificar como o perfil de expressão gênica identificado em nosso trabalho se relacionava com o perfil gerado em outros trabalhos da literatura com os mesmos genes, realizamos duas análises comparativas: (i) análise dos genes identificados em um trabalho sobre câncer de colo do útero publicado por outros autores; (ii) análise dos genes identificados em vários outros trabalhos sobre câncer.

i) Análise dos genes identificados em um trabalho sobre câncer de colo do útero publicado por outros autores

Analizamos nossos genes em outro trabalho publicado recentemente de câncer de colo do útero (Scotto et al., 2008) para verificar se o comportamento desses genes nesse trabalho equivale ao que encontramos nas nossas análises. Utilizamos esse trabalho pois além dos dados estarem disponíveis, ele possui amostras de tecido normal e tumoral, sendo estes com o mesmo estadió e tipo histológico que as nossas amostras. Realizamos uma análise de comparação entre as classes tumor e normal com a nossa lista de genes no trabalho. Dos 810 genes, 725 tinham dados de expressão no estudo de Scotto e colaboradores. Desses 725 genes, 502 (69,2%) foram também significantes nesse trabalho ($p < 0,05$). Analizamos esses genes para verificar se havia concordância entre os estudos. Dentre os 502 genes, 459 mostraram expressão aumentada ou diminuída na mesma classe nos dois estudos (genes concordantes) e 43 genes apresentaram expressão diferentes nos estudos (genes discordantes) (Figura 10 e Tabela 2). A análise estatística mostrou concordância altamente significativa ($p < 0,0001$) entre os dois estudos.

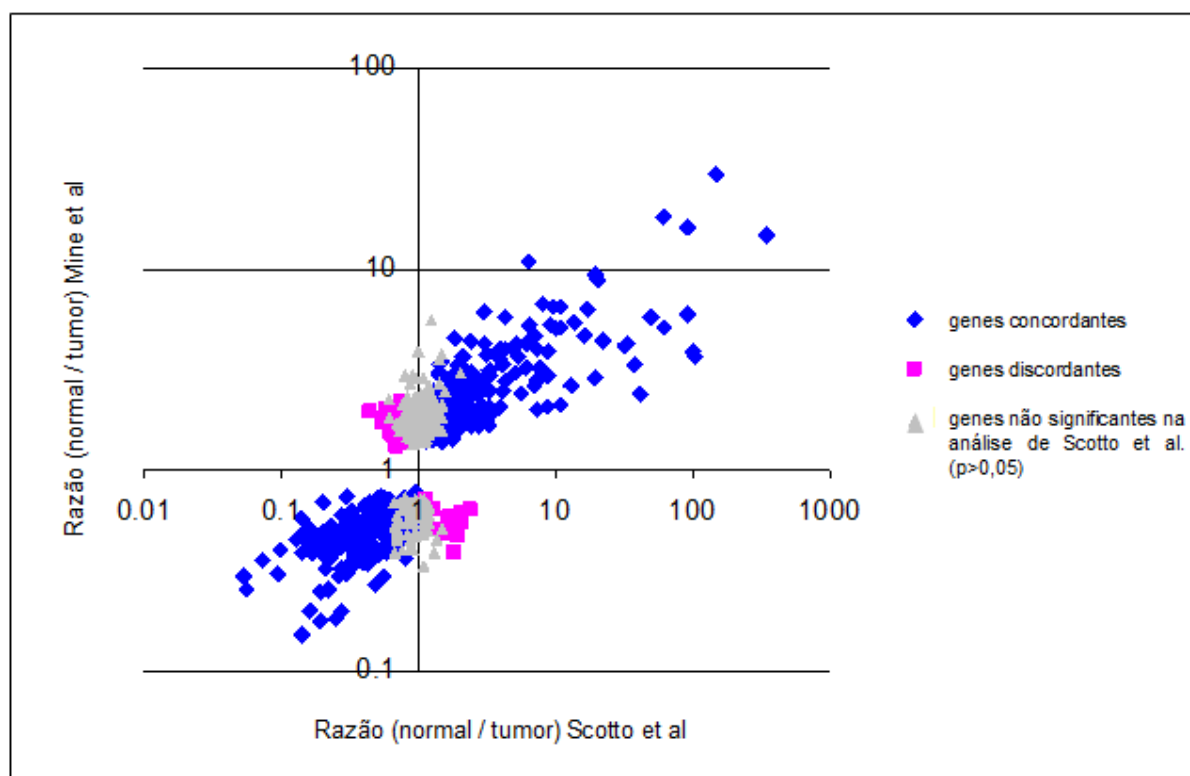


Figura 10. Distribuição da razão da expressão das amostras de tecido normal e tumoral dos 725 genes da nossa lista encontrados no estudo de Scotto et al. Os losangos azuis representam os genes significantes ($p < 0,05$) no estudo de Scotto et al. e com a expressão concordante (expressão aumentada ou diminuída na mesma classe) com a do nosso estudo (459 genes). Os quadrados rosa representam os genes significantes mas com a expressão discordante da nossa (43 genes) e os triângulos cinza indicam os genes não significantes no trabalho de Scotto et al. (223 genes).

Tabela 2. Concordância da alteração de expressão de 502 genes, entre o nosso trabalho e o de Scotto et al (2008) .

Mine et al.	Scotto et al.		Total
	Aumentada	Diminuída	
Aumentada	229	13	242
Diminuída	30	230	260
Total	259	243	502

Teste exato de Fisher: $p < 0,001$

(ii) Análise dos genes identificados em outros trabalhos de câncer

Para essa análise, como descrito nos métodos (item 3.5.6), utilizamos 11 estudos em câncer (Quadro 3). Para cada trabalho obtivemos os valores da razão normal/tumor para os 810 genes da nossa lista. Em cada estudo o número de genes encontrados dos 810 diferiram devido as diferentes plataformas de *microarray* utilizadas e o número de genes que passaram pelos filtros iniciais (Quadro 16).

Em seguida, realizamos um agrupamento hierárquico dos estudos utilizando a razão normal/tumor dos genes (Figura 11). O estudo de Wong et al. (colo uterino) não agrupou com nenhum estudo, o que pode ter ocorrido devido ao número de genes que foi reconhecido nesse trabalho (448 dos 810 genes). Em um segundo nível de agrupamento verificamos que com exceção do estudo de Haslinger et al (leucemia) todos os outros estudos se agruparam. Dentro deste agrupamento, formaram-se outros onde observamos que os estudos de câncer de estômago (Hippo et al.) de pulmão (Wachi et al.) e mesotelioma pleural (Gordon et al.) formaram um agrupamento indicando uma proximidade entre eles e, os estudos de câncer de colo uterino (Biewenga et al., Pyeon et al., Mine et al., Scotto et al.) e de câncer de esôfago (Kimchi et al.) formaram um outro agrupamento também indicando uma proximidade entre os estudos.

Quadro 16. Número de genes dos nossos 810 diferencialmente expressos entre tumor e normal encontrados em cada um dos 11 trabalhos da literatura.

Localização ou tipo do tumor	Referência do estudo	Número e porcentagem de genes dos nossos 810 encontrados no estudo
Colo do útero	Wong et al., 2003	448 (55,3%)
Colo do útero	Pyeon et al., 2007	774 (95,6%)
Colo do útero	Biewenga et al., 2008	759 (93,7%)
Colo do útero	Scotto et al., 2008	725 (89,5%)
Estômago	Hippo et al., 2002	345 (42,6%)
Leucemia	Haslinger et al., 2004	554 (68,4%)
Mesotelioma pleural	Gordon et al., 2005	682 (84,2%)
Pâncreas	Ishikawa et al., 2005	707 (87,2%)
Esôfago	Kimchi et al., 2005	682 (84,2%)
Melanoma	Talantov et al., 2005	682 (84,2%)
Pulmão	Wachi et al., 2005	682 (84,2%)

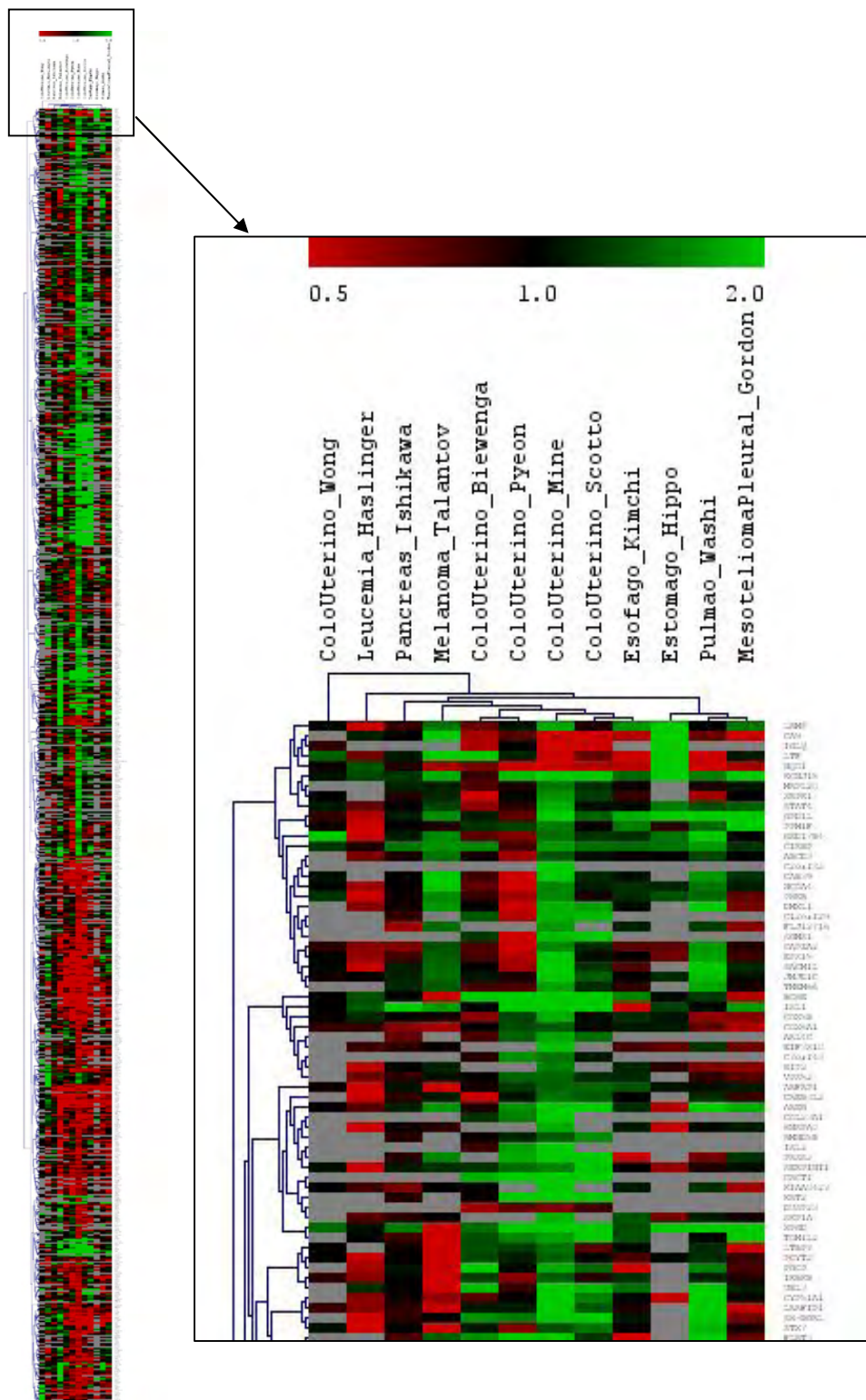


Figura 11. Agrupamento hierárquico dos 11 trabalhos da literatura e o nosso (ColoUterino_Mine) de acordo com a razão da expressão das amostras de normal e tumor dos 810 genes identificados no nosso estudo. A cor vermelha indica expressão no grupo tumor (razão menor que 1) e a cor verde no normal (razão maior que 1). A cor cinza indica que o gene não foi identificado no estudo em questão.

5 DISCUSSÃO

5.1. Busca de um perfil de expressão gênica capaz de prever a resposta ao tratamento

Em muitos países em desenvolvimento o câncer de colo do útero ainda causa um grande número de mortes anualmente. Em geral, os programas de rastreamento não são efetivos e a maioria dos casos são detectados em estádios mais avançados (Waggoner, 2003; Sankaranarayanan, Ferlay 2006). Durante muito tempo a radioterapia (RT) foi o tratamento padrão para os casos localmente avançados. Em 1999, uma série de estudos clínicos randomizados relataram uma melhora da sobrevida com a adição de quimioterapia (QT) concomitante a RT e agora esse esquema é considerado o tratamento padrão (Rose, 2002). No entanto, apesar da melhora na sobrevida, ainda cerca de 35% das pacientes não respondem ao tratamento (Santin et al., 2005; Gaffney et al., 2007). Um teste que pudesse identificar o grupo das pacientes que não respondem ao tratamento dentre as que respondem poderia ser utilizado para adequar o tratamento em cada caso, por exemplo dando uma terapia mais agressiva para o primeiro grupo e menos agressiva para o segundo. Assim, para buscar um perfil molecular que pudesse identificar tumores que se comportam de maneira clinicamente distinta após o tratamento, realizamos a análise simultânea da expressão de milhares de genes com a metodologia do *microarray*.

Nas análises utilizamos 23 lâminas hibridizadas com tecido tumoral de colo do útero. Dentre elas, 13 correspondiam a pacientes R e 10, a NR. Na análise para identificação de um preditor molecular, verificamos que apenas 10 (43,5%) das 23 amostras analisadas foram corretamente classificadas. Isso nos levou a supor que a correção que realizamos nos dados, devido ao uso de diferentes tipos de lâminas, poderia ter retirado o sinal de genes importantes relacionados com a resposta ao tratamento. Para tentar verificar isso, realizamos a mesma análise com apenas 17 lâminas do mesmo tipo. Embora a porcentagem de amostras classificadas corretamente tenha aumentado (64,7%) esse resultado não foi satisfatório sendo que apenas uma das seis amostras do grupo de NR foi corretamente classificada.

A busca por uma assinatura capaz de prever a resposta após o tratamento em câncer do colo do útero também já foi realizada por outros autores. Dentre esses, quatro estudos identificaram um grupo de genes associados com a evolução pós-

tratamento analisando amostras coletadas antes do tratamento (Kitahara et al., 2002; Wong et al., 2003; Wong et al., 2006b; Klopp et al., 2008). No entanto, a sobreposição entre os genes encontrados nos diferentes estudos, quando houve, foi baixa (Klopp, Eifel, 2006). Para que os genes encontrados nesses trabalhos pudessem ser considerados para utilização na prática clínica, seria necessária, em primeiro lugar, a validação dos resultados em um grupo grande e independente. No entanto, até hoje não há relatos de validação para quaisquer dos grupos de genes identificados.

Em três dos trabalhos acima citados (Kitahara et al., 2002; Wong et al., 2003; Wong et al., 2006b), o tratamento das pacientes era apenas RT, e no trabalho mais recente (Klopp et al., 2008) o tratamento foi RT+QT. No nosso estudo utilizamos tanto amostras de pacientes tratadas apenas com RT quanto com RT+QT. Apesar de não termos verificado uma associação significativa entre o tipo de tratamento e a evolução pós-tratamento na nossa casuística, pensamos que talvez a inclusão de pacientes submetidas a diferentes esquemas terapêuticos possa ter prejudicado a identificação de um preditor. No entanto, recentemente um estudo, analisando amostras pré-tratamento de 22 pacientes submetidas a um mesmo esquema terapêutico, também não encontrou uma assinatura gênica que pudesse prever a evolução pós-tratamento (Weidhaas et al., 2009). O fato de não identificarmos um perfil de expressão gênica para prever a resposta ao tratamento utilizando amostras pré-tratamento concorda com este trabalho. Além disso, é interessante notar que no trabalho de Weidhaas et al. foi identificada uma assinatura de sete genes capaz de prever a resposta ao tratamento, quando foi comparada a expressão entre amostras pré-tratamento e durante tratamento. Este resultado, baseado na mudança de expressão gênica entre amostras coletadas antes e durante o tratamento, embora ainda não validado e realizado com 13 pares de amostras, traz novas perspectivas na elaboração de estudos para busca de perfis moleculares preditores da evolução ou resposta após o tratamento.

5.2. Identificação de genes e grupos funcionais de genes com expressão alterada em amostras de tumor de colo do útero

O exame citopatológico utilizado nos programas de rastreamento para prevenção do câncer de colo de útero tem reduzido a incidência desse câncer onde os programas são efetivos. No entanto, devido à alta taxa de falsos positivos e negativos novos métodos estão sendo abordados para melhorar o rastreamento, tais como o teste de detecção de DNA do HPV e marcadores moleculares (Baldwin et al., 2003; Nijhuis et al., 2006). O advento de tecnologias de análise em larga escala trouxe novas perspectivas nos estudos para a descoberta de marcadores e melhor entendimento dos mecanismos envolvidos na carcinogênese, o que poderá ajudar no rastreamento, detecção precoce e tratamento (Martin et al., 2007). Neste trabalho, utilizando a análise de *microarray* buscamos identificar um perfil de genes com expressão alterada no tecido tumoral em relação ao normal. Em seguida, utilizamos a lista de genes encontrada para realizar análises de vias e redes gênicas para entender possíveis mecanismos moleculares envolvidos no câncer do colo do útero.

Foram estudadas 34 lâminas de *microarray* com tecido tumoral de colo do útero e 20, com tecido normal. Embora 13 amostras de tecidos tumoral e normal fossem da mesma paciente, não realizamos uma análise pareada pois dentre as amostras analisadas muitas não eram pareadas e, além disso, algumas amostras foram agrupadas em pares e hibridizadas na mesma lâmina, diminuindo as amostras pareadas individuais. Para assegurar a identificação de uma lista de genes confiável, inicialmente dividimos as amostras em dois grupos e selecionamos somente os genes com expressão diferencial em ambos os grupos. Além disso, realizamos uma análise com os genes selecionados apenas nas lâminas que foram hibridizadas com amostras de tecido normal cuja referência era um *pool* de amostras de tumor. Optamos por fazer uma análise mais conservadora para diminuir o número de resultados falso-positivos. Após essas análises selecionamos 810 genes diferencialmente expressos entre tumor e normal, sendo 469 com baixa e 341 com alta expressão no tumor.

Em seguida, realizamos a análise de grupos funcionais de genes com as listas obtidas.

Inicialmente, realizamos análises de vias nas listas de genes obtidas e identificamos 13 vias representadas. Destas, destacamos nove que possuíam um número razoável de genes na via: *Cell Cycle*, *p53 Signaling*, *Protein Ubiquitination Pathway*, *Antigen*

Processing and Presentation, Cell Adhesion Molecules, Fatty Acid Metabolism, MAPKinase Signaling Pathway, Oxidative Phosphorylation e Ribosome.

Cell Cycle, p53 Signaling e Protein Ubiquitination Pathway. Junto com a via do ciclo celular e da p53 outra via já com conhecida relação com o câncer de colo uterino é a via de ubiquitinação de proteínas. Essas três vias podem ter uma participação conjunta no processo de carcinogênese. As oncoproteínas E6 e E7 dos HPVs de alto risco ligam-se, respectivamente, às proteínas supressoras de tumor p53 e retinoblastoma (Rb), que estão envolvidas na regulação do ciclo celular (Kim, Zhao, 2005). A via mediada pela ubiquitina participa da degradação de proteínas como a p53 e outras regulatórias do ciclo celular (Mani, Gelmann, 2005). O DNA do HPV de alto risco é encontrado em quase todos os casos de câncer de colo do útero, sendo os tipos 16 e 18 os mais prevalentes (Doorbar, 2006; Boulet et al., 2007; Castellsagué, 2008).

As proteínas oncogênicas do HPV, E6 e E7, são essenciais para induzir e manter a transformação celular devido à interferência com o controle do ciclo celular e apoptose (Narisawa-Saito, Kiyono, 2007). A p53 é um supressor de tumor que regula a interrupção do crescimento e apoptose após dano do DNA. Assim, a inativação de p53 mediada pela E6 ou evita a interrupção do ciclo celular, devido ao dano no DNA, levando ao acúmulo de mutações, ou inibe a apoptose (Narisawa-Saito, Kiyono, 2007). A proteína E7 liga-se ao produto do gene supressor de tumor retinoblastoma, pRb, e membros de sua família. No estado hipofosforilado, proteínas da família do pRb podem ligar-se a fatores de transcrição, tais como membros da família E2F, e reprimir a transcrição de genes envolvidos na síntese de DNA e progressão do ciclo celular. Como E7 é capaz de ligar-se ao pRb não fosforilado, ele pode prematuramente induzir as células a entrarem na fase S, por romper o complexo pRb-E2F (Narisawa-Saito, Kiyono, 2007). Adicionalmente, a degradação de pRb, mediada por ubiquitina, induzida por E7, parece ser essencial para evitar a interrupção do ciclo celular (Boulet et al., 2007). No nosso estudo, as vias do ciclo celular e da ubiquitinação de proteínas foram obtidas a partir da lista de genes com alta expressão no tumor. Isso está de acordo com o descrito anteriormente, no sentido que a contínua progressão do ciclo celular e degradação de certas proteínas mediada pela ubiquitina parecem ter papéis importantes na carcinogênese do colo uterino.

A via de sinalização da p53 também estava representada na lista de genes com alta expressão no tumor. Esta via é composta por vários genes sendo que muitos deles são compartilhados pelas vias interligadas a ela, como a do ciclo celular. Dos sete genes

pertencentes a nossa lista que fazem parte da via de sinalização da p53, quatro são componentes do ciclo celular, por isso são genes com alta expressão no tumor. Os outros genes participam de diferentes funções nessa via como inibição de apoptose, vias de sinalização e respiração mitocondrial (informações obtidas no site do Entrez Gene, NCBI, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>). Apesar de alguns genes da nossa lista presentes nessas vias, como CDKN2A (p16INK4a), BIRC5 (ou *survivin*), MCM2 e MCM5, serem relatados em vários trabalhos como marcadores promissores em câncer de colo do útero (revistos em Martin et al., 2007), verificamos que muitos genes das vias do ciclo celular e da via de ubiquitinação de proteínas ainda não foram estudados nesse câncer (Quadro 14).

Antigen Processing and Presentation e Cell Adhesion Molecules. Como apenas uma pequena porcentagem de mulheres infectadas por HPV de alto risco desenvolvem o câncer outros fatores, como a resposta imune, podem estar envolvidos na progressão da neoplasia. Há crescentes evidências que o sistema imune tem um papel central em determinar o resultado da infecção pelo HPV (Sheu et al., 2007). Assim como o HPV, as células carcinogênicas transformadas precisam evadir a vigilância imune para sobreviver. Mas, ainda que os vírus e as células neoplásicas tenham desenvolvido mecanismos para evitar o ataque imune (Tindle, 2002; Kanodia et al., 2007) é observado nos tumores de colo uterino infiltrado de linfócitos (Sheu et al., 2007; Piersma et al., 2008). Nos nossos resultados, os genes da via de apresentação e processamento de antígenos estavam com expressão aumentada no tumor. Na literatura, no entanto, os trabalhos indicam que há uma diminuição da expressão de componentes da maquinaria de processamento de antígenos no câncer de colo do útero (Ritz et al., 2001; Vermeulen et al., 2007; Mehta et al., 2008). Mas essa diminuição de expressão não ocorre em todos os casos analisados e a maioria dos trabalhos analisam proteínas. Ahn e colaboradores (2004) utilizando análise de *microarray* também identificaram aumento de expressão de genes relacionados com a resposta imune em amostras de câncer de colo uterino. Dos 12 genes da nossa lista pertencentes a via de processamento e apresentação de antígenos, sete ainda não foram estudados neste câncer (Quadro 14). Estudos adicionais poderiam ajudar a melhor entender os papel desses genes no câncer de colo uterino. Na via das moléculas de adesão dois genes foram estudados em relação ao câncer de colo do útero, onde sugere-se também uma aumentada expressão no tumor (D'Anna et al., 2001; Lee et al., 2005).

Fatty Acid Metabolism. Alguns estudos, de cerca de 50 anos atrás, mostraram que a biossíntese de ácidos graxos endógenos em tecidos tumorais ocorrem em taxas muito altas (Menendez, Lupu, 2006). Há duas fontes de ácidos graxos para o metabolismo animal, os exógenos derivados da dieta e os sintetizados *de novo*, endógenamente. No entanto, foi estabelecido que a síntese *de novo* é responsável por quase todos os ácidos graxos na célula tumoral, mesmo com um suprimento nutricional adequado (Menendez, Lupu, 2006). Essa biossíntese é catalisada pela proteína lipogênica *Fatty Acid Synthase* (FAS). FAS é altamente expressa em muitos carcinomas humanos e está associada com pior prognóstico em câncer de mama e próstata (Kuhajda, 2006). Três artigos relacionam essa via com o câncer de colo uterino (Louw et al, 1998a; Engelbrecht et al, 1998; Louw et al, 1998b). Estes estudos, realizados pelo mesmo grupo de pesquisadores, comparam a composição de ácidos graxos de tecido normal, de lesão intra-epitelial e de carcinoma de colo uterino. Eles observaram uma deficiência de alguns ácidos graxos essenciais e um alto nível de ácido oléico no câncer. A deficiência de ácidos graxos essenciais pode resultar em muitos mecanismos celulares defeituosos pois eles são associados com eventos como a peroxidação de lipídios, transdução de sinal e resposta imune. O alto nível de ácido oléico em células tumorais pode ativar a *Protein Kinase C* (PKC) e contribuir para um contínuo estímulo do crescimento das células malignas (Engelbrecht et al, 1998). A obtenção dessa via na análise da lista de genes com baixa expressão no tumor poderia estar relacionada com uma menor degradação e o conseqüente acúmulo dos ácidos graxos. Pois, dos oito genes da nossa lista que fazem parte dessa via, cinco participam de processos, principalmente da beta oxidação, que levam à degradação dos ácidos graxos (informações obtidas no site do Entrez Gene, NCBI, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene>). Assim, a baixa expressão desses genes poderia contribuir para o acúmulo de ácidos graxos nas células tumorais.

MAPKinase Signaling Pathway. A via da *Mitogen-activated protein kinase* (MAPK) regula diversos processos abrangendo desde proliferação e diferenciação até apoptose. Cinco famílias de MAPKs têm sido definidas em células de mamíferos: *extracellular signal-regulated kinases* (ERK1 and ERK2), *Jun N-terminal kinases* (JNK1, JNK2 and JNK3), *p38 kinase isozymes* (p38 α , p38 β , p38 γ and p38 δ), ERK3/ERK4 e ERK5 (Qi, 2005). Diversos estudos sugerem um envolvimento de ERK1/2 na proteção contra a morte celular pela regulação transcripcional e/ou inibição direta da maquinaria da apoptose (Bonni et al, 1999; Hetman, Gozdz, 2004). Em contraste, JNK e p38 são

estimulados por condições de estresse e associados com apoptose (Ho et al, 2006; Kim et al, 2006). Alguns estudos sugerem que ERK e p38 regulam opostamente a apoptose e que o balanço entre essas vias pode determinar o destino da célula (Xia et al, 1995; Kim et al, 2002). Estudos usando células de câncer cervical mostraram que p38 e JNK são ativados em resposta a alguns tratamentos e estão envolvidos na morte celular por apoptose (Boldt et al, 2002; Brozovic et al, 2004; Kang, Lee, 2008a; Kang, Lee, 2008b). Embora ERK, em associação com Ras, tenha sido implicado na progressão do tumor, ERK, em associação com JNK, foi referido como indutor de apoptose (Lin, 2003; Ruffels et al, 2004). Nos nossos resultados, essa via foi obtida na lista de genes com baixa expressão no tumor. Na literatura há resultados contraditórios em relação a essa via, todos com estudo de proteínas. Matsuura e colaboradores (2006) observaram que em carcinomas cervicais invasivos quase todas as células eram positivas para Ki-67 (um marcador de proliferação celular) mas negativas para p-ERK, sugerindo que, ao contrário de outros tipos de tumores, a via de sinalização de ERK estaria diminuída em carcinoma cervical invasivo. Em acordo com esse resultado, Engelbrecht e colaboradores (2006) observaram que o estado de fosforilação de ERK estava significativamente diminuído em carcinoma de colo do útero. Eles também encontraram uma fosforilação diminuída do p38-MAPK e JNK no grupo de carcinoma comparado com tecido normal. Em oposição a esses trabalhos, outros observaram aumento ou de ERK2, phospho-MAPK/ERK1/2 ou ERK1 em carcinomas invasivos (Branca et al, 2004; Chen et al, 2007; Kapetanios et al, 2008). Os resultados contraditórios desses trabalhos indicam a necessidade de mais estudos, talvez envolvendo análise da expressão gênica juntamente com análise protéica, para tentar melhor esclarecer a relação dessa via com o câncer de colo do útero.

Oxidative Phosphorylation. As células podem obter energia pela via dependente de oxigênio da fosforilação oxidativa (*Oxidative Phosphorylation* – OXPHOS) e da via independente de oxigênio da glicólise. Como a OXPHOS é mais eficiente em gerar ATP do que a glicólise, é reconhecido que a presença de oxigênio resulta na ativação da OXPHOS e inibição da glicólise (“Efeito Pasteur”) (López-Lázaro, 2008). Contudo, Otto Warburg verificou que células tumorais utilizam a glicólise ao invés da OXPHOS, mesmo na presença de oxigênio (“Efeito Warburg”) (Warburg, 1956). Apesar de controversa por anos, a base molecular para o Efeito Warburg está surgindo de estudos genéticos e farmacológicos que demonstram que determinadas mutações em oncogenes e genes supressores de tumor regulam diretamente a glicólise e OXPHOS

(Shaw, 2006). Além disso, evidências crescentes sugerem que a ativação persistente da glicólise em células tumorais tem um papel crucial no desenvolvimento do câncer. A inibição da aumentada capacidade glicolítica das células malignas pode representar uma estratégia anti-tumoral chave (López-Lázaro, 2008). Na nossa análise, a via da OXPHOS estava representada na lista de genes com baixa expressão no tumor, o que poderia estar relacionado com o Efeito Warburg. Realizamos uma busca de trabalhos que relacionam *Oxidative Phosphorylation* e câncer de colo do útero e dos seis artigos encontrados, nenhum estuda essa via diretamente em amostras desse câncer. Além disso verificamos, na análise realizada com a ferramenta MILANO, que dos nove genes da nossa lista que fazem parte da via ainda nenhum foi estudado em câncer de colo uterino (Quadro 14).

Ribosome. A biogênese do ribossomo e a síntese protéica são processos críticos ligados ao crescimento celular e à proliferação. Estima-se que células em proliferação gastem mais que 80% de sua energia celular na síntese, transporte, processamento e montagem das subunidades ribossomais (Opferman, Zambetti, 2006). Tem sido mostrado que diversos proto-oncogenes e supressores de tumor regulam diretamente a produção de ribossomo, a iniciação da tradução de proteína, ou ambos. Esses achados levantam a possibilidade que alterações em componentes da maquinaria de síntese protéica e regulação aberrante da tradução possam promover a transformação maligna (Ruggero, Pandolfi, 2003). Alterações na expressão de proteínas ribossomais parecem estar consistentemente associadas com tumorigênese. Muitas dessas proteínas que pertencem às subunidades ribossomais possuem uma alta expressão em linhagens de células e tecidos tumorais (Ferrari et al, 1990; Loging, Reisman, 1999; Kondoh et al, 2001). Ao contrário, entretanto, os genes da nossa lista estão com a expressão diminuída no tumor. No entanto, resultados de estudos em câncer colo-retal tem revelado que a regulação das proteínas ribossomais no carcinoma diferem da encontrada em adenoma ou na mucosa normal, com algumas proteínas ribossomais sendo reguladas positivamente enquanto outras, negativamente (Lai, Xu, 2007). Isso poderia ser explicado pelo fato das proteínas ribossomais, além do seu papel na síntese protéica, possuírem outras funções extra-ribossomais que poderiam estar envolvidas na tumorigênese. Todos os 14 genes da nossa lista que fazem parte dessa via são genes que codificam proteínas ribossomais. Embora não exista trabalhos específicos publicados sobre esses genes em relação ao câncer de colo do útero (Quadro 14), verificamos a presença desses genes alterados em um

estudo, semelhante ao nosso, que utilizou *microarray* para analisar amostras de carcinoma de células escamosas do colo do útero e amostras de colo normal (Pérez-Plasencia et al., 2007). Ao consultar a lista de genes identificados como significativamente aumentados ou diminuídos no tumor, disponível como material suplementar, identificamos 14 genes do ribossomo alterados. Desses, quatro genes eram comuns aos nossos e, um deles, ao contrário, tinha expressão aumentada no tumor. No entanto, dos 14 genes da lista de Pérez-Plasencia e colaboradores, 11 tinham a expressão diminuídas no tumor, o que está em acordo com o perfil dos nossos genes do ribossomo.

Em resumo, verificamos que: (i) dentre as vias identificadas algumas já possuem alguma participação descrita em câncer de colo do útero, enquanto que outras mesmo com relação descrita em outros tipos de câncer, ainda não tinham sido descritas em câncer de colo do útero; (ii) as vias que foram descritas em outros tipos de câncer, mas não em colo uterino, como a da fosforilação oxidativa, assim como vias cuja participação é controversa nesse câncer, como a da MAPK, podem ser vias interessantes para estudos futuros; (iii) mesmo as vias já estudadas, como a do ciclo celular, e da ubiquitinação de proteínas, possuem genes presente na nossa lista, ainda não estudados em câncer de colo uterino; (iv) a via do ribossomo pode ser uma candidata interessante para outros estudos visto que foi pouco estudada em câncer de colo do útero e há 14 genes alterados na nossa lista.

Após essa análise de vias, realizamos a análise de redes gênicas. Como na análise de redes utilizamos 14 categorias de anotação diferentes para gerar a ligação entre os genes, esperávamos encontrar algumas redes com anotação diferente das vias já identificadas. Identificamos 23 sub-redes formadas entre os genes das duas listas. Após a anotação funcional dessas sub-redes verificamos que sete delas também tinham sido identificadas na análise de vias: *Cell cycle*, *p53 signaling pathway*, *Protein Ubiquitination Pathway*, *Cell Adhesion Molecules*, *MAPK signaling Pathway*, *Ribosome*, *Oxidative Phosphorylation*. Entre as sub-redes encontradas em cada lista, alta e baixa expressão no tumor, identificamos várias com a mesma anotação nas duas listas, *Transcription*, *RNA related mechanisms*, *Ubiquitin-dependent protein catabolic process*, *Membrane* e *Phosphatase activity*. Isso não foi inesperado pois muitos processos envolvem a presença de genes com baixa e alta expressão, indicando desregulação geral desses processos.

Dentre as sub-redes identificadas na lista de genes com baixa expressão no tumor, destacamos a sub-rede 5 (Figura 9), que parece ser uma rede de genes importante relacionada ao câncer de colo uterino. Essa sub-rede foi anotada como “*Proteolysis*” de acordo com o GO. Ao verificarmos quais genes fazem parte dessa rede notamos que, dos 15 genes, oito são genes da família das *Kallikrein-related peptidases* (KLKs). As KLKs têm sido implicadas em diversos processos relacionados ao câncer e possuem algumas aplicabilidades clínicas como biomarcador, sendo estudadas em vários tipos de câncer, como de próstata, ovário e mama (Zhang et al., 2006; Paliouras et al., 2007; Shan et al., 2007; Emami, Diamandis, 2008). O *Prostate-specific antigen* (PSA ou KLK3), a KLK mais conhecida é um marcador útil para câncer de próstata. As proteínas codificadas pelos genes KLKs são serino-proteases com diversos padrões de expressão e papéis fisiológicos. Evidências recentes relacionam essas proteínas com muitos processos relacionados ao câncer, como regulação do crescimento celular, angiogênese, invasão e metástase (Borgoño, Diamandis, 2004). O padrão de expressão das KLKs, expressão aumentada ou diminuída, varia dependendo da KLK e do tipo de câncer (Paliouras et al., 2007). Três KLKs já foram estudadas em câncer de colo do útero, KLK 7, 8 e 10 (Cane et al., 2004; Santin et al., 2004; Tian et al., 2004; Zhang et al., 2006). A KLK7 mostrou expressão aumentada tanto em adenocarcinoma como em carcinoma de células escamosas (Santin et al., 2004; Tian et al., 2004). Em nossos dados, apesar da KLK7 estar presente na nossa lâmina, o gene foi eliminado nos filtros iniciais antes das análises. Há apenas um estudo com a KLK8 que mostra uma alta expressão no tumor (Cane et al., 2004), ao contrário do nosso resultado. Para a KLK10 também há apenas um estudo que mostra uma diminuição da expressão no tumor (Zhang et al., 2006), como observamos nos nossos resultados. Apesar de não haver outros trabalhos específicos com as KLKs em câncer de colo de útero, observamos na lista de genes publicada no artigo de Wong e colaboradores (2006a) a presença das KLKs. Nesse trabalho os autores realizaram um estudo parecido com o nosso, analisando amostras de carcinoma de células escamosas do colo uterino e de colo normal utilizando *microarray*. Entre os genes com baixa expressão no tumor em relação ao normal estavam presentes cinco genes das KLKs, sendo quatro iguais ao nosso, inclusive o da KLK8. Esses resultados reforçam nossos achados e suportam nossa hipótese de um importante papel desses genes no câncer de colo do útero.

Além do estudo de vias e redes gênicas, realizamos análises para verificar como o nosso perfil de expressão gênica se relacionava com o perfil gerado em outros trabalhos. Para isso, realizamos duas análises: a primeira comparando nossos resultados com os de outro trabalho de câncer de colo uterino, cujas amostras possuíam características semelhantes às nossas (Scotto et al., 2008), e, a segunda, comparando com os dados de trabalhos realizados com outros tumores e com casos de câncer de colo do útero em estádios diferentes dos da nossa amostra. Ao comparar nossos dados com os de Scotto e colaboradores, em relação à nossa lista de genes, verificamos uma concordância altamente significativa entre os dois estudos. Isso indica que o perfil de genes que identificamos não é característico apenas do nosso estudo, mas que é um perfil que pode ser generalizado para outras pacientes de câncer de colo uterino. Além disso, é interessante relatar que as amostras utilizadas por Scotto e colaboradores foram microdissecadas para enriquecê-las com células tumorais ou células epiteliais, no caso das amostras de tecido normal. Assim, a concordância observada sugere, como esperado, que nossas amostras, mesmo não sendo microdissecadas, também contem grande número de células tumorais. Na segunda análise, em que comparamos nossos dados com outros tumores e outros trabalhos de câncer de colo uterino, verificamos que o padrão de expressão dos genes da nossa lista também é parecido com o de outros trabalhos de câncer de colo uterino mesmo com estádios diferentes (Figura 11). Apenas um trabalho de colo uterino não agrupou com os demais (Wong et al., 2003), talvez devido ao fato deste trabalho ter tido dados para apenas 448 (55,3%) genes dentre os nossos 810. Interessantemente, verificamos que dentro do agrupamento dos estudos de câncer de colo uterino também encontra-se um trabalho de câncer do esôfago (Kimchi et al., 2005). Embora o tipo mais frequente de câncer de esôfago também seja o carcinoma epidermóide escamoso (INCA, 2009), no trabalho acima mencionado foi estudado o adenocarcinoma. No câncer de esôfago o adenocarcinoma geralmente está relacionado como uma condição denominada esôfago de Barrett, que é uma metaplasia que ocorre nas células da porção inferior do esôfago, onde o epitélio normal escamoso é substituído por um epitélio colunar (Kimchi et al., 2005). Greenawalt e colaboradores (2007) realizaram um estudo de expressão gênica utilizando *microarray* em 128 amostras representando os principais tipos histológicos do câncer de esôfago, ou seja, carcinoma de células escamosas (SCC) e adenocarcinoma (ADC), bem como a lesão precursora do adenocarcinoma, esôfago de Barrett (BE), e o epitélio normal de esôfago. Neste estudo eles identificaram uma

grande sobreposição entre os genes nas comparações SCC vs normal, ADC vs normal, BE vs normal. Além disso, observaram que o SCC e ADC têm uma regulação coordenada de genes compartilhando anotações funcionais específicas. SCC e ADC têm um aumento de expressão de genes envolvidos na carcinogênese, incluindo reguladores do ciclo celular, efetores da matriz extracelular e genes da resposta imune. Foi observado também uma diminuição de expressão de genes de inibidores de serino proteases, das KLKs, ligação de cálcio-íon e *gap junction* em comparação ao epitélio normal do esôfago. Assim, a semelhança no perfil de expressão entre os dois tipos histológicos de câncer de esôfago indicado nesse trabalho pode explicar o agrupamento encontrado no nosso resultado do adenocarcinoma de esôfago com o carcinoma de colo do útero.

Em resumo, neste trabalho realizamos análise de expressão gênica em larga escala para o estudo de eventos relacionados ao câncer de colo do útero. Em relação à evolução clínica após o tratamento, não identificamos nenhum perfil consistente capaz de prever a resposta ao tratamento, o que pode ter ocorrido devido a utilização de amostras pré-tratamento. Na busca por genes que poderiam estar relacionados com a patogênese do câncer de colo do útero identificamos uma lista de genes diferencialmente expressos entre tecido tumoral e normal. Na análise de vias e redes gênicas, destacamos alguns mecanismos que parecem ser importantes nesse câncer e que são candidatos promissores para estudos posteriores como, por exemplo, a via do ribossomo e genes da família das KLKs. Além disso, nossos resultados sugerem que o perfil de genes identificado não é característico apenas da nossa população, mas que também se aplica a casos de câncer de colo de útero estudados em outros trabalhos.

Os genes pertencentes às vias e redes identificadas neste trabalho são candidatos a serem explorados em estudos futuros. Utilizando análise de expressão, os genes poderiam ser estudados como marcadores a serem utilizados no diagnóstico precoce do câncer de colo do útero. Além disso, estes estudos poderiam avaliar se a variação na expressão dos genes que nós identificamos estaria relacionada a fatores genéticos, o que poderia indicar candidatos a marcadores de susceptibilidade, ou a fatores ambientais, relacionando novos mecanismos moleculares com já conhecidos fatores de risco de câncer de colo uterino. Adicionalmente, estudos envolvendo a relação entre expressão gênica e alterações cromossômicas (perdas e ganhos de

regiões cromossômicas), evento frequente no câncer de colo do útero (Scotto et al., 2008; Wilting et al., 2008), também seriam importantes para um melhor entendimento das alterações genéticas que ocorrem nesse câncer.

6 CONCLUSÕES

1. As alterações na expressão gênica do tumor de colo do útero avaliadas antes do tratamento e que são relacionadas com a resposta ao tratamento, ou são pequenas ou inexistentes. Assim, a análise do transcriptoma do tumor antes do tratamento não possibilita a geração de um perfil gênico capaz de prever a resposta ao tratamento.

2. Identificamos vias e redes gênicas, na análise de grupos funcionais de genes, com relação previamente conhecida em câncer de colo do útero bem como algumas ainda não descritas nesse câncer. Esses resultados podem contribuir para o entendimento de mecanismos moleculares envolvidos na patogênese do câncer de colo do útero.

3. Os genes pertencentes aos grupos funcionais identificados poderiam ser explorados como possíveis candidatos a marcadores a serem utilizados no diagnóstico precoce, assim como, se polimórficos, servir como marcadores genéticos de susceptibilidade ao câncer de colo de útero.

7 ANEXOS

Anexo 1. Carta de aprovação do Comitê de Ética em Pesquisa



Universidade Federal de São Paulo
Escola Paulista de Medicina

Comitê de Ética em Pesquisa
Hospital São Paulo

São Paulo, 16 de setembro de 2005.
CEP 0640/05

Ilmo(a). Sr(a).
Pesquisador(a) KARINA LUMI MINE
Co-Investigadores: Maria Gerbase de Lima
Disciplina/Departamento: Alergia, Imuno, Reumatologia/Pediatria da Universidade Federal de São Paulo/Hospital São Paulo
Patrocinador: Recursos Próprios.

PARECER DO COMITÊ DE ÉTICA INSTITUCIONAL

Ref: Projeto de pesquisa intitulado: "Análise genômica e transcriptômica em câncer de colo uterino".

CARACTERÍSTICA PRINCIPAL DO ESTUDO: Intervenção diagnóstica.

RISCOS ADICIONAIS PARA O PACIENTE: Sem risco, desconforto moderado, envolvendo coleta de sangue e biópsia de tecidos..

OBJETIVOS: Identificar genes que conferem susceptibilidade ao carcinoma de colo uterino. Elaborar um perfil de expressão gênica que prediga a resposta ao tratamento. Avaliar o papel de transcritos não codantes no carcinoma de colo uterino.

RESUMO: Serão estudados cerca de 100 pacientes portadoras de carcinoma espinocelular cervical em estadiamento II e III. Serão coletados fragmentos de tumor e de tecido normal adjacente por ocasião da realização de biópsia com finalidade diagnóstica, que serão congelados em nitrogênio líquido. Nesta ocasião serão colhidas também amostras de sangue. O grupo controle consistirá de 100 indivíduos sadios, dos quais será coletado sangue para extração de DNA.

FUNDAMENTOS E RACIONAL: Estudo fundamentado, analisando expressão gênica em sangue e biópsia de tecido de pacientes com câncer de colo uterino..

MATERIAL E MÉTODO: Estão descritas as técnicas empregadas, sendo utilizado material já estocado e também envolvendo nova coleta..

TCLE: Apresenta 3 TCLEs para coleta de sangue de voluntários sadios e pacientes com carcinoma cervical, e para coleta de biópsia. Estão de acordo com a Res 196/96.

DETALHAMENTO FINANCEIRO: Sem financiamento externo - R\$ 142 963,53.

CRONOGRAMA: 48 meses.

OBJETIVO ACADÊMICO: Doutorado.

ENTREGA DE RELATÓRIOS PARCIAIS AO CEP PREVISTOS PARA: **11/09/2006 e 06/09/2007.**

O Comitê de Ética em Pesquisa da Universidade Federal de São Paulo/Hospital São Paulo **ANALISOU e APROVOU** o projeto de pesquisa referenciado.

1. Comunicar toda e qualquer alteração do projeto e termo de consentimento livre e esclarecido. Nestas circunstâncias a inclusão de pacientes deve ser temporariamente interrompida até a resposta do Comitê, após análise das mudanças propostas.

2. Comunicar imediatamente ao Comitê qualquer evento adverso ocorrido durante o desenvolvimento do estudo.

3. Os dados individuais de todas as etapas da pesquisa devem ser mantidos em local seguro por 5 anos para possível auditoria dos órgãos competentes.

Atenciosamente

Prof. Dr. José Osmar Medina Pestana
Coordenador do Comitê de Ética em Pesquisa da
Universidade Federal de São Paulo/ Hospital São Paulo

Anexo 2. Termo de Consentimento Livre e Esclarecido

TERMO DE CONSENTIMENTO PÓS INFORMADO DE PARTICIPAÇÃO EM ESTUDO CLÍNICO

O câncer de colo de útero é uma doença comum em nosso meio, sendo que além das infecções genitais (especialmente pelo papiloma vírus humano) os fatores genéticos podem contribuir para o desenvolvimento dessa doença. Sendo assim, nosso laboratório está desenvolvendo o projeto “Análise genômica e transcriptômica em câncer de colo uterino”, com o objetivo de contribuir para o melhor entendimento dos mecanismos deste tipo de câncer, assim como para seu diagnóstico e prognóstico.

Neste estudo, o material que será utilizado é proveniente de tecidos obtidos por ocasião da realização de biópsia com finalidade diagnóstica, onde alguns fragmentos serão congelados em nitrogênio líquido. Portanto, esta pesquisa não apresenta nenhum risco às pacientes envolvidas. Não será utilizado nenhum outro procedimento não rotineiro.

Não há benefícios diretos para as pacientes envolvidas neste estudo. Porém no final deste estudo talvez seja possível identificar alguns marcadores relacionados com este tipo de câncer, podendo futuramente servir para diagnóstico e prognóstico da doença. O material restante que não for utilizado neste projeto será armazenado para possíveis estudos futuros sob a responsabilidade da diretora do laboratório de Imunogenética da UNIFESP, Profa Dra Maria Gerbase de Lima. Todos os projetos desenvolvidos com este material serão, antes de tudo, submetidos à aprovação do Comitê de Ética em Pesquisa.

Informamos que estamos à disposição para qualquer esclarecimento a respeito deste estudo e que a paciente poderá retirar o seu consentimento a qualquer momento e deixar de participar do estudo, sem qualquer prejuízo à continuidade de seu tratamento na Instituição. Garantimos ainda que a paciente não será identificada em qualquer divulgação dos resultados deste trabalho e que poderá obter informações atualizadas sobre os resultados à medida que os mesmos venham sendo obtidos. Os custos dos testes realizados serão absorvidos pelo orçamento da pesquisa. Qualquer dúvida pode ser esclarecida com a orientadora deste projeto, Dra. Maria Gerbase de Lima, que pode ser encontrada no Laboratório de Imunogenética da UNIFESP, à rua Loefgreen, número 1235, CEP 04040-031, telefone 2101-7100. Se você tiver alguma

consideração ou dúvida sobre a ética da pesquisa, entre em contato com o Comitê de Ética em Pesquisa (CEP) – Rua Botucatu, 572 -1º andar cj.14 – telefone (11) 5571-1062, Fax (011)5539-7162.

Acredito ter sido suficientemente informado a respeito das informações que li ou que foram lidas para mim, descrevendo este estudo.

Ficaram claros para mim quais são os propósitos do estudo, os procedimentos a serem realizados, as garantias de confidencialidade e de esclarecimentos permanentes. Ficou claro também que minha participação é isenta de despesas e que tenho garantia do acesso a tratamento hospitalar quando necessário. Concordo voluntariamente em participar deste estudo e poderei retirar o meu consentimento a qualquer momento, antes ou durante o mesmo, sem penalidades ou prejuízo ou perda de qualquer benefício que eu possa ter adquirido, ou no meu atendimento neste Serviço.

Assinatura do paciente/representante legal

Data ____ / ____ / ____

Assinatura da testemunha

Data ____ / ____ / ____

(Somente para o responsável do projeto)

Declaro que obtive de forma apropriada e voluntária o Consentimento Livre e Esclarecido deste paciente ou representante legal para a participação neste estudo.

Assinatura do responsável pelo estudo

Data ____ / ____ / ____

Anexo 3. Dados clínicos das pacientes incluídas na análise de resposta ao tratamento

Amostras	Idade (anos)	Estadio	Tratamento	Classificação	Tipo de lâmina
1	31	IIb	RT	R	HsaC
2*	37	IIb	RT + QT	R	HsaB
3*	67	IIb	RT + QT	R	HsaB
4	44	IIb	RT	R	HsaC
5	45	IIb	RT + QT	R	HsaC
6	47	IIb	RT	R	HsaC
7	48	IIIb	RT + QT	R	HsaB
8	51	IIb	RT + QT	R	HsaC
9	55	IIIb	RT + QT	R	HsaC
10	59	IIIb	RT	R	HsaC
11	60	IIIb	RT + QT	R	HsaC
12	66	IIb	RT	R	HsaC
13	69	IIIb	RT	R	HsaC
14	75	IIb	RT	R	HsaC
15	30	IIIb	RT	NR	HsaB
16	33	IIb	RT + QT	NR	HsaB
17	40	IIb	RT + QT	NR	HsaC
18	41	IIb	RT + QT	NR	HsaC
19	41	IIIb	RT + QT	NR	HsaC
20	51	IIIb	RT	NR	HsaC
21	55	IIIb	RT	NR	HsaB
22	56	IIIb	RT + QT	NR	HsaB
23	57	IIb	RT + QT	NR	HsaC
24	74	IIIb	RT	NR	HsaC

(*) indicam as duas amostras que foram hibridizadas juntas na mesma lâmina de *microarray*

Tratamento: radioterapia (RT); radioterapia com quimioterapia (RT+QT)

Classificação: respondedoras ao tratamento (R); não respondedoras ao tratamento (NR)

Anexo 4. *Gene Symbol* e razão N/T (razão da expressão da amostra normal pela tumoral) dos 341 genes com alta expressão no tumor.

Gene symbol	Razão N/T	Gene symbol	Razão N/T	Gene symbol	Razão N/T
ACYP1	0.69657	COL4A1	0.52187	GAPDH	0.65166
AGRN	0.48733	COL4A2	0.47434	GAS1	0.47682
AKR1B1	0.66300	COPE	0.63336	GNB1	0.64798
ALDH3A1	0.50831	COX17	0.39618	GOLPH2	0.49776
AMOTL2	0.53938	CPSF1	0.54266	GPSN2	0.56173
ANAPC5	0.73753	CTSZ	0.43145	GPX4	0.52074
ANTXR1	0.48402	CXCL10	0.29715	GRM7	0.57734
AP2M1	0.57354	CXCL9	0.18189	GZMA	0.50422
APOBEC3B	0.48734	CYR61	0.46896	H2AFV	0.47828
APOBEC3G	0.55327	DAD1	0.54220	H2AFX	0.53610
ARL6IP1	0.65537	DBF4	0.41655	HARS2	0.65949
ARPP-19	0.52820	DCDC2	0.64070	HCG2P7	0.56094
ATP5J	0.61050	DDOST	0.63862	HDAC6	0.69974
AURKA	0.41870	DDR2	0.47926	HDGF	0.66643
AZ11	0.59940	DDX11	0.64783	HERC5	0.45505
B3GNT1	0.65660	DFNA5	0.42565	HEY1	0.32358
B4GALT4	0.55453	DKFZP586K1520	0.67291	HIRA	0.64698
BCAR1	0.70036	DLC1	0.50415	HIST1H2BE	0.70734
BCL7B	0.54606	DLG5	0.49018	HLA-A	0.51641
BIRC5	0.52949	DNAJA1	0.63062	HLA-B	0.40357
BST2	0.58805	DNAJB11	0.56415	HLA-DRA	0.38785
BUB3	0.62214	DNMT1	0.49619	HLA-DRB3	0.47823
C10orf22	0.69920	DNTTIP1	0.64296	HLA-F	0.52165
C1orf104	0.69838	DUSP12	0.72648	HMGB2	0.55208
C1RL	0.60219	DUSP23	0.67681	HNF4G	0.53691
C4A	0.55603	DUSP9	0.41023	HOMER3	0.62029
C9orf114	0.67599	E2F3	0.41423	HOXB4	0.62994
CA9	0.24604	EBI3	0.46745	HSD17B6	0.55287
CBS	0.37423	EEF2	0.46385	HSP90AB1	0.59115
CBX1	0.60538	EFNA1	0.53054	HSPB1	0.39686
CCL11	0.27547	ELF4	0.53118	HSPC152	0.66346
CCNB1	0.62075	ENO1	0.50817	ICMT	0.66257
CCNB2	0.46235	EXOSC3	0.58291	IFI44L	0.29676
CCT2	0.72677	EXT1	0.65392	IFI6	0.46833
CD74	0.52793	FADS2	0.35880	IGKC	0.19954
CDC20	0.60100	FAM126B	0.62109	IGL@	0.36593
CDC25C	0.42959	FAM64A	0.48108	IL15RA	0.61299
CDC45L	0.48340	FANCA	0.53123	IL27RA	0.57140
CDCA4	0.60754	FANCL	0.51490	IL2RB	0.60671
CDCA8	0.44187	FARP1	0.56686	INDO	0.19969
CDH3	0.42183	FLJ10241	0.71885	ING1	0.68618
CDK2	0.54638	FLJ10292	0.49090	INPPL1	0.68362
CDK4	0.55144	FLJ33790	0.45931	IQGAP2	0.51992
CDKN2A	0.15098	FLNB	0.48525	IRF1	0.60161
CDKN3	0.54227	FMO1	0.56228	ISG15	0.40394
CENPF	0.46399	FNDC3B	0.51353	ITGB4	0.62411
CEP55	0.38354	FOSL2	0.56705	KIAA0020	0.68905
CFB	0.40983	FOXD1	0.30022	KIAA0226	0.60675
CKS2	0.47961	FTL	0.48746	KIAA0406	0.56682
CLDN1	0.45088	FUCA1	0.68854	KIAA0474	0.36731

continua...

... continuação e conclusão

Gene symbol	Razão N/T	Gene symbol	Razão N/T	Gene symbol	Razão N/T	Gene symbol	Razão N/T
KIAA0515	0.61382	NCAPD3	0.44073	PTTG2	0.45627	TCF3	0.68363
KIAA0586	0.68410	NCAPG2	0.37357	PXDN	0.40226	TEAD4	0.43197
KIAA0746	0.48623	NCOR1	0.69155	RACGAP1	0.43644	TERF2	0.63251
KIF22	0.61402	NDC80	0.46169	RAPGEF5	0.57817	TK1	0.50771
KIF2C	0.40837	NDUFA13	0.49475	RBBP5	0.73743	TKT	0.54559
KIF4A	0.49154	NDUFS6	0.40245	RBMX	0.66262	TMEM29	0.73811
KLHL8	0.63643	NEDD4L	0.47084	RCC1	0.45180	TMEM97	0.52592
KPNB1	0.67112	NEK6	0.40824	REC8	0.63357	TMSB10	0.53462
KRT17	0.49557	NIPA2	0.59682	RFC4	0.32124	TNK2	0.65287
KRT8	0.57309	NLGN2	0.49119	RFX5	0.60951	TNNT1	0.37085
KRT8P12	0.48581	NOLC1	0.69520	RNASEH1	0.63655	TPR	0.58636
LAMC1	0.63558	NQO1	0.40262	RNASEH2A	0.38374	TRAPPC3	0.72733
LAMP3	0.50900	NUDT4	0.50040	RRP1	0.65258	TRIP13	0.49974
LASP1	0.56050	NUP107	0.63668	RRP1B	0.47771	TRMT12	0.57452
LDLRAD2	0.69046	NUP62CL	0.51794	RUNX3	0.54057	TTC1	0.66490
LHX2	0.34944	NVL	0.74262	SCO2	0.61012	TUBA1B	0.45525
LIG1	0.55905	OLFM1	0.52715	SEMA3E	0.62396	TUBG1	0.53573
LOC401565	0.52211	PAQR4	0.65303	SESTD1	0.44319	TXNL4B	0.54401
LOC51145	0.66852	PAX6	0.46747	SETD5	0.63274	TYMS	0.35356
LOC90586	0.67841	PCNA	0.47769	SEZ6L2	0.43568	UBB	0.49115
LRRC8D	0.66698	PDPN	0.45342	SFRS1	0.65788	UBD	0.25252
LSM4	0.65286	PELP1	0.60400	SFRS6	0.68507	UBE1	0.70459
LSR	0.60364	PIGT	0.47794	SGPL1	0.54707	UBE2C	0.50902
LTF	0.42148	PIH1D1	0.65576	SIX1	0.17802	UBE2M	0.69786
LUZP2	0.63599	PIK3R3	0.61014	SLC12A7	0.49039	UBE2S	0.35140
MAGEA6	0.33711	PLK1	0.50589	SLC14A2	0.63658	UBE2T	0.51022
MARCKSL1	0.52671	PML	0.60700	SLC1A5	0.60989	UCN	0.50065
MARS	0.73296	PNKD	0.53569	SLC41A3	0.64207	VCAM1	0.42275
MCM2	0.39516	PODXL	0.54536	SLC8A2	0.53750	VCAN	0.50386
MCM5	0.46001	POLA2	0.69154	SMAD6	0.58224	WDR55	0.62420
MCM7	0.50000	POLR2H	0.70036	SNAPC4	0.65848	WDR8	0.63324
MDK	0.39632	POLS	0.67866	SNRPA	0.70010	WFDC2	0.45087
MED28	0.65781	POM121	0.71561	SNRPA1	0.65204	WWC3	0.63352
MELK	0.39356	PPAP2C	0.45759	SNRPB	0.64919	XRCC2	0.65367
MEST	0.33268	PPAT	0.59644	SOX4	0.39789	ZC3HC1	0.63485
MGAT4B	0.53784	PPIH	0.72884	SPAG5	0.42018	ZHX2	0.41483
MMD	0.54504	PPT1	0.58051	SRRM2	0.61367	ZNF148	0.52870
MMP11	0.36533	PRC1	0.46724	STAG3	0.37648	ZNF428	0.64128
MMP12	0.30229	PRDX4	0.49660	STC1	0.39137	ZNF793	0.65819
MOCOS	0.25914	PRPF8	0.68381	STIL	0.41355	ZRSR2	0.65798
MORC2	0.71279	PRPS1	0.50150	STX1A	0.64005	ZWINT	0.43741
MPHOSPH1	0.56713	PRPS2	0.57824	STX4	0.57657		
MRPL22	0.68699	PSMB2	0.62478	SULF1	0.30696		
MRPS17	0.56820	PSMB4	0.73550	SYCP2	0.35490		
MX1	0.50503	PSMB9	0.53600	SYT17	0.42349		
MYBL2	0.44379	PSMC4	0.78100	TAF15	0.39749		
MYLIP	0.56371	PSMD14	0.56892	TAF5L	0.54614		
MYO1B	0.57555	PSME1	0.57158	TAP1	0.56113		
NAB2	0.63737	PTHLH	0.53142	TAPBP	0.64347		
NASP	0.63272	PTTG1	0.49829	TCF15	0.48508		

Anexo 5. *Gene Symbol* e razão N/T (razão da expressão da amostra normal pela tumoral) dos 469 genes com baixa expressão no tumor.

Gene symbol	Razão N/T	Gene symbol	Razão N/T	Gene symbol	Razão N/T
ABCD3	1.43949	C1orf144	1.89045	CRYAB	5.52234
ABHD5	2.60078	C20orf108	2.08784	CRYL1	2.08381
ABI1	1.86163	C20orf111	1.93823	CSNK1A1	1.81726
ABI3BP	4.17012	C20orf42	2.50483	CST4	1.60328
ACAA1	1.98078	C2orf32	2.04059	CSTA	6.24083
ACADS	1.85299	C3orf28	2.65849	CTNND1	2.34931
ACBD3	1.69953	C7orf27	1.54443	CTSH	1.92882
ACBD5	2.03720	C7orf43	1.57768	CUL4B	1.93441
ACPP	2.14225	C9orf95	3.31944	CYP3A5	5.23651
ACSL1	2.62890	CAB39	2.06585	CYP4F12	1.55331
ADD3	2.10751	CACNA2D3	4.01072	CYP51A1	2.03936
AIG1	1.90260	CAPZA1	1.52048	DBI	2.08143
AIM1L	2.63483	CAPZA2	1.67376	DDT	1.80931
ALDH3B2	1.76284	CAST	1.55852	DDX50	1.64617
ALG5	1.66623	CCNDBP1	1.76148	DHRS9	2.56087
ALOX12	5.12611	CCNG2	1.68119	DHX32	2.04286
ALOX12B	2.41257	CCPG1	2.89412	DLD	2.19527
ALOX15B	4.08253	CD164	2.59993	DLEU1	1.75338
AMMECR1	2.71596	CD24	3.85042	DMXL1	1.86497
ANGPTL1	2.86550	CD40	1.59678	DOCK4	2.85441
ANKMY1	1.58353	CDA	3.36568	DOCK9	3.40155
ANKRD17	1.36125	CDH13	1.92839	DSC2	2.82100
ANKRD28	1.72461	CDYL	2.96040	DSG1	3.96331
ANKRD46	1.97787	CEACAM5	3.16879	DUSP22	1.44122
ANKRD50	2.05407	CEACAM7	6.67059	ECM1	2.69045
ANXA1	6.08063	CEBPA	2.15416	EEF1B2	2.47408
ANXA9	3.58310	CG018	1.80172	EGFR	1.43239
AREG	2.77359	CGRRF1	1.55178	EHBP1	1.79467
ARFRP1	1.50161	CHSY1	1.64409	EHD3	2.01717
ARG2	2.95017	CIRBP	1.78624	EIF3S10	1.94057
ARHGEF4	2.35364	CLCA4	5.16451	EIF4G2	2.39058
ARIH2	1.44945	CLK1	2.16656	ELF1	1.84342
ARL4A	2.68644	CLTB	2.46777	EML1	1.92467
ARL4C	1.82187	CMPK	1.66356	EMP1	10.80272
ARL5A	2.01500	CNN3	2.23430	ENDOD1	2.26903
ARNTL	1.59836	COCH	2.24429	EPB41L3	2.94444
ASGR1	1.91144	COIL	1.57540	EPS15	2.11096
ATP5E	1.87699	COL23A1	2.17293	EPS8L1	3.07052
ATP5G1	1.79402	COPS4	1.69121	ERBB2IP	1.71956
ATP6V1D	1.84737	COQ3	1.67807	ERCC8	1.63567
ATRX	1.98649	COX5B	1.84785	ERO1L	2.61675
AURKC	2.79536	COX6A1	1.46610	ESR1	4.60143
B3GALT4	1.64149	COX6B1	1.67448	ESRRB	1.66347
BACH1	1.85290	CPEB3	2.18503	ETFDH	2.16226
BCHE	2.16090	CRADD	1.68412	ETNK1	1.69439
BEXL1	2.98322	CRCT1	18.17583	EVA1	2.18928
BLVRB	1.67179	CREB3L2	1.57573	EXOC2	1.85261
C10orf61	1.35800	CREBL2	2.41980	FAM135A	2.10210
C12orf29	1.88000	CRISP3	29.50520	FAM18B	1.82543
C18orf8	1.68243	CRNN	14.81011	FAM32A	1.47999

continua...


... continuação

Gene symbol	Razão N/T	Gene symbol	Razão N/T	Gene symbol	Razão N/T
FAM45A	1.92631	KCNG1	2.17705	MED4	1.65667
FAM49A	2.13726	KCNJ15	2.78220	MEIS1	2.08698
FBXL11	1.45384	KCTD12	1.85534	METTL9	1.73147
FBXO28	1.49781	KIAA0240	1.64178	MFHAS1	1.92595
FCER1A	4.03641	KIAA0423	1.99308	MGC10334	1.63005
FGFBP1	4.04741	KLF4	2.43763	MKKNK2	1.88580
FLJ12716	1.75236	KLF8	2.75309	MLLT4	1.93893
FLJ20273	1.97350	KLHL18	2.17775	MPST	2.27359
FLRT3	3.56770	KLK10	5.19449	MRPL20	1.58443
FOXP1	1.76445	KLK11	6.28353	MRPL35	1.67264
FRAG1	1.63551	KLK12	5.83160	MRPL48	1.49681
FUSIP1	1.55069	KLK13	4.27582	MRPS16	1.66177
FUT3	2.04317	KLK5	2.62285	MRPS18B	1.47170
FZD7	1.72723	KLK6	5.34741	MTMR1	1.91361
GALNT7	1.71138	KLK8	3.30449	MTMR6	1.83007
GAPVD1	2.21364	KLK9	3.28387	MTPN	1.87255
GATA3	2.12320	KPNA3	1.75350	MYCBP2	2.41297
GATM	5.79686	KRT1	3.66275	MYO5B	2.19410
GCNT3	3.76101	KRT10	3.03796	NACA	1.48735
GCSH	2.06273	KRT2	2.09558	NAG8	1.53737
GLUD1	2.02618	LANCL1	2.39860	NCKAP1	2.00475
GM2A	1.85579	LARS2	1.48334	NCOA4	1.65590
GOLGA4	1.90040	LGALS3	3.06326	NDRG2	2.00704
GPD1L	2.08984	LMBR1L	1.51868	NDRG4	3.64370
GRHPR	1.62288	LMO2	2.22261	NDST2	2.18100
GSPT1	1.30070	LOC171220	1.94235	NDUFC2	1.85220
GTF2A2	1.58226	LOC57228	1.77149	NFE2L2	2.33296
GTF2B	1.48047	LOC642250	1.53843	NFU1	1.78846
hCG_2024094	2.48554	LOC642741	1.67325	NLRP1	2.06240
HEBP2	2.61526	LOC644820	2.23214	NOP5/NOP58	2.08434
HECTD1	2.59429	LOC646912	1.99485	NP	2.94520
HES2	2.30679	LOC653447	1.76006	NPEPPS	1.69283
HIGD1A	2.49323	LOC731536	1.38980	NPHP4	3.89260
HIP2	1.62959	LPIN1	3.34650	NR1D1	2.39733
HIPK1	1.86818	LRMP	1.84770	NUP54	1.75658
HNRPA0	1.85959	LRP10	2.08891	NUPL2	1.47769
HPGD	6.54533	LRRFIP1	1.72936	OSBPL2	1.60002
HSD17B4	1.55707	LSM1	1.39847	OSTF1	1.64284
HSD17B8	1.58949	LTBP3	1.57117	OVOL1	2.03249
ICHTHYIN	3.05910	LYPLA1	1.84783	P2RY5	2.07011
ICOSLG	1.85553	MAFB	2.43813	PABPC1	2.58700
IDUA	1.52387	MAL	16.29842	PCLO	5.55308
IKBKB	1.46602	MALT1	2.40434	PCYT2	1.48516
IL1RN	4.02507	MANSC1	1.89071	PDCD2	1.75736
ISL1	4.34445	MAP3K4	1.78476	PDE2A	2.26929
ISL2	2.18395	MAP4K3	1.77058	PDS5A	1.84275
ITPKC	1.75174	MAPK11	2.30255	PGCP	1.85620
JMJD1B	1.36063	MARCH7	3.47581	PGM2	2.63063
JMJD1C	2.08970	MBNL1	1.82962	PHKB	1.67594
KCMF1	2.99713	ME1	2.06293	PIK3C3	1.56490

continua...

... continuação e conclusão


Gene symbol	Razão N/T	Gene symbol	Razão N/T	Gene symbol	Razão N/T	Gene symbol	Razão N/T
PIM1	2.38011	RRAGC	1.75241	SRPK1	1.72278	UROD	1.49443
PLEKHM1	2.51812	RSAD1	2.02793	SRPK2	1.68865	USP25	1.76578
PLK2	1.66373	RTCD1	1.75342	SRRM1	1.55157	USP6NL	1.90348
PLS3	2.56301	S100A9	3.78569	STARD9	1.98230	USP8	1.66177
PMPCB	1.59458	SACM1L	2.07707	STAT4	1.86914	VAPA	1.90910
POLR1D	2.30976	SAPS2	1.74223	STK24	1.54427	VPS52	1.51227
PPA1	1.88785	SAR1B	1.99191	STK40	1.84841	VPS54	2.22505
PPFIBP2	1.98243	SBDS	1.60883	STX7	1.40800	WDR47	2.54567
PPL	4.70656	SCNN1B	3.15005	STYXL1	1.39983	WIPI2	1.60122
PPM1F	1.61972	SDC2	1.70052	SUMO2	1.47700	YIPF1	1.39813
PPP1CB	2.13701	SEC31A	2.06070	SYNGR1	1.83863	YIPF6	1.52885
PPP2R5C	1.43512	SEC63	2.08077	TANC1	2.45822	YTHDF2	2.09251
PRDM10	1.83536	SEMA4D	1.44080	TBC1D4	1.52245	ZC3H7A	2.02447
PRELID1	1.44361	SEPT7	1.86880	TBX3	1.99284	ZDHH3	1.52036
PRPF40A	2.18345	SERGEF	1.74821	TFB1M	1.53796	ZNF185	3.40908
PRPF4B	1.98243	SERPINB2	3.21229	TGFA	2.10330	ZNF701	1.52680
PRSS2	4.31829	SERPINB3	4.42487	TGM1	3.88185	ZNF706	1.43296
PRSS3	2.65594	SERPINB5	2.40257	TGM3	4.22309	ZNF791	1.53351
PSCA	3.03300	SERPINI1	1.65771	TIAM1	3.69144	ZRANB2	1.79034
PSMC6	1.79674	SERTAD2	2.28607	TK2	1.74497		
PTEN	1.64786	SF3B14	1.75359	TM9SF4	1.56383		
PTS	1.82623	SFRS10	2.33052	TMED3	1.88259		
RAB10	2.31291	SFRS11	1.47840	TMEM109	1.77278		
RAB4A	1.56482	SFRS2B	1.40991	TMEM40	3.17736		
RAB9A	1.91192	SGMS1	1.93280	TMEM50B	1.66507		
RAPGEF2	2.16164	SH3BGRL	1.97422	TMEM57	1.91701		
RASA4	2.01602	SH3D19	1.84349	TMEM66	1.79910		
RASAL1	1.87791	SIPA1L2	1.78971	TMPRSS11E	8.77543		
RBL2	2.00413	SKAP2	1.78627	TOM1L2	1.80957		
RCBTB2	2.68782	SKP1A	1.88923	TOMM7	1.62245		
RG9MTD1	1.51814	SLC11A2	1.71726	TP53I3	2.56584		
RIN1	1.85631	SLC22A3	1.86001	TP53TG3	3.37315		
RMND5B	1.45299	SLC2A5	1.73936	TPP2	1.44178		
RNF111	1.63901	SLC35C1	1.76235	TPT1	2.01755		
RNGTT	1.83262	SLC37A4	1.75920	TRIM16	2.54637		
RORA	2.57359	SLK	2.80582	TRIT1	2.25891		
RPL13A	1.68052	SLURP1	6.01967	TSG101	1.78978		
RPL14	1.89447	SMAD4	1.72805	TSPAN6	2.46976		
RPL19	1.57665	SNX14	1.70647	TSPAN7	2.45700		
RPL3	1.92614	SNX3	1.67678	TUBA1A	1.88243		
RPL30	2.15784	SNX9	1.96748	TXN2	2.14365		
RPL32	1.85038	SOSTDC1	4.69609	TXNDC9	2.21551		
RPL7	1.95265	SP3	1.77725	TXNIP	1.58094		
RPLP1	2.32155	SPAG9	2.20130	TYRO3	1.92795		
RPLP2	1.87469	SPARCL1	2.95182	UBE2D3	1.99958		
RPS13	1.60656	SPCS1	1.74145	UBE2G1	1.92779		
RPS15A	1.62510	SPRR1B	6.50861	UBL3	3.05201		
RPS17	1.45994	SPRR2C	2.84707	UPK1A	4.11612		
RPS18	1.68336	SPRR2D	4.39971	UPP1	2.46572		
RPS27	1.67549	SPRR3	9.30974	UQCRC2	1.67899		

Anexo 6. Pôster apresentado no 54^o Congresso Brasileiro de Genética, Salvador, BA.


GENOME-WIDE PROFILING REVEALS NEW MOLECULAR PATHWAYS IN CERVICAL CANCER

Mine KL¹, Sanson GFO¹, Morgun A³, Silva ID², Brenna SMF², Carvalho CRN², Ribalta JCL², Gerbase-DeLima M¹, Shulzhenko N³

¹Immunogenetics Division, Pediatrics Department and ²Gynecology Department, Federal University of de Sao Paulo, SP, Brazil
³Ghost Lab, NIAID, NIH, Bethesda, MD, United States



Introduction

• Cervical cancer is the second most frequent cancer in women worldwide and the majority of cases occur in developing countries.

• Although infection with high-risk human papillomavirus (HPV) is considered to be the initial event of tumorigenesis in most of the case, only a small number of women infected with HPV actually develop cervical cancer. Therefore additional factors or cellular events are thought to be required for progression to cancer.

• In the present study we aimed to compare genome-wide gene expression between cervical tumor and normal adjacent tissue, using microarrays, to identify genes and functional categories associated with cervical cancer.

Material & Methods

• We hybridized 34 tumor and 20 normal adjacent tissue samples against a reference sample in chips containing 70-mer oligonucleotides (~12,000 human genes).

• Analysis steps:

1. Identification of differentially expressed genes (DEGs) between the tumor and normal tissues. (Software: BRB Array Tools)
2. Identification of functional categories (pathways) over-represented in the DEGs. (Software: DAVID, Fatigo, and Ingenuity Pathway Analysis (IPA))
3. Verification of how our findings relate to the published literature. (Software: MILANO)

Results

1. We identified **810 DEGs** (false discovery rate <10%):

341 were up-regulated in tumor samples

469 were down-regulated in tumor samples

2. Among the DEGs we identified **13 over-represented pathways**

• Most of these pathways have been previously implicated in the pathogenesis of cervical cancer.

• The finding of two down-regulated categories in cancer was quite surprising: **'Ribosome'** (protein synthesis) and **'Urea Cycle and Metabolism of Amino Groups'** (involved in the production of polyamines that are related with cell growth and transformation).

3. A literature search showed that:

• In the up-regulated DEGs, the majority of the genes from the **'Pentose Phosphate'** and **'Protein Ubiquitination'** pathways had not been related to cervical cancer in other studies.

• Among the down-regulated DEGs, none of **'Ribosome'** and **'Urea Cycle and Metabolism of Amino Groups'** genes had been previously described to be associated with cervical cancer, even the 'Ribosome' that is one of the largest functional categories, comprising 14 genes.

Pathways	Software			Gene expression in tumor samples
	DAVID	Fatigo	IPA	
Cell Cycle				Up-regulated
Antigen Processing and Presentation				Up-regulated
Protein Ubiquitination Pathway				Up-regulated
p53 Signaling				Up-regulated
Pentose Phosphate Pathway				Up-regulated
Methionine Metabolism				Up-regulated
Cell Adhesion Molecules (CAMs)				Up-regulated
Fatty Acid Metabolism				Down-regulated
Oxidative Phosphorylation				Down-regulated
Ribosome				Down-regulated
MAPKase Signaling Pathway				Down-regulated
Riboflavin Metabolism				Down-regulated
Urea cycle and metabolism of amino groups				Down-regulated

Pathways	Gene expression in tumor samples	Pathway gene number	Numbers and percentages of genes with published articles			
			Cervical Cancer	Cervical Cancer Cells	Squamous Cell Cancer	Cancer
Ribosome	Down-regulated	14	0	0	4 (29%)	13 (93%)
Oxidative Phosphorylation	Down-regulated	9	0	0	1 (11%)	8 (87%)
MAPKase Signaling Pathway	Down-regulated	7	1 (24%)	1 (14%)	2 (29%)	7 (100%)
Urea Cycle and metabolism of amino groups	Down-regulated	3	0	0	0	3 (100%)
Riboflavin Metabolism	Down-regulated	3	0	0	0	3 (100%)
Pentose Phosphate Pathway	Up-regulated	3	0	0	1 (33%)	3 (100%)
Fatty Acid Metabolism	Down-regulated	8	2 (25%)	0	2 (25%)	7 (88%)
Protein Ubiquitination Pathway	Up-regulated	18	6 (33%)	5 (28%)	7 (39%)	17 (94%)
Antigen Processing and Presentation	Up-regulated	12	5 (42%)	3 (25%)	6 (50%)	12 (100%)
Cell cycle	Up-regulated	22	12 (55%)	11 (50%)	16 (73%)	22 (100%)
Cell Adhesion Molecules (CAMs)	Up-regulated	9	5 (56%)	3 (33%)	6 (67%)	8 (89%)
p53 Signaling	Up-regulated	7	5 (71%)	5 (71%)	5 (71%)	7 (100%)
Methionine Metabolism	Up-regulated	3	3 (100%)	1 (33%)	3 (100%)	3 (100%)

Conclusion

Using genome-wide transcriptome analysis, we found previously known (e.g. 'Cell Cycle' and 'Protein Ubiquitination Pathway') and new molecular pathways ('Ribosome' and 'Urea Cycle and Metabolism of Amino Groups') in cervical cancer.

These results might help to draw a more general picture of the cervical cancer process, as well as to open novel research avenues.

Supported by FAPESP

Anexo 7. Pôster apresentado na *Molecular Medicine Tri-Conference 2009, San Francisco, EUA.*



PATHWAYS AND GENE NETWORKS ANALYSIS REVEALS MOLECULAR EVENTS INVOLVED IN CERVICAL CANCER

Mine KL¹, Shulzhenko N^{1,3}, Yambartsev A⁴, Sanson GFO¹, Silva ID², Brenna SMF², Carvalho CRN², Ribalta JCL², Gerbase-DeLima M¹, Morgun A^{1,3}

¹Immunogenetics Division, Pediatrics Department and ²Gynecology Department, Federal University of de Sao Paulo, SP, Brazil
³Ghost Lab, NIAID, NIH, Bethesda, MD, United States

⁴Institute of Mathematics and Statistics, Department of Statistics, University of Sao Paulo, SP, Brazil



Introduction

❖ Cervical cancer is the second most frequent cancer in women worldwide and the majority of cases occur in developing countries.

❖ Although infection with high-risk human papillomavirus (HPV) is considered to be the initial event of tumorigenesis in most of the cases, only a small number of women infected with HPV actually develop cervical cancer. Therefore, additional factors are thought to be required for progression to cancer.

In order to search for molecular events involved in the cervical cancer pathogenesis, we aimed to identify pathways and gene networks represented in the genes differentially expressed (DEGs) between cervical tumor and normal adjacent tissue.

Material & Methods

❖ We hybridized 34 tumor and 20 normal adjacent tissue samples against a reference sample in chips containing 70-mer oligonucleotides (~12,000 human genes).

❖ Analysis steps:

1. Identification of DEGs between the two tissue types through microarray analysis. (Software: BRB Array Tools).

2a. Identification of pathways over-represented in the DEGs. (Software: DAVID, PathGO, and Ingenuity Pathway Analysis (IPA))

2b. Identification of gene networks in the DEGs.

We used two types of interaction measures between the genes:

- I. kappa statistics (DAVID), as a measure of the annotation co-occurrence of any given gene pair, to identify related genes with similarity of their annotation profiles;
- II. relative co-occurrences of each gene pair within the same article estimated by the MILANO tool. (Software: DAVID, MILANO, CFinder, Cytoscape and GeneCodis)

Results

1. We identified **810 DEGs** (false discovery rate <10%): **341 genes were up-regulated (Up)** and **469 were down-regulated (Dn)** in tumor samples.

2a. We identified **13 over-represented pathways** in the DEGs:

Up-regulated gene list	Pathways	Down-regulated gene list
<ul style="list-style-type: none"> ★ Cell cycle ★ Protein Ubiquitination Pathway ★ Antigen Processing and Presentation ★ Cell Adhesion Molecules (CAMs) ★ Pentose Phosphate Pathway ★ p53 Signaling ★ Methionine Metabolism 	<ul style="list-style-type: none"> ★ Ribosome ★ Urea Cycle and metabolism of amino groups ★ Oxidative Phosphorylation ★ MAPKine Signaling Pathway ★ Riboflavin Metabolism ★ Fatty Acid Metabolism 	<ul style="list-style-type: none"> ★ Ribosome ★ Urea Cycle and metabolism of amino groups ★ Oxidative Phosphorylation ★ MAPKine Signaling Pathway ★ Riboflavin Metabolism ★ Fatty Acid Metabolism

★ Pathways also found in the gene network analysis.

2b. In the gene network analysis we found **23 sub-networks** using **kappa statistics** (Fig.1 and Fig.2) and **2 sub-networks** using **MILANO results**.

Sub-networks Functional Annotation	
kappa statistics: Up-regulated gene list <ol style="list-style-type: none"> 1. Transcription 2. RNA related mechanisms 3. Cell Cycle/ p53 signaling pathway/ DNA replication 4. Protein Ubiquitination Pathway 5. Cell adhesion molecules (CAMs) 6. Glycan Structures – biosynthesis 1/ N- Glycan biosynthesis/ Glycan structures – biosynthesis 2 7. Phosphatase activity 8. Unknown 9. Ubiquitin mediated proteolysis 	kappa statistics: Down-regulated gene list <ol style="list-style-type: none"> 1. Transcription 2. MAPK signaling pathway 3. Ribosome / Translation 4. Membrane 5. Kallikrein-related peptidase (KLK)/ Cytokine-cytokine receptor interaction 6. GTP binding and GTPase activity 7. RNA related mechanisms 8. Ion transport 9. Glycan Structures – biosynthesis 1 10. Phosphatase activity 11. Keratinization 12. Ubiquitin-dependent protein catabolic process 13. Oxidative phosphorylation 14. Membrane
MILANO results <ol style="list-style-type: none"> 1. Kallikrein-related peptidase (KLK) 2. Unknown 	

★ Sub-networks also found in the pathways analysis.
★ Sub-network found in both network analyses.

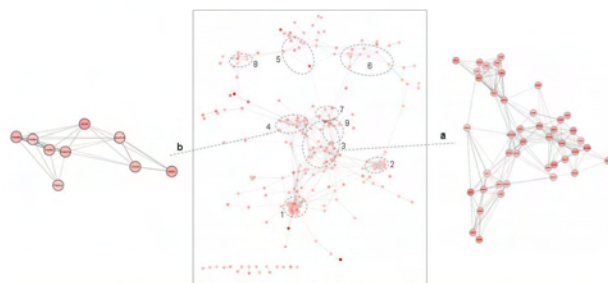


Figure 1. Gene network from the Up-regulated gene list found using kappa statistics. The numbers indicate the sub-networks listed in the table above. (a) Sub-network related to Cell Cycle/p53 signaling pathways/ DNA replication and (b) Sub-network related to Protein Ubiquitination Pathway. The intensity of node color is proportional to the gene expression ratio between tumor and normal samples and the intensity of edge color is proportional to the kappa value between a gene pair.

❖ We found previously known (e.g. 'Cell Cycle' and 'Protein Ubiquitination Pathway') and unknown molecular pathways ('Ribosome' and 'Urea Cycle and Metabolism of Amino Groups') in cervical cancer.

❖ The sub-networks related to **Ribosome** (Fig.2a) and **KLKs genes** (Fig.2b) have not been previously identified in cervical cancer tissue and might be interesting candidates for future studies.

❖ The **ribosomal proteins**, besides their role in the protein synthesis, have others extra-ribosomal functions that could be involved in the tumorigenesis and **human tissue kallikreins** have been implicated in several cancer-related processes and have some clinical applicability as cancer biomarkers.

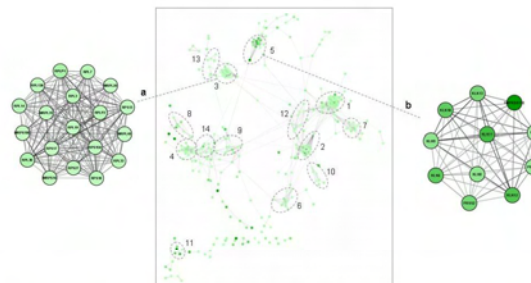


Figure 2. Gene network from the Dn-regulated gene list found using kappa statistics. The numbers indicate the sub-networks listed in the table above. (a) Sub-network related to Ribosome and (b) Sub-network related to KLKs genes. The intensity of node color is proportional to the gene expression ratio between normal and tumor samples and the intensity of edge color is proportional to the kappa value between a gene pair.

Conclusion

Using combination of pathway enrichment and network analysis based on whole transcriptome results we have found some previously known and revealed some new molecular pathways in cervical cancer.

Supported by: Fundacao de Amparo a Pesquisa do Estado de Sao Paulo (FAPESP), Brazil

8 REFERÊNCIAS

- Acevedo Rocha CG, Alvarez E, Zafra de la Rosa G, Alvarez Navarro M, Gariglio P. [Cervical cancer and DNA microarrays: tumour marker identification]. *Ginecol Obstet Mex.* 2007;75(4):205-13. Spanish.
- Adamcsek B, Palla G, Farkas IJ, Derényi I, Vicsek T. CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics.* 2006;22(8):1021-3.
- Ahn WS, Bae SM, Lee JM, Namkoong SE, Han SJ, Cho YL, et al. Searching for pathogenic gene functions to cervical cancer. *Gynecol Oncol.* 2004;93(1):41-8.
- Al-Shahrour F, Minguéz P, Vaquerizas JM, Conde L, Dopazo J. BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. *Nucleic Acids Res.* 2005;33(Web Server issue):W460-4.
- Alon U, Barkai N, Notterman DA, Gish K, Ybarra S, Mack D, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc Natl Acad Sci U S A.* 1999;96(12):6745-50.
- Arantes AM. Análise do perfil de expressão gênica em pacientes submetidos a transplante alogênico HLA-idêntico de células tronco hematopoéticas [tese]. São Paulo: Universidade Federal de São Paulo; 2009.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000;25(1):25-9.
- Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics.* 2003;4:2.
- Baldwin P, Laskey R, Coleman N. Translational approaches to improving cervical screening. *Nat Rev Cancer.* 2003;3(3):217-26.
- Benedet JL, Bender H, Jones H 3rd, Ngan HY, Pecorelli S. FIGO staging classifications and clinical practice guidelines in the management of gynecologic cancers. FIGO Committee on Gynecologic Oncology. *Int J Gynaecol Obstet.* 2000;70(2):209-62

Benito M, Parker J, Du Q, Wu J, Xiang D, Perou CM, et al. Adjustment of systematic microarray data biases. *Bioinformatics*. 2004;20(1):105-14.

Biewenga P, Buist MR, Moerland PD, Ver Loren van Themaat E, van Kampen AH, ten Kate FJ, et al. Gene expression in early stage cervical cancer. *Gynecol Oncol*. 2008;108(3):520-6.

Boldt S, Weidle UH, Kolch W. The role of MAPK pathways in the action of chemotherapeutic drugs. *Carcinogenesis*. 2002;23(11):1831-8.

Bonni A, Brunet A, West AE, Datta SR, Takasu MA, Greenberg ME. Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. *Science*. 1999;286(5443):1358-62.

Borgoño CA, Diamandis EP. The emerging roles of human tissue kallikreins in cancer. *Nat Rev Cancer*. 2004;4(11):876-90.

Bosch FX, Lorincz A, Muñoz N, Meijer CJ, Shah KV. The causal relation between human papillomavirus and cervical cancer. *J Clin Pathol*. 2002;55(4):244-65.

Boulet G, Horvath C, Vanden Broeck D, Sahebali S, Bogers J. Human papillomavirus: E6 and E7 oncogenes. *Int J Biochem Cell Biol*. 2007;39(11):2006-11.

Barbera L, Thomas G. Management of early and locally advanced cervical cancer. *Semin Oncol*. 2009;36(2):155-69.

Branca M, Ciotti M, Santini D, Bonito LD, Benedetto A, Giorgi C, et al. Activation of the ERK/MAP kinase pathway in cervical intraepithelial neoplasia is related to grade of the lesion but not to high-risk human papillomavirus, virus clearance, or prognosis in cervical cancer. *Am J Clin Pathol*. 2004;122(6):902-11.

Brasil. Ministério da Saúde. Instituto Nacional de Câncer. INCA. Estimativas 2008: Incidência de câncer no Brasil. Rio de Janeiro: INCA, 2007. Disponível em: <http://www.inca.gov.br/>

Brasil. Ministério da Saúde. Instituto Nacional de Câncer. INCA, 2009. Câncer de esôfago [texto na internet]. Disponível em : http://www.inca.gov.br/conteudo_view.asp?id=328

Brozovic A, Fritz G, Christmann M, Zisowsky J, Jaehde U, Osmak M, et al. Long-term activation of SAPK/JNK, p38 kinase and fas-L expression by cisplatin is attenuated in human carcinoma cells that acquired drug resistance. *Int J Cancer*. 2004;112(6):974-85.

Cane S, Bignotti E, Bellone S, Palmieri M, De las Casas L, Roman JJ, et al. The novel serine protease tumor-associated differentially expressed gene-14 (KLK8/Neuropilin/Ovasin) is highly overexpressed in cervical cancer. *Am J Obstet Gynecol*. 2004;190(1):60-6.

Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol*. 2007;8(1):R3.

Castellsagué X, Sanjose S, Aguado T, Louie K.S, Bruni L, Munoz J, et al. HPV and Cervical Cancer in the World: 2007 Report. *Vaccine*. 2007;25 Suppl 3:C1-C230.

Castellsagué X. Natural history and epidemiology of HPV infection and cervical cancer. *Gynecol Oncol*. 2008;110(3 Suppl 2):S4-7.

Chao A, Wang TH, Lai CH. Overview of microarray analysis of gene expression and its applications to cervical cancer investigation. *Taiwan J Obstet Gynecol*. 2007;46(4):363-73.

Chen TP, Chen CM, Chang HW, Wang JS, Chang WC, Hsu SI, et al. Increased expression of SKP2 and phospho-MAPK/ERK1/2 and decreased expression of p27 during tumor progression of cervical neoplasms. *Gynecol Oncol*. 2007;104(3):516-23.

Chen Y, Miller C, Mosher R, Zhao X, Deeds J, Morrissey M, et al. Identification of cervical cancer markers by cDNA and tissue microarrays. *Cancer Res*. 2003;63(8):1927-35.

Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. *Mol Syst Biol*. 2007;3:140.

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, et al. Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc*. 2007;2(10):2366-82.

D'Anna R, Le Buanec H, Bizzini B, Burny A, Giannouli C, Zagury JF, et al. Human papillomavirus-16-E7 oncoprotein enhances the expression of adhesion molecules in cervical endothelial cells but not in human umbilical vein endothelial cells. *J Hum Virol.* 2001;4(2):85-95.

de Araujo Souza PS, Villa LL. Genetic susceptibility to infection with human papillomavirus and development of cervical cancer in women in Brazil. *Mutat Res.* 2003;544(2-3):375-83.

Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* 2003;4(5):P3.

Doorbar J. Molecular biology of human papillomavirus infection and cervical cancer. *Clin Sci (Lond).* 2006;110(5):525-41.

Emami N, Diamandis EP. Utility of kallikrein-related peptidases (KLKs) as cancer biomarkers. *Clin Chem.* 2008 ;54(10):1600-7.

Engelbrecht AM, Louw L, Cloete F. Comparison of the fatty acid compositions in intraepithelial and infiltrating lesions of the cervix: part II, free fatty acid profiles. *Prostaglandins Leukot Essent Fatty Acids.* 1998;59(4):253-7.

Engelbrecht AM, Gebhardt S, Louw L. Ex vivo study of MAPK profiles correlated with parameters of apoptosis during cervical carcinogenesis. *Cancer Lett.* 2006;235(1):93-9.

Ferrari S, Manfredini R, Tagliafico E, Rossi E, Donelli A, Torelli G, et al. Noncoordinated expression of S6, S11, and S14 ribosomal protein genes in leukemic blast cells. *Cancer Res.* 1990;50(18):5825-8.

Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. *Science.* 1991;251(4995):767-73.

Gaffney DK, Winter K, Dicker AP, Miller B, Eifel PJ, Ryu J, et al. A Phase II study of acute toxicity for Celebrex (celecoxib) and chemoradiation in patients with locally advanced cervical cancer: primary endpoint analysis of RTOG 0128. *Int J Radiat Oncol Biol Phys.* 2007;67(1):104-9.

Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999;286(5439):531-7.

Gordon GJ, Rockwell GN, Jensen RV, Rheinwald JG, Glickman JN, Aronson JP, et al. Identification of novel candidate oncogenes and tumor suppressors in malignant pleural mesothelioma using large-scale transcriptional profiling. *Am J Pathol*. 2005;166(6):1827-40.

Greenawalt DM, Duong C, Smyth GK, Ciavarella ML, Thompson NJ, Tiang T, et al. Gene expression profiling of esophageal cancer: comparative analysis of Barrett's esophagus, adenocarcinoma, and squamous cell carcinoma. *Int J Cancer*. 2007;120(9):1914-21.

Guzman VB, Silva ID, Brenna SM, Carvalho CR, Ribalta JC, Gerbase-Delima M. High levels of granzyme B expression in invasive cervical carcinoma correlates to poor response to treatment. *Cancer Invest*. 2008;26(5):499-503.

Harima Y, Togashi A, Horikoshi K, Imamura M, Sougawa M, Sawada S, et al. Prediction of outcome of advanced cervical cancer to thermoradiotherapy according to expression profiles of 35 genes selected by cDNA microarray analysis. *Int J Radiat Oncol Biol Phys*. 2004;60(1):237-48.

Haslinger C, Schweifer N, Stilgenbauer S, Döhner H, Lichter P, Kraut N, et al. Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. *J Clin Oncol*. 2004;22(19):3937-49.

Hetman M, Gozdz A. Role of extracellular signal regulated kinases 1 and 2 in neuronal survival. *Eur J Biochem*. 2004;271(11):2050-5.

Hildesheim A, Wang SS. Host and viral genetics and risk of cervical cancer: a review. *Virus Res*. 2002;89(2):229-40.

Hippo Y, Taniguchi H, Tsutsumi S, Machida N, Chong JM, Fukayama M, et al. Global gene expression analysis of gastric cancer by oligonucleotide microarrays. *Cancer Res*. 2002;62(1):233-40.

Ho TC, Yang YC, Cheng HC, Wu AC, Chen SL, Chen HK, et al. Activation of mitogen-activated protein kinases is essential for hydrogen peroxide -induced apoptosis in retinal pigment epithelial cells. *Apoptosis*. 2006;11(11):1899-908.

Huang da W, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, et al. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol*. 2007;8(9):R183.

Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44-57.

Ishigami T, Uzawa K, Higo M, Nomura H, Saito K, Kato Y, et al. Genes and molecular pathways related to radioresistance of oral squamous cell carcinoma cells. *Int J Cancer*. 2007;120(10):2262-70.

Ishikawa M, Yoshida K, Yamashita Y, Ota J, Takada S, Kisanuki H, et al. Experimental trial for diagnosis of pancreatic ductal carcinoma based on gene expression profiles of pancreatic ductal cells. *Cancer Sci*. 2005;96(7):387-93.

Kallioniemi O. Medicine: profile of a tumour. *Nature*. 2004;428(6981):379-82.

Kang YH, Lee SJ. Role of p38 MAPK and JNK in enhanced cervical cancer cell killing by the combination of arsenic trioxide and ionizing radiation. *Oncol Rep*. 2008a;20(3):637-43.

Kang YH, Lee SJ. The role of p38 MAPK and JNK in Arsenic trioxide-induced mitochondrial cell death in human cervical cancer cells. *J Cell Physiol*. 2008b;217(1):23-33.

Kanodia S, Fahey LM, Kast WM. Mechanisms used by human papillomaviruses to escape the host immune response. *Curr Cancer Drug Targets*. 2007;7(1):79-89.

Kapetanios V, Lazaris AC, Bogris P, Kouneli S, Nonni A, Arvaniti H, et al. Extracellular regulated kinase-2 immunoreactivity increases in parallel with cervical intraepithelial neoplasia grade in cervical neoplasia. *Int J Gynecol Cancer*. 2008;18(3):540-5.

Khatri P, Voichita C, Kattan K, Ansari N, Khatri A, Georgescu C, et al. Onto-Tools: new additions and improvements in 2006. *Nucleic Acids Res.* 2007;35(Web Server issue):W206-11.

Kihara C, Tsunoda T, Tanaka T, Yamana H, Furukawa Y, Ono K, et al. Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. *Cancer Res.* 2001;61(17):6474-9.

Kim BJ, Ryu SW, Song BJ. JNK- and p38 kinase-mediated phosphorylation of Bax leads to its activation and mitochondrial translocation and to apoptosis of human hepatoma HepG2 cells. *J Biol Chem.* 2006;281(30):21256-65.

Kim SJ, Ju JW, Oh CD, Yoon YM, Song WK, Kim JH, et al. ERK-1/2 and p38 kinase oppositely regulate nitric oxide-induced apoptosis of chondrocytes in association with p53, caspase-3, and differentiation status. *J Biol Chem.* 2002;277(2):1332-9.

Kim YT, Zhao M. Aberrant cell cycle regulation in cervical carcinoma. *Yonsei Med J.* 2005;46(5):597-613.

Kimchi ET, Posner MC, Park JO, Darga TE, Kocherginsky M, Karrison T, et al. Progression of Barrett's metaplasia to adenocarcinoma is associated with the suppression of the transcriptional programs of epidermal differentiation. *Cancer Res.* 2005;65(8):3146-54.

Kitahara O, Katagiri T, Tsunoda T, Harima Y, Nakamura Y. Classification of sensitivity or resistance of cervical cancers to ionizing radiation according to expression profiles of 62 genes selected by cDNA microarray analysis. *Neoplasia.* 2002;4(4):295-303.

Klopp AH, Eifel PJ. Gene expression profiling in cervical cancer: state of the art and future directions. *Cancer J.* 2006;12(3):170-4.

Klopp AH, Jhingran A, Ramdas L, Story MD, Broadus RR, Lu KH, et al. Gene expression changes in cervical squamous cell carcinoma after initiation of chemoradiation and correlation with clinical outcome. *Int J Radiat Oncol Biol Phys.* 2008;71(1):226-36.

Kondoh N, Shuda M, Tanaka K, Wakatsuki T, Hada A, Yamamoto M. Enhanced expression of S8, L12, L23a, L27 and L30 ribosomal protein mRNAs in human hepatocellular carcinoma. *Anticancer Res.* 2001;21(4A):2429-33.

Kordi Tamandani MK, Sobti RC, Shekari M, Mukesh M, Suri V. Expression and polymorphism of IFN-gamma gene in patients with cervical cancer. *Exp Oncol.* 2008;30(3):224-9.

Kuhajda FP. Fatty acid synthase and cancer: new application of an old pathway. *Cancer Res.* 2006;66(12):5977-80.

Lai MD, Xu J. Ribosomal proteins and colorectal cancer. *Curr Genomics.* 2007;8(1):43-9.

Lee JW, Lee SJ, Seo J, Song SY, Ahn G, Park CS, et al. Increased expressions of claudin-1 and claudin-7 during the progression of cervical neoplasia. *Gynecol Oncol.* 2005;97(1):53-9.

Li X, Quigg RJ, Zhou J, Gu W, Nagesh Rao P, Reed EF. Clinical utility of microarrays: current status, existing challenges and future outlook. *Curr Genomics.* 2008;9(7):466-74.

Lin A. Activation of the JNK signaling pathway: breaking the brake on apoptosis. *Bioessays.* 2003;25(1):17-24.

Loging WT, Reisman D. Elevated expression of ribosomal protein genes L37, RPP-1, and S2 in the presence of mutant p53. *Cancer Epidemiol Biomarkers Prev.* 1999;8(11):1011-6.

López-Lázaro M. The warburg effect: why and how do cancer cells activate glycolysis in the presence of oxygen? *Anticancer Agents Med Chem.* 2008;8(3):305-12.

Louw L, Engelbrecht AM, Cloete F. Comparison of the fatty acid compositions in intraepithelial and infiltrating lesions of the cervix: part I, total fatty acid profiles. *Prostaglandins Leukot Essent Fatty Acids.* 1998a;59(4):247-51.

Louw L, Engelbrecht AM, Cloete F. Comparison of the fatty acid compositions in intraepithelial and infiltrating lesions of the cervix: part III, saturated and unsaturated fatty acid profiles. *Prostaglandins Leukot Essent Fatty Acids*. 1998b;59(4):259-64.

Maciag PC, Villa LL. Genetic susceptibility to HPV infection and cervical cancer. *Braz J Med Biol Res*. 1999;32(7):915-22.

Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*. 2005;21(16):3448-9.

Mani A, Gelmann EP. The ubiquitin-proteasome pathway and its role in cancer. *J Clin Oncol*. 2005;23(21):4776-89.

Martin CM, Kehoe L, Spillane CO, O'Leary JJ. Gene discovery in cervical cancer : towards diagnostic and therapeutic biomarkers. *Mol Diagn Ther*. 2007;11(5):277-90.

Matsuura K, Nohno Y, Hijiya N, Uchida T, Tsukamoto Y, Moriyama M. Extracellular signal-regulated protein kinase is activated in cervical intraepithelial neoplasms but inactivated in invasive cervical carcinoma. *Pathol Int*. 2006;56(7):368-74.

Mehta AM, Jordanova ES, Kenter GG, Ferrone S, Fleuren GJ. Association of antigen processing machinery and HLA class I defects with clinicopathological outcome in cervical carcinoma. *Cancer Immunol Immunother*. 2008;57(2):197-206.

Menendez JA, Lupu R. Oncogenic properties of the endogenous fatty acid metabolism: molecular pathology of fatty acid synthase in cancer cells. *Curr Opin Clin Nutr Metab Care*. 2006;9(4):346-57.

Mine KL, Sanson GFO, Morgun A, Silva ID, Brenna SMF, Carvalho CRN, et al. Genome-wide profiling reveals new molecular pathways in cervical cancer [resumo]. [Apresentado no 54º Congresso Brasileiro de Genética; 2008 Set 16-19; Salvador, BA].

Mine KL, Shulzhenko N, Yambartsev A, Sanson GFO, Silva ID, Brenna SMF, et al. Pathways and gene networks analysis reveals molecular events involved in cervical cancer [resumo]. [Apresentado na Molecular Medicine Tri-Conference; 2009 Fev 24-27; San Francisco, EUA]

Miyatake T, Ueda Y, Nakashima R, Yoshino K, Kimura T, Murata T, et al. Down-regulation of insulin-like growth factor binding protein-5 (IGFBP-5): novel marker for cervical carcinogenesis. *Int J Cancer*. 2007;120(10):2068-77.

Monk BJ, Tewari KS, Koh WJ. Multimodality therapy for locally advanced cervical carcinoma: state of the art and future directions. *J Clin Oncol*. 2007;25(20):2952-65.

Morgun A, Shulzhenko N, Perez-Diez A, Diniz RV, Sanson GF, Almeida DR, et al. Molecular profiling improves diagnoses of rejection and infection in transplanted organs. *Circ Res*. 2006;98(12):e74-83.

Muñoz N, Castellsagué X, de González AB, Gissmann L. Chapter 1: HPV in the etiology of human cancer. *Vaccine*. 2006;24 Suppl 3:S3/1-10.

Narisawa-Saito M, Kiyono T. Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: roles of E6 and E7 proteins. *Cancer Sci*. 2007;98(10):1505-11.

Nguyen C, Rocha D, Granjeaud S, Baldit M, Bernard K, Naquet P, et al. Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics*. 1995;29(1):207-16.

Nijhuis ER, Reesink-Peters N, Wisman GB, Nijman HW, van Zanden J, Volders H, et al. An overview of innovative techniques to improve cervical cancer screening. *Cell Oncol*. 2006;28(5-6):233-46.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 1999;27(1):29-34.

Opferman JT, Zambetti GP. Translational research? Ribosome integrity and a new p53 tumor suppressor checkpoint. *Cell Death Differ*. 2006;13(6):898-901.

Paliouras M, Borgono C, Diamandis EP. Human tissue kallikreins: the cancer biomarker family. *Cancer Lett*. 2007;249(1):61-79.

Palla G, Derényi I, Farkas I, Vicsek T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*. 2005;435(7043):814-8.

Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin.* 2005;55(2):74-108.

Perez-Diez A, Morgun A, Shulzhenko N. Microarrays for cancer diagnosis and classification. *Adv Exp Med Biol.* 2007;593:74-85.

Pérez-Plasencia C, Vázquez-Ortiz G, López-Romero R, Piña-Sanchez P, Moreno J, Salcedo M. Genome wide expression analysis in HPV16 Cervical Cancer: identification of altered metabolic pathways. *Infect Agent Cancer.* 2007;2:16.

Piersma SJ, Welters MJ, van der Hulst JM, Kloth JN, Kwappenberg KM, Trimbos BJ, et al. Human papilloma virus specific T cells infiltrating cervical cancer and draining lymph nodes show remarkably frequent use of HLA-DQ and -DP as a restriction element. *Int J Cancer.* 2008;122(3):486-94.

Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R, et al. Genomic signatures to guide the use of chemotherapeutics. *Nat Med.* 2006;12(11):1294-300.

Pyeon D, Newton MA, Lambert PF, den Boon JA, Sengupta S, Marsit CJ, et al. Fundamental differences in cell cycle deregulation in human papillomavirus-positive and human papillomavirus-negative head/neck and cervical cancers. *Cancer Res.* 2007;67(10):4605-19.

Qi M, Elion EA. MAP kinase pathways. *J Cell Sci.* 2005;118(Pt 16):3569-72.

Ritz U, Momburg F, Pilch H, Huber C, Maeurer MJ, Seliger B. Deficient expression of components of the MHC class I antigen processing machinery in human cervical carcinoma. *Int J Oncol.* 2001;19(6):1211-20.

Rojas-Espaillet LA, Rose PG. Management of locally advanced cervical cancer. *Curr Opin Oncol.* 2005;17(5):485-92.

Rose PG. Chemoradiotherapy for cervical cancer. *Eur J Cancer.* 2002;38(2):270-8.

Rubinstein R, Simon I. MILANO--custom annotation of microarray results using automatic literature searches. *BMC Bioinformatics.* 2005;6:12.

Ruffels J, Griffin M, Dickenson JM. Activation of ERK1/2, JNK and PKB by hydrogen peroxide in human SH-SY5Y neuroblastoma cells: role of ERK1/2 in H₂O₂-induced cell death. *Eur J Pharmacol.* 2004;483(2-3):163-73.

Ruggero D, Pandolfi PP. Does the ribosome translate cancer? *Nat Rev Cancer.* 2003;3(3):179-92.

Sankaranarayanan R, Ferlay J. Worldwide burden of gynaecological cancer: the size of the problem. *Best Pract Res Clin Obstet Gynaecol.* 2006;20(2):207-25.

Santin AD, Cane' S, Bellone S, Bignotti E, Palmieri M, De Las Casas LE, et al. The serine protease stratum corneum chymotryptic enzyme (kallikrein 7) is highly overexpressed in squamous cervical cancer cells. *Gynecol Oncol.* 2004;94(2):283-8.

Santin AD, Zhan F, Bignotti E, Siegel ER, Cané S, Bellone S, et al. Gene expression profiles of primary HPV16- and HPV18-infected early stage cervical cancers and normal cervical epithelium: identification of novel candidate molecular markers for cervical cancer diagnosis and therapy. *Virology.* 2005;331(2):269-91.

Savli H, Szendrői A, Romics I, Nagy B. Gene network and canonical pathway analysis in prostate cancer: a microarray study. *Exp Mol Med.* 2008;40(2):176-85.

Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science.* 1995;270(5235):467-70.

Schiffman M, Castle PE, Jeronimo J, Rodriguez AC, Wacholder S. Human papillomavirus and cervical cancer. *Lancet.* 2007;370(9590):890-907.

Scotto L, Narayan G, Nandula SV, Arias-Pulido H, Subramaniam S, Schneider A, et al. Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. *Genes Chromosomes Cancer.* 2008;47(9):755-65.

Shan SJ, Scorilas A, Katsaros D, Diamandis EP. Transcriptional upregulation of human tissue kallikrein 6 in ovarian cancer: clinical and mechanistic aspects. *Br J Cancer.* 2007;96(2):362-72.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498-504.

Shaw RJ. Glucose metabolism and cancer. *Curr Opin Cell Biol.* 2006;18(6):598-608.

Sheu BC, Chang WC, Lin HH, Chow SN, Huang SC. Immune concept of human papillomaviruses and related antigens in local cancer milieu of human cervical neoplasia. *J Obstet Gynaecol Res.* 2007;33(2):103-13.

Shim C, Zhang W, Rhee CH, Lee JH. Profiling of differentially expressed genes in human primary cervical cancer by complementary DNA expression array. *Clin Cancer Res.* 1998;4(12):3045-50.

Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst.* 2003;95(1):14-8.

Simon R, Lam A. BRB-ArrayTools User's Guide Release 3.6. National Cancer Institute. 2007. Disponível na página: <http://linus.nci.nih.gov/brb>.

Slodkowska EA, Ross JS. MammaPrint 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Expert Rev Mol Diagn.* 2009;9(5):417-22.

Song JY, Lee JK, Lee NW, Jung HH, Kim SH, Lee KW. Microarray analysis of normal cervix, carcinoma in situ, and invasive cervical cancer: identification of candidate genes in pathogenesis of invasion in cervical cancer. *Int J Gynecol Cancer.* 2008;18(5):1051-9.

Steinau M, Rajeevan MS, Lee DR, Ruffin MT, Horowitz IR, Flowers LC, et al. Evaluation of RNA markers for early detection of cervical neoplasia in exfoliated cervical cells. *Cancer Epidemiol Biomarkers Prev.* 2007;16(2):295-301.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43):15545-50.

Talantov D, Mazumder A, Yu JX, Briggs T, Jiang Y, Backus J, et al. Novel genes associated with malignant melanoma but not benign melanocytic lesions. *Clin Cancer Res.* 2005;11(20):7234-42.

Tambaro R, Scambia G, Di Maio M, Pisano C, Barletta E, Iaffaioli VR, et al. The role of chemotherapy in locally advanced, metastatic and recurrent cervical cancer. *Crit Rev Oncol Hematol.* 2004;52(1):33-44.

Tian X, Shigemasa K, Hirata E, Gu L, Uebaba Y, Nagai N, et al. Expression of human kallikrein 7 (hK7/SCCE) and its inhibitor antileukoprotease (ALP/SLPI) in uterine endocervical glands and in cervical adenocarcinomas. *Oncol Rep.* 2004;12(5):1001-6.

Tindle RW. Immune evasion in human papillomavirus-associated cervical cancer. *Nat Rev Cancer.* 2002;2(1):59-65.

Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics.* 2001;17(6):520-5.

Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A.* 2001;98(9):5116-21.

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature.* 2002;415(6871):530-6.

Vermeulen CF, Jordanova ES, ter Haar NT, Kolkman-Uljee SM, de Miranda NF, Ferrone S, et al. Expression and genetic analysis of transporter associated with antigen processing in cervical carcinoma. *Gynecol Oncol.* 2007;105(3):593-9.

Wachi S, Yoneda K, Wu R. Interactome-transcriptome analysis reveals the high centrality of genes differentially expressed in lung cancer tissues. *Bioinformatics.* 2005;21(23):4205-8.

Waggoner SE. Cervical cancer. *Lancet.* 2003;361(9376):2217-25.

Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, et al. Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol.* 1999;189(1):12-9.

Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM. High-fidelity mRNA amplification for gene profiling. *Nat Biotechnol.* 2000;18(4):457-9.

Wang SS, Hildesheim A. Chapter 5: Viral and host factors in human papillomavirus persistence and progression. *J Natl Cancer Inst Monogr.* 2003;(31):35-40.

Wang SS, Bratti MC, Rodríguez AC, Herrero R, Burk RD, Porras C, et al. Common variants in immune and DNA repair genes and risk for human papillomavirus persistence and progression to cervical cancer. *J Infect Dis.* 2009;199(1):20-30.

Warburg O. On the origin of cancer cells. *Science.* 1956;123(3191):309-14.

Watanabe T, Komuro Y, Kiyomatsu T, Kanazawa T, Kazama Y, Tanaka J, et al. Prediction of sensitivity of rectal cancer cells in response to preoperative radiotherapy by DNA microarray analysis of gene expression profiles. *Cancer Res.* 2006;66(7):3370-4.

Weidhaas JB, Li SX, Winter K, Ryu J, Jhingran A, Miller B, et al. Changes in gene expression predicting local control in cervical cancer: results from Radiation Therapy Oncology Group 0128. *Clin Cancer Res.* 2009;15(12):4199-206.

Werner T. Bioinformatics applications for pathway analysis of microarray data. *Curr Opin Biotechnol.* 2008;19(1):50-4.

Wilting SM, de Wilde J, Meijer CJ, Berkhof J, Yi Y, van Wieringen WN, et al. Integrated genomic and transcriptional profiling identifies chromosomal loci with altered gene expression in cervical cancer. *Genes Chromosomes Cancer.* 2008;47(10):890-905.

Wong DJ, Nuyten DS, Regev A, Lin M, Adler AS, Segal E, et al. Revealing targeted therapy for human cancer by gene module maps. *Cancer Res.* 2008;68(2):369-78.

Wong YF, Selvanayagam ZE, Wei N, Porter J, Vittal R, Hu R, et al. Expression genomics of cervical cancer: molecular classification and prediction of radiotherapy response by DNA microarray. *Clin Cancer Res.* 2003;9(15):5486-92.

Wong YF, Cheung TH, Tsao GS, Lo KW, Yim SF, Wang VW, et al. Genome-wide gene expression profiling of cervical cancer in Hong Kong women by oligonucleotide microarray. *Int J Cancer.* 2006a;118(10):2461-9.

Wong YF, Sahota DS, Cheung TH, Lo KW, Yim SF, Chung TK, et al. Gene expression pattern associated with radiotherapy sensitivity in cervical cancer. *Cancer J*. 2006b;12(3):189-93.

Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. *Bioinformatics*. 2003;19(18):2448-55.

Xia Z, Dickens M, Raingeaud J, Davis RJ, Greenberg ME. Opposing effects of ERK and JNK-p38 MAP kinases on apoptosis. *Science*. 1995;270(5240):1326-31.

Yu KJ, Rader JS, Borecki I, Zhang Z, Hildesheim A. CD83 polymorphisms and cervical cancer risk. *Gynecol Oncol*. 2009;114(2):319-22.

Zhai Y, Kuick R, Nan B, Ota I, Weiss SJ, Trimble CL, et al. Gene expression analysis of preinvasive and invasive cervical squamous cell carcinomas identifies HOXC10 as a key mediator of invasion. *Cancer Res*. 2007;67(21):10163-72.

Zhang Y, Bhat I, Zeng M, Jayal G, Wazer DE, Band H, et al. Human kallikrein 10, a predictive marker for breast cancer. *Biol Chem*. 2006;387(6):715-21.

Zhao Y, Simon R. BRB-ArrayTools Data Archive for Human Cancer Gene Expression: A Unique and Efficient Data Sharing Resource. *Cancer Inform*. 2008;6:9-15.

zur Hausen H. Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer*. 2002;2(5):342-50.

Abstract

Cervical cancer is the second most frequent cancer in women worldwide. Approximately 80% of cases occur in developing countries and the majority of cases are diagnosed in advanced stages. The **goals** of this study were: (i) to identify a gene expression profile to predict the treatment response, since about 35% of patients with locally advanced cervical cancer will not respond to treatment; (ii) to identify genes and gene groups, pathways and gene networks, with different expression between cervical tumor and normal adjacent tissue, in order to identify genes related to cervical cancer pathogenesis and to better understand the molecular mechanisms involved in this cancer. **Methods:** To perform the genome-wide gene expression analysis we used microarrays containing oligonucleotides corresponding to ~14,000 genes. We used 23 slides hybridized with tumor samples for the treatment response analysis and for the comparison analysis between tumor and normal tissues we used 34 slides with tumor samples and 20 with normal samples. For data analysis we used the following programs: BRB Array Tools to search the molecular predictor and to obtain gene list; DAVID, Babelomics e Ingenuity Pathway Analysis to identify the over-represented pathways in the gene list; MILANO to verify how our findings relate to the published literature; DAVID, CFinder e Cytoscape for gene networks analysis. **Results:** (i) we did not identify a gene expression profile that could predict the treatment response; (ii) we found 810 differentially expressed genes between tumor and normal tissues, 341 were up-regulated and 469 were down-regulated in tumor samples. We identify 13 over-represented pathways and among them, we found previously known (e.g. 'Cell Cycle' and 'p53 signaling') and unknown molecular pathways ('Oxidative Phosphorylation' and 'Ribosome') in cervical cancer. Several pathway genes, even in the previously known pathways related to cervical cancer, have not been studied in this cancer. In the gene network analysis, we found 23 subnetworks. Among them we highlighted the subnetwork that contains genes from the kallikrein family. Our results also suggest that our gene profile can also be applied for cervical cancer cases from the other studies and for esophagus cancer. **Conclusions:** Using genome-wide transcriptome analysis we identified genes and gene groups, pathways and gene networks, involved in cervical cancer. These results might help to understand some molecular mechanisms involved in the cervical cancer pathogenesis and the genes might be interesting candidates to further susceptibility and biomarker studies in cervical cancer.

Bibliografia consultada

Rother ET, Braga MER. Como elaborar sua tese: estrutura e referências. 2a ed. rev. e ampl. São Paulo; 2005.