ALTERAÇÕES GENÔMICAS EM PACIENTES E SEUS FAMILIARES COM A SÍNDROME DO CÂNCER COLORRETAL HEREDITÁRIO

ROLANDO ANDRÉ RIOS VILLACIS

Tese apresentada à Fundação Antônio Prudente para a obtenção do Título de Doutor em Ciências

Área de concentração: Oncologia

Orientadora: Profa. Dra. Silvia Regina Rogatto

São Paulo

2015

FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Villacis Rolando André Rios

Alterações genômicas em pacientes e seus familiares com a síndrome do câncer colorretal hereditário/ Rolando André Rios Villacis - São Paulo, 2015.
91p.

Tese (Doutorado)-Fundação Antônio Prudente.
Curso de Pós-Graduação em Ciências - Área de concentração: Oncologia.
Orientadora: Silvia Regina Rogatto.

Descritores: 1. NEOPLASIAS COLORRETAIS/genética. 2. SÍNDROME DE LYNCH. 3. ANÁLISE DE MICROARRANJOS. 4. POLIMORFISMO DE NUCLEOTÍDEO ÚNICO.

"Muitas coisas não ousamos empreender por parecerem difíceis; entretanto, são difíceis porque não ousamos empreende-las."

Sêneca

DEDICATÓRIA

Aos meus pais, Rolando Rios e Silvia Villacis, pelo amor e apoio incondicional em todos os momentos da minha vida. Não há palavras para descrever o quão importante vocês são para mim.

Aos meus irmãos, Danielli e Angel Alfredo, que apesar da distância sempre se preocuparam comigo. Obrigado pelas conversas e risadas pelo Skype e por tornarem cada fim de ano inesquecível.

À minha noiva, Melina Kumoto, por todo seu amor, paciência e compreensão ao longo destes últimos oito anos maravilhosos. Sem você eu não teria chegado ao fim desta jornada. Espero que sejamos muito felizes nesta nova etapa que estamos iniciando juntos.

Aos pais (Sérgio e Christiane), irmã (Monike), tias (Alessandra e Katen) e avós (Orlando e Ivani) da minha noiva; que eu também considero a minha família. Sou muito grato por todo amor e carinho que me deram e dão desde que nos conhecemos.

AGRADECIMENTOS

À minha orientadora, Sílvia Regina Rogatto, por todo o apoio, paciência e ensinamentos nestes últimos oito anos. Muitas vezes foi como uma mãe para mim, me aconselhando para tomar as melhores decisões. Termino uma etapa da minha vida no seu lab, mas sempre lhe serei grato por tudo que me ensinou, seja no âmbito da ciência ou da vida em si.

Aos meus colegas do Lab Neogene por todos os últimos oito anos de convivência. Além de colegas de trabalho, também são meus amigos. Em especial queria agradecer ao Fábio Marchi, Mateus Camargo, Juan Muñoz e Prisicla Mayrink pelos papos descontraídos, risadas, conselhos e pela grande ajuda na finalização da tese aqui apresentada. Sentirei saudades de todos vocês.

Ao meu amigo de longa data, Luis Bicalho, com o qual dividi o mesmo apartamento em todos estes anos em São Paulo. Obrigado pela ótima convivência que tivemos e por toda a tua ajuda, seja ela de caráter emocional ou financeira.

Aos pesquisadores (Dr. Benedito Rossi, Dra. Érika Monteiro e Dra. Maria Isabel Achatz) e enfermeiras (Amanda França e Bruna Katyn) que de uma forma ou outra tornaram possível a realização deste projeto.
Às funcionárias do Banco de Tumores do A. C. Camargo Cancer Center, em especial Eloísa R. Olivieri e Louise Motta, que sempre estiveram dispostas a me ajudar com a maior rapidez possível.

Aos funcionários do SAME do A. C. Camargo Cancer Center; Luciano, Paulo, Odirlei, André e Alexandre; que antes do advento dos prontuários eletrônicos me ajudaram enormemente na coleta de dados clínicos.

Às funcionárias da pós-graduação e do CEP do A. C. Camargo Cancer Center; Vanuza, Luciana e Rúbia; que sempre estiveram dispostas a me ajudar na resolução

de qualquer problema de ordem burocrática.

À FAPESP pela bolsa de doutorado concedida e ao Instituto Nacional de Ciência e Tecnologia em Oncogenômica – INCiTO (FAPESP e CNPq) pelo suporte financeiro para a realização da minha tese.

RESUMO

Villacis RAR. **Alterações genômicas em pacientes e seus familiares com a Síndrome do câncer colorretal hereditário.** São Paulo; 2015. [Tese de Doutorado-Fundação Antônio Prudente]

O câncer colorretal (CCR) é a terceira neoplasia mais comum no mundo. Estima-se que 35% dos CCRs são hereditários, no entanto, apenas 5% destes casos são explicados por mutações patogênicas em genes de alta penetrância, A Síndrome de Lynch (SL) é a doença hereditária mais comum associada com o risco de CCR estando associada com mutações germinativas nos genes de reparo a erros de pareamento (*Mismatch repair genes* - MMR). Neste estudo foram avaliadas as alterações genômicas em 54 pacientes com a SL (critérios clínicos de Amsterdam ou Bethesda) e negativos para mutações patogênicas nos genes MMR (*MLH1*, *MSH2*, *MSH6* e *PMS2*), *TP53* ou *CHEK2* (100delC); com o objetivo de identificar novos genes que poderiam estar associados com a predisposição ao CCR. As variações no número de cópias (*Copy number variations* - CNVs) foram avaliadas em todos os casos utilizando a plataforma de microarranjos (*microarray*) Agilent 4X180K, enquanto 26 casos também foram reanalisados com a plataforma Affymetrix CytoScan HD. Em 21 casos foram identificadas apenas CNVs comuns, enquanto 33 pacientes apresentaram 58 CNVs raras. Destas, 43 CNVs raras cobriram genes e foram detectadas em 28 pacientes. Nos casos avaliados com as ambas as plataformas foi possível validar 9 CNVs raras. A análise *in silico* revelou que 52 dos 81 genes afetados por CNVs raras foram associados com câncer, dos quais 26 estavam relacionados com o CCR. Entre estas alterações foi detectada uma deleção intrônica no gene *ROBO1* em uma paciente jovem com CCR e sem história de câncer na família. Coincidentemente, deleções idênticas no mesmo gene também foram detectadas em duas pacientes não relacionadas com história familiar de câncer de mama, fornecendo assim evidência adicional do potencial patogénico dessa CNV. Foram também identificados ganhos (*GALNT6* e *GALNT11*) e perdas (*CENPP* e *SEMA3C*) em genes pertencentes às mesmas famílias de genes recentemente

associados com o risco de desenvolvimento ao CCR em famílias negativas para mutação nos genes de predisposição conhecidos. Um estudo de segregação de algumas CNVs raras foi realizado em 14 membros de três famílias distintas. Os resultados sugeriram o envolvimento de *CHODL*, *GPR39*, *GALNT11* e *KMT2C* na predisposição ao CCR. Além das CNVs, também foram avaliados 23 SNPs (*Single nucleotide polymorphisms*) descritos como associados ao risco de desenvolvimento do CCR. Dez SNPs foram genotipados em 47 casos utilizando a metodologia TaqMan (Applied Biosystems), enquanto 13 loci adicionais foram genotipados em 26 casos com a plataforma CytoScan HD. A frequência dos alelos de risco foi superior a 50% em 11 SNPs genotipados. Considerando os 24 pacientes com 23 SNPs genotipados, foi observada uma média de 21 alelos de risco (de um total de 46). Cinco dos 10 SNPs genotipados pelo método TaqMan (rs961253, rs3802842, rs4444235, rs4779584, rs4939827, rs6983267, rs9929218, rs10411210, rs10795668 e rs16892766), previamente associados com o risco familiar de CCR, foram detectados com frequência superior a 50% na nossa casuística. Quando considerados os 44 casos genotipados para todos os 10 SNPs foi observada uma média de 10 alelos de risco (6 a 14 por caso). Nenhuma das CNVs raras detectadas se sobrepôs as regiões contendo os SNPs ou afetou genes associados com estes SNPs. Tomados em conjunto, os resultados sugerem que CNVs raras e SNPs de risco podem contribuir para um aumento no risco de desenvolvimento do CCR. A identificação de novos genes de predisposição ao CCR implicará em estratégias mais eficientes de diagnóstico e aconselhamento genético em famílias com um alto risco de desenvolver neoplasias.

SUMMARY

Villacis RAR. [**Genomic alterations in patients and their relatives with unexplained familial or early-onset colorectal cancer]**. São Paulo; 2015. [Tese de Doutorado-Fundação Antônio Prudente]

Colorectal Cancer (CRC) is the third most common cancer worldwide. It is estimated that 35% of CRCs are hereditary. However, only 5% of these cases are explained by pathogenic mutations in high-penetrance genes. Lynch Syndrome (LS) is the most common hereditary disease related to CRC risk being associated with germline mutations in the mismatch repair genes (MMR). In this study, we evaluated genomic alterations in 54 LS patients (Amsterdam or Bethesda clinical criteria) negative for pathogenic mutations in MMR genes (*MLH1*, *MSH2*, *MSH6* and *PMS2*), *TP53* or *CHEK2* (100delC), aiming to identify new genes that might be associated with CRC predisposition. Copy number variations (CNVs) were assessed in all cases using Agilent 4x180K microarray platform, while 26 cases were also reanalyzed by Affymetrix CytoScan HD platform. Twenty-one cases presented only common CNVs, while 33 patients presented 58 rare CNVs. From these, 43 rare CNVs covered genes and were detected in 28 patients. Nine rare CNVs were validated in patients evaluated by both platforms. In silico analysis revealed that 52 of the 81 genes affected by rare CNVs have been associated with cancer, of these 26 were related to CRC. Among these alterations, an intronic deletion in *ROBO1* gene was detected in a young patient with CRC with no family history of cancer. Coincidentally, identical deletion was found in two unrelated patients with family history of breast cancer, thus providing further evidence of the pathogenic effect of this CNV. Gains (*GALNT6* and *GALNT11*) and losses (*CENPP* and *SEMA3C*) were also found in our cases affected genes belonging to the same gene families recently associated with CRC developing risk in families negative for mutations in known predisposing genes. A segregation analysis was performed in 14 relatives from three different families. Our results suggested the involvement of *CHODL*, *GPR39*, *GALNT11* and *KMT2C* in CRC predisposition. In addition to CNVs, we also evaluated 23 single

nucleotide polymorphisms (SNPs) associated with CRC risk. Ten SNPs were genotyped in 47 cases using the TaqMan methodology (Applied Biosystems), while 13 additional loci were genotyped in 26 cases by CytoScan HD platform. Risk alleles frequency was higher than 50% in 11 genotyped SNPs. Considering the 24 patients with 23 genotyped SNPs, an average of 21 risk alleles was observed (from a total of 46). Five of 10 SNPs genotyped by TaqMan method (rs961253, rs3802842, rs4444235, rs4779584, rs4939827, rs6983267, rs9929218, rs10411210, rs10795668 and rs16892766), previously related to familial CRC risk, were detected with a frequency higher than 50% in our samples. When considering the 44 cases genotyped for all 10 SNPs, was observed an average of 10 risk alleles (6 to 14 per case). None of detected rare CNVs overlapped regions containing SNPs or genes associated with these SNPs. Taken together, the results suggest that rare CNVs and risk SNPs may contribute to an increased risk of developing CRC. The identification of new CRC predisposing genes implies in more efficient strategies of diagnosis and genetic counseling of families with high risk of developing neoplasms.

LISTA DE SIGLAS E ABREVIATURAS

aCGH*	*Array comparative genomic hybridization*
CCR	Câncer colorretal
CNV*	*Copy number variation*
DGV*	*Database of Genomic Variants*
FAP*	*Familial adenomatous polyposis*
FCCTX*	*Familial Colorectal Cancer Type X*
GWAS*	*Genome Wide Association Studies*
HMPS*	*Hereditary mixed polyposis syndrome*
IHC*	*Immunohistochemistry*
INCA	Instituto Nacional de Câncer
JPS*	*Juvenile polyposis syndrome*
MAP*	*MUTHY-associated polyposis*
MMR*	*Mismatch repair*
MSI*	*Microsatellite instability*
MSI-H*	*High microsatellite instability*
MSS*	*Microsatellite stable*
NAHR*	*Nonallelic homologous recombinant*
NCI*	*National Cancer Institute*
NHEJ*	*Nonhomologous end joining*
PHTS*	*PTEN hamartoma tumor syndrome*
PJS*	*Peutz-Jeghers syndrome*
SL	Síndrome de Lynch
SNP*	*Single nucleotide polymorphism*

*abreviaturas derivadas do inglês

ÍNDICE

1 INTRODUÇÃO

O câncer colorretal (CCR) é a terceira neoplasia mais prevalente no mundo com aproximadamente 1,4 milhões de novos casos registrados em 2012 (TORRE et al. 2015). Sua incidência é maior nos países desenvolvidos, onde representa a segunda principal causa de mortalidade por câncer (TORRE et al. 2015). No entanto, estudos recentes mostraram um aumento tanto nas taxas de incidência como de mortalidade em países em desenvolvimento (América do Sul e Europa) (BOSETTI et al. 2013; CHATENOUD et al. 2014). No Brasil, o Instituto Nacional de Câncer-INCA estimou 32.600 novos casos de CCR para os anos de 2014/2015, sendo 15.070 em homens e 17.530 em mulheres (Ministério da Saúde 2014). Excluindo-se os tumores de pele não melanoma, o CCR representa a segunda neoplasia mais frequente na região Sudeste considerando ambos os sexos (Ministério da Saúde 2014).

Embora a maioria dos casos de CCR seja de natureza esporádica (~70%); decorrente da interação entre fatores ambientais e mutações somáticas ao longo do tempo, um grande estudo com gêmeos monozigóticos e dizigóticos estimou que aproximadamente 35% dos tumores colorretais apresentam uma predisposição genética associada ao seu desenvolvimento (LICHTENSTEIN et al. 2000; GOODENBERGER e LINDOR 2011). Recentemente, foi relatado que mais de um terço dos casos de tumores colorretais de pacientes com idade igual ou inferior a 35 anos está associado com síndromes hereditárias de câncer (MORK et al. 2015). Mutações germinativas em genes de alta penetrância explicam apenas 5% dos casos

de CCR suspeitos de terem um componente hereditário, ou seja, aqueles que se desenvolvem em famílias com história prévia de CCR e/ou em pacientes jovens (JASPERSON et al. 2010; SAMADDER et al. 2015). A herdabilidade dos demais tumores colorretais (30%) ainda é desconhecida, podendo ser causada por variantes raras ou pelo efeito combinado de alelos comuns de baixa penetrância (VALLE 2014; WHIFFIN e HOULSTON 2014).

Atualmente, ao menos sete síndromes associadas ao CCR foram identificadas e apresentam mutações germinativas em genes específicos, incluindo a Síndrome de Lynch (SL), a polipose adenomatosa familiar (FAP - *Familial adenomatous polyposis*), a polipose associada ao gene *MUTYH* (MAP - *MUTHY-associated polyposis*), a Síndrome de polipose hereditária mista (HMPS - *Hereditary mixed polyposis syndrome)*, a Síndrome de Peutz-Jeghers (PJS - *Peutz-Jeghers syndrome*), Síndrome de polipose juvenil (JPS - *Juvenile polyposis syndrome*) e a Síndrome do tumor *PTEN* - hamartoma (PHTS - *PTEN hamartoma tumor syndrome*) (VALLE 2014; SAMADDER et al. 2015). O entendimento da biologia molecular associada à herdabilidade do CCR propiciará uma melhor identificação de pacientes em risco e, por conseguinte melhores estratégias de seguimento e diagnóstico (SAMADDER et al. 2015).

1.1 SÍNDROME DE LYNCH

O primeiro relato de uma família suspeita de ter a SL (“Família G”), numa época que tumores hereditários eram praticamente desconhecidos, foi realizado por WHARTIN (1913). O autor estudou numerosos indivíduos com tumores colorretais,

câncer gástrico e neoplasias endometriais. Posteriormente, LYNCH et al. (1966) estudaram duas outras famílias ("Famílias N e M") com vários indivíduos com tumores colorretais em idade jovem, assim como mulheres com carcinomas de endométrio e ovário. A partir destes dois estudos pioneiros foi hipotetizada a existência de uma nova síndrome hereditária de câncer associada ao desenvolvimento de CCR, a qual mais tarde seria conhecida como SL (LYNCH et al. 2009; COLAS et al. 2012). A etiologia molecular associada com a SL começou a ser desvendada em 1993, com dois estudos que detectaram loci de suscetibilidade nos cromossomos 2p e 3p por meio de análises genéticas de ligação (LINDBLOM et al. 1993; PELTOMÄKI et al. 1993). Mais tarde foram identificadas mutações nos genes *MLH1* (3p21.3) e *MSH2* (2p21) como principais envolvidos na predisposição à SL (LYNCH et al. 2009)

A SL é a doença hereditária mais comum associada com o CCR. É herdada de forma autossômica dominante e causada por mutações germinativas em quatro genes do sistema de reparo a erros de pareamento no DNA (MMR - *Mismatch Repair Genes*): *MLH1*, *MSH2*, *MSH6* (2p16) e *PMS2* (7p22) (VALLE 2014). Recentemente foi confirmado o envolvimento de um quinto gene como associado com a SL: *EPCAM/TACSTD1* (2p21) (KOVACS et al. 2009; LIGTENBERG et al. 2009).

Aproximadamente 70% dos pacientes com a SL apresentam mutações nos genes *MLH1* e *MSH2*, enquanto mutações em *MSH6* e *PMS2* são identificadas em 14 e 15% dos pacientes com a SL, respectivamente (PELTOMÄKI e VASEN et al. 2004; PALOMAKI et al. 2009). Mais recentemente foi estimado que deleções no *EPCAM* acometem 20% dos casos de SL sem mutação nos genes MMR (TUTLEWSKA et al. 2013). Em indivíduos com a SL, uma cópia do gene é

constitutivamente inativada em todas as células, enquanto o outro alelo é inativado no tumor por mutações somáticas, perda de heterozigose ou silenciamento epigenético (VALLE 2014). A maioria das mutações germinativas que acomete estes genes é do tipo em ponto ou pequenas deleções/inserções. Contudo, grandes rearranjos genômicos representam aproximadamente 30% do total de mutações identificadas (BOZZAO et al. 2011; PÉREZ-CARBONERO et al. 2011; COLAS et al. 2012).

Além dos cinco genes associados com a Síndrome de Lynch, há pelo menos outros cinco genes relacionados ao MMR: *PMS1*, *MSH3*, *MLH3*, *MSH4* e *MSH5* (JIRICNY 2013). Apesar de estes genes não terem sido confirmados como associados à SL, mutações nos genes *PMS1*, *MSH3* e *MLH3* já foram detectadas em famílias com CCR, dessa forma não é possível exlcuir completamente o papel destes genes na tumorigênese colorretal (NICOLAIDES et al. 1994; WANG et al. 1999; HIENONEN et al. 2003; DURATURO et al. 2011). Adicionalmente, outros estudos também relataram o envolvimento dos genes *EXO1*, *RPA*, *RFC* e *PCNA* no sistema MMR (MASTROCOLA e HEINEN et al. 2010; SMITH et al. 2013).

Os genes MMR codificam diversas proteínas nucleares que desempenham um papel importante na manutenção da estabilidade genética e fidelidade do DNA duplicado, uma vez que corrigem os erros que ocorrem durante a fase S da duplicação celular (HASSEN et al. 2011). As proteínas MMR funcionam em heterodímeros, sendo os complexos MLH1/PMS2 e MSH2/MSH6 os mais importantes (COLAS et al. 2012). Os heterodímeros formados pela proteína MSH2 reconhecem os erros de pareamento de bases (MSH2/MSH6) e os *loops* de inserções/deleções (MSH2/MSH3), enquanto os complexos formados pela proteína

MLH1 com PMS1 e PMS2 são responsáveis por ativar endonucleases, incluindo EXO1, as quais removem os segmentos de DNA incorretos (GEORGE e ALANI 2012; MJELLE et al. 2015). Por outro lado, o heterodímero MSH4/MSH5; cuja expressão está limitada aos ovários e testículos; não está associado com o a correção de erros de pareamento e sim com a estabilização das junções de *Holliday* durante a meiose, permitindo a recombinação gênica (SNOWDEN et al. 2004). O gene *EPCAM* codifica uma proteína de membrana implicada em diversas funções celulares como adesão, migração, proliferação e diferenciação (TUTLEWSKA et al. 2013). Deleções nos últimos éxons do *EPCAM* causam a hipermetilação do promotor do gene *MSH2* e consequentemente seu silenciamento, levando assim à deficiência no sistema de reparo MMR (KUIPER et al. 2011; TUTLEWSKA et al. 2013). Os genes *MSH2* e *EPCAM* estão mapeados no cromossomo 2 a uma distância de apenas 17Kb (KEMPERS et al. 2011).

A instabilidade de microssatélites (MSI - *Microsatellite instability*), representada por inserções ou deleções em sequências curtas e repetidas de DNA, é a manifestação mais extrema do aumento da taxa de mutações nos genes MMR e é considerado um excelente marcador para identificar pacientes suspeitos de apresentar a SL (AALTONEN et al. 1994; BOLAND et al. 1998). O NCI (do inglês, *National Cancer Institute*) recomenda a análise de um painel composto por cinco marcadores de microssatélites (painel Bethesda) para detectar MSI, envolvendo repetições de dois mononucleotídeos (BAT25 e BAT26) e três dinucleotídeos (D2S123, D5S346 e D17S250) (BOLAND et al. 1998). Um tumor apresenta alta MSI (MSI-H) quando dois ou mais marcadores do painel Bethesda mostrar instabilidade, enquanto a estabilidade de microssatélites (MSS - *Microsatellite stable*) é definida quando

houver ausência de instabilidade para todos os marcadores testados (UMAR et al. 2004).

Aproximadamente 90% dos CCRs associados à SL apresentam MSI-H, tornando esta análise muito sensível, embora sua especificidade seja baixa, uma vez que 15% dos CCRs esporádicos também apresentam MSI-H. Neste caso, a presença de MSI-H em CCRs esporádicos é decorrente da metilação somática bialélica da região promotora do gene *MLH1* e acomete pessoas mais velhas, sem história familiar de CCR (JASPERSON et al. 2010; BALAGUER et al. 2011). Os casos de CCR esporádicos com MSI-H também estão associados com mutações no códon V600E do gene *BRAF* e estas características (metilação do *MLH1* e/ou mutação no *BRAF*) podem ser utilizadas para diferenciar carcinomas decorrentes da SL das neoplasias esporádicas (PARSONS et al. 2012). NEWTON et al. (2014) avaliaram por pirosequenciamento tanto a metilação do *MLH1* como a mutação V600E no gene *BRAF* em 71 tumores de pacientes com mutações germinativas em *MLH1* e 73 carcinomas esporádicos com perda da metilação em *MLH1*. Os autores relataram que a análise de metilação do gene *MLH1* foi uma metodologia que revelou ter maior sensibilidade (95%) quando comparada com a de mutação do gene *BRAF* (66%) na identificação de pacientes com a SL.

A SL é responsável por 2% a 4% de todos os tumores colorretais (HAMPEL et al. 2008). Pacientes com a SL apresentam um risco estimado de 50-80% de desenvolver CCR ao longo da vida (STOFFEL et al. 2009). Mulheres portadoras de mutações nos genes do sistema MMR também apresentam um risco elevado de desenvolver câncer de endométrio (40-60%) (QUEHENBERGER et al. 2005; STOFFEL et al. 2009). Mutações nos genes *MSH6* e *PMS*2 predispõem a um risco

substancialmente menor de desenvolver tanto CCR como câncer endometrial (PLASCHKE et al. 2004; SENTER et al. 2008; BAGLIETTO et al. 2010). Outros tumores associados com a SL incluem intestino delgado, pâncreas, estômago, ovário, cérebro, trato urinário e do trato hepatobiliar (SAMADDER et al. 2015). O risco de câncer de mama para pacientes com a SL ainda é incerto, porém, estudos brasileiros mostraram esta associação (OLIVEIRA FERREIRA et al. 2004; DA SILVA et al. 2010). Adicionalmente, um estudo australiano verificou que 51% (18/35) dos tumores de mama detectados em pacientes com mutações conhecidas nos genes MMR também apresentaram alterações quando analisados por IHQ e MSI, sugerindo assim que neoplasias malignas de mama também fazem parte do espectro de tumores da SL (WALSH et al. 2010). Mais recentemente, foi sugerido que mulheres com mutações em *MLH1* apresentam um risco moderado de desenvolver câncer de mama (HARKNESS et al. 2015).

1.2 DIAGNÓSTICO CLÍNICO E MOLECULAR DA SL

Tradicionalmente, os métodos de triagem clínica utilizados para identificar as famílias com SL incluem os critérios de Amsterdam I e Amsterdam II (BARTLEY et al. 2012). Os Critérios de Amsterdam I estabelecem que as famílias devem apresentar três membros com CCR, dos quais um deve ser parente de primeiro grau dos outros dois, diagnóstico de CCR em pelo menos duas gerações sucessivas, exclusão diagnóstica de FAP e um membro da família afetado com menos de 50 anos de idade (VASEN et al. 1991). Posteriormente, estes critérios foram revisados para

incluir os tumores extra-colônicos no fenótipo da SL, dando origem aos critérios de Amsterdam II (VASEN et al. 1999).

Para melhor caracterizar e identificar os indivíduos suspeitos de ter a SL foi proposto o critério de Bethesda (RODRIGUEZ-BIGAS et al. 1997; UMAR et al. 2004), o qual apresenta características menos rigorosas de enquadramento de famílias na SL (menos rigorosos). Estas diretrizes estabelecem que os indivíduos devem ser testados para MSI nas seguintes situações: (1) CCR diagnosticado com idade inferior a 50 anos; (2) presença de CCR sincrônico ou metacrônico ou outros tumores associados com a SL, independente da idade; (3) CCR com característica histológica de MSI-H, diagnosticado em idade inferior a 60 anos; (4) CCR diagnosticado em paciente com um ou mais parentes de primeiro grau com tumores associados com a SL, um dos quais em idade inferior a 50 anos; (5) CCR diagnosticado em paciente com dois ou mais parentes de primeiro ou segundo grau com tumores associados com a SL, independente da idade (UMAR et al. 2004). Entretanto, utilizando os critérios de Amsterdam e Bethesda nem sempre é possível identificar um inúmeros casos de SL.

Os critérios de Amsterdam apresentam uma alta especificidade, porém, uma baixa sensibilidade. Por outro lado, os critérios de Bethesda são mais sensíveis, mas também falham em identificar um grande número de portadores de mutações nos genes MMR (BARTLEY et al. 2012; LASTELLA et al. 2011; PÉREZ-CARBONELL et al. 2012). Em adição, aproximadamente 25% dos indivíduos com a SL não preenchem os critérios de Amsterdam ou de Bethesda (HAMPEL et al. 2008). As abordagens baseadas exclusivamente em critérios familiares subestimam a

verdadeira prevalência de pacientes com SL (HAMPEL et al. 2008; LASTELLA et al. 2011).

Nos últimos anos, tornou-se evidente que um número considerável de famílias categorizadas pelos Critérios de Amsterdam (~50%) não apresentavam mutações germinativas patogênicas nos genes MMR e seus tumores eram microssatélite estáveis (MSS), ou seja, não apresentaram instabilidade para nenhum dos marcadores de microssatélite utilizados, 10 deles no estudo de LINDOR et al. (2005) e cinco (Painel Bethesda) no estudo de LLOR et al. (2005). Estes casos têm sido referidos como "CCR familial do tipo X" (*Familial Colorectal Cancer Type X* - FCCTX) devido à sua etiologia desconhecida (JASPERSON et al. 2010; DOMINGUEZ-VALENTIN et al. 2015). Comparado com a SL, famílias com o FCCTX têm um menor risco de desenvolver câncer, os tumores aparecem em idade mais avançada, há predomínio de CCR no lado esquerdo do cólon, os tumores colorretais são prevalentemente mucinosos, ocorrem poucos casos de múltiplos tumores e as neoplasias extra-colônicos são raras (BOZZAO et al. 2011; COLAS et al. 2012; KLARSKOV et al. 2012; SÁNCHEZ-TOMÉ et al. 2015). As famílias com FCCTX constituem um grupo bastante heterogêneo, incluindo casos que se ajustam melhor ao modelo monogênico ou poligênico de doenças (VALLE 2014). Foi sugerido que uma grande parcela da agregação familiar observada nesses casos é devido à interação com fatores ambientais. Contudo, diversos genes candidatos foram associados com o desenvolvimento de CCR nestas famílias e incluem entre outros *GALNT12*, *CDH18*, *GREM1*, *RPS20* e *FAN1* (VENKATACHALAM et al. 2011; CLARKE et al. 2012; NIEMINEN et al. 2014; VALLE 2014; SEGUÍ et al. 2015). Um estudo recente sugeriu que mutações patogênicas no gene de alta

penetrância *BRCA2* também poderiam explicar a presença da doença em algumas famílias (GARRE et al. 2015).

Atualmente, os testes genéticos disponíveis para identificar pacientes com a SL envolvem uma combinação de sequenciamento completo do DNA e análise de rearranjos gênicos a fim de rastrear e identificar mutações em ponto, deleções, duplicações e inserções envolvendo os genes *MLH1*, *MSH2*, *MSH6*, *PMS2* e *EPCAM* (WEISSMAN et al. 2012). O número de famílias diagnosticadas com a SL foi aumentando ao longo do tempo à medida que foram desenvolvidos novos procedimentos e técnicas diagnósticas (COLAS et al. 2012). Estudos mais recentes também preconizam a avaliação de mutações nos genes *POLE*/*POLD1* na ausência de mutações nos genes MMR (BELLIDO et al. 2015; ELSAYED et al. 2015).

Em muitas instituições é pratica padrão utilizar testes de imunoistoquímica (IHC - *Immunohistochemistry*) e/ou MSI em amostras de tumores colorretais como um primeiro passo na identificação de pacientes suspeitos de ter SL (VASEN et al. 1999; UMAR et al. 2004; PALOMAKI et al. 2009). Esses testes visam identificar a falta de expressão protéica de um dos genes MMR (*MLH1*, *MSH2*, *MSH6* e *PMS2*) e/ou uma elevada instabilidade de microsatélites, ambos indicativos de SL. Tumores MSI podem ser classificados como hereditários ou esporádicos dependendo do achado de mutações germinativas nos genes MMR ou hipermetilação somática no promotor do gene *MLHI*, respectivamente (BOUZOURENE et al. 2010). Quando houver ausência de expressão de MLH1 e PMS2 por IHC, é necessário realizar testes de metilação no promotor de *MLH1* ou da mutação V600E do gene *BRAF*, uma vez que estas alterações só acometem tumores esporádicos (COLAS et a. 2012; VALLE 2014).

As análises das amostras tumorais por IHC e MSI são sensíveis e específicas para identificar pacientes e famílias com a SL (PALOMAKI et al. 2009). Como mencionado anteriormente, um quarto dos indivíduos com a SL não preenchem os critérios clínicos da síndrome. Portanto, limitar os testes de IHC e MSI a indivíduos que preenchem esses critérios leva a perda da identificação de um número considerável de casos com mutações germinativas nos genes MMR (HAMPEL et al. 2008). Para evitar esse viés na triagem dos pacientes, é recomendável que os testes de IHC e MSI sejam realizados em todos os tumores colorretais e endometriais, independente da idade ao diagnóstico ou da história pessoal e familiar de câncer (LASTELLA et al. 2011; VALLE 2014). Apesar da significativa sensibilidade e especificidade dos testes de IHC e MSI em identificar tumores associados à SL, a única forma de confirmar o diagnóstico da síndrome é pela identificação de mutações germinativas em um dos genes MMR (WEISSMAN et al. 2012; COLAS et al. 2012).

1.3 IDENTIFICAÇÃO DE NOVOS GENES DE PREDISPOSIÇÃO AO CCR

1.3.1 Variação no Número de Cópias (CNVs)

Originalmente, as variações no número de cópias genômicas (CNVs - *Copy Number Variations*) foram definidas como alterações envolvendo segmentos de DNA de 1 kb ou maiores quando comparados com um genoma de referência (IAFRATE et al. 2004; REDON et al. 2006), porém, a limitação de 1 kb parece ser reflexo das primeiras tecnologias utilizadas para identificar estas variações (principalmente BAC-*arrays*) e não um limiar biológico ou funcional (CONRAD et

al. 2010; ALMAL e PADH 2012). Uma estimativa recente baseada na última atualização do DGV (*Database of Genomic Variants*, www.dgv.tcag.ca/dgv/app/home) sugere que até 22% do genoma humano é coberto por CNVs (PARK et al. 2015). A fração do genoma afetada por CNVs é comparativamente maior do que a contabilizada para os polimorfismos de base única (SNPs - *Single Nucleotide Polymorphisms*), assim, as CNVs tem um papel significativo na variabilidade genômica e fenotípica observada nos seres humanos (CONRAD et al. 2010; MILLS et al. 2011).

O mecanismo exato da formação das CNVs não é claramente compreendido (STANKIEWICZ e LUPSKI 2010). Duplicações, deleções, inserções, inversões e translocações representam alguns dos processos que levam à formação de novas CNVs (STANKIEWICZ e LUPSKI 2010; ALMAL e PADH 2012). Outros rearranjos genômicos mais complexos incluem a transposição, recombinação homóloga não alélica (NAHR - *Nonallelic homologous recombination*), associada com a similaridade de sequências de longos trechos em torno de pontos de quebra e união terminal não-homóloga (NHEJ - *Nonhomologous end joining*), associada com o mecanismo de reparo do DNA (KIDD et al. 2008; CONRAD et al. 2010; MILLS et al. 2011).

Na última década, a detecção de CNVs foi realizada principalmente por tecnologias de microarranjos. As técnicas de aCGH (*Array comparative genomic hybridization*) e SNP-*array* utilizam sequências de DNA plotadas em ‘‘chips’’. Os dois métodos podem detectar facilmente alterações cromossômicas com uma resolução muito mais fina do que o limite de 5Mb do cariótipo convencional (MULLEY e MEFFORD 2011). Os mecanismos patogênicos das CNVs estão

associados à dosagem de determinado gene, interrupção e fusão de genes, efeitos de posição, desmascaramento de alelos recessivos ou polimorfismos funcionais (STANKIEWICZ e LUPSKI 2010). Diversos estudos associaram as CNVs com doenças autoimunes, infecciosas, neuropsiquiátricas, cardiovasculares e ao desenvolvimento de tumores (SHLIEN e MALKIN 2010; KREPISCHI et al. 2012a; GELB e CHUNG 2014; HOLLOX e HOH 2014; KIROV 2015). Em particular, CNVs germinativas raras foram associadas ao risco de câncer em indivíduos com diagnóstico clínico da Síndrome de Li-Fraumeni (critérios de Chompret), mas sem mutações patogênicas em *TP53* e em pacientes com critérios para síndromes hereditárias de câncer de mama, negativos para mutações em *BRCA1*/*BRCA2* (KREPISCHI et al. 2012b; PYLKÄS et al. 2012; KUUSISTO et al. 2013; MASSON et al. 2014).

Recentemente, três estudos identificaram CNVs germinativas raras como potenciais candidatas à predisposição ao CCR em pacientes com história familiar de câncer e/ou tumores em idade jovem, mas sem mutações nos genes MMR. O Quadro 1 resume os principais achados desses três estudos. VENKATACHALAM et al. (2011) avaliaram 41 pacientes com a plataforma de microarranjos Affymetrix SNP 6.0 e detectaram CNVs germinativas novas em seis deles. Entre essas CNVs, quatro afetaram genes previamente associados com a carcinogênese colorretal (*GREM1*, *PTPRJ*, *BCR* e *CDH18*) em pacientes distintos, fornecendo evidência adicional de sua provável patogenicidade. Utilizando a mesma plataforma do estudo anterior, YANG et al. (2014) detectaram um nova deleção em 12p12.3, englobando o gene *RERGL*, em três casos de CCR de uma casuística de 384 casos familiares de CCR e 1262 controles avaliados. Adicionalmente, utilizando o método TaqMan, os autores

identificaram, mais uma deleção envolvendo o gene *RERGL* em um caso de CCR em 2520 pacientes e 5015 controles analisados. Este gene é predito como pertencente à superfamília Ras de GTPases, assim como o gene *RERG*, e pode atuar como supressor tumoral em CRC. Além disso, os autores também identificaram um aumento significativo no número total de CNVs (duplicações e deleções) nos casos comparados com os controles, sugerindo o envolvimento de alterações estruturais na suscetibilidade ao CCR.

Utilizando a plataforma de microarranjos de alta resolução Affymetrix Cyto 2.7M, MASSON et al. (2013) avaliaram 125 pacientes que preenchiam os critérios de Amsterdam ou Bethesda e eram negativos para mutações nos genes MMR. Os autores identificaram 207 CNVs germinativas novas associados com 317 genes (MASSON et al. 2013). Interessantemente, 60 desses genes foram afetados por CNVs em mais de um indivíduo, incluindo genes relacionados ao CCR; como *LCP1*, *CTNNA3* e *IGSF11*; sugerindo que podem contribuir no desenvolvimento de CCR nessas famílias. Além disso, uma análise de enriquecimento também revelou vias moleculares associados com o CRC nesses 317 genes, incluindo vias metabólicas, a via das junções de oclusão e a de sinalização da neurotrofina.

1.3.2 Mutações em Ponto

Nos últimos cinco anos, estudos utilizando tecnologias de sequenciamento, como o exoma, também identificaram genes candidatos à predisposição ao CCR. CLARKE et al. (2012) identificaram duas mutações deletérias envolvendo o gene *GALNT12* em 4/118 pacientes com história familiar de CCR, sugerindo o potencial envolvimento deste gene na predisposição ao CCR. Contudo, mais recentemente a

avaliação em 103 pacientes com a FCCTX de outra população não revelou nenhuma mutação patogênica nesse gene (SEGUÍ et al. 2014), sugerindo a necessidade de novos estudos para avaliar a função do *GALNT12* como um gene de baixo ou moderado risco ao desenvolvimento do CCR.

Mutações em ponto no domínio da exonuclease de duas DNAs polimerases, *POLE* e *POLD1*, foram associadas ao risco de CCR e múltiplos adenomas (PALLES et al. 2013). Estas mutações possivelmente levam a uma falha na função de ambas as polimerases em corrigir bases pareadas erradas durante o processo de duplicação do DNA. Recentemente foram identificadas mutações patogênicas no gene *POLE* em 4% de casos familiares de CCR (3/77) com etiologia desconhecida, sugerindo que alterações neste gene devem ser avaliadas em casos com história de CCR em famílias sem mutações nos genes conhecidos de predisposição (SPIER et al. 2015).

GYLFE et al. (2013) sequenciaram o exoma de 96 pacientes com história familial de CCR com etiologia desconhecida e identificaram 3654 mutações truncadas. Após diversos procedimentos de filtragem para excluir os polimorfismos e falsos positivos, os autores identificaram 14 variantes germinativas em 11 genes candidatos à predisposição ao CCR (*UACA*, *SFXN4*, *TWSG1*, *PSPH*, *NUDT7*, *ZNF490*, *PRSS37*, *CCDC18*, *PRADC1*, *MRPL3* e *AKR1C4*) em pelo menos duas famílias distintas. A análise de perda de heterozigose nos tumores desses pacientes confirmou a deleção do alelo normal para os genes *UACA*, *TWSG1*, *PSPH* e *ZNF490* em sete casos. Além disso, análises de segregação utilizando parentes de primeiro grau com CCR mostraram que as alterações detectadas em *SFXN4*, *PRADC1* e *AKR1C4* segregavam com a doença. Estes resultados forneceram evidência adicional do possível envolvimento destes genes na tumorigênese colorretal.

O sequenciamento completo do exoma em 40 pacientes de 16 famílias revelou 32 variantes *nonsense* ou em sítios de *splicing*, 375 *missense* e 50 *indels* (DERYCKE et al. 2013). Algumas mutações que originam proteínas truncadas foram detectadas em mais de um indivíduo com câncer nestas famílias e incluem genes relacionados com o controle da mitose (*CDC27*, *CENPE*, *TACC2* e *KIF23*), regulação da transcrição (*CTBP2*, *IRF5*, *MED12* e *TRIP4*) e resposta imune (*BTNL2*, *BAGE*, *MASP1* e *NLRP8*). Duas variantes *missense* dos genes *CENPE* e *KIF23* foram detectadas em regiões previamente implicadas com o CCR (CICEK et al. 2012). Adicionalmente, estas mutações foram validadas e replicadas apenas em parentes com câncer, tornando estes dois genes bons candidatos de risco ao CCR.

ZHANG J. et al. (2015) sequenciaram o exoma de 15 famílias chinesas e identificaram 24 mutações raras/novas potencialmente deletérias. Estas mutações afetaram 19 genes, incluindo uma variante idêntica no gene *EIF2AK4* presente em sete diferentes pacientes, e três variantes em genes codificadores de receptores de riaonodina (*RYR2* e *RYR3*), os quais regulam os níveis de cálcio intracelular (LANNER et al. 2010). Por outro lado, SCHULZ et al. (2014) detectaram três mutações germinativas no gene *SEMA4A* em pacientes com a FCCTX. Análises funcionais com uma das variantes mostraram que o alelo mutado confere uma maior capacidade de proliferação, aumentando o número de células na fase S da mitose. Os autores concluíram que alterações neste gene tem potencial para estarem associadas com o risco de desenvolvimento do CCR. NIEMINEN et al. (2014) sequenciaram o exoma de 4 indivíduos com CCR de uma mesma família e detectaram uma mutação truncada em todos eles no gene *RPS20*. Análises de *Northern blot* mostraram que a mutação ocasiona um acúmulo do pré-rRNA 21S e uma diminuição do rRNA 18S.

Os autores especularam que a haploinsuficiência deste gene pode interferir na predisposição ao CCR tanto pela disrupção da biogênese de ribossomos como pela alteração da dosagem do *TP53*.

Recentemente, ESTEBAN-JURADO et al. (2015) utilizaram o sequenciamento do exoma para avaliar 43 pacientes de 29 famílias com história de CCR; sem mutações nos genes *MMR*, *APC* ou *MUTYH*. Os autores identificaram alterações potencialmente deletérias em 10 genes previamente implicados com o desenvolvimento de CCR, como *POLE*, *MYC*, *BMPR1A* e AKR1C4. Foram também relatadas variantes patogênicas em genes de reparo de DNA (*XRCC4*, *BARD1*, *NSMCE2*, *RAD52*) e que poderiam estar associados ao risco de CCR nessas famílias. SEGUÍ et al. (2015) identificaram uma mutação patogênica no gene *FAN1* em três indivíduos com tumores proficientes em MMR e pertencentes a uma família categorizada pelos critérios de Amsterdam. A atividade nuclease deste gene é utilizada no sitema MMR de reparo a danos no DNA. Os autores também avaliaram 176 pacientes categorizados pelo critério de Amsterdam e identificaram quatro mutações patogênicas em *FAN1*. Os 11 parentes com tumores avaliados também apresentaram mutações nesse gene. Estes resultados sugerem o envolvimento do *FAN1* na predisposição ao CCR em famílias Amsterdam sem alterações nos genes MMR.

O Quadro 1 resume os principais achados de estudos com microarranjos e sequenciamento que identificaram genes candidatos de predisposição ao CCR em famílias negativas para mutações nos genes MMR.

Quadro 1 - Resumo dos principais relatos publicados em literatura que utilizaram as metodologias de microarranjos e sequenciamento para investigar pacientes com a Síndrome de Lynch (critérios de Amsterdam ou Bethesda) e que não apresentavam mutações nos genes de predisposição conhecidos como associados com a doença.

Referência	Técnica	Plataforma	Número de pacientes avaliados	Principais achados	Principais genes candidatos
Venkatachalam et al. 2011	Microarranjo	Affymetrix SNP 6.0	40	7 CNVs novas em 6 pacientes	*GREM1*, *PTPRJ*, *BCR*, *CDH18*
Clarke et al. 2012	Sequenciamento direto	—	118	Mutações deletérias numa transferase envolvida com a glicosilação em 4 pacientes	*GALNT12*
Palles et al. 2013	Sequenciamento do genoma inteiro	Illumina HiSeq2000	20 (15 famílias)	Detecção de mutações patogênicas em duas DNAs polimerases	*POLE* e *POLD1*
DeRycke et al. 2013	Sequenciamento do exoma	Illumina HiSeq2000	40 (16 famílias)	Mutações truncadas em genes associados com motilidade celular, mitose e regulação da transcrição	*CENPE* e *KIF23*
Gylfe et al. 2013	Sequenciamento do exoma	Illumina GAII/HiSeq	96	14 mutações truncadas em 11 genes	*UACA*, *SFXN4*, *TWSG1*, *PSPH*, *NUDT7*, *ZNF490*, *PRSS37*, *CCDC18*, *PRADC1*, *MRPL3* e *AKR1C4*
Masson et al. 2013	Microarranjo	Affymetrix Cyto 2.7M	125	CNVs novas englobaram 60 genes presentes em mais do que um paciente/Análise *in silico* revelou vias associadas com câncer em 317 genes associados com 207 CNVs raras	—
Yang et al. 2014	Microarranjo	Affymetrix SNP 6.0	384	Deleção nova em 12p12.3 identificada em 3 pacientes	*RERGL*
Nieminen et al. 2014	Sequenciamento do exoma	Illumina HiSeq2000	26	Mutações patogênicas num gene associado com a biogênese ribossômica em 4 indivíduos com câncer de uma mesma família	*RPS20*
Schulz et al. 2014	Sequenciamento do exoma	Illumina HiSeq2000	57 (3 indivíduos pertencentes a uma mesma família)	Mutações patogênicas em um dos genes codificadores de receptores de semaforinas	*SEMA4A*
Zhang J. et al. 2014	Sequenciamento do exoma	Illumina HiSeq2000	23 (21 famílias)	24 variantes novas potencialmente patogênicas em 19 genes	*EIF2AK4*, *BUB1*, *MAX*, *ETV4*, *LIG3*, *PRDM1*, *TSC2*, *RYR2*, *RYR3*
Esteban-Jurado et al. 2015	Sequenciamento do exoma	Illumina HiSeq2000	43 (29 famílias)	Mutações patogênicas em genes associados com a replicação e reparo do DNA e na via do *BRCA1*	*CDKN1B, XRCC4, EPHX1, NFKBIZ, SMARCA4* e *BARD1*
Seguí et al. 2015	Sequenciamento do exoma	—	179 (3 indivíduos pertencentes a uma mesma família)	Mutações patogênicas numa nuclease associada com Anemia de Fanconi	*FAN1*

1.3.3 Polimorfismos de Base Única (SNPs)

Até o momento, estudos de associação genômica ampla (GWAS - *Genome-Wide Association Studies*) identificaram aproximadamente 60 SNPs, os quais são considerados como alterações envolvendo genes de baixa penetrância associados ao risco de desenvolvimento do CCR em populações européias e asiáticas (PITTMAN et al. 2009; REAL et al. 2014; WHIFFIN et al. 2014; ZHANG B et al. 2014; ZHANG K et al. 2014; AL-TASSAN et al. 2015) (Quadro 2). A maioria destes SNPs está associada com vias já conhecidas; como transdução de sinal, controle do ciclo celular e instabilidade gênomica. Contudo, um número significativo destes SNPs se encontra em regiões sem genes conhecidos e seu mecanismo de ação quanto à susceptibilidade ao CCR permanece incerto (ZHANG K et al. 2014).

Quadro 2 - Lista de SNPs associados com a predisposição ao CCR descritos em literatura.

SNP	Cromossomo	Gene
rs7524102	1p36.12	*ZBTB40*
rs72647484	1p36.2	*CDC42* e *WNT4*
rs10911251	1q25.3	*LAMC1*
rs16823149	1q31	*C1orf21*
rs6691170	1q41	*DUSP10*
rs6687758	1q41	*DUSP10*
rs4574118	2q12	*PLGLA*
rs11903757	2q32.3	*NABP1*
rs10936599	3q26.2	*MYNN*
rs4140904	4p15.3	*NCAPC*
rs3987	4q26	*NDST3*
rs647161	5q31.1	*PITX1*
rs1321311	6p21	*CDKN1A*
rs7758229	6q26-q27	*SLC22A3*
rs12701937	7p14.1	*GLI3* e *INHBA*
rs886774	7q31	*LAMB1*
rs16892766*	8q23.3	*EIF3H*
rs7014348	8q24	*POU5FIP1*
rs7837328	8q24	*LOC101930033*
rs6983267*	8q24.21	*MYC*
rs2209907	9q21.3	*TLE4*
rs11014993	10p12.1	*MYO3A*

Cont/ Quadro 2

SNP	Cromossomo	Gene
rs10795668*	10p14	*KRT8P16 e TCEB1P3*
rs4925386	10q13.33	*LAMA5*
rs704017	10q22.3	*ZMIZ1-AS1*
rs1035209	10q24.2	*ABCC2* e *MRP2*
rs11196172	10q25.2	*TCF7L2*
rs174537	11q12.2	*MYRF*
rs3824999	11q13.4	*POLD3*
rs3802842*	11q23.1	*POU2AF1*
rs10849432	12p13.31	*CD9*
rs10774214	12p13.32	*CCND2*
rs3217810	12p13.32	*CCND2*
rs3217901	12p13.32	*CCND2*
rs11169552	12q13.13	*DIP2*
rs7136702	12q13.13	*ATF1*
rs59336	12q24.21	*TBX3*
rs9548988	13q13.3	*COG6* e *FOXO1*
rs4444235*	14q22.2	*BMP4*
rs1957636	14q22.2	*BMP4*
rs16969681	15q13.3	*GREM1*
rs4779584*	15q13.3	*GREM1*
rs11632715	15q13.3	*GREM1*
rs1728785	16q22.1	*CDH1*
rs9929218*	16q22.1	*CDH1*
rs16941835	16q24.1	*FOXL1*
rs78378222	17p13	*TP53*
rs12603526	17p13.3	*NXN*
rs12953717	18q21.1	*SMAD7*
rs4464148	18q21.1	*SMAD7*
rs4939827*	18q21.1	*SMAD7*
rs58920878*	18q21.1	SMAD7
rs10411210*	19q13.1	*RHPN2*
rs1800469	19q13.2	*TGFB1*
rs961235*	20p12.3	*BMP2*
rs4813802	20p12.3	*BMP2*
rs2423279	20p12.3	*PLCB1*
rs6038071	20p13	*CASNK2A1*
rs6017342	20q13.12	*HNF4A*
rs5934683	Xp22.2	*SHR00M2*

*SNPs associados com o risco de desenvolver CCR em famílias sem mutações nos genes MMR.

Cada SNP apresenta uma elevada frequência nas populações (> 5%) e têm um efeito modesto no risco de câncer (geralmente com a razão de chances - do inglês, *odds ratio* - < 1,5) (ESTEBAN-JURADO et al. 2014). Contudo, foi sugerido um efeito cumulativo dos SNPs, sendo o risco de CCR maior com o aumento do número de alelos de risco (HOULSTON et al. 2012). Um total de 31 SNPs comuns de risco foi estimado como respondendo por aproximandamente 7,5% da herdabilidade

observada no CRC, sugerindo que outros SNPs de risco ainda podem ser descobertos (JIAO et al. 2014). A utilização de dados de 27 SNPs de risco juntamente com outros fatores como a história familiar e de endoscopias, pode aumentar a acurácia nas estimativas de desenvolver CCR (HSU et al. 2015).

Variantes comuns de baixo risco também modificam o risco no desenvolvimento do CCR em pacientes com diagnóstico molecular de SL. WIJNEN et al. (2009) identificaram uma associção significativa dos SNPs rs38022842 e rs16892766 com aumento da predisposição ao CCR em 675 pacientes pertencentes a 127 famílias com a SL, principalmente em mulheres carreadoras de mutações nos genes MMR. Posteriormente estes resultados foram replicados utilizando pacientes de outras populações (TALSETH-PALMER et al. 2011, 2013). Contudo, um estudo françês não encontrou associação entre o risco de CCR e estes dois SNPs em 748 indivíduos com SL. Esta discrepância entre os relatos pode ser explicada pela heterogeneidade genética das populações e dos pacientes com a SL (HOULLE et al. 2011).

Em torno de 11 SNPs (rs4939827, rs10411210, rs4444235, rs9929218, rs4779584, rs961253, rs10795668, rs6983267, rs3802842, rs16892766 e rs58920878) também foram associados ao risco de câncer em pacientes com história familiar de CCR, tumores em idade jovem e/ou múltiplos tumores, mas sem etiologia conhecida (MIDDELDORP et al. 2009; NIITTYMÄKI et al. 2010; GIRÁLDEZ et al. 2012). Um aumento significativo do número de alelos de risco foi relatado em indivíduos com história familiar em comparação com casos esporádicos de CCR, assim como em pacientes com diagnóstico precoce de CCR comparados a casos com tumores em idade tardia (NIITTYMÄKI et al. 2010; GIRÁLDEZ et al. 2012). Dez

dos 11 alelos de risco previamente mencionados podem explicar aproximadamente 9% da variância no risco familial de CCR (NIITTYMÄKI et al. 2010).

Um estudo recente com 1029 pacientes e 350 controles confirmou a associação de quatro SNPs (rs16892766, rs4779584, rs4939827 e rs58920878) com o risco ao CCR e mostrou um efeito cumulativo significativo no risco de câncer para pacientes com ao menos dois alelos ou genótipos de risco (razão de chances > 2) (BAERT-DESURMONT et al. 2015). As razões de chance podem chegar a 3,88 e 6,21 para pacientes com quatro alelos e três genótipos de risco, respectivamente.

2 OBJETIVOS

2.1 OBJETIVO PRINCIPAL

Identificar CNVs raras envolvidas com potenciais novos genes de predisposição ao CCR em pacientes com critérios clínicos para a SL, mas sem mutações patogênicas nos genes MMR, assim como avaliar a contribuição de um subconjunto de alelos comuns na predisposição ao CCR.

2.2 OBJETIVOS SECUNDÁRIOS

1. Comparar as sequências alteradas com informações disponíveis em bancos de dados e em amostras de indivíduos controle (obtidos previamente pelo grupo);
2. Identificar a função e o envolvimento dos genes selecionados em vias e redes biológicas;
3. Avaliar os mecanismos de segregação de CNVs em famílias específicas.

3 ARTIGO 1 - RESEARCH ARTICLE

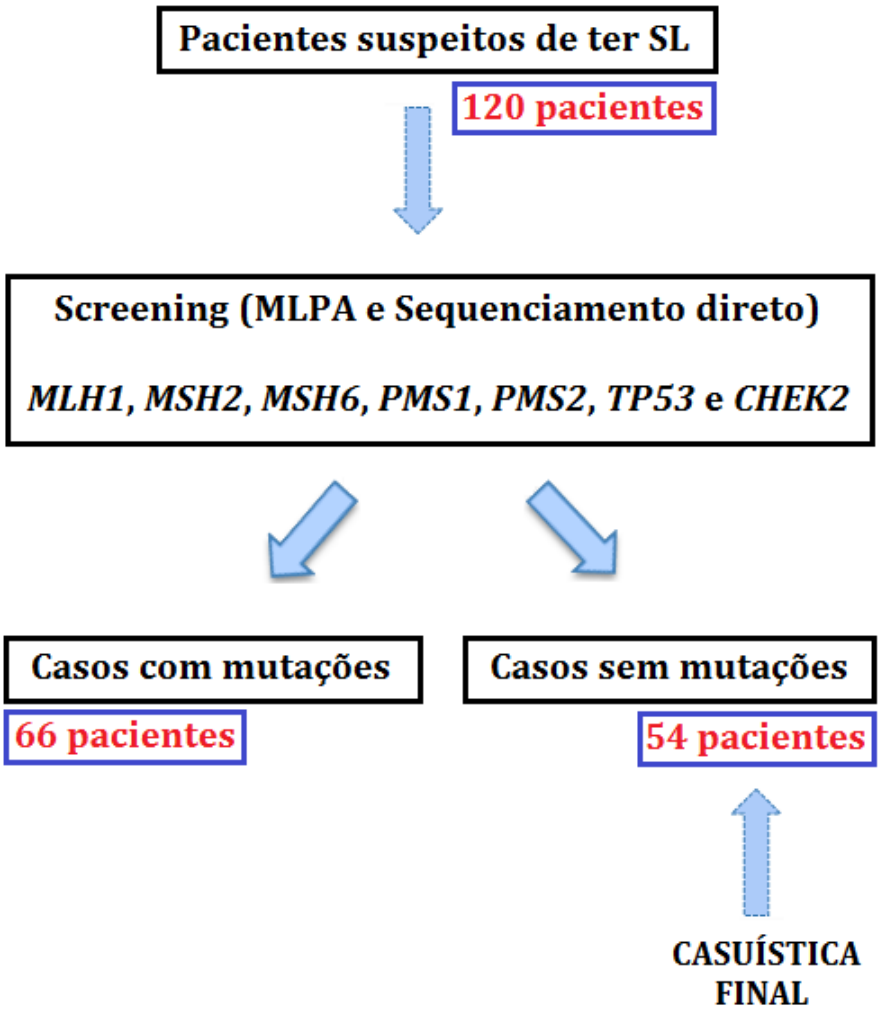

Figura 1 - Etapas realizadas na análise dos pacientes que preenchem os critérios clínicos da Síndrome de Lycnh utilizados no primeiro manuscrito descrito a seguir.

Contribution of rare germline copy number variations and common susceptibility loci in Lynch Syndrome patients negative for mutations in the mismatch repair genes

Rolando A. R. Villacis[1], Priscila M. Miranda[1], Erika M. M. Santos[2], Dirce M. Carraro[1], Maria I. Achatz[3], Benedito M. Rossi[2], Silvia R. Rogatto[1,4]*

[1] International Research Center (CIPE), A.C. Camargo Cancer Center, São Paulo, SP, Brazil

[2] Oncology Center, Sírio-Libanês Hospital, São Paulo, SP, Brazil

[3] Department of Oncogenetics, A.C. Camargo Cancer Center, São Paulo, SP, Brazil

[4] Department of Urology, Faculty of Medicine, University of Sao Paulo State (UNESP), Botucatu, SP, Brazil

*Correspondence to:
Silvia Regina Rogatto
Department of Urology, Faculty of Medicine (UNESP), CEP: 18618-970, Botucatu, SP - Brazil, Phone: +55-14-38116436, Fax: +55-14- 38116271. E-mail: rogatto@fmb.unesp.br
and International Research Center (CIPE), A. C. Camargo Cancer Center, Rua Taguá 440, Liberdade, CEP: 01508-010, São Paulo, SP - Brazil, Phone: +55-11-21895152. Email: silvia.rogatto@cipe.accamargo.org.br

The molecular etiology of a significant number of cases suspected of having Lynch Syndrome remains unclear. In this study, the authors investigated copy number variations and single nucleotide polymorphisms associated with colorectal cancer (CRC) risk in a cohort of patients with no mutations in the mismatch repair genes. Copy number alterations and risk alleles with the potential to increase the risk of hereditary CRC were reported.

Abstract

In colorectal carcinoma (CRC), 35% of cases are known to have a hereditary component, while a lower proportion (~5%) can be explained by known genetic

factors. In this study, copy number variations (CNVs) were evaluated in 54 unrelated patients with clinical hypothesis of Lynch Syndrome (Amsterdam or Bethesda criteria); negative for *MLH1*, *MSH2*, *MSH6*, *PMS2*, *CHEK2**1100delC and *TP53* pathogenic mutations; aiming to reveal new predisposing genes. Analyses with two different microarray platforms (Agilent 180K and Affymetrix CytoScan HD) revealed common CNVs in 21 patients and 58 rare CNVs in 33 patients, of which 28 subjects presented 43 rare CNVs covering 81 known genes. Gains (*GALNT6* and *GALNT11*) and losses (*CENPP* and *SEMA3C*) involving the same gene families related to CRC susceptibility were found among the rare CNVs. Segregation analysis performed on 14 relatives from three families suggested the involvement of *CHODL*, *GPR39*, *GALNT11* and *KMT2C* in those at risk of developing CRC. Notably, *in silico* molecular analysis revealed that 64% (52/81) of the genes covered by rare CNVs were associated with cancer, mainly colorectal (26 genes). Twenty-three common SNPs, previously associated with CRC, were genotyped in 49 patients, of which 26 showed more than 50% of risk alleles. None of the SNPs were covered by CNVs, suggesting an independent effect of each alteration in cancer susceptibility. In conclusion, rare germline CNVs and common SNPs may contribute to an increased risk for hereditary CRC in patients with mismatch repair proficiency.

Keywords

Colorectal cancer - Hereditary cancer - Copy number variation - Lynch Syndrome - Cancer predisposition - SNP

Introduction

Colorectal cancer (CRC) is the third most common neoplasm worldwide and the fourth leading cause of cancer death for both genders.[1] In the last two decades, a tendency towards increased CRC mortality rates has been reported for most Latin American countries, including Brazil.[2] Among the most prevalent neoplasms, CRC presents a significant case burden (~35%), with its development having been associated to a genetic predisposition.[3] However, only a minor proportion (~5%) can be explained by germline mutations in genes related to well-known hereditary diseases, such as Lynch syndrome (LS) (Mismatch Repair Genes - MMR: *MLH1*, *MSH2*, *MSH6*, and *PMS2*) and familial adenomatous polyposis (*APC*).[4] Approximately 11% of germline alterations found in LS patients are associated with *MSH2* deletions.[5] Moreover, *EPCAM* deletions leading to epigenetic silencing of *MSH2* have also been reported in LS.[6]

Copy number variations (CNVs) and single-nucleotide polymorphisms (SNPs), which are responsible for the vast majority of variations observed in the human genome, have been reported as being associated with hereditary cancers (family history of cancer, multiple tumors and/or early-onset neoplasms).[7,8] Recent studies have demonstrated the involvement of rare germline CNVs with cancer risk in Li-Fraumeni patients negative for *TP53* mutations, as well as for Hereditary Breast and Ovarian Cancer syndrome negative for *BRCA1*/*BRCA2* mutations.[9-11] Moreover, rare germline CNVs that cover genes potentially associated to an increased risk of CRC development, including *GREM1*, *CDH18* and *RERGL*, have been reported in LS patients.[12-14]

In addition to the MMR mutations and CNVs reported, genome-wide association studies (GWAS) have identified 60 common SNPs associated with an increased risk of CRC in European and Asian populations,[15-20] of which 11 (rs4939827, rs10411210, rs4444235, rs9929218, rs4779584, rs961253, rs10795668, rs6983267, rs3802842, rs16892766, rs58920878) were associated with familial susceptibility and early-onset in MMR mutation-negative CRC.[21-23] Ten of these 11 low-penetrance variants (excluding rs58920878) may explain up to 9% of the heritability observed for these families.[21] Although each SNP contributes modestly to CRC susceptibility,

the combined effect of these loci has been reported as being involved in an increased disease risk as much as higher the number of risk alleles.[24,25]

In this study, germline CNVs were analyzed in a cohort of 54 Brazilian patients, who fulfilled the clinical criteria for LS (Amsterdam or Bethesda), without mutations in the most common genes related to LS. A subset of 23 common low-risk SNPs, previously associated with CRC susceptibility, was also evaluated. The main aim was to identify genomic alterations with the potential to increase the risk of hereditary CRC.

Material and Methods

Patients

Fifty-four unrelated patients that fulfilled the Amsterdam I/II criteria (13 cases) or revised Bethesda guidelines (41 cases) were evaluated.[26,27] All patients were previously screened for germline mutations in genes related to hereditary cancer (*MLH1*, *MSH2*, *MSH6*, *PMS1*, *PMS2, CHEK2**1100delC and *TP53*) using Sanger sequencing and multiplex ligation-dependent probe amplification (MLPA), with no patient presenting pathogenic variants (data partially reported in Valentin et al.[28]). Fourteen relatives from three index patients were also evaluated. The Human Research Ethics Committee of the A.C. Camargo Cancer Center (CEP 1175/08) approved the study and all subjects provided written informed consent prior to sample collection.

The clinical and pathological information, as well as the procedures used for each case, are detailed in Supplementary Table S1. Five of 54 patients had colonic adenomas with high-grade dysplasia (SL-4, SL-10, SL-17, SL-52 and SL-53), while four presented malignant tumors other than CRC (SL-2, SL-7, SL-8 and SL-50). However, all of them are suspected to have LS and, hence, were enrolled in the study.

Copy number variation analysis

Genomic DNA was isolated from peripheral blood (index patients) and saliva (relatives) using the Gentra Puregene Blood Kit (Qiagen, Valencia, CA, USA) and *prepIT-L2P* kit (*DNA Genotek*, Ontario, Canada), respectively. All 54 index patients were evaluated for chromosomal imbalances using the Human CGH 4x180K microarray platform, following the manufacturer's recommendations. Genomic data were extracted by *Feature Extraction 10.1.1.1* (Agilent) and analyzed using the CytoGenomics *3.0.1.1* software (Agilent) (algorithm ADM-2). A threshold of 6.0, at least four altered probes, log2 ratio > 0.35 for gains and < -0.35 for losses were used as CNV defining criteria. The fuzzy zero correction was applied, with a detailed visual analysis being performed for all alterations, excluding regions with poor data quality.

A high-density microarray platform (CytoScan HD; Affymetrix, Santa Clara, CA, USA) was used to reassess 26 index patients (according to DNA quality and availability). Moreover, 14 relatives from three families that had accepted to participate in the study were evaluated. The procedures were carried out according to the manufacturer's protocols. Data were analyzed using the Chromosome Analysis Suite (ChAS) software 3.0 (Affymetrix) considering at least 25 markers for losses, 50 markers for gains and cnLOHs with a minimum of 5 Mb. All data were also visually inspected to confirm the alterations.

The CNVs obtained in the analysis were compared with the Database of Genomic Variants (DGV, http://dgv.tcag.ca/dgv/app/home, updated in July 2015), containing almost 67,000 samples; 100 healthy Brazilian individuals assessed by the 180K Agilent platform,[29] and the Affymetrix Database of Variants (2,421 phenotypically healthy individuals evaluated using the CytoScan HD platform). The common and rare alterations were classified according to size, type (gain or loss) and presence in the reference databases. Only rare CNVs, found in less than 10 subjects of the CNV datasets, were evaluated in detail. Rare CNVs not found in any of the three reference databases were considered new.

Single nucleotide polymorphism genotyping analysis

In 47 index patients, the standard TaqMan® SNP Genotyping Assay 5' nuclease technology protocol (Applied Biosystems, Foster City, CA, USA) was used to genotype 10 SNPs (rs961253, rs3802842, rs4444235, rs4779584, rs4939827, rs6983267, rs9929218, rs10411210, rs10795668 and rs16892766), which had been previously associated with familial CRC risk in the European population [20]. PCR amplification was carried out in duplicate using automated QIAgility pipetting (Qiagen, Courtaboeuf, France). Allelic discrimination was performed by the 7900HT Fast Real-Time PCR System using the Sequence Detection System software and auto-calling algorithm.

An additional 13 loci (rs12701937, rs11014993, rs59336, rs4813802, rs10774214, rs5934683, rs7524102, rs4574118, rs4140904, rs9548988, rs3987, rs12603526 and rs2241714) were genotyped in 26 patients using the CytoScan HD platform, containing polymorphic probes to interrogate the alleles of almost 750,000 SNPs. Two of 26 cases were genotyped exclusively using the microarray platform (SL-20 and SL-52). Thus, a total of 49 index patients were genotyped using the TaqMan assay and/or CytoScan HD platform. Two loci (rs6983267 and rs10795668), genotyped by the TaqMan assay, were also investigated in the microarray platform and used as a reference to confirm the results obtained by both procedures.

Results

Detection of CNVs with two different microarray platforms

Analysis by the Agilent 4x180K microarray platform for the 54 index patients revealed 245 CNVs (121 gains and 124 losses), with a median of 4.5 per case. Two patients (SL-4 and SL-13) presented no CNVs and 31 had common CNVs only. The remaining 21 cases showed 25 rare germline CNVs (one or two per case) with 23 harboring genes (Supplementary Table S2).

A subgroup of 26 cases was re-evaluated using the CytoScan HD platform including two cases with no CNVs, 16 cases with common CNVs only and eight cases with nine rare CNVs detected by the Agilent platform. Analysis with the high-resolution platform revealed 338 CNVs (median of 13 per case), while the Agilent platform revealed 100 CNVs (median of 3.8 per case) for the same cases (Supplementary Table S3). Six patients (SL-1, SL-12, SL-18, SL-23, SL-25 and SL-41), evaluated by both microarray platforms, presented common CNVs only. The nine rare CNVs that were initially detected by the Agilent platform were confirmed using the CytoScan HD platform. An additional 33 rare CNVs (20 harboring genes) were identified in 17 patients, exclusively by the Affymetrix platform, including two cases with no CNVs and five with rare CNVs that had been previously identified by the Agilent platform (Supplementary Table S3). The Supplementary Figure S1 summarizes the results obtained by both platforms used in this analysis.

The analysis by both microarray platforms (Agilent and Affymetrix) revealed common CNVs in 21 cases, while 58 rare CNVs were found in 33 patients. Forty-three rare CNVs overlapping genes were detected in 28 index patients (Table 1), of which, interestingly, 22 were new. Fifteen rare CNVs not covering known genes, seven of which were new, were identified in 10 index patients (Supplementary Table S4), including five cases (SL-2, SL-4, SL-6, SL-14 and SL-37) that also showed rare CNVs harboring genes. Two cases (SL-34 and SL-37) presented rare CNVs covering different regions mapped at intron 1 of the *CNTNAP2* gene.

To gain further insight into the function and pathways related to the 81 genes covered by the rare CNVs, an annotation enrichment analysis was performed using the Ingenuity Pathway Analysis software (IPA, QIAGEN Redwood City, www.qiagen.com/ingenuity). A significant enrichment of cancer-associated genes was detected (52 genes), 26 of which were associated with colonic and rectal adenocarcinomas (Table 1). Rare CNVs identified in three patients (SL-13, SL-46 and SL-53) covered two or more genes associated with CRC predisposition. The function of each gene related to CRC susceptibility is shown in Table 2.

Segregation Analysis

The segregation of rare CNVs overlapping genes related to cancer was investigated in 14 relatives belonging to three index patient families: two relatives from patient SL-2, four from SL-11 and five from SL-47 (Figure 1). No cancer-free relative of patient SL-2 presented the same rare CNV mapped in 21q21.1 encompassing the *CHODL* and *TMPRSS15* genes.

Two of four relatives from patient SL-11 reported a history of malignant tumors, both having presented the same rare duplication, partially covering the *GALNT11* and *KM2TC* genes, which were also detected in the index case. In addition, none of the four relatives presented the same rare deletion harboring the *DNMT3A* gene that was detected in the index patient.

Two relatives (sons), currently cancer-free, from patient SL-47 presented the same rare duplication, partially harboring the *GPR39* gene.

Genotyping of risk SNPs for CRC

A high genotype rate (> 95%) was obtained for the 10 SNPs evaluated with the TaqMan method. In 13 loci genotyped exclusively with the CytoScan HD platform, a call rate of <92% was detected for one SNP only (rs4574118: 88%, 23/26 cases). Two loci (rs6983267 and rs10795668) genotyped by both the CytoScan HD platform and TaqMan assay, performed in 24 index patients, presented corroborating results in 100% of cases. The SNPs, number of subjects genotyped and frequency of risk alleles are described in Table 3.

Eleven risk alleles (6 genotyped by the Affymetrix platform) were detected in more than 50% of cases. In addition, 26 cases presented a significant percentage of risk loci considering their genotyped SNPs, which ranged from 50% to 70% (Supplementary Table S5). Five of 10 risk loci previously related to CRC risk, (rs4939827, rs10411210, rs9929218, rs107955668 and rs6983267) were evaluated by the TaqMan assay and presented a frequency higher than 50%. A median of 10 risk alleles (6 to 14 risk alleles per case) were detected when considering the cases

genotyped with the TaqMan method, only, excluding three cases without the 100% call rate (total of 44 cases) (Table 4). Of the 33 cases with rare CNVs, 30 were genotyped for the aforementioned 10 SNPs using the TaqMan method, with 18 having 10 or more risk alleles (Supplementary Table 6).

Discussion

In the last decade, rare CNVs have been associated with cancer susceptibility, providing an explanation for part of the missing heritability observed in ~30% of familial and early-onset CCR cases.[8,30] Rare CNVs harboring genes mostly involved with cancer were, herein, described as candidates for being associated with colorectal carcinoma predisposition in a subset of patients suspected to have LS (Amsterdam or Bethesda criteria). When considering the clinical interpretation of CNVs, one must take into account not only their rarity, but also their type (gain and loss), size, location, and gene content, as well as data already reported in the literature.[31]

Constitutive genomic alterations were screened in 54 Brazilian patients, of which 26 were assessed by two platforms, Agilent 180K and a high-density platform (CytoScan HD, Affymetrix). In addition to confirming the rare CNVs detected by the Agilent platform, the second study allowed the identification of a higher number of CNVs, better estimating the chromosomal breakpoints and unmasking new CNVs potentially associated with CRC predisposition.

Approximately 52% (28/54) of cases presented rare CNVs (N=43) overlapping genes, of which 22 were new and hence not described in any reference dataset used in this study. All alterations were unique. Recent studies have described different rare and new CNVs, when compared with our data.[12-14] These differences may be explained by the platform resolution, the reference population used, as well as the criteria adopted to analyze the data and define a rare CNV. Venkatachalam et al.[12] identified seven new CNVs in 6 of 41 patients, involving genes previously related to CRC development, including *CDH18*, *GREM1*, *BCR* and *PTPRJ*. An identical rare deletion at 12p12.3, encompassing *RERGL*, has been described in two patients from 371 familial CRC cases.[14] Moreover, Masson et al. evaluated 125 MMR gene mutation-negative patients that met the Amsterdam or Bethesda criteria and, when

compared to a control group of 40 individuals, identified 207 unique CNVs overlapping genes, which included two gains covering *CNTNAP2*.[13] When comparing the data from this study with previous reports, two deletions harboring distinct regions of *CNTNAP2* on intron 1 were revealed for our patients.

Interestingly, functional *in silico* analysis (IPA) revealed cancer as the main disease associated with 52 of the 81 genes covered by the 43 rare CNVs. Remarkably, 50% of the cancer genes (26/52) were related to colorectal carcinogenesis, supporting its potential involvement in cancer risk. Despite the CNVs harboring intronic regions being found in seven of these CRC related-genes (*ROBO1*, *XRCC4*, *HCN1*, *CADM1*, *CNTNAP2*, *GALNT6* and *CENPP*), their potential pathogenic effect could not be excluded since intronic alterations may affect the splicing process leading to a loss of gene function, as previously reported for *MLH1* and *MSH2* genes.[32] In addition, our group has previously detected identical *ROBO1* intronic deletions (as presented in this study) in two cases of family breast cancer, providing further evidence for their potential pathogenicity (Villacis et al., submitted).

Seven (*DOCK8*, *KANK1*, *TEK*, *CADM1*, *CNTNAP2*, *ECM2* and *OMD*) of the 26 genes related to CRC development regulate the cell adhesion process. Recently, an increased level of TEK protein, whose gene was involved in gains in the present study, was identified in the plasma of 36 CRC patients when compared to the same number of healthy controls using ELISA (enzyme-linked immunosorbent assay) test, suggesting TEK as a potential biomarker for CRC.[33] In addition, *CADM1* (losses in our study) inactivation has been described as a recurrent event in colorectal tumorigenesis.[34] Genes involved in epigenetic regulation (methylation and deacetylation), such as *TRPS1*, *USP7* and *KMT2C*, were also described as associated with CRC development. An increased expression of the TRPS1 protein (gains in our study) has been reported as being related to lymph node metastasis and an advanced stage in CRC patients.[35] Losses involving *USP7* corroborate its function as a CRC tumor suppressor, with its inactivation having been reported as a critical event in colorectal carcinogenesis.[36,37] Overall, these previously reported findings give additional support for the involvement of *CADM1*, *TEK, TRPS1* and *USP7* in familial CRC cases. Moreover, duplications and deletions covering *GALNT6*,

GALNT11, *CENPP* and *SEMA3C*, belonging to the same gene families previously associated with CRC predisposition, were identified. Recently, the potential pathogenic effect of germline mutations in *CENPE*, *GALNT12* and *SEMA4A* to cancer risk in unexplained familial CRC has been described.[38-40]

Segregation analysis of the rare CNVs was performed for three unrelated families. The partial duplication, covering *CHODL*, was detected in the index patient SL-2, exclusively, and was absent in two relatives without cancer. A strong positivity for the CHODL protein, assessed by immunohistochemistry, has been reported as being related to poor prognosis in lung cancer patients.[41] Small interfering RNA experiments have shown that *CHODL* acts as an oncogene in lung carcinogenesis, regulating cellular growth and invasion.[42] Five relatives of patient SL-47 were tested, with two being cancer-free and showing the deletion partially covering *GPR39*. Despite being healthy and young (28 and 30 years old), the possibility of them develop cancer in the near future cannot be discarded. Recently, *GPR39* was described as acting at the tight junction barrier integrity in ulcerative colitis.[42] Patients with ulcerative colitis have a higher risk of developing CRC; thus, *GPR39* seems to be a good candidate for CRC susceptibility.[43] Furthermore, four of SL-11 relatives, two of whom had cancer, were evaluated for genomic imbalances. None of the relatives presented the deletion harboring *DNMT3A,* however, the partial duplication covering *GALNT11* and *KMT2C* was identified in three relatives, including the two with a diagnosis of cancer, suggesting an association between these alterations and CRC predisposition. Increased *GALNT11* expression has been reported for leukemia, while *KMT2C* has been associated with tumorigenesis in a family with a history of both CRC and leukemia.[44,45] To our knowledge, there was no history of leukemia for the studied family. However, one of the relatives tested (nephew) developed multiple myeloma at the age of 39, presenting *GALNT11* and *KMT2C* gains. These results pinpoint putative candidates for CRC predisposition.

Twenty-one of the 54 cases analyzed revealed the absence of rare CNVs, suggesting that another type of alterations may contribute to the CRC predisposition observed for these families. Recently, exome sequencing studies have revealed new candidate genes for CRC predisposition in MMR mutation-negative familiar CRC cases.[46,47]

Mutations in genes related to CRC development (*POLE*, *MYC*, *BMPR1A* and *AKR1C4*), as well as in DNA repair genes (*XRCC4*, *BARD1* and *RAD52*) were detected in 43 patients belonging to 29 families.[46] Interestingly, in this study a *XRCC4* deletion and a gain involving the *RAD54L* gene, which belongs to same *RAD52* family, were detected in two different patients. Three relatives of the same Amsterdam family presented a potential deleterious mutation of the *FAN1* gene, which codes for a nuclease involved in the MMR system.[47] New mutations in *FAN1* were found in 4/176 patients that met the Amsterdam criteria, suggesting the involvement of this gene in CRC risk.[47]

In addition to rare CNVs, common low-susceptibility SNPs may account for a fraction of the unknown heritability noted in complex diseases, including cancer.[30,48] The effect of the SNPs tested by genome wide association studies for Brazilian individuals to evaluate the risk of CRC development is unknown. For this reason, the effect of the 26 SNPs tested, herein, was compared with previous studies carried out in European and Asian subjects, justified by the fact that the Brazilian population is extremely heterogeneous.[49,50] Reports using ancestry-informative markers have concluded that at an individual level in Brazil, race, as determined by physical evaluation, is a poor predictor of genomic ancestry.[49,50] It is worth mentioning that not all SNPs reported as associated with CRC risk have been replicated in other studies and, at a certain frequency, new potential risk alleles are described in the literature.

A cumulative effect of the SNPs in CRC predisposition has been reported, in as much as a higher number of risk alleles.[24] In 24 patients with 23 SNPs genotyped by both TaqMan and CytoScan HD methods, a median of 21 risk alleles was detected, even when not all SNPs presented a 100% call rate. Recently, 11 SNPs have been related to an increased risk of familial and early-onset CRC of unexplained etiology, with 10 having been genotyped in 47 patients using the TaqMan assay.[21-23] Niittymäki et al.[21] evaluated the same 10 SNPs and estimated an elevated risk of familial cancer with an increase of number of risk alleles (OR = 1.16 per allele). Moreover, the authors identified that 52% of family cases presented 11 or more risk alleles compared to 36.7% of sporadic cases. Considering the 44 cases with a call

rate of 100% for these 10 SNPs, we identified that 21 cases (48%) presented 11 to 14 risk alleles. Recent data have demonstrated that the presence of at least two of three risk alleles (rs16892766, rs4779584 and rs58920878) cause a significant increase in CRC risk in unexplained familial and/or early-onset cancer.[23] Here, two SNPs (rs16892766 and rs4779584) were genotyped for 47 patients using the TaqMan method. A frequency of 12% was revealed for the risk allele for the SNP rs16892766. Thirteen of 47 cases showed no risk alleles for rs16892766 and rs4779584, while 16 index patients had two risk alleles, two cases (SL-19 and SL-26) having three risk alleles with only one case (SL-10) having four risk alleles. These findings suggested an increase of risk alleles in the study population that may contribute to CRC susceptibility.

In this study, the SNPs evaluated were mapped at different regions when compared with the rare CNVs, suggesting that both types of alteration have the potential to contribute to CRC predisposition, both in an independent and cumulative manner. Of the 33 patients with rare CNVs, 30 were also genotyped for the 10 SNPs reported as being involved in CRC familial risk, using the TaqMan method, of which 18 cases presented 10 or more risk alleles. Overall, these results suggest that cases with rare CNV have a significant number of risk alleles, which may further increase cancer risk. Patients with a large number of risk alleles have a threefold increase of developing CRC compared to cases with a low proportion of risk alleles.[24]

In summary, we identified germline CNVs that may contribute to cancer risk in unexplained Brazilian familial and early-onset CRC cases. Interestingly, the rare CNVs detected were enriched by cancer gene involved in colorectal tumorigenesis, hence suggesting their pathogenic role. Segregation analyses of three families also provided further evidence of the involvement of candidate genes in CRC predisposition. Additionally, the SNPs investigated revealed a significant number of patients with a higher number of risk alleles, which may contribute to CRC predisposition, since the cumulative effect of these loci in cancer risk has been shown.

Acknowledgements

The authors would like to thank the patients who agreed to participate in this study, the Nucleic Acid Bank and Department of Pathology of A.C. Camargo Cancer Center (São Paulo, Brazil). We also thank Felipe Carneiro da Silva and Bruna Katyn for their contribution during the study and Juan Cruz Cuevas who gently provided the Affymetrix Database of Variants. This study was supported by grants from the National Institute of Science and Technology in Oncogenomics (INCITO FAPESP 2008/57887-9 and CNPq 573589/08-9) and FAPESP (2011/07742-7).

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin* 2015;**65**:87-108.

2. Chatenoud L, Bertuccio P, Bosetti C, Malvezzi M, Levi F, Negri E, La Vecchia C. Trends in mortality from major cancers in the Americas: 1980-2010. *Ann Oncol* 2014;**25**:1843-53.

3. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;**343**:78-85.

4. Valle L. Genetic predisposition to colorectal cancer: where we stand and future perspectives. *World J Gastroenterol* 2014;**20**:9828-49.

5. Romero A, Garre P, Valentin O, Sanz J, Pérez-Segura P, Llovet P, Díaz-Rubio E, de la Hoya M, Caldés T. Frequency and variability of genomic rearrangements on MSH2 in Spanish Lycnh Syndrome families. *PLoS One* 2013;**8**:e72195.

6. Kuiper RP, Vissers LE, Venkatachalam R, Bodmer D, Hoenselaar E, Goossens M, Haufe A, Kamping E, Niessen RC, Hogervorst FB, Gille JJ, Redeker B, et al. Recurrence and variability of germline EPCAM deletions in Lynch syndrome. *Hum Mutat* 2011;**32**:407-14.

7. Fletcher O, Houlston RS. Architecture of inherited susceptibility to common cancer. *Nat Rev Cancer* 2010;**10**:353-61.

8. Krepischi AC, Pearson PL, Rosenberg C. Germline copy number variations and cancer predisposition. *Future Oncol* 2012;**8**:441-50.

9. Aury-Landas J, Bougeard G, Castel H, Hernandez-Vargas H, Drouet A, Latouche JB, Schouft MT, Férec C, Leroux D, Lasset C, Coupier I, Caron O, et al. Germline copy number variation of genes involved in chromatin remodelling in families suggestive of Li-Fraumeni syndrome with brain tumours. *Eur J Hum Genet* 2013;**21**:1369-76.

10. Kuusisto KM, Akinrinade O, Vihinen M, Kankuri-Tammilehto M, Laasanen SL, Schleutker J. copy number variation analysis in familial BRCA1/2-negative Finnish breast and ovarian cancer. *PLoS One* 2013;**8**:e71802.

11. Krepischi AC, Achatz MI, Santos EM, Costa SS, Lisboa BC, Brentani H, Santos TM, Gonçalves A, Nóbrega AF, Pearson PL, Vianna-Morgante AM, Carraro DM, et al. Germline DNA copy number variation in familial and early-onset breast cancer. *Breast Cancer Res* 2012; **14**:R24.

12. Venkatachalam R, Verwiel ET, Kamping EJ, Hoenselaar E, Görgens H, Schackert HK, van Krieken JH, Ligtenberg MJ, Hoogerbrugge N, van Kessel AG, Kuiper RP. Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. *Int J Cancer* 2011;**129**:1635-42.

13. Masson AL, Talseth-Palmer BA, Evans TJ, Grice DM, Duesing K, Hannan GN, Scott RJ. Copy number variation in hereditary non-polyposis colorectal cancer. *Genes (Basel)* 2013;**4**:536-55.

14. Yang R, Chen B, Pfütze K, Buch S, Steinke V, Holinski-Feder E, Stöcker S, von Schönfels W, Becker T, Schackert HK, Royer-Pokora B, Kloor M, et al. Genome-wide analysis associates familial colorectal cancer with increases in copy number variations and a rare structural variation at 12p12.3. *Carcinogenesis* 2014;**35**:315-23.

15. Pittman AM, Naranjo S, Webb E, Broderick P, Lips EH, van Wezel T, Morreau H, Sullivan K, Fielding S, Twiss P,Vijayakrishnan J, Casares F, et al. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. *Genome Res* 2009;**19**:987-93.

16. Real LM, Ruiz A, Gayán J, González-Pérez A, Sáez ME, Ramírez-Lorca R, Morón FJ, Velasco J, Marginet-Flinch R, Musulén E, Carrasco JM, Moreno-Rey C, et al. A colorectal cancer susceptibility new variant at 4q26 in the Spanish population identified by genome-wide association analysis. *PLoS One* 2014;**9**:e101178.

17. Zhang B, Jia WH, Matsuda K, Kweon SS, Matsuo K, Xiang YB, Shin A, Jee SH, Kim DH, Cai Q, Long J, Shi J, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet* 2014;**46**:533-42.

18. Zhang K, Civan J, Mukherjee S, Patel F, Yang H. Genetic variations in colorectal cancer risk and clinical outcome. *World J Gastroenterol* 2014;**20**:4167-77.

19. Whiffin N, Hosking FJ, Farrington SM, Palles C, Dobbins SE, Zgaga L, Lloyd A, Kinnersley B, Gorman M, Tenesa A, Broderick P, Wang Y, et al. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum Mol Genet* 2014;**23**:4729-37.

20. Al-Tassan NA, Whiffin N, Hosking FJ, Palles C, Farrington SM, Dobbins SE, Harris R, Gorman M, Tenesa A, Meyer BF, Wakil SM, Kinnersley B, et al. A new GWAS and meta-analysis with 1000Genomes imputation dentifies novel risk variants for colorectal cancer. *Sci Rep* 2015;**5**:10442.

21. Niittymäki I, Kaasinen E, Tuupanen S, Karhu A, Järvinen H, Mecklin JP, Tomlinson IP, Di Bernardo MC, Houlston RS, Aaltonen LA. Low-penetrance susceptibility variants in familial colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2010;**19**:1478-83.
22. Giráldez MD, López-Dóriga A, Bujanda L, Abulí A, Bessa X, Fernández-Rozadilla C, Muñoz J, Cuatrecasas M, Jover R, Xicola RM, Llor X, Piqué JM, et al. Susceptibility genetic variants associated with early-onset colorectal cancer. *Carcinogenesis* 2012;**33**:613-19.

23. Baert-Desurmont S, Charbonnier F, Houivet E, Ippolito L, Mauillon J, Bougeard M, Abadie C, Malka D, Duffour J, Desseigne F, Colas C, Pujol P. Clinical relevance of 8q23, 15q13 and 18q21 SNP genotyping to evaluate colorectal cancer risk. *Eur J Hum Genet* 2015. doi:10.1038/ejhg.2015.72 [Epub ahead of print]

24. Houlston RS; members of COGENT. COGENT (COlorectal cancer GENeTics) revisited. *Mutagenesis* 2012;27:143-51.

25. Esteban-Jurado C, Garre P, Vila M, Lozano JJ, Pristoupilova A, Beltrán S, Abulí A, Muñoz J, Balaguer F, Ocaña T, Castells A, Piqué JM, et al. New genes emerging for colorectal cancer predisposition. *World J Gastroenterol* 2014;**20**:1961-71.

26. Vasen HFA, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative Group on HNPCC. *Gastroenterology* 1999;**116**:1453-6.

27. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R,Hamilton SR, Hiatt RA, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 2004;**96**:261-8.

28. Valentin MD, da Silva FC, dos Santos EM, Lisboa BG, de Oliveira LP, Ferreira Fde O, Gomy I, Nakagawa WT, Aguiar Junior S, Redal M, Vaccaro C, Valle AD, et al. Characterization of germline mutations of MLH1 and MSH2 in unrelated south American suspected Lynch syndrome individuals. *Fam Cancer* 2011;10:641-7.

29. Krepischi AC, Achatz MI, Santos EM, Costa SS, Lisboa BC, Brentani H, Santos TM, Gonçalves A, Nóbrega AF, Pearson PL, Vianna-Morgante AM, Carraro DM, et al. Germline DNA copy number variation in familial and early-onset breast cancer. *Breast Cancer Res* 2012;**14**:R24.

30. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, et al. Finding the missing heritability of complex diseases. *Nature* 2009;**461**:747-53

31. Hehir-Kwa JY, Pfundt R, Veltman JA, de Leeuw N. Pathogenic or not? Assessing the clinical relevance of copy number variants. *Clin Genet* 2013;**84**:415-21.

32. Petersen SM, Dandanell M, Rasmussen LJ, Gerdes AM, Krogh LN, Bernstein I, Okkels H, Wikman F, Nielsen FC, Hansen TV. Functional examination of MLH1, MSH2, and MSH6 intronic mutations identified in Danish colorectal cancer patients. *BMC Med Genet* 2013;**14**:103.

33. Engin H, Üstündağ Y, Tekin İÖ, Gökmen A, Ertop Ş, İlikhan SU. Plasma concentrations of angiopoietin-1, angiopoietin-2 and Tie-2 in colon cancer. *Eur Cytokine Netw* 2012;**23**:68-71.

34. Chen K, Wang G, Peng L, Liu S, Fu X, Zhou Y, Yu H, Li A, Li J, Zhang S, Bai Y, Zhang Y. CADM1/TSLC1 inactivation by promoter hypermethylation is a frequent event in colorectal carcinogenesis and correlates with late stages of the disease. *Int J Cancer* 2011;128:266-73.

35. Hong J, Sun J, Huang T. Increased expression of TRPS1 affects tumor progression and correlates with patients' prognosis of colon cancer. *Biomed Res Int* 2013;**2013**:454085.

36. Becker K, Marchenko ND, Palacios G, Moll UM. A role of HAUSP in tumor suppression in a human colon carcinoma xenograft model. *Cell Cycle* 2008;**7**:1205-13.

37. Yang Z, Huo S, Shan Y, Liu H, Xu Y, Yao K, Li X, Zhang X. STAT3 repressed USP7 expression is crucial for colon cancer development. *FEBS Lett* 2012;**586**:3013-7.

38. Clarke E, Green RC, Green JS, Mahoney K, Parfrey PS, Younghusband HB, Woods MO. Inherited deleterious variants in GALNT12 are associated with CRC susceptibility. *Hum Mutat* 2012;**33**:1056-8.

39. DeRycke MS, Gunawardena SR, Middha S, Asmann YW, Schaid DJ, McDonnell SK, Riska SM, Eckloff BW, Cunningham JM, Fridley BL, Serie DJ, Bamlet WR, et al. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. *Cancer Epidemiol Biomarkers Prev* 2013;**22**:1239-51.

40. Schulz E, Klampfl P, Holzapfel S, Janecke AR, Ulz P, Renner W, Kashofer K, Nojima S, Leitner A, Zebisch A, Wölfler A, Hofer S, et al. Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X. *Nat Commun* 2014;**5**:5191.

41. Masuda K, Takano A, Oshita H, Akiyama H, Tsuchiya E, Kphno N, Nakamura Y, Daigo Y. Chondrolectin is a novel diagnostic biomarker and a therapeutic target for lung cancer. *Clin Cancer Res* 2011;**17**:7712-22.

42. Cohen L, Sekler I, Hershfinkel M. The zinc sensing receptor, ZnR/GPR39, controls proliferation and differentiation of colonocytes and thereby tight junction formation in the colon. *Cell Death Dis* 2014;**5**:e1307.

43. Hata K, Kishikawa J, Anzai H, Shinagawa T, Kazama S, Ishii H, Nozawa H, Kawai K, Kiyomatsu T, Tanaka J, Tanaka T, Nishikawa T, et al. Surveillance colonoscopy for colitis-associated dysplasia and cancer in ulcerative colitis patients. *Dig Endosc* 2015. doi: 10.1111/den.12505 [Epub ahead of print]

44. Li WD, Li QR, Xu SN, Wei FJ, Ye ZJ, Cheng JK, Chen JP. Exome sequencing identifies an MLL3 gene germ line mutation in a pedigree of colorectal cancer and acute myeloid leukemia. *Blood* 2013;**121**:1478-9.

45. Libisch MG, Casás M, Chiribao M, Moreno P, Cayota A, Osinaga E, Oppezzo P, Robello C. GALNT11 as a new molecular marker in chronic lymphocytic leukemia. *Gene* 2014;**533**:270-9.

46. Esteban-Jurado C, Vila-Casadesús M, Garre P, Lozano JJ, Pristoupilova A, Beltran S, Muñoz J, Ocaña T, Balaguer F, López-Cerón M, Cuatrecasas M, Franch-Expósito S, et al. Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. *Genet Med* 2015;**17**:131-42.

47. Seguí N, Mina LB, Lázaro C, Sanz-Pamplona R, Pons T, Navarro M, Bellido F, López-Doriga A, Valdés-Mas R, Pineda M, Guinó E, Vidal A, et al. Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. *Gastroenterology* 2015. doi: 10.1053/j.gastro.2015.05.056 [Epub ahead of print]

48. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 2010;**11**:446-50

49. Sans M. Admixture studies in Latin America: from 20th to the 21st century. *Hum Biol* 2000;**72**:155–77.

50. Salzano FM. Molecular variability in Amerindians: wide-spread but uneven information. *An Acad Bras Sci* 2002;**74**:223-63.

Table 1. Rare CNVs harboring genes detected in 28 out of 54 patients analyzed by two different microarray platforms (Agilent and Affymetrix).

Case	Type	Agilent 4x180K (hg19)	Affymetrix CytoScan HD (hg19)	Genes covered by the CNV[a]	Cancer-related genes[d]
SL-2	Gain*	ND	21q21.1 (19613367-20024179)	*CHODL*[b], *TMPRSS15*	*CHODL*, ***TMPRSS15***
SL-4	Gain*	ND	2p21 (45147359-45180958)	*SIX3*	*SIX3*
	Gain*	ND	8q23.3 (116670346-116682184)	*TRPS1*[b]	***TRPS1***
	Loss*	ND	14q24.2 (71950576-72026183)	*LOC145474*, *SIPA1L1*[b]	***SIPA1L1***
SL-6	Loss	3p12.3 (78908078-78945207)	3p12.3 (78913131-78955823)	*ROBO1*[c]	***ROBO1***
	Loss*	ND	5q14.2 (82375769-82385492)	*XRCC4*[c]	*XRCC4*
	Gain	9p21.2 (27218901-27273012)	9p21.2 (27190654-27294778)	*TEK*[b], *LINC00032*, *EQTN*[b]	***TEK***, *EQTN*
SL-10	Loss	ND	4q31.3 (151916165-152013101)	*LRBA*[b]	***LRBA***
SL-11	Loss	ND	2p23.3 (25226859-25288451)	*DNAJC27-AS1*[b], *EFR3B*[b]	*EFR3B*
	Loss*	2p23.3 (25490089-25754371)	2p23.3 (25439308-25760891)	*DNMT3A*, *MIR1301*, *DTNB*[b]	*DNMT3A*, *DTNB*
	Loss	ND	5p12 (45526349-45574986)	*HCN1*[c]	*HCN1*
	Gain	ND	7q36.1 (151795403-151895090)	*GALNT11*[b], *KMT2C*[b]	*GALNT11*, ***KMT2C***
SL-13	Gain	9p24.3 (434742-587418)	9p24.3(435188-586909)	*DOCK8*[b], *KANK1*[b]	***DOCK8***, ***KANK1***
SL-14	Gain*	4q31.1 (140046328-140166873)	4q31.1 (140043256-140169116)	*ELF2*[b]	***ELF2***
SL-17	Gain*	6q21 (107021184-107130595)	NE	*RTN4IP1*[b], *QRSL1*	***QRSL1***
SL-19	Gain	1p13.2 (112982660-113005539)	NE	*CTTNBP2NL*[b]	***CTTNBP2NL***
	Loss*	6q25.3 (160596664-160643290)	NE	*SLC22A2*[b]	*SLC22A2*
SL-20	Loss*	ND	7q11.22 (70180371-70236724)	*AUTS2*[b]	*AUTS2*
	Loss*	7q21.11 (80411532-80440032)	7q21.11 (80402154-80446014)	*SEMA3C*[b]	*SEMA3C*
SL-24	Gain*	2p24.1 (20401961-20424892)	NE	*SDC1*[b]	*SDC1*
SL-26	Loss	ND	2p23.3 (26372934-26400039)	*GAREML*[b]	*GAREML*
	Loss*	4q31.21 (145893086-145932678)	4q31.21 (145875635-145934750)	*ANAPC10*[b]	*ANAPC10*
SL-27	Loss*	16p13.2 (8987391-9206804)	NE	*USP7*[b], *C16orf72*[b]	***USP7***, *C16orf72*
SL-29	Gain	ND	3q13.33 (121494030-121514161)	*IQCB1*[b]	___
	Loss*	ND	8q23.3 (117231919-117276655)	*LINC00536*[b]	___
SL-31	Gain*	5q14.1 (80224419-80376556)	NE	*RASGRF2*[b]	***RASGRF2***
	Gain	13q12.12 (25186782-25313751)	NE	*ATP12A*	***ATP12A***
SL-32	Gain*	16p13.12 (12643340-12792035)	NE	*SNX29*[b], *CPPED1*[b]	*SNX29*, *CPPED1*
SL-34	Loss	ND	7q35 (145997480-146092958)	*CNTNAP2*[c]	***CNTNAP2***
	Loss*	ND	11q23.3 (115307680-115343279)	*CADM1*[c]	***CADM1***
SL-36	Loss	ND	16q12.1 (47351018-47390600)	*ITFG1*[c]	___
SL-37	Loss	ND	7q35 (146190711-146196724)	*CNTNAP2*[c]	***CNTNAP2***

SL-39	Loss	15q14 (34105933-34217569)	NE	*RYR3*[b], *AVEN*[b]	***RYR3***, *AVEN*
SL-43	Loss	ND	8q24.3 (146218161-146295771)	*ZNF252P*[b], *TMED10P1*, *ZNF252P-AS1*, *C8orf33*	*ZNF252P*, *C8orf33*
	Loss	ND	10p11.1 (38670606-38751908)	*SEPT7P9*, *LOC399744*	___
SL-44	Loss	2p22.2 (38400961-38492621)	NE	*CYP1B1-AS1*[b]	___
	Gain*	21q22.3 (43554325-43724162)	NE	*UMODL1*[b], *ABCG1*	***UMODL1***[b], *ABCG1*
SL-46	Gain	1p34.1 (46618671-46750887)	NE	*TSPAN1*, *POMGNT1*, *LURAP1*, *RAD54L*, *LRRC41*[b]	***RAD54L***, ***POMGNT1***, *LRRC41*
SL-47	Gain	2q21.2 (133177414-133355884)	NE	*GPR39*[b]	*GPR39*
SL-48	Gain*	ND	19q13.33 (50477562-50612095)	*SIGLEC16*, *VRK3*, *ZNF473*, *FLJ26850*, *SNAR-A9*, *SNAR-A5*, *SNAR-A10*, *SNAR-A14*, *SNAR-A3*, *SNAR-A6*, *SNAR-A7*, *SNAR-A11*, *SNAR-A8*, *SNAR-A4*	*SIGLEC16*, *VRK3*, ***ZNF473***
SL-49	Gain*	4q34.1 (173411914-173469892)	NE	*GALNT6*[c]	*GALNT6*
SL-51	Loss*	7p21.3 (9614342-9717761)	7p21.3 (9595215-9730096)	*PER4*	___
SL-53	Loss	9q22.31 (95146549-95290156)	NE	*CENPP*[c], *OGN*, *OMD*, *ASPN*, *ECM2*[b]	***CENPP***, ***ASPN***, ***OMD***, ***ECM2***

ND: not detected. NE: not evaluated

* Unique alterations, not detected in the CNVs databases used for comparison

[a] Based on the chromosomal breakpoints estimated by the CytoScan HD when the CNV was detected by both microarray platforms

[b] Genes partially involved (CNV covered exon regions)

[c] Genes partially involved (CNV covered only an intron region)

[d] According to the analysis with the IPA software. In bold genes related specifically with colon or rectal carcinogenesis

Table 2. Function of the 26 genes covered by the rare CNVs and associated with CRC carcinogenesis (IPA software).

Alteration	Genes	Function of the protein*
Gain	*ATP12A*	Transport of molecules/ATP hydrolysis/potassium absorption
	CTTNBP2NL	Membrane transport
	DOCK8	Cell adhesion/Apoptosis
	KANK1	Cell adhesion
	KMT2C	Histone methylation/Chromatin organization
	POMGNT1	Glycoprotein metabolism
	QRSL1	Mitochondrial translation
	RAD54L	DNA repair
	RASGRF2	Regulation of apoptotic process
	TEK	Cell adhesion/Nuclear transport
	TMPRSS15	Regulation of proteolysis
	TRPS1	Regulation of histone deacetylation and extrinsic apoptosis
	UMODL1	Regulation of apoptotic process and gene expression.
	ZNF473	Regulation of gene expression
Loss	*ASPN*	Extracellular matrix organization
	CADM1	Cell adhesion
	CENPP	Organização da cromatina/Regulação da mitose
	CNTNAP2	Cell adhesion
	ECM2	Cell adhesion/ Extracellular matrix organization
	ELF2	Cell differentiation
	LRBA	Transport of vesicles
	OMD	Cell adhesion
	ROBO1	Cell adhesion
	RYR3	Regulation of intracelular calcium
	SIPA1L1	Cytoskeleton reorganization
	USP7	Histone deubiquitination/Maintenance of DNA methylation

*Based on the information deposited in two databases: Gene (http://www.ncbi.nlm.nih.gov/gene) and GeneCards (http://www.genecards.org/)

Table 3. Single nucleotide polymorphisms genotyped using the TaqMan method and the CytoScan HD microarray platform

SNP	Chromosome	Risk Allele	Frequency of the risk allele	Call rate
TaqMan Genotyping assay[a] (47 cases)				
rs16892766	8q23.3	C	0.12	100% (47/47)
rs6983267*	8q24.21	G	0.64	100% (49/49)
rs10795668*	10p14	G	0.74	100% (49/49)
rs3802842	11q23.1	C	0.30	100% (47/47)
rs4444235	14q22.2	C	0.41	98% (46/47)
rs4779584	15q13.3	T	0.49	100% (47/47)
rs9929218	16q22.1	G	0.67	96% (45/47)
rs4939827	18q21.1	T	0.59	98% (46/47)
rs10411210	19q13.1	C	0.81	100% (47/47)
rs961253	20p12.3	A	0.35	98% (46/47)
CytoScan HD microarray platform (26 cases)[b]				
rs7524102	1p36.2	A	0.87	100% (26/26)
rs4574118	2q12	C	0.22	88% (23/26)
rs4140904	4p15.3	G	0.27	100% (26/26)
rs3987	4q26	C	0.52	100% (26/26)
rs12701937	7p14.1	T	0.35	100% (26/26)
rs1101937	10p12.1	C	0.19	100% (26/26)
rs10774214	12p13.32	T	0.38	100% (26/26)
rs59336	12q24.21	T	0.58	96% (25/26)
rs9548988	13q13.3	T	0.52	96% (25/26)
rs12603526	17p13.3	C	0.04	100% (26/26)
rs2241714	19q13.2	C	0.65	92% (24/26)
rs4813802	20p12.3	G	0.40	100% (26/26)
rs5934683	Xp22.2	C	0.58	100% (26/26)

*SNPs also genotyped with the CytoScan array platform. In this case, the genotyping was performed in 49 (two cases were genotyped only with the microarray platform).

[a]Loci genotyped with this method were related with hereditary CRC risk in cases with no mutation in the MMR genes.

[b]24 out of 26 patients were also genotyped with the TaqMan method.

Table 4. Number of risk alelles per case considering only the 10 SNPs genotyped with the TaqMan method and the cases with 100% call rate (44 cases)

Number of risk alleles (maximum of 20)	Number of cases (%)
6	2 (4.5)
7	4 (9.1)
8	7 (15.9)
9	3 (6.8)
10	7 (15.9)
11	5 (11.4)
12	8 (18.2)
13	6 (13.7)
14	2 (4.5)
Total	44

Figure legend

Figure 1. Pedigree charts of the three index patient (black arrows) who presented relatives (red arrows) evaluated for germline genomic alterations. In the right side of each pedigree are represented the rare CNVs (covering genes associated with cancer) identified in the probands of each family with the Affymetrix CytoScan HD platform. Deletions and duplications represented by red and blue bars, respectively. The positive and negative symbols means that the relatives tested have and have no rare CNVs detected in the respective index patient of each family, respectively. The age at diagnosis of the index patients and their healthy relatives are indicated above each case. CRC: colorectal cancer, BC: breast cancer, KC: kidney cancer; TC: thyroid cancer, LC: lung cancer; MM: multiple myeloma; PC: prostate cancer; HNC: head and neck cancer.

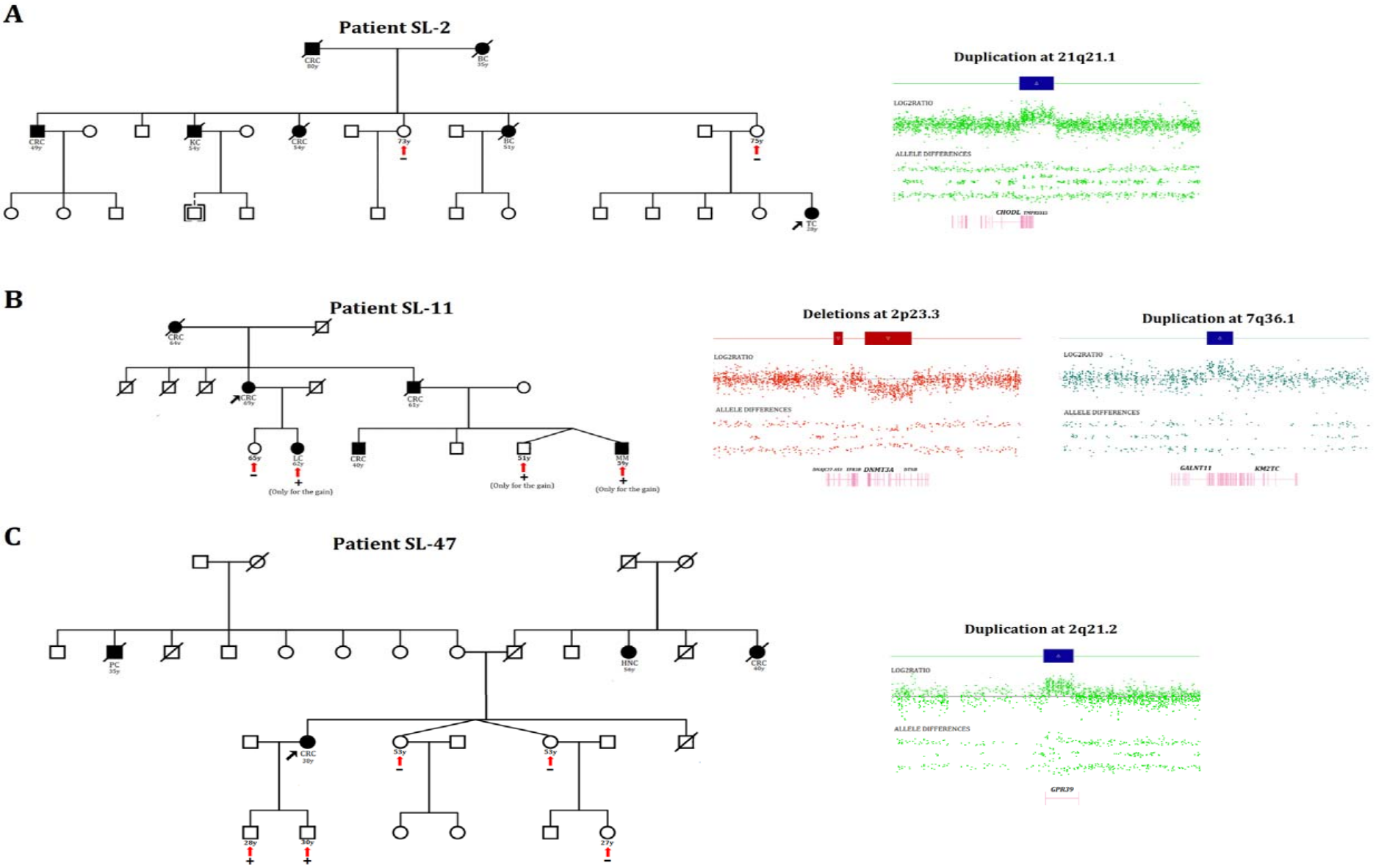
A
Patient SL-2
Duplication at 21q21.1
LOG2RATIO
ALLELE DIFFERENCES
B
Patient SL-11
Deletions at 2p23.3
Duplication at 7q36.1
(Only for the gain)
C
Patient SL-47
Duplication at 2q21.2

Figure 1

Additional files

Supplementary Figure 1. Flowchart showing the CNVs detected using the 4x180K (Agilent Technologies) and CytoScan HD (Affymetrix) microarray platforms.

Supplementary Table S1. Clinico-pathological characteristics of 54 patients with no pathogenic mutations in the most common genes related to hereditary colorectal cancer

Supplementary Table S2. Number of CNVs detected in 54 index patients with the Agilent 4x180K microarray platform

Supplementary Table S3. Comparison of genomic alterations found in 26 cases analyzed by two different microarray platforms

Supplementary Table S4. List of rare CNVs covering no genes and detected both by the Agilent and Affymetrix microarray platforms

Supplementary Table S5. Total number of alleles genotyped for each case, including the proportion of risk alleles

4 ARTIGO 2 - RESEARCH ARTICLE

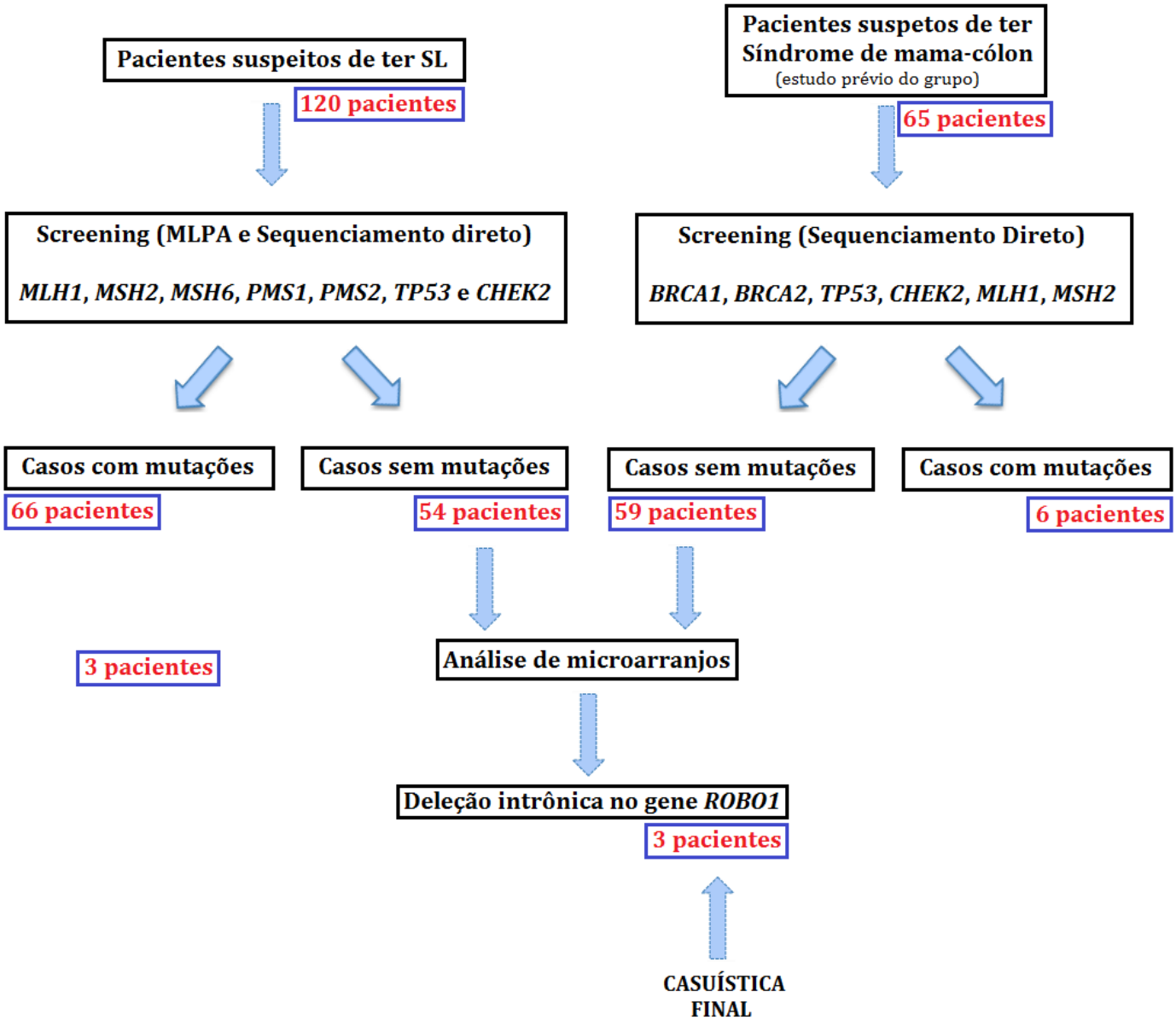

Figura 2 - Etapas realizadas para a descrição de pacientes portadores da deleção do gene ROBO1.

ROBO1 deletion as a novel germline alteration in breast and colorectal cancer patients

Rolando A. R. Villacis[a,*], Francine B. Abreu[a,*], Priscila M. Miranda[a], Maria A. C. Domingues[b], Dirce M. Carraro[a], Erika M. M. Santos[c], Benedito M. Rossi[c], Maria I. Achatz[d], Silvia R. Rogatto[a,e]

[a]International Research Center (CIPE), A.C. Camargo Cancer Center, São Paulo, SP, Brazil
[b]Department of Pathology, Faculty of Medicine, University of São Paulo State - UNESP, Botucatu, São Paulo, SP, Brazil
[c]Oncology Center, Sírio-Libanês Hospital, São Paulo, SP, Brazil
[d]Department of Oncogenetics, A.C. Camargo Cancer Center, São Paulo, SP, Brazil
[e]Department of Urology, Faculty of Medicine, University of São Paulo State (UNESP), Botucatu, SP, Brazil

*R.A.R.V and F.B.A. contributed equally.

Correspondence: Silvia Regina Rogatto, PhD

Department of Urology, Faculty of Medicne (UNESP), CEP: 18618-970, Botucatu - SP,Brazil, Phone: +55-11-38116436, Fax: +55-11- 38116271 (e-mail: rogatto@fmb.unesp.br)

International Research Center (CIPE), A. C. Camargo Cancer Center, Rua Taguá 440, Liberdade, CEP: 01508-010, São Paulo - SP, Brazil, Phone: +55-11-21895152 (e-mail: silvia.rogatto@cipe.accamargo.org.br)

Abstract

Despite one-third of breast (BC) and colorectal cancer (CRC) cases having a hereditary component, only a small proportion can be explained by germline mutations. The aim of this study was to identify genomic alterations with the potential of being related to cancer predisposition. Copy number variations (CNVs) were evaluated (Agilent 4x180K and Affymetrix CytoScan HD) in three index patients and six relatives, selected from 113 unrelated cases that fulfilled the criteria for hereditary BC/CRC and presented non-pathogenic mutations in *BRCA1*, *BRCA2*, *MLH1*, *MSH2*, *TP53* and *CHEK2* genes. We used direct sequencing to screen mutations in *ROBO1*, qPCR to confirm the deletion, and immunohistochemistry to evaluate the protein expression in an independent set of 351 breast and colorectal cancer cases. An identical germline deep intronic deletion of *ROBO1* was identified, via two microarray platforms and qPCR, in three patients. Microarray analyses confirmed a co-segregation of the *ROBO1* deletion with the occurrence of cancer in two families. Pathogenic *ROBO1* point mutations were not detected. Although not statically significant, the absence of immunoexpression of the ROBO1 protein was more frequent in CRC and BC groups with a poor outcome and shorter survival. Co-segregation analysis revealing the presence of disease in affected individuals, as well as the rarity of this alteration, suggest a pathogenic effect of the *ROBO1* deletion in cancer predisposition.

Keywords

breast carcinoma; colorectal cancer; germline deletion; hereditary cancer; *ROBO1*

Introduction

Colorectal cancer (CRC) and breast cancer (BC) are two of the most prevalent neoplasms worldwide, with over 2.5 million new cases diagnosed each year [1]. Although a large population-based twin study estimated that hereditary factors contribute to a considerable proportion of CRC (35%) and BC (27%) susceptibility, only 5% of familial and early-onset CRC/BC cases are due to germline mutations in well-known high penetrance genes [2-4]. Pathogenic germline mutations of the *BRCA1* and *BRCA2* genes are the most common alterations associated with familial BC, while mutations in the DNA mismatch repair (MMR: *MLH1*, *MSH2*, *MSH6* and *PMS2*) and *APC* genes explain the majority of hereditary CRC [3-4]. However, mutations in other high or moderate penetrance alleles, such as *TP53* (Li-Fraumeni Syndrome), *STK11* (Peutz-Jeghers Syndrome), *PTEN* (Cowden Syndrome), *CDH1* (Hereditary Diffuse Gastric Cancer), *CHEK2* and *PALB2,* may also increase the risk of developing these tumors [3-4].

Despite recent genome-wide association studies (GWAS) having identified 48 and 72 common single nucleotide polymorphisms (SNPs) associated with CRC and BC inheritance, respectively [4-5], the etiology of the vast majority of cases remains unclear, since the combined effect of these variants accounts for only a small fraction of cancer risk. Some of the remaining unexplained heritability may be related to copy number variations (CNVs), which are structural alterations (gains or losses) involving DNA fragments of 1 kb or larger [6]. CNVs cover approximately 12% of the genome, and have been associated to phenotypic variation in healthy individuals, as well as to several diseases, including neuropsychiatric conditions (autism, schizophrenia and learning disorders), obesity, Crohn's disease, systemic lupus erythematosis and several types of cancer [6-7].

Two recent studies have suggested the involvement of rare germline CNVs in patients negative for the *BRCA1* and *BRCA2* mutations with a family history of BC [8-9]. In addition, Masson et al. [10] identified germline CNVs in *WWOX* and *FHIT* genes, which was suggested as being associated with BC risk. In familial CRC patients without MMR gene mutations, an overall increase in CNV number was identified when compared to healthy controls [11]. Rare genomic rearrangements covering nine different chromosomes (5p14.3, 7q11.21, 8p23.1, 9p21.3, 11p11,

12p12.3, 15q13.3, 20q13.33, 22q11.23) have been associated to CRC predisposition [11-13].

The *ROBO1/DUTT1* gene, mapped on 3p12.3, is a member of the immunoglobulin superfamily transmembrane receptors that interacts with *SLIT2* and regulates a wide spectrum of biological functions, including axon guidance and neuronal migration [14]. *ROBO1* contains 31 exons spanning approximately 1,171 kb and 11 splice variants (www.ensembl.org; hg19). While *ROBO1* overexpression has been reported in hepatocellular and colorectal carcinoma [15, 16], other studies have demonstrated its potential role as a tumor suppressor gene, inactivated by deletion/methylation in several neoplasms, including lung, kidney, breast and pancreas [17-19]. Je et al. [20] reported five frameshift mutations in mononucleotide repeats within the *ROBO1* coding exons in CRC and gastric cancer samples with microsatellite instability.

The presence of multiple primary cancers, a positive family history of cancer and cancer onset at an early age strongly suggests a hereditary predisposition within these families [21]. In this study, an identical novel germline deep intronic deletion in the *ROBO1* gene in three unrelated patients and their relatives affected by cancer, yet negative for the most significant genes associated with hereditary BC/CRC, is described.

Materials and methods

Patients

The index patients (P1, P2 and P3) were selected from 113 unrelated cases that met criteria for hereditary BC/CRC, and previously evaluated by array CGH. All of them have no pathogenic mutations in the most common genes associated with these tumors. Written informed consent was obtained prior to sample collection. The Human Research Ethics Committee of the A.C. Camargo Cancer Center approved the study (Number 1175/08). The sequencing results of *MLH1*, *MSH2*, *MSH6*, *PMS1*, *PMS2, TP53, CHEK2* (1100delC), *BRCA1* and *BRCA2* for the index patients are described in Supplemental Table 1. Two patients (P2 and P3) fulfilled the criteria for Hereditary Breast and Colorectal Cancer (HBCC) [22], while also meeting the criteria for Hereditary Breast and Ovarian Cancer Syndrome (HBOC), as per the

National Comprehensive Cancer Network (NCCN) Guidelines (http://www.ncccn.org). In addition, one patient (P1) fulfilled the Bethesda Criteria described for hereditary CRC [23].

Patient P1 (II.1) developed early onset CRC at 34 years old, with no family history of cancer (Fig. 1A). Patient P2 (III.1) presented CRC at 52 years of age, and bilateral BC at 73 and 76 years of age, while also reporting several neoplasms, mainly breast tumors, within her family (Fig. 1B). Patient P3 (III.1) had BC at 48 years of age, as well as a family history of several neoplasms (Fig. 1C). Complete information regarding the neoplasm and treatments received by each index patient are presented in Table 1. Four relatives from patient P2: one daughter (P2-2/IV-3), two sons (P2-3/IV.1 and P2-4/IV.2) and one brother (P2-5/III.7) were included in this study. Similarly, two sisters of the patient P3 (P3-2/III.5 and P3-3/III.6) also agreed to be included in this study. The relatives of P2-2 and P3-2 were diagnosed with BC at 55 years of age and with endometrial cancer at age 58, respectively. In addition, an independent set of 192 BC and 159 CRC samples from patients with or without a family history of these tumors was included to evaluate ROBO1 protein expression.

Genomic Alterations

Genomic DNA was extracted using the Gentra Puregene Blook Kit (Qiagen, Valencia, CA, USA). Genomic imbalances were assessed using the SurePrint G3 Human CGH 4x180K (G4449A, Agilent Technologies, Santa Clara, CA, USA) array platform, according to the manufacturer's recommendations. Commercial genomic DNA (Promega, Madison, WI, USA) was used as a reference. Microarray data were extracted with Feature Extraction software 10.1.1.1 (Agilent Technologies) and analyzed with Genomic Workbench software 5.0.14 (Agilent Technologies) for CNV analysis (reference genome build hg18, ADM2 algorithm, threshold of 6.7, three consecutives probes and absolute log ratio filter of $-0.3 \leq \log_2 \geq 0.3$). In four individuals (P1, P2-2, P2-4 and P3), a higher density platform was also used (CytoScan HD, Affymetrix, Santa Clara, CA, USA) according to the manufacturer's instructions. Data were analyzed with the Chromosome Analysis Suite (ChAS) software 3.0 (Affymetrix) using NetAffx Build 33 (hg19) with 25 markers for losses

and 50 markers for gains considered for significance. The CNVs detected were compared with the Database of Genomic Variants (DGV, http://dgv.tcag.ca/dgv/app/home, data version: October 2014), 100 healthy Brazilian individuals [8], and 2,421 phenotypically normal individuals evaluated by the CytoScan HD platform (Affymetrix Database of Variants). Moreover, we searched for long noncoding RNAs (lncRNAs) in the lncRNAdb v2.0 database (http://www.lncrnadb.org/) [24].

Quantitative Real Time PCR (qPCR)

Two index patients (P2 and P3) and four relatives from patient P2 were evaluated by qPCR. Two primer sets were designed using the Primer-Blast software (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). A set of primers (S1) flanking one of the probes (Agilent probe name A_16_P16292303) and mapped within the intronic deletion, as well as one primer pair (S2) covering the probe (Agilent probe name A_14_P133797) located on exon 4 of the *ROBO1* gene (NM_002941.3/ENST00000464233) were designed for validation by qPCR. The PCR amplifications were prepared with the QIAgility automated pipetting system (Qiagen, Courtaboeuf, France) using Power SYBR Green (Applied Biosystems, Foster City, CA, USA) to a final volume of 12 µL, and run in duplicate on a 7500 Real Time PCR System (Applied Biosystems). DNA samples from five healthy individuals were used as a reference. The copy number calculation was carried out according to the $2^{-\Delta\Delta CT}$ method [25], using *GAPDH* and *HPRT1* as reference genes. The gene dosage interval was defined as one copy loss for a ratio of <0.55 and as a gain for a ratio of >1.35. Primer sequences for both target and reference genes are provided in Supplemental Table 2.

ROBO1 Sanger Sequencing Analysis

The patients were screened for germline mutations in the long *ROBO1* variant (NM_002941.3). PCR primers were designed to cover all 31 exons of the *ROBO1* gene (Supplemental Table 3). The PCR fragments were amplified using PCR Master Mix, 2X (Promega) and purified with EXO-SAP IT (USB Corporation, Cleveland, OH, USA), as per the manufacturer's recommendations. Sequencing reactions were

performed in both forward and reverse, and were run on an ABI Prism 3130xl Genetic Analyzer (Life Technologies, Foster City, CA, USA). The chromatogram traces were analyzed using the CLC Main Workbench software 6.0 (CLC Bio, Aarhus, Denmark). Genetic variant nomenclature followed the Human Genome Variation Society (HGVS) guidelines (http://www.hgvs.org/mutnomen/).

ROBO1 Protein Expression Analysis

Immunohistochemistry (IHC) assays were performed to evaluate ROBO1 protein expression in 351 (192 BC and 159 CRC) formalin fixed paraffin-embedded (FFPE) samples arranged in two tissue microarrays (TMAs), constructed as described previously [26]. The IHC reactions were achieved using rabbit polyclonal anti-ROBO1 (ab7279, Abcam Inc., Cambridge, MA, USA). One positive (lung tissue) and two negative references (one that did not use the primary antibody and the second that removed the secondary antibody), were evaluated as controls. IHC scores were estimated according to the intensity of cytoplasm staining (0: no staining; 1+: weak; 2+: moderate; and 3+: strong). Tumors with IHC scores of 0/+1 and +2/+3 were considered negative and positive for ROBO1 immunoexpression, respectively. Fisher's exact test was used to determine the association between the categorical variables, with a value of $P < 0.05$ considered significant. The overall survival (OS) curves were calculated using the Kaplan-Meier method and compared by log-rank test. IBM SPSS Statistical software version 20 (SPSS Inc., Chicago, IL, USA) was used for the statistical analysis.

Results

Identification of a Novel Germline *ROBO1* Deletion

The search for genomic alterations in the three unrelated index patients using the Agilent 4x180K platform revealed a novel heterozygous germline deletion on 3p12.3, within the intron 4 of *ROBO1* (NM_002941.3/ENST00000464233). The identical CNV was identified by four consecutive probes and spanned 37.470 kb (78,990,568-79,028,038 hg18) (Fig. 2A). For CNV confirmation, as well as a more precise identification of chromosome breakpoints, the CytoScan HD microarray platform (Affymetrix) was used to evaluate two index patients (P1 and P3). While

Agilent 4x180K platform contains 76 probes, the Affymetrix CytoScan HD microarray has 1,885 probes (1,604 for CNV detection and 281 for SNP genotyping) to assess *ROBO1* gene alteration. Analysis with a higher-resolution platform confirmed the deletion mapped on intron 4 of *ROBO1 (*48 non-polymorphic probes with an estimated 50.477 Kb: 78,905,346-78,955,823; hg19) (Fig. 2A).

The distal breakpoint (78,955,823) of the deletion was identified as being 31.928 kb away from exon 4 of the long *ROBO1* transcript, whose coordinates ranged from 78,987,751 to 78,988,977 (hg19) according to the Ensembl database. Exon 5 is approximately 109.453 kb away from the CNV detected on intron 4. Furthermore, the deletion was detected in the intronic region of four of six alternatively spliced variants of *ROBO1* that codes for proteins. Currently, no lncRNAs were identified in the deleted region (lncRNAdb v2.0).

Alterations involving the *ROBO1* gene were investigated in three different CNV databases. The Brazilian dataset [8], as well as the Affymetrix Database of Variants, revealed the absence of this CNV, hence confirming its rarity. According to the DGV, four cases with partial deletions involving this region (accession number nsv834744, nsv470672 and dgv8393n54) have been found in healthy individuals (Fig. 2B).

Table 2 shows the rare CNVs described in less than 1% of the reference population, which was identified in the patients using the Agilent 4x180K and CytoScan HD platforms. Only three rare alterations, including the *ROBO1* deletion, were detected and subsequently confirmed by both microarray platforms. Virtually all regions that were identified by only the CytoScan HD platform presented one or two probes on the Agilent platform, not permitting its detection at high resolution.

Co-segregation Study of the *ROBO1* Deletion

Four relatives from patient P2 and two from P3 were also evaluated for segregation of the *ROBO1* deletion. The CNV that was co-segregated in relatives with cancer revealed two relatives (P2-2 and P3-2) with a previous history of cancer that presented the same deletion (4x180K, Agilent Technologies). Three relatives (two of them without cancer: P2-4 and P3-3) were also evaluated using the Affymetrix

CytoScan HD platform, with the presence of the deletion being confirmed in one relative with a previous diagnosis of cancer (P2-2).

Relatedness Analysis

The relatedness of the three families was verified using the Mendelian error checking analysis (ChAS 3.0 software), which takes into account the SNP data (750,000 probes) provided by Affymetrix CytoScan HD platform. At least one member of each family (P1, P2-2, P2-4 and P3) was evaluated with this array platform. The duo analysis confirmed the unrelatedness among the individuals of each family (Higher Error Rate and Role Validity was zero), as well as the close genetic relatedness between P2-2 and P2-4 (Lower Error Rate and Role Validity was 1) (Supplemental Table 4).

CNV Validation by qPCR

The *ROBO1* deletion (primers S1) was confirmed by qPCR in two index patients (P2 and P3) and in one relative with cancer (P2-2). The relatives without cancer (P2-3, P2-4 and P2-5) presented no alteration within the same region. The set of primers S2 flanking exon 4, revealed no alteration in all evaluated cases, which was consistent with the microarray results (Fig. 2C).

Evaluation of *ROBO1* Mutations by Sanger Sequencing

Non-pathogenic germline variants for the *ROBO1* gene were found in the index patients and in six relatives. While three silent mutations were identified in the relatives of patient P2 (Fig. 2D), four relatives (P2-2, P2-3, P2-4 and P2-5) presented a silent alteration in exon 13 (c.1739G>A; p.S580S). In addition, relatives P2-2/P2-3 and P2-4/P2-5 also presented silent mutations in exons 16 (c.2210A>T; p.V737V) and 19 (c.2792C>A; p.T931T), respectively.

ROBO1 Protein Expression

IHC analysis was performed in an independent set of breast and colorectal carcinomas, obtained from patients with (47) and without (304) a family history of

malignancy aiming to evaluate the involvement of ROBO1 in breast and colorectal tumorigenesis. Seventy-four of 140 (53%) and 101 of 164 (62%) sporadic CRC and BC showed negative expression of ROBO1, respectively. In addition, 17 of 28 (61%) familial BC and 12 of 19 (63%) familial CRC cases presented negative expression for ROBO1, respectively (Fig. 3). No association between clinical pathological variables and the ROBO1 expression was found for either, familial or sporadic BC and CRC samples (Supplemental Table 5). However, negative ROBO1 expression was detected in more than half of the cases with pathologic features associated to a worse prognosis, such as histologic grade (II-III), TNM staging (III-IV) and lymph node metastasis.

The median follow-up was 9.3 and 7.4 years for BC and CRC patients, respectively. No statistical differences were found in 10-years OS curves considering the ROBO1 expression and both group of tumors (Supplemental Fig. 1). However, the log-rank test revealed a marginal significance for shorter survival in BC ($P = 0.099$) and CRC ($P = 0.086$) patients with negative ROBO1 expression.

Discussion

The clinical interpretation of CNVs (benign, uncertain significance or pathogenic) should be performed taking into account size, type (gain or loss), gene content, inheritance pattern and frequency in healthy individuals [27]. We identified an identical germline deletion of the *ROBO1* gene, which was confirmed by an independent microarray platform and qPCR analysis, in three unrelated patients with BC and/or CRC. The *ROBO1* deletion identified comprises only an intronic region (intron 4) of four protein-coding splice variants of *ROBO1*. Although unable to obtain tumor samples from the patients, with all of them having been surgically treated in different institutions, our results suggest a pathogenic effect of CNV, since the segregation analysis confirmed the presence of the alteration in patients affected by cancer only. Furthermore, the Mendelian error checking analysis using the SNP data in cases evaluated by Affymetrix plataform revealed no kinship among the members of these three families.

Interestingly, the heterozygous deletion on intron 4 of the long *ROBO1* transcript is within the homozygous deletion in the short *ROBO1* transcript

[NM_133631/ENST00000495273; hg17 spanning from intron 3 to intron 4 in the long transcript, converted by hg19 using Batch Coordinate Conversion-liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOvert)] reported by Angeloni et al. [28]. The authors also identified two lncRNAs on intron 4 of the long transcript (intron 2 of the short transcript) with the potential to modulate *ROBO1* expression. These lncRNAs are located in different regions than the alteration herein described. To our knowledge no other lncRNA or miRNA have been mapped in this intron (lncRNAdb v2.0).

In addition, no identical deletion of *ROBO1* was found in the CNV databases derived from healthy Brazilian individuals [8], and the normal reference provided by the Affymetrix dataset. Four deletions (4 of 17,959; 0.02%) that overlapped partially the *ROBO1* deletion detected in our patients were described in the DGV. However, these CNVs were found using low-resolution platforms (BAC and Illumina arrays) as opposed to CytoScan HD. Besides being extremely difficult to obtain clinical and follow-up information in cases deposited in DGV, we can not exclude the possibility of these individuals having personal or family history of cancer. In addition, the comparison with external databases should be performed with caution, since lower-density platforms may under- or overestimate the real size of the alterations. Furthermore, CNVs reported in public databases should be carefully analyzed, since experimental validation and clinical data is not provided in whole [27].

According to Kearney et al. [27] small CNVs that affect only intronic regions have no apparent effect on gene function. However, some examples showing intronic alterations associated with human conditions have been described [29-32]. In most cases, pathogenic intronic deletions lead to abnormal RNA splicing via the creation/removal of splice sites, minimizing intron size [29-32]. In cancer, deep intronic mutations had been related to hereditary retinoblastoma and melanoma[33-34]. In terms of BC and CRC susceptibility specifically, Clendenning et al. [35] reported an intronic mutation in *MSH2* associated with Lynch Syndrome; while intronic deletions at *EPHA3*, *CSMD1*, *ERRB4* and *BARD1* genes have been reported as potential CNVs related to BC risk [36-37].

Recently, a higher overall CNV burden has been identified in familial CRC patients compared to healthy individuals, suggesting its importance in CRC

susceptibility [11, 13]. Venkatachalam et al. [12] reported six CNVs in 41 familial or early-onset CRC cases with unexplained etiology, including a duplication encompassing the entire *GREM1* gene, a deletion affecting *CDH18* (exon 2) and two deletions encompassing two microRNAs (*MIR491* and *MIR646*). A rare intergenic gain on 7q11.21 was also identified in 29 of 96 CRC patients with mutations in one of the MMR genes, as well as a novel deletion on the 12p12.3 (*RERGL* gene) in two of 371 familial CRC cases negative for MMR gene mutations [11, 13]. None of these genes were covered by CNVs in our patients.

For hereditary BC, two studies performed in more than 100 familial or early-onset BC patients from the Finnish and Brazilian populations, negative for *BRCA1*/*BRCA2* gene mutations, identified an increase of rare CNVs compared to healthy individuals in the same populations [8-9]. Interestingly, in the Finnish BC cases, network analysis revealed that rare CNVs were related to the *TP53* and beta-estradiol pathways [9]. More recently, 129 familial BC patients without *BRCA1*/*BRCA2* mutations demonstrated 67 genes associated with novel CNVs that may contribute to BC risk [10]. The authors focused on CNVs encompassing the *WWOX* gene in two unrelated patients and a deletion affecting the *FHIT* gene in one patient. A detailed analysis of our cases did not reveal any common or rare CNVs associated with the *SLIT2/ROBO1* pathway or *WWOX* and *FHIT* genes.

In this study, other rare CNVs in the index patients and relatives were also identified. However, only a gain encompassing the last 14 exons of the *TEK* gene and a gain covered the first four exons of the *RCBTB1* gene were detected in both platforms for patients P1 and P2, and two of P2 relatives (P2-3 and P2-4), respectively. The differences in CNV detection may be explained, in part, by technical procedures for each platform, resolution and location of the probes, the reference used to perform comparison, and the bioinformatics analyses used to identify CNVs (algorithm and software). Both *TEK* and *RCBTB1* genes were related to malignant tumors, mainly leukemia [38, 39], and may contribute to increase the risk of cancer predisposition.

The ROBO1 protein expression in familial and non-familial BC/CRC cases was negative in approximately 60% of the samples. Although no significant, it was observed that more than 50% of these cases with negative expression of ROBO1

presented poor prognosis features, particularly lymph node metastasis, as well as shorter OS. These results suggest the involvement of *ROBO1* in breast and colorectal carcinogenesis. These findings are corroborated by the association described between low ROBO1 expression and poor prognosis in intrahepatic cholangiocarcinoma [40], and a ROBO1 protein loss-of-function in several neoplasms [18, 19, 28].

Conclusions

We identified a novel deep intronic *ROBO1* germline deletion in 2.6% (3/113) in patients negative for pathogenic mutations in the major genes related to hereditary BC/CRC. The pathogenic effect of this deletion is suggested by the familial co-segregation with the disease and the absence of identical alterations in public CNV databases. Since we were unable to obtain RNA and tumor from these patients, further studies are needed to clarify the effect of the deletion at the transcriptional and protein levels, as well as the potential role of this CNV in cancer predisposition.

Acknowledgements The authors would like to thank the Nucleic Acid Bank and Department of Pathology of A.C. Camargo Cancer Center (São Paulo, Brazil) and the patients and families who agreed to participate in this study. We are also grateful for the assistance with the sequencing data given by Dr. Felipe C. da Silva.

Funding This study was supported by grants from the National Institute of Science and Technology in Oncogenomics (INCITO FAPESP 2008/57887-9 and CNPq 573589/08-9), FAPESP (2010/15901-5 and 2011/07742-7).

Conflict of interest The authors declare that they have no conflict of interest.

References

1 Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61:69-90.

2 Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of

cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343:78-85.

3 Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. Gastroenterology. 2010;138:2044-58.

4 Ghoussaini M, Pharoah PD, Easton DF. Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning? Am J Pathol. 2013;183:1038-51.

5 Zhang K, Civan J, Mukherjee S, Patel F, Yang H. Genetic variations in colorectal cancer risk and clinical outcome. World J Gastroenterol. 2014;20:4167-77.

6 Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. Annu Rev Med. 2010;61:437-55.

7 Krepischi AC, Pearson PL, Rosenberg C. Germline copy number variations and cancer predisposition. Future Oncol. 2012;8:441-50.

8 Krepischi AC, Achatz MI, Santos EM, Costa SS, Lisboa BC, Brentani H, et al. Germline DNA copy number variation in familial and early-onset breast cancer. Breast Cancer Res. 2012;14:R24.

9 Pylkäs K, Vuorela M, Otsukka M, Kallioniemi A, Jukkola-Vuorinen A, Winqvist R. Rare copy number variants observed in hereditary breast cancer cases disrupt genes in estrogen signaling and TP53 tumor suppression network. PLoS Genet 2012;8:e1002734.

10 Masson AL, Talseth-Palmer BA, Evans TJ, Grice DM, Hannan GN, Scott RJ. Expanding the genetic basis of copy number variation in familial breast cancer. Hered Cancer Clin Pract. 2014;12:15.

11 Yang R, Chen B, Pfütze K, Buch S, Steinke V, Holinski-Feder E, et al. Genome-wide analysis associates familial colorectal cancer with increases in copy number variations and a rare structural variation at 12p12.3. Carcinogenesis. 2014;35:315-23.

12 Venkatachalam R, Verwiel ET, Kamping EJ, Hoenselaar E, Görgens H, Schackert HK, et al. Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. Int J Cancer. 2011;129:1635-42.

13 Talseth-Palmer BA, Holliday EG, Evans TJ, McEvoy M, Attia J, Grice DM, et al. Continuing difficulties in interpreting CNV data: lessons from a genome-wide CNV association study of Australian HNPCC/lynch syndrome patients. BMC Med Genomics. 2013;6:10.

14 Brose K, Bland KS, Wang KH, Arnott D, Henzel W, Goodman CS. Slit proteins bind Robo receptors and have an evolutionarily conserved role in repulsive axon guidance. Cell. 1999;96:795-806.

15 Gröne J, Doebler O, Loddenkemper C, Hotz B, Buhr HJ, Bhargava S. Robo1/Robo4: differential expression of angiogenic markers in colorectal cancer. Oncol Rep. 2006;15:1437-43.

16 Ito H, Funahashi S, Yamauchi N, Shibahara J, Midorikawa Y, Kawai S, et al. Identification of ROBO1 as a novel hepatocellular carcinoma antigen and a potential therapeutic and diagnostic target. Clin Cancer Res. 2006;12:3257-64.

17 Sundaresan V, Chung G, Heppell-Parton A, Xiong J, Grundy C, Roberts I, et al. Homozygous deletions at 3p12 in breast and lung cancer. Oncogene. 1998;17:1723-9.

18 Dallol A, Forgacs E, Martinez A, Sekido Y, Walker R, Kishida T, et al. Tumour specific promoter region methylation of the human homologue of the Drosophila Roundabout gene DUTT1 (ROBO1) in human cancers. Oncogene. 2002;21:3020-8.

19 Xian J, Aitchison A, Bobrow L, Corbett G, Pannell R, Rabbitts T, et al. Targeted disruption of the 3p12 gene, Dutt1/Robo1, predisposes mice to lung adenocarcinomas and lymphomas with methylation of the gene promoter. Cancer Res. 2004;64:6432-7.

20 Je EM, Gwak M, Oh H, Choi MR, Choi YJ, Lee SH, et al. Frameshift mutations of axon guidance genes ROBO1 and ROBO2 in gastric and colorectal cancers with microsatellite instability. Pathology. 2013;45:645-50.

21 Cybulski C, Nazarali S, Narod SA. Multiple primary cancers as a guide to heritability. Int J Cancer. 2014;135:1756-63.

22 Naseem H, Boylan J, Speake D, Leask K, Shenton A, Lalloo F, ET AL. Inherited association of breast and colorectal cancer: limited role of CHEK2 compared with high-penetrance genes. Clin Genet. 2006;70:388-95.

23 Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst. 2004;96:261-8.

24 Quek XC, Thomson DW, Maag JL, Bartonicek N, Signal B, Clark MB, et al. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. Nucleic Acids Res. 2015;43:D168-73.

25 Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25:402-8.

26 Nagai MA, Fregnani JH, Netto MM, Brentani MM, Soares FA. Down-regulation of PHLDA1 gene expression is associated with breast cancer progression. Breast Cancer Res Treat. 2007;106:49-56.

27 Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST; Working Group of the American College of Medical Genetics Laboratory Quality Assurance Committee. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med. 2011;13:680-5.

28 Angeloni D, ter Elst A, Wei MH, van der Veen AY, Braga EA, Klimov EA, et al. Analysis of a new homozygous deletion in the tumor suppressor region at 3p12.3 reveals two novel intronic noncoding RNA genes. Genes Chromosomes Cancer. 2006;45:676-91.

29 Bovolenta M, Neri M, Fini S, Fabris M, Trabanelli C, Venturoli A, et al. A novel custom high density-comparative genomic hybridization array detects common rearrangements as well as deep intronic mutations in dystrophinopathies. BMC Genomics. 2008;9:572.

30 Spielman M, Klopocki E. CNVs of noncoding cis-regulatory elements in human disease. Curr Opin Genet Dev. 2013;23:249-56.

31 Szafranski P, Yang Y, Nelson MU, Bizzarro MJ, Langston C, Stankiewicz P. Novel FOXF1 deep intronic deletion causes lethal lung developmental disorder, alveolar capillary dysplasia with misalignment of pulmonary veins. Hum Mutat. 2013;34:1467-71.

32 Khelifi MM, Ishmukhametova A, Khau Van Kien P, Thorel D, Méchin D, Perelman S, et al. Pure intronic rearrangements leading to aberrant pseudoexon inclusion in dystrophinopathy: a new class of mutations? Hum Mutat. 2011;32:467-75

33 Harland M, Mistry S, Bishop DT, Bishop JA. A deep intronic mutation in CDKN2A is associated with disease in a subset of melanoma pedigrees. Hum Mol Genet. 2001;10:2679-86.

34 Dehainault C, Michaux D, Pagès-Berhouet S, Caux-Moncoutier V, Doz F, Desjardins L, et al. A deep intronic mutation in the RB1 gene leads to intronic sequence exonisation. Eur J Hum Genet. 2007;15:473-7.

35 Clendenning M, Buchanan DD, Walsh MD, Nagler B, Rosty C, Thompson B, et al. Mutation deep within an intron of MSH2 causes Lynch syndrome. Fam Cancer. 2011;10:297-301.

36 Rouleau E, Jesson B, Briaux A, Nogues C, Chabaud V, Demange L, et al. Rare germline large rearrangements in the BRCA1/2 genes and eight candidate genes in 472 patients with breast cancer predisposition. Breast Cancer Res Treat. 2012;133:1179-90.

37 Kuusisto KM, Akinrinade O, Vihinen M, Kankuri-Tammilehto M, Laasanen SL, Schleutker J. Copy number variation analysis in familial BRCA1/2-negative Finnish breast and ovarian cancer. PLoS One. 2013;8:e71802.

38 Mabuchi H, Fujii H, Calin G, Alder H, Negrini M, Rassenti L, et al. Cloning and characterization of CLLD6, CLLD7, and CLLD8, novel candidate genes for leukemogenesis at chromosome 13q14, a region commonly deleted in B-cell chronic lymphocytic leukemia. Cancer Res. 2001;61:2870-7.

39 Tyner JW, Rutenberg-Schoenberg ML, Erickson H, Willis SG, O'Hare T, Deininger MW, et al. Functional characterization in acute myeloid leukemia: a cellular context-dependent activating mutation. Leukemia. 2009;23:1345-8.

40 Mano Y, Aishima S, Fukuhara T, Tanaka Y, Kubo Y, Motomura T, et al. Decreased roundabout 1 expression promotes development of intrahepatic cholangiocarcinoma. Hum Pathol. 2013;44:2419-26.

Table 1. List of neoplasms and treatments received by each index patient

Case	First neoplasm (age at diagnosis)	Treatment	Second neoplasm (age at diagnosis)	Treatment	Third neoplasm (age at diagnosis)	Treatment
P1	Rectosigmoid adenocarcinoma - T3N0M0 (34)	Surgery	-	-	-	-
P2	Colon cancer (57)	Surgery Adjuvant QT	*In situ* lobular carcinoma of the right breast (73)	Surgery Adjuvant TMX	*In situ* ductal carcinoma of the left breast (76)	Surgery Adjuvant RT (50.4 Gy) Adjuvant TMX
P3	Breast cancer - T1N1M0 (48 years)	Neoadjuvant QT Surgery Adjuvant RT Adjuvant QT	Colonic tubular adenoma (53)	Surgery	Pituitary microadenoma (59)	Clinically asymptomatic

Gy: Gray, QT: Chemotherapy, RT: Radiotherapy, TMX: tamoxifen, TNM: Tumor-node-metastasis.

Table 2. List of rare CNVs detected in the three index unrelated patients and six relatives

Case	Agilent 4x180K (hg18) *		Affymetrix CytoScan HD (hg19) *	
	Gain	**Loss**	**Gain**	**Loss**
P1	**9p21.2 (27208901-27263012) *TEK, LINC00032***	**3p12.3 (78990568-79028038) *ROBO1***	**9p21.2 (27190654-27294778) *TEK, LINC00032, EQTN***	**3p12.3 (78905346-78955823) *ROBO1*** 5q14.2 (82375769-82385492) *XRCC4*[a] 16q21 (63583651-63623497)[a]
P2	13q14.3 (49039910-49062297) *RCBTB1*	3p12.3 (78990568-79028038) *ROBO1*	Not evaluated	
P2-2		**3p12.3 (78990568-79028038) *ROBO1***		**3p12.3 (78905346-78955823) *ROBO1***
P2-3	13q14.3 (49039910-49062297) *RCBTB1*	11q4.1 (81259541-81298056)	Not evaluated	
P2-4	13q14.3 (49039910-49062297) *RCBTB1*		**13q14.2 (50140480-50163808) *RCBTB1***	11p13 (31351172-31382950) *DCDC1*[a] 11p14.3 (21816401-21848316)[a]
P2-5	No rare alterations		Not evaluated	
P3		**3p12.3 (78990568-79028038) *ROBO1***	16q24.2 (88190403-88301984) *LOC101928880*[a]	2p21 (45314720-45346666)[a] **3p12.3 (78905346-78955823) *ROBO1*** 16q23.3 (82669957-82704660) *CDH13*[b]
P3-2		3p12.3 (78990568-79028038) *ROBO1*	Not evaluated	
P3-3	No rare alterations		Not evaluated	

* The rare CNVs confirmed by both array platforms are represented in bold.
[a] Regions with only one or two probes in the Agilent platform.
[b] Region with four probes in the Agilent platform
[a] Regions with only one or two probes in the Agilent platform.
[b] Region with four probes in the Agilent platform

Figure legends

Fig. 1 Pedigree charts of the Brazilian families with germline deep intronic mutations in the *ROBO1* gene: P1 (A), P2 (B) and P3 (C). Black arrows identify index patients, while grey arrows indicate the relatives tested by microarray.

Fig. 2 *ROBO1* copy number alterations by microarrys, qPCR and Sanger sequencing. (A) Genomic deletion on intron 4 of the *ROBO1* gene, mapped on 3p12.3, identified both by Agilent 4x180K and Affymetrix CytoScan HD microarray platforms. (B) Representation of the three deletions present in DGV that partially overlapped the deletion found in our patients. (C) Graphical analysis of qPCR results. Female cases include P2, P3 and P2-2, while male cases are represented by P2-3, P2-4 and P2-5. Primer pair S1 mapped in intron 4 (a) and primer S2 covered exon 4 of *ROBO1* gene (b). The dashed lines represent the limit for losses (<0.55) and gains (>1.35). (D) Electropherograms showing the silent *ROBO1* mutations identified in all relatives of patient P2 (P2-2, P2-3, P2-4 and P2-5) (a), in P2-2 and P2-3 only (b) and in P2-4 and P2-5 only (c).

Fig. 3 Examples of cytoplasmic expression of ROBO1 protein in breast (A) and colorectal (B) carcinoma cells. Score 0 = no staining (a/e). Score +1 = weak staining (b/f). Score +2 = moderate staining (c/g). Score 3 = strong staining (d/h). Scores 0/+1 were considered negative, while scores +2/+3 represented positive ROBO1 expression.

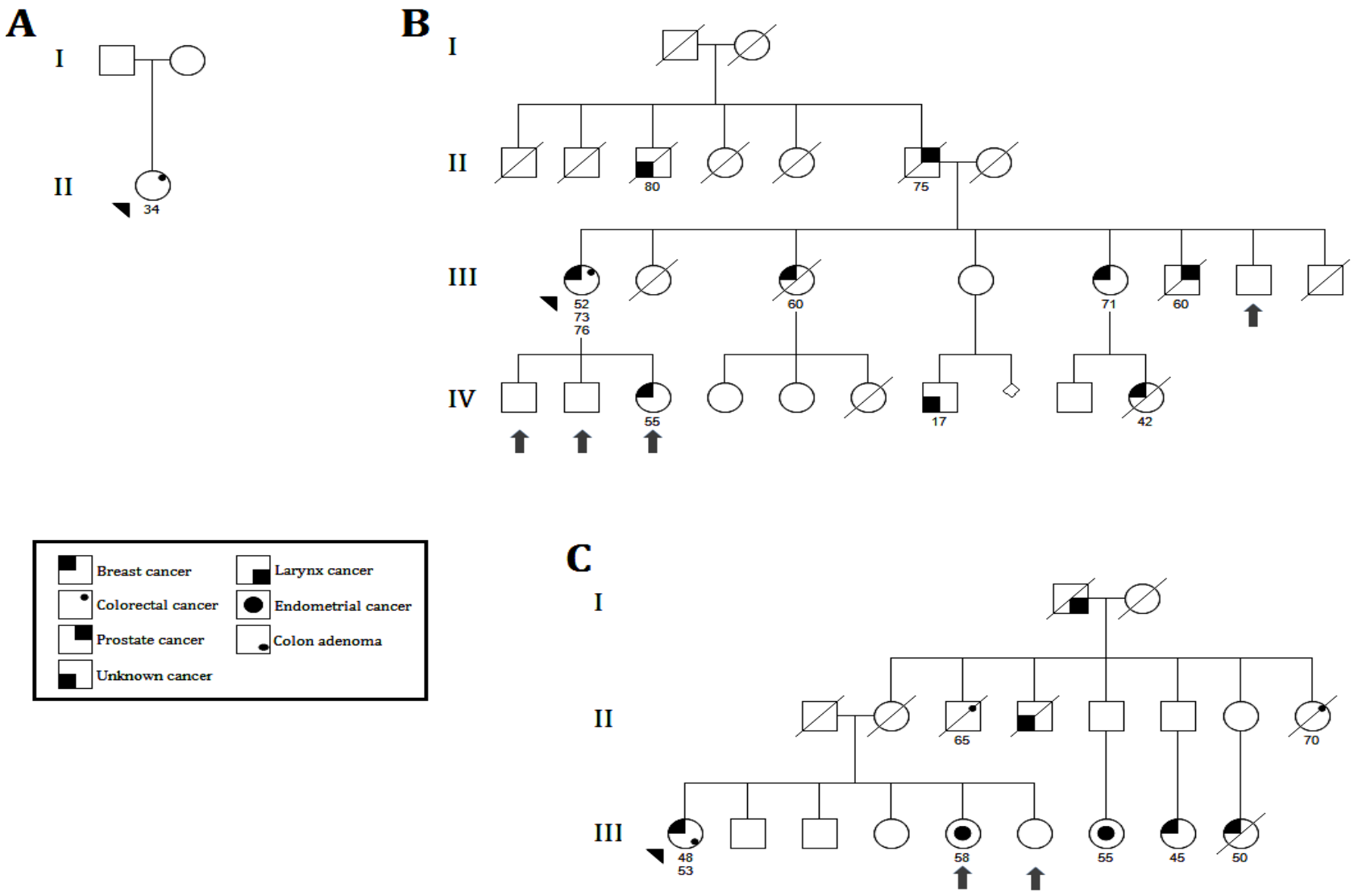
A
I
II
34
B
I
II
80
75
III
52
73
76
60
71
60
IV
55
17
42
Breast cancer
Colorectal cancer
Prostate cancer
Unknown cancer
Larynx cancer
Endometrial cancer
Colon adenoma
C
I
II
65
70
III
48
53
58
55
45
50

Figure 1

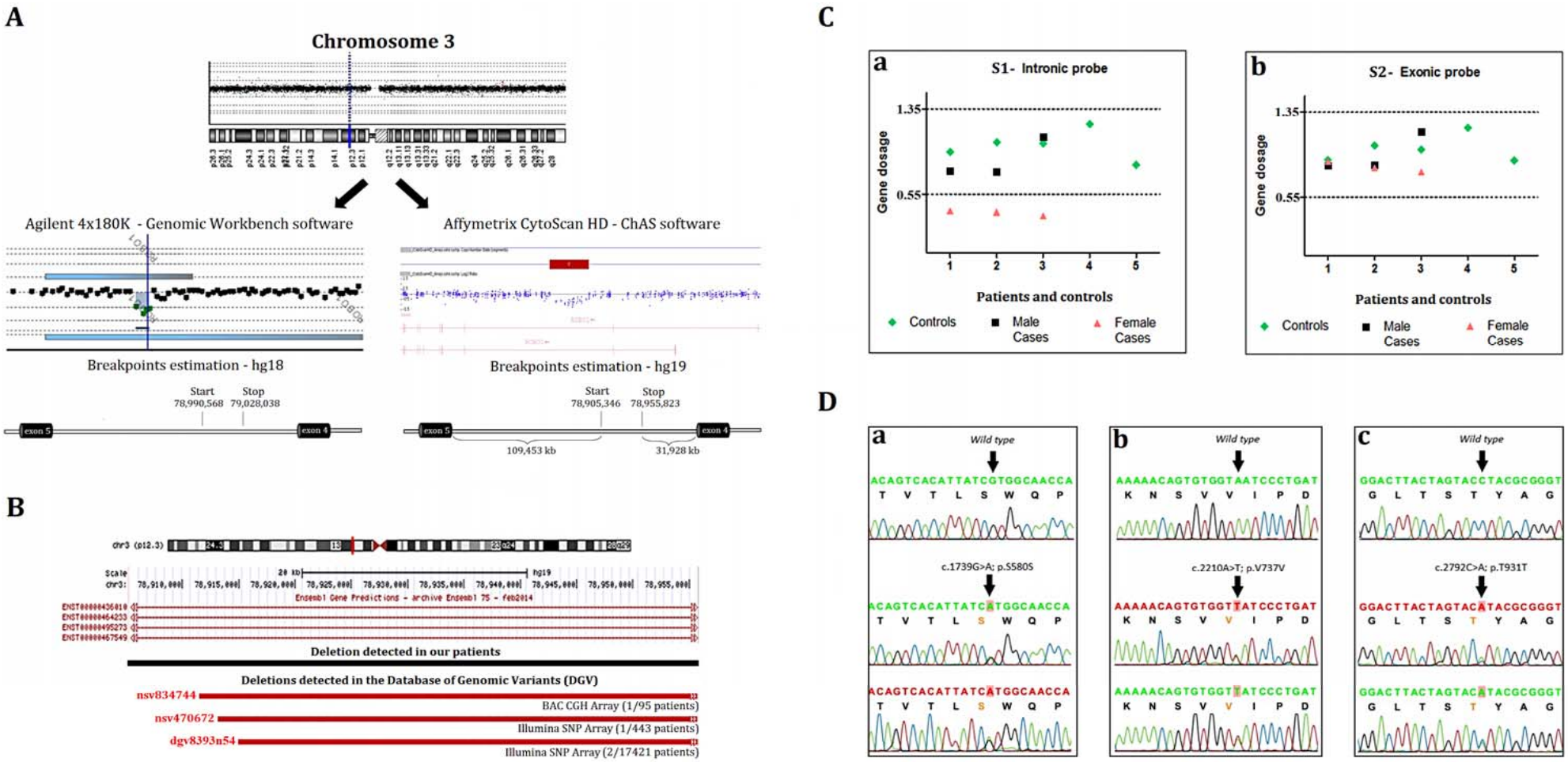
A
Chromosome 3
Agilent 4x180K - Genomic Workbench software
Affymetrix CytoScan HD - ChAS software
Breakpoints estimation - hg18
Start 78,990,568 Stop 79,028,038
Breakpoints estimation - hg19
Start 78,905,346 Stop 78,955,823
exon 5
exon 4
109,453 kb
31,928 kb
B
Deletion detected in our patients
Deletions detected in the Database of Genomic Variants (DGV)
nsv834744
BAC CGH Array (1/95 patients)
nsv470672
Illumina SNP Array (1/443 patients)
dgv8393n54
Illumina SNP Array (2/17421 patients)
C
a
S1- Intronic probe
Gene dosage
1.35
0.55
Patients and controls
Controls
Male Cases
Female Cases
b
S2- Exonic probe
D
a
Wild type
ACAGTCACATTATCGTGGCAACCA
T V T L S W Q P
c.1739G>A; p.S580S
ACAGTCACATTATCATGGCAACCA
b
Wild type
AAAAACAGTGTGGTAATCCCTGAT
K N S V V I P D
c.2210A>T; p.V737V
AAAAACAGTGTGGTTATCCCTGAT
c
Wild type
GGACTTACTAGTACCTACGCGGGT
G L T S T Y A G
c.2792C>A; p.T931T
GGACTTACTAGTACATACGCGGGT

Figure 2

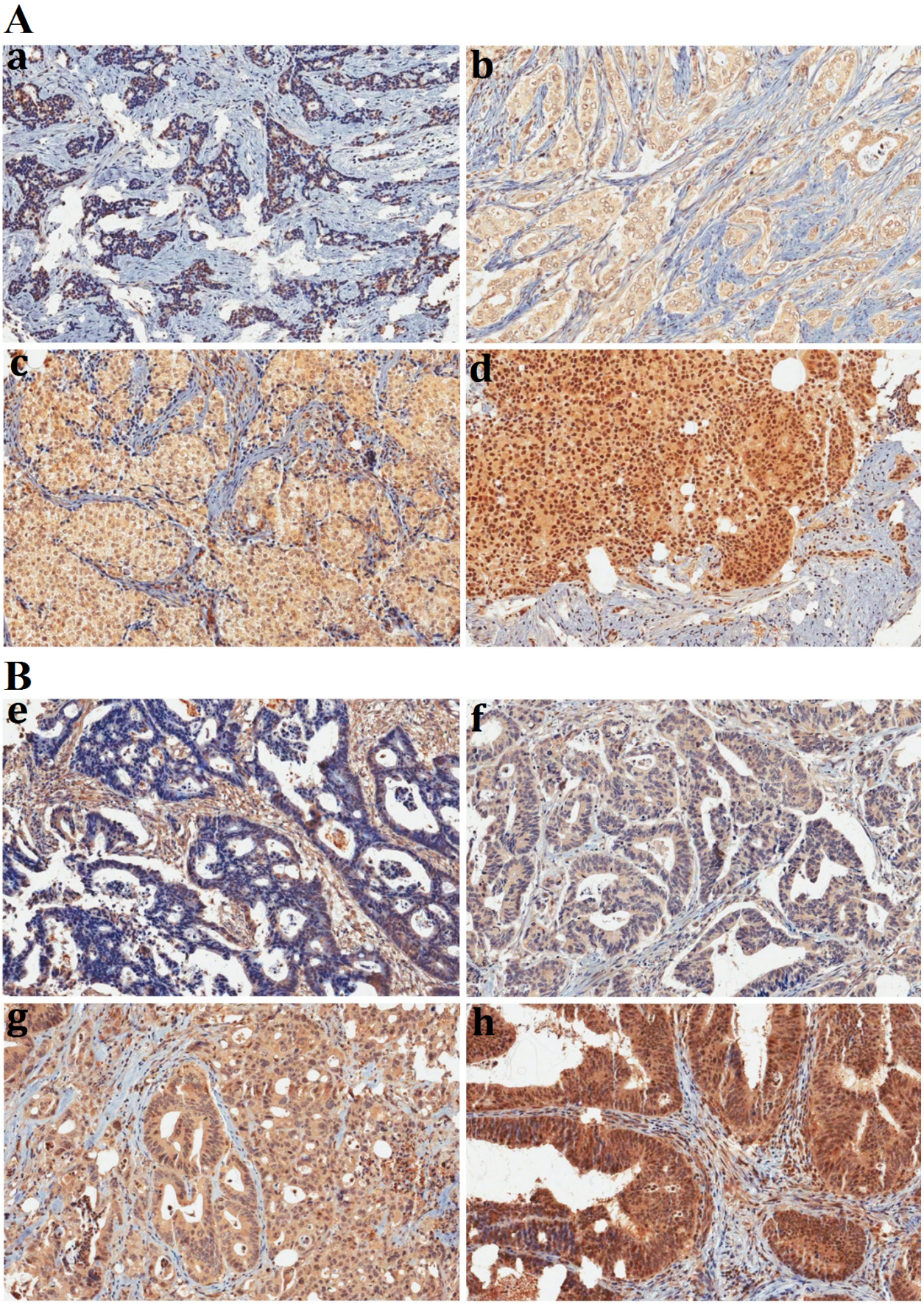

Figure 3

Supplemental files

Supplemental Table 1 List of alterations identified in genes associated with hereditary colorectal and breast cancers

Supplemental Table 2 Primer sequences used in the qPCR experiments

Supplemental Table 3 List of primers used for Sanger sequencing of ROBO1 gene

Supplemental Table 4 Results from the Mendelian Error Checking Analysis (.xlsx)

Supplemental Table 5 Correlation between the expression of ROBO1 protein and clinicopathologic features of BC and CRC samples from patients with and without family history of cancer

Supplemental Fig. 1 Ten-years overall survival curves for BC and CRC patients according to ROBO1 protein expression

5 CONCLUSÕES

1. A análise genômica em 54 pacientes com CCR em idade jovem e/ou história familiar de CCR, mas sem mutações patogênicas nos principais genes conhecidos de predisposição a este tipo tumoral, permitiu identificar CNVs raras englobando novos genes candidatos ao risco de desenvolvimento destes tumores.

2. A plataforma de microarranjos de alta-resolução Affymetrix CytoScan HD confirmou (por meio das sondas de SNPs e de CNVs) as 9 CNVs raras detectadas com a plataforma Agilent 4x180K, além de identificar 33 novas alterações genômicas em 17 pacientes.

3. Foram identificadas 43 CNVs raras afetando pelo menos um gene em 28 pacientes utilizando tanto a plataforma da Agilent como da Affymetrix. A análise funcional *in silico* mostrou que mais da metade dos genes (52/81) afetados pelas CNVs raras estavam associadas com câncer, principalmente com a carcinogênese colorretal (26 genes). Estes resultados sugerem fortemente que a disrupção de um ou mais genes contidos nestas CNVs podem interferir no risco ao CCR observado nestas famílias.

4. Alguns dos genes afetados pela CNVs raras fazem parte de uma mesma família de genes previamente associados com o risco aumentado de CCR,

como no caso do *GALNT6*, *GALNT11*, *CENPP* e *SEMA3C*. Estes achados dão evidência adicional à possível patogenicidade das CNVs identificadas.

5. A análise de segregação de CNVs raras em três famílias mostrou que as CNVs englobando os genes *CHODL*, *GPR39*, *GALNT11* e *KMT2C* tem o potencial para estarem envolvidas com o desenvolvimento de neoplasias nessas famílias.

6. Foi identificada uma deleção intrônica no gene *ROBO1* em uma paciente com CCR em idade jovem. Esta mesma deleção foi detectada previamente em duas pacientes com história familiar de câncer de mama em um estudo prévio do nosso grupo. A análise de segregação em parentes de duas pacientes mostrou que a alteração só está presente em pacientes com neoplasias. Tomados em conjunto, estes resultados sugerem o envolvimento do *ROBO1* com a predisposição ao CCR e câncer de mama.

7. A genotipagem de 23 SNPs associados ao risco de CCR permitiu verificar que 26 pacientes testados apresentaram mais de 50% de alelos de risco para os SNPs genotipados.

8. A análise de SNPs mostrou que uma parcela de pacientes apresenta um número elevado de alelos de risco, o que poderia contribuir com a predisposição ao CCR, uma vez que já foi demonstrado o efeito cumulativo destes loci no risco de desenvolvimento de câncer.

9. Os dois SNP que foram genotipados em 24 pacientes utilizando tanto o método TaqMan como a plataforma da Affymetrix apresentaram resultados 100% concordantes. Este resultado mostra a confiabilidade dos dados gerados pelo nosso estudo.

10. O fato das CNVs raras detectadas não estarem mapeadas nas mesmas regiões dos SNPs de risco ou cobrindo genes associados com estes SNPs, sugere que estas alterações contribuem para a predisposição ao CCR. Além disso, nossos resultados também sugerem que mais de uma CNV rara poderia atuar na predisposição ao CCR familial, do mesmo modo que já foi descrito para SNPs comuns de baixa penetrância.

6 REFERÊNCIAS BIBLIOGRÁFICAS

Aaltonen LA, Peltomäki P, Mecklin JP, et al. Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. **Cancer Res** 1994; 54:1645-8.

Almal SH, Padh H. Implications of gene copy-number variation in health and diseases. **J Hum Genet** 2012; 57:6-13.

Al-Tassan NA, Whiffin N, Hosking FJ, et al. A new GWAS and meta-analysis with 1000Genomes imputation identifies novel risk variants for colorectal cancer. **Sci Rep** 2015; 5:10442.

Baert-Desurmont S, Charbonnier F, Houivet E, et al. Clinical relevance of 8q23, 15q13 and 18q21 SNP genotyping to evaluate colorectal cancer risk. **Eur J Hum Genet** 2015; doi: 10.1038/ejhg.2015.72.

Baglietto L, Lindor NM, Dowty JG, et al. Risks of Lynch syndrome cancers for MSH6 mutation carriers. **J Natl Cancer Inst** 2010; 102:193-201.

Balaguer F, Moreira L, Lozano JJ, et al. Colorectal cancers with microsatellite instability display unique miRNA profiles. **Clin Cancer Res** 2011; 17:6239-49.

Bartley AN, Luthra R, Saraiya DS, Urbauer DL, Broaddus RR. Identification of cancer patients with Lynch syndrome: clinically significant discordances and problems in tissue-based mismatch repair testing. **Cancer Prev Res (Phila)** 2012; 5:320-7.

Bellido F, Pineda M, Aiza G, et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. **Genet Med** 2015; doi: 10.1038/gim.2015.75.

Boland CR, Thibodeau SN, Hamilton SR, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. **Cancer Res** 1998; 58:5248-57.

Bosetti C, Bertuccio P, Malvezzi M, et al. Cancer mortality in Europe, 2005-2009, and an overview of trends since 1980. **Ann Oncol** 2013; 24:2657-71.

Bouzourene H, Hutter P, Losi L, Martin P, Benhattar J. Selection of patients with germline MLH1 mutated Lynch syndrome by determination of MLH1 methylation and BRAF mutation. **Fam Cancer** 2010; 9:167-72.

Bozzao C, Lastella P, Stella A. Anticipation in lynch syndrome: where we are where we go. **Curr Genomics** 2011; 12:451-65.

Chatenoud L, Bertuccio P, Bosetti C, Malvezzi M, Levi F, Negri E, La Vecchia C. Trends in mortality from major cancers in the Americas: 1980-2010. **Ann Oncol** 2014; 25:1843-53.

Clarke E, Green RC, Green JS, et al. Inherited deleterious variants in GALNT12 are associated with CRC susceptibility. **Hum Mutat** 2012; 33:1056-8.

Colas C, Coulet F, Svrcek M, et al. Lynch or not Lynch? Is that always a question? **Adv Cancer Res** 2012; 113:121-66.

Conrad DF, Pinto D, Redon R, et al. Origins and functional impact of copy number variation in the human genome. **Nature** 2010; 464:704-12.

da Silva FC, de Oliveira LP, Santos EM, et al. Frequency of extracolonic tumors in Brazilian families with Lynch syndrome: analysis of a hereditary colorectal cancer institutional registry. **Fam Cancer** 2010; 9:563-70.

DeRycke MS, Gunawardena SR, Middha S, et al. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. **Cancer Epidemiol Biomarkers Prev** 2013; 22:1239-51.

Dominguez-Valentin M, Therkildsen C, Da Silva S, Nilbert M. Familial colorectal cancer type X: genetic profiles and phenotypic features. **Mod Pathol** 2015; 28:30-6.

Duraturo F, Liccardo R, Cavallo A, De Rosa M, Grosso M, Izzo P. Association of low-risk MSH3 and MSH2 variant alleles with Lynch syndrome: probability of synergistic effects. **Int J Cancer** 2011; 129:1643-50.

Elsayed FA, Kets CM, Ruano D, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. **Eur J Hum Genet** 2015; 23:1080-4.

Esteban-Jurado C, Garre P, Vila M, t al. New genes emerging for colorectal cancer predisposition. **World J Gastroenterol** 2014; 20:1961-71.

Esteban-Jurado C, Vila-Casadesús M, Garre P, et al. Whole-exome sequencing identifies rare pathogenic variants in new predisposition genes for familial colorectal cancer. **Genet Med** 2015; 17:131-42.

Garre P, Martín L, Sanz J, et al. BRCA2 gene: a candidate for clinical testing in familial colorectal cancer type X. **Clin Genet** 2015; 87:582-7.

Gelb BD, Chung WK. Complex genetics and the etiology of human congenital heart disease. **Cold Spring Harb Perspect Med** 2014; 4:a013953.

George CM, Alani E. Multiple cellular mechanisms prevent chromosomal rearrangements involving repetitive DNA. **Crit Rev Biochem Mol Biol** 2012; 47:297-313.

Giráldez MD, López-Dóriga A, Bujanda L, et al. Susceptibility genetic variants associated with early-onset colorectal cancer. **Carcinogenesis** 2012; 33:613-9.

Goodenberger M, Lindor NM. Lynch syndrome and MYH-associated polyposis: review and testing strategy. **J Clin Gastroenterol** 2011; 45:488-500.

Gylfe AE, Katainen R, Kondelin J, et al. Eleven candidate susceptibility genes for common familial colorectal cancer. **PLoS Genet** 2013; 9:e1003876.

Hampel H, Frankel WL, Martin E, et al. Feasibility of screening for Lynch syndrome among patients with colorectal cancer. **J Clin Oncol** 2008; 26:5783-8.

Harkness EF, Barrow E, Newton K, et al. Lynch syndrome caused by MLH1 mutations is associated with an increased risk of breast cancer: a cohort study. **J Med Genet** 2015; pii: jmedgenet-2015-103216.

Hassen S, Boman BM, Ali N, et al. Detection of DNA mismatch repair proteins in fresh human blood lymphocytes--towards a novel method for hereditary non-polyposis colorectal cancer (Lynch syndrome) screening. **J Exp Clin Cancer Res** 2011; 30:100.

Hienonen T, Laiho P, Salovaara R, et al. Little evidence for involvement of MLH3 in colorectal cancer predisposition. **Int J Cancer** 2003; 106:292-6.

Hollox EJ, Hoh BP. Human gene copy number variation and infectious disease. **Hum Genet** 2014; 133:1217-33.

Houlle S, Charbonnier F, Houivet E, et al. Evaluation of Lynch syndrome modifier genes in 748 MMR mutation carriers. **Eur J Hum Genet** 2011; 19:887-92.

Houlston RS; members of COGENT. COGENT (COlorectal cancer GENeTics) revisited. **Mutagenesis** 2012; 27:143-51.

Hsu L, Jeon J, Brenner H, et al. A model to determine colorectal cancer risk using common genetic susceptibility Loci. **Gastroenterology** 2015; 148:1330-9.

Iafrate AJ, Feuk L, Rivera MN, et al. Detection of large-scale variation in the human genome. **Nat Genet** 2004; 36:949-51.

Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. **Gastroenterology** 2010; 138:2044-58.

Jiao S, Peters U, Berndt S, et al. Estimating the heritability of colorectal cancer. **Hum Mol Genet** 2014; 23:3898-905.

Jiricny J. Postreplicative mismatch repair. **Cold Spring Harb Perspect Biol** 2013; 5:a012633.

Kempers MJ, Kuiper RP, Ockeloen CW, et al. Risk of colorectal and endometrial cancers in EPCAM deletion-positive Lynch syndrome: a cohort study. **Lancet Oncol** 2011; 12:49-55.

Kidd JM, Cooper GM, Donahue WF, et al. Mapping and sequencing of structural variation from eight human genomes. **Nature** 2008; 453:56-64.

Kirov G. CNVs in neuropsychiatric disorders. **Hum Mol Genet** 2015; pii: ddv253.

Klarskov L, Holck S, Bernstein I, Nilbert M. Hereditary colorectal cancer diagnostics: morphological features of familial colorectal cancer type X versus Lynch syndrome. **J Clin Pathol** 2012; 65:352-6.

Kovacs ME, Papp J, Szentirmay Z, Otto S, Olah E. Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome. **Hum Mutat** 2009; 30:197-203.

Krepischi AC, Pearson PL, Rosenberg C. Germline copy number variations and cancer predisposition. **Future Oncol** 2012a; 8:441-50.

Krepischi AC, Achatz MI, Santos EM, et al. Germline DNA copy number variation in familial and early-onset breast cancer. **Breast Cancer Res** 2012b; 14:R24.

Kuiper RP, Vissers LE, Venkatachalam R, et al. Recurrence and variability of germline EPCAM deletions in Lynch syndrome. **Hum Mutat** 2011; 32:407-14.

Kuusisto KM, Akinrinade O, Vihinen M, Kankuri-Tammilehto M, Laasanen SL, Schleutker J. Copy number variation analysis in familial BRCA1/2-negative Finnish breast and ovarian cancer. **PLoS One** 2013; 8:e71802.

Lanner JT, Georgiou DK, Joshi AD, Hamilton SL. Ryanodine receptors: structure, expression, molecular details, and function in calcium release. **Cold Spring Harb Perspect Biol** 2010; 2:a003996.

Lastella P, Patruno M, Forte G, et al. Identification and surveillance of 19 Lynch syndrome families in southern Italy: report of six novel germline mutations and a common founder mutation. **Fam Cancer** 2011; 10:285-95.

Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. **N Engl J Med** 2000; 343:78-85.

Ligtenberg MJ, Kuiper RP, Chan TL, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. **Nat Genet** 2009; 41:112-7.

Lindblom A, Tannergard P, Werelius B, et al. Genetic mapping of a second locus predisposing to hereditary nonpolyposis colorectal cancer. **Nat Genet** 1993; 5:279-82.

Lindor NM, Rabe K, Petersen GM, et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. **JAMA** 2005; 293:1979-85.

Llor X, Pons E, Xicola RM, et al. Differential features of colorectal cancers fulfilling Amsterdam criteria without involvement of the mutator pathway. **Clin Cancer Res** 2005; 11:7304-10.

Lynch HT, Shaw MW, Magnuson CW, et al. Hereditary factors in cancer: study of two large Midwestern kindreds. **Arch Intern Med** 1966; 117:206-12.

Lynch HT, Lynch PM, Lanspa CL, et al. Review of the Lycnh syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. **Clin Genet** 2009; 76:1-18.

Masson AL, Talseth-Palmer BA, Evans TJ, et al. Copy number variation in hereditary non-polyposis colorectal cancer. **Genes (Basel)** 2013; 4:536-55.

Masson AL, Talseth-Palmer BA, Evans TJ, Grice DM, Hannan GN, Scott RJ. Expanding the genetic basis of copy number variation in familial breast cancer. **Hered Cancer Clin Pract** 2014; 12:15.

Mastrocola AS, Heinen CD. Nuclear reorganization of DNA mismatch repair proteins in response to DNA damage. **DNA Repair (Amst)** 2010; 9:120-33.

Middeldorp A, Jagmohan-Changur S, van Eijk R, et al. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. **Cancer Epidemiol Biomarkers Prev** 2009; 18:3062-7.

Mills RE, Walter K, Stewart C, et al. Mapping copy number variation by population-scale genome sequencing. **Nature** 2011; 470:59-65.

Ministério da Saúde. Instituto Nacional do Câncer José Alencar Gomes Da Silva. **Estimativa/2014 incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2014.

Mjelle R, Hegre SA, Aas PA, et al. Cell cycle regulation of human DNA repair and chromatin remodeling genes. **DNA Repair (Amst)** 2015; 30:53-67.

Mork ME, You YN, Ying J, et al. High prevalence of hereditary cancer syndromes in adolescents and young adults with colorectal cancer. **J Clin Oncol 2015;** pii: JCO.2015.61.4503.

Mulley JC, Mefford HC. Epilepsy and the new cytogenetics. **Epilepsia** 2011; 52:423-32.

Newton K, Jorgensen NM, Wallace AJ, et al. Tumour MLH1 promoter region methylation testing is an effective prescreen for Lynch Syndrome (HNPCC). **J Med Genet** 2014; 51:789-96.

Nicolaides NC, Papadopoulos N, Liu B, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. **Nature** 1994; 371:75-80.

Nieminen TT, O'Donohue MF, Wu Y, et al. Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. **Gastroenterology** 2014; 147:595-598.e5.

Niittymäki I, Kaasinen E, Tuupanen S, et al. Low-penetrance susceptibility variants in familial colorectal cancer. **Cancer Epidemiol Biomarkers Prev** 2010; 19:1478-83.

Oliveira Ferreira F, Napoli Ferreira CC, Rossi BM, et al. Frequency of extra-colonic tumors in hereditary nonpolyposis colorectal cancer (HNPCC) and familial colorectal cancer (FCC) Brazilian families: An analysis by a Brazilian Hereditary Colorectal Cancer Institutional Registry. **Fam Cancer** 2004; 3:41-7.

Palles C, Cazier JB, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. **Nat Genet** 2013; 45:136-44.

Palomaki GE, McClain MR, Melillo S, Hampel HL, Thibodeau SN. EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. **Genet Med** 2009; 11:42-65.

Park RW, Kim TM, Kasif S, Park PJ. Identification of rare germline copy number variations over-represented in five human cancer types. **Mol Cancer** 2015; 14:25.

Parsons MT, Buchanan DD, Thompson B, Young JP, Spurdle AB. Correlation of tumour BRAF mutations and MLH1 methylation with germline mismatch repair (MMR) gene mutation status: a literature review assessing utility of tumour features for MMR variant classification. **J Med Genet** 2012; 49:151-7.

Peltomäki P, Aaltonen L, Sistonen P. Genetic mapping of a locus predisposing to human colorectal cancer. **Science** 1993; 260:810-2.

Peltomäki P, Vasen H. Mutations associated with HNPCC predisposition - Update of ICG-HNPCC/INSiGHT mutation database. **Dis Markers** 2004; 20:269-76.

Pérez-Cabornero L, Infante Sanz M, Velasco Sampedro E, et al. Frequency of rearrangements in Lynch syndrome cases associated with MSH2: characterization of a new deletion involving both EPCAM and the 5' part of MSH2. **Cancer Prev Res (Phila)** 2011; 4:1556-62.

Pérez-Carbonell L, Ruiz-Ponte C, Guarinos C, et al. Comparison between universal molecular screening for Lynch syndrome and revised Bethesda guidelines in a large population-based cohort of patients with colorectal cancer. **Gut** 2012; 61:865-72.

Pittman AM, Naranjo S, Webb E, et al. The colorectal cancer risk at 18q21 is caused by a novel variant altering SMAD7 expression. **Genome Res** 2009; 19:987-93.

Plaschke J, Engel C, Krüger S, et al. Lower incidence of colorectal cancer and later age of disease onset in 27 families with pathogenic MSH6 germline mutations compared with families with MLH1 or MSH2 mutations: the German Hereditary Nonpolyposis Colorectal Cancer Consortium. **J Clin Oncol** 2004; 22:4486-94.

Pylkäs K, Vuorela M, Otsukka M, Kallioniemi A, Jukkola-Vuorinen A, Winqvist R. Rare copy number variants observed in hereditary breast cancer cases disrupt genes in estrogen signaling and TP53 tumor suppression network. **PLoS Genet** 2012; 8:e1002734.

Quehenberger F, Vasen HF, van Houwelingen HC. Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. **J Med Genet** 2005; 42:491-6.

Real LM, Ruiz A, Gayán J, et al. A colorectal cancer susceptibility new variant at 4q26 in the Spanish population identified by genome-wide association analysis. **PLoS One** 2014; 9:e101178.

Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. **Nature** 2006; 444:444-54.

Rodriguez-Bigas MA, Boland CR, Hamilton SR, et al. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. **J Natl Cancer Inst** 1997; 89:1758-62.

Samadder NJ, Jasperson K, Burt RW. Hereditary and common familial colorectal cancer: evidence for colorectal screening. **Dig Dis Sci** 2015; 60:734-47.

Sánchez-Tomé E, Rivera B, Perea J, et al. Genome-wide linkage analysis and tumoral characterization reveal heterogeneity in familial colorectal cancer type X. **J Gastroenterol** 2015; 50:657-66.

Schulz E, Klampfl P, Holzapfel S, et al. Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X. **Nat Commun** 2014; 5:5191.

Seguí N, Pineda M, Navarro M, et al. GALNT12 is not a major contributor of familial colorectal cancer type X. **Hum Mutat** 2014; 35:50-2.

Seguí N, Mina LB, Lázaro C, et al. Germline Mutations in FAN1 Cause Hereditary Colorectal Cancer by Impairing DNA Repair. **Gastroenterology** 2015; pii: S0016-5085(15)00783-0.

Senter L, Clendenning M, Sotamaa K, et al. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. **Gastroenterology** 2008; 135:419-28.

Shlien A, Malkin D. Copy number variations and cancer susceptibility. **Curr Opin Oncol** 2010; 22:55-63.

Smith CE, Mendillo ML, Bowen N, et al. Dominant mutations in S. cerevisiae PMS1 identify the Mlh1-Pms1 endonuclease active site and an exonuclease 1-independent mismatch repair pathway. **PLoS Genet** 2013; 9:e1003869.

Snowden T, Acharya S, Butz C, Berardini M, Fishel R. hMSH4-hMSH5 recognizes Holliday Junctions and forms a meiosis-specific sliding clamp that embraces homologous chromosomes. **Mol Cell** 2004; 15:437-51.

Spier I, Holzapfel S, Altmüller J, et al. Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. **Int J Cancer** 2015; 137:320-31.

Stankiewicz P, Lupski JR. Structural variation in the human genome and its role in disease. **Annu Rev Med** 2010; 61:437-55.

Stoffel E, Mukherjee B, Raymond VM, et al. Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome. **Gastroenterology** 2009; 137:1621-7.

Talseth-Palmer BA, Brenne IS, Ashton KA, et al. Colorectal cancer susceptibility loci on chromosome 8q23.3 and 11q23.1 as modifiers for disease expression in Lynch syndrome. **J Med Genet** 2011; 48:279-84.

Talseth-Palmer BA, Wijnen JT, Brenne IS, et al. Combined analysis of three Lynch syndrome cohorts confirms the modifying effects of 8q23.3 and 11q23.1 in MLH1 mutation carriers. **Int J Cancer** 2013; 132:1556-64.

Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. **CA Cancer J Clin** 2015; 65:87-108.

Tutlewska K, Lubinski J, Kurzawski G. Germline deletions in the EPCAM gene as a cause of Lynch syndrome - literature review. **Hered Cancer Clin Pract** 2013; 11:9.

Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. **J Natl Cancer Inst** 2004; 96:261-8.

Valle L. Genetic predisposition to colorectal cancer: where we stand and future perspectives. **World J Gastroenterol** 2014; 20:9828-49.

Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC. **Dis Colon Rectum** 1991; 34:424-5.

Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. **Gastroenterology** 1999; 116:1453-6.

Venkatachalam R, Verwiel ET, Kamping EJ, et al. Identification of candidate predisposing copy number variants in familial and early-onset colorectal cancer patients. **Int J Cancer** 2011; 129:1635-42.

Walsh MD, Buchanan DD, Cummings MC, et al. Lynch syndrome-associated breast cancers: clinicopathologic characteristics of a case series from the colon cancer family registry. **Clin Cancer Res** 2010; 16:2214-24.

Wang Q, Lasset C, Desseigne F, et al. Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and hMSH6 genes in 75 French kindreds with nonpolyposis colorectal cancer. **Hum Genet** 1999; 105:79-85.

Weissman SM, Burt R, Church J, et al. Identification of individuals at risk for Lynch syndrome using targeted evaluations and genetic testing: National Society of Genetic Counselors and the Collaborative Group of the Americas on Inherited Colorectal Cancer joint practice guideline. **J Genet Couns** 2012; 21:484-93.

Warthin AS. Heredity with reference to carcinoma as shown by the study of the cases examined in the pathological laboratory of the University of Michigan, 1895–1913. **Arch Intern Med** 1913; 12:546-55.

Whiffin N, Hosking FJ, Farrington SM, et al. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. **Hum Mol Genet** 2014; 23:4729-37.

Whiffin N, Houlston RS. Architecture of inherited susceptibility to colorectal cancer: a voyage of discovery. **Genes (Basel)** 2014; 5:270-84.

Wijnen JT, Brohet RM, van Eijk R, et al. Chromosome 8q23.3 and 11q23.1 variants modify colorectal cancer risk in Lynch syndrome. **Gastroenterology** 2009; 136:131-7.

Yang R, Chen B, Pfütze K, et al. Genome-wide analysis associates familial colorectal cancer with increases in copy number variations and a rare structural variation at 12p12.3. **Carcinogenesis** 2014; 35:315-23.

Zhang B, Jia WH, Matsuda K, et al. Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. **Nat Genet** 2014; 46:533-42.

Zhang JX, Fu L, de Voer RM, et al. Candidate colorectal cancer predisposing gene variants in Chinese early-onset and familial cases. **World J Gastroenterol** 2015; 21:4136-49.

Zhang K, Civan J, Mukherjee S, Patel F, Yang H. Genetic variations in colorectal cancer risk and clinical outcome. **World J Gastroenterol** 2014; 20:4167-77.

Anexo 1 - Carta de aprovação do Comitê de Ética em Pesquisa-CEP

Hospital A.C.Camargo
Centro de Tratamento, Ensino e Pesquisa em Câncer

Comitê de Ética em Pesquisa - CEP

São Paulo, 29 de Julho de 2011.

À
Dra. Silvia Regina Rogatto

Ref.: Projeto de Pesquisa nº. 1175/08-C
"Variações do número de cópias (Copy Number Variations – CNVs) em Síndrome de Câncer Hereditário e Agregação Familial: Carcinomas colorretal (Síndrome de Lynch) – Carcinomas de mama e ovário – Síndrome de Li Fraumeni".

Os membros do Comitê de Ética em Pesquisa em Seres Humanos da Fundação Antonio Prudente – Hospital do Câncer - A.C. Camargo/SP, em sua última reunião de 12/07/2011, **tomaram conhecimento e aprovaram** a solicitação de dispensa da apreciação do estudo intitulado **"Alterações genômicas em pacientes e seus familiares com a Síndrome do Câncer Colorretal Hereditário",** em razão deste estudo estar inserido em projeto supracitado, previamente aprovado por este comitê.

Atenciosamente,

Dr. Alexandre Sá de Andrade
2º Vice-Coordenador do Comitê de Ética em Pesquisa

Fundação Antonio Prudente – CNPJ/MF N. 60.961.968/0001-06
Rua Prof. Antônio Prudente, 211 – Liberdade – São Paulo, SP – 01509-900
Telefone: (11) 2189-5000
www.accamargo.org.br

Human Pathology (2011) **42**, 1221–1229

ELSEVIER

Human PATHOLOGY

www.elsevier.com/locate/humpath

Original contribution

Cortactin is associated with perineural invasion in the deep invasive front area of laryngeal carcinomas☆

Eliane Papa Ambrosio PhD[a,b], Fabíola Encinas Rosa PhD[a,b], Maria Aparecida Custódio Domingues MD, PhD[c], Rolando André Rios Villacis MSc[b], Renata de Almeida Coudry MD, PhD[d], José Vicente Tagliarini MD, PhD[e], Fernando Augusto Soares MD, PhD[d], Luiz Paulo Kowalski MD, PhD[f], Silvia Regina Rogatto PhD[b,g,*]

[a]*Institute of Biosciences, UNESP-São Paulo State University, Botucatu, São Paulo, Brazil*
[b]*NeoGene Laboratory, AC Camargo Cancer Treatment and Research Center, São Paulo, Brazil*
[c]*Department of Pathology, UNESP-São Paulo State University, Botucatu, São Paulo, Brazil*
[d]*Department of Anatomic Pathology, AC Camargo Hospital, São Paulo, São Paulo, Brazil*
[e]*Department of Ophtalmology and Otorhinolaryngology, Faculty of Medicine, UNESP-São Paulo State University, Botucatu, São Paulo, Brazil*
[f]*Department of Head and Neck Surgery and Otorhinolaryngology, AC Camargo Hospital, São Paulo, São Paulo, Brazil*
[g]*Department of Urology, Faculty of Medicine, UNESP-São Paulo State University, Botucatu, São Paulo, Brazil*

Received 4 March 2010; revised 24 May 2010; accepted 26 May 2010

Keywords:
Deep invasive front;
Laryngeal squamous cell carcinomas;
Cortactin;
TMA;
qRT-PCR

Summary The cortactin gene, mapped at 11q13, has been associated with an aggressive clinical course in many cancers because of its function of invasiveness. This study evaluated CTTN protein and its prognostic value in the deep invasive front and superficial areas of laryngeal squamous cell carcinomas. The transcript expression levels were evaluated in a subset of cases. Overexpression of CTTN cytoplasmatic protein (80% of cases in both the deep invasive front and superficial areas) and transcript (30% of samples) was detected in a significant number of cases. In more than 20% of cases, observation verified membrane immunostaining in the deep invasive front and superficial areas. Perineural invasion was significantly associated with N stage and recurrence (P = .0058 and P = .0037, respectively). Higher protein expression levels were correlated with perineural invasion (P = .004) in deep invasive front cells, suggesting that this area should be considered a prognostic tool in laryngeal carcinomas. Although most cases had moderate to strong CTTN expression on the tumor surface, 2 sets of cases revealed a differential expression pattern in the deep invasive front. A group of cases with absent to weak expression of CTTN in the deep invasive front showed good prognosis parameters, and a second group with moderate to strong expression of CTTN were associated with an unfavorable prognosis, suggesting an association with worse outcome. Taken together, these results suggest that the deep

☆ This research was supported by FAPESP and CNPq.

* Corresponding author. NeoGene Laboratory, Fundação Antonio Prudente, Hospital AC Camargo, Rua Professor Antonio Prudente, 211, São Paulo, Brazil 01509-010.

E-mail addresses: rogatto@fmb.unesp.br, silvia.rogatto@hcancer.org.br (S. R. Rogatto).

0046-8177/$ – see front matter
doi:10.1016/j.humpath.2010.05.030

invasive front might be considered a grading system in laryngeal carcinomas and that cortactin is a putative marker of worse outcome in the deep invasive front of laryngeal carcinomas.

1. Introduction

Amplification is a frequent event in cancer and is often located in regions that harbor genes implicated in tumor initiation or progression. In head and neck carcinomas, amplification at 11q13 is one of the most prevalent genetic alterations and has been correlated with tumor grade, lymph node metastasis, tumor recurrence, and decreased overall survival (OS) [1].

It has been demonstrated that the 11q13 amplicon harbors several genes with oncogenic potential. The cortactin gene (*CTTN/EMS1*), identified as a candidate within this amplicon, is frequently overexpressed in breast and head and neck cancers [2]. In tumor cell lines, the down-regulation of cortactin decreases cellular motility and ability to migrate, whereas overexpression results in increased invasive potential [3,4]. Supporting the hypothesis that cortactin promotes tumor invasion and systemic spread, some studies demonstrated that overexpression of this protein in cancer cells increases the development of metastasis in nude mouse models [5,6].

In laryngeal squamous cell carcinomas (LSCCs), cortactin was considered the best predictor for decreased survival [7] as compared with cyclin D1 and FADD (Fas [TNFRSF6]-associated via death domain), also mapped at the 11q13 amplicon. Recently, Rodrigo et al [8] reported that *CTTN* amplification was concomitantly accompanied by increased mRNA and protein expression in LSCC. Furthermore, CTTN overexpression was strongly correlated with lymph node metastasis and reduced disease-specific survival. According to the authors, *CTTN* amplification showed a major impact on prognosis and survival in patients with laryngeal tumors.

Cortactin overexpression is reported to inhibit the ubiquitination-mediated degradation of the epidermal growth factor receptor (EGFR), resulting in sustained ligand-induced epidermal growth factor receptor activity [9]. Its overexpression promotes resistance to the EGFR kinase inhibitor gefitinib [10], indicating that cortactin affects not only invasive capacity but also therapeutic responsive properties. Recently, Clark et al [11] demonstrated that cortactin expression modulates multiple cellular traits that may permit survival in a tumor environment, suggesting that the frequent overexpression of cortactin in tumors is not an epiphenomenon but rather promotes tumor aggressiveness.

Taken together with the biological markers, it is believed that the most useful prognostic information can be deduced from the deep invasive front of tumors, where the deepest and presumably most aggressive cells reside [12-14]. It has been also postulated that many crucial molecular interactions that enhance or inhibit tumor progression occur at the tumor-host interface, including aberrant expression of molecular markers involved in cell adhesion [15]. The prognostic significance of the tumor front in oral squamous cell carcinomas has been recognized in several studies [16-20]. Jakobsson [21] reported a pilot study in 230 glottis carcinomas of the larynx treated by radiotherapy using 8 morphologic criteria including mode of invasion, stage of invasion, vascular invasion, and cellular response. After multivariate analysis, the author demonstrated that the mode of invasion was one of the most important factors of prediction of the disease in 5 years. To the best of our knowledge, this is the first study concerning LSCC that considers the deep invasive front in the evaluation of candidate genes and proteins as prognostic markers.

The aim of this study was to determine the clinical significance of *CTTN* in the deep invasive front and superficial regions of LSCC in a large series of samples.

2. Material and methods

2.1. Patients and tumor specimens

LSCC specimens were obtained from AC Camargo Hospital, São Paulo, Brazil. Forty-seven frozen samples were used to investigate transcript expression by quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR) and 151 formalin-fixed, paraffin-embedded cases to evaluate protein expression by immunohistochemistry (IHC) analysis in a tissue microarray (TMA). Patients were followed up prospectively with a mean follow-up of 35 ± 28 months (3-115 months) and 76 ± 101 months (1-280 months) for qRT-PCR and IHC analysis, respectively. All samples were from untreated patients before surgery. The eligibility criteria included previously untreated patients submitted to treatment with curative intent in the institution. The medical records of all patients were examined to obtain detailed demographic data (age, sex, and race), information regarding lifestyle (alcoholic beverage and tobacco consumption), and clinicopathologic data (clinical stage, lymph node involvement, histologic grade, angiolymphatic invasion, perineural invasion, and deep invasive front; Supplemental Table 1). The histologic classification system for HNSCC (head and neck squamous cell carcinomas) is based on differentiation grade: well differentiated, moderately differentiated, and poorly differentiated. The invasion front

is based on growth pattern, according to Bryne et al [22]. Angiolymphatic invasion was classified according to the presence or absence of neoplasic cells, located both in the wall and upon examination of the blood or lymphatic vessels; perineural infiltration was considered present when the tissue adjacent to the peri- and/or intratumoral nerves was surrounded by neoplasic cells. All patients were advised of the procedures and provided written informed consent, as approved by the institution's ethics committee.

2.2. Quantitative real-time RT-PCR

Total RNA was extracted from pulverized frozen tumor tissue using Trizol reagent (Invitrogen Life Technologies Inc, Carlsbad, CA), according to the manufacturer's instructions. Forty-seven laryngeal carcinomas and 6 normal adjacent larynx tissue samples were digested with Dnase I Amplification Grade (Life Technologies, Rockville, MD) and reverse transcribed using SuperScript II reverse transcriptase (Invitrogen Life Technologies Inc), as previously described [23]. The cDNA was stored at −70°C.

PCR amplification was performed in a StepOnePlus Real-Time PCR System (v2.0; Applied Biosystems, Foster City, CA). Primers for *CTTN* and *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) control reference genes were designed using the Primer Express software (v2.0; Applied Biosystems) as 5′-GGGCCACTATCCCGCAGA-3′ and 5′-CCGTCGCCCTGTACGACTAC-3′ for the *CTTN* gene and 5′-GGCCTCCAAGGAGTAAGACC-3′ and 5′-AGGGGTCTACATGGCAACTG-3′ for *GAPDH*. Quantitative data were analyzed using the Sequence Detection System software (v1.0; Applied Biosystems). PCR reactions were conducted in a total volume of 10 μL using Power SYBR Green PCR Master Mix (Applied Biosystems), according to the manufacturer's instructions, and an ABI Prism 7000 Sequence Detection System (Applied Biosystems). Dissociation and standard curves for all primers were constructed. PCR efficiency (E) was calculated according to the following equation: $E = 10^{(-1/\text{slope})} - 1$. Gene relative quantification (RQ) was calculated according to Pfaffl [24]. The transcript levels were considered up-regulated (RQ ≥2.0) or down-regulated (RQ ≤0.5).

2.3. Immunohistochemistry

A TMA was constructed containing sections from the deep invasive front and superficial regions of 151 LSCC paraffin-embedded samples: 101 paired superficial and deep invasive front, 17 tumor surface, and 33 tumor invasion front nonpaired. Core biopsies were extracted from previously defined areas using a Tissue Microarrayer (Beecher Instruments, Silver Spring, MD). Tissue cores with a dimension of 1.0 mm from each specimen were punched and arrayed in duplicate on a recipient paraffin block. Each core was spaced 0.2 mm apart. The paraffin-embedded LSCCs were freshly cut (3 μm) and mounted on conventional slides with organosilane (3-aminopropyl triethoxy-silane) (Sigma-Aldrich Co, St Louis, MO). Immunohistochemical reactions were performed using the primary antibody anti-CTTN (Becton, Dickinson and Company Biosciences, Franklin Lakes, NJ) (dilution, 1:200). After 30 minutes of incubation, the sections were washed in phosphate-buffered saline, incubated for 30 minutes with secondary antibody (Advanced TM HRP Link, K0690; DakoCytomation, Denmark), followed by incubation with the polymer detection system (Advanced TM HRP Link, DakoCytomation) for 30 minutes. Reactions were developed with a solution containing 0.6 mg/mL of 3,3′-diaminobenzidine tetrahydrochloride (Sigma) and 0.01% H_2O_2. Positive and negative controls were included in all reactions, in accordance with the manufacturer's recommendations.

The IHC reactions were performed in duplicate on different TMA levels, representing a 4-fold redundancy for

Table 1 *CTTN* transcript expression compared with clinicopathologic features in 47 LSCC

Variables	Categories	No. (%)	*CTTN* expression (range) [a]	*P* [c]
Age (y)	≤60	20 (43)	1.91 (0.1-10.3)	
	>60	27 (57)	3.87 (0.3-48.8)	.3777
Sex	Female	11 (23)	5.22 (0.1-48.8)	
	Male	36 (77)	2.37 (0.1-13.9)	.1558
Alcohol consumption	Yes	31 (66)	1.84 (0.1-10.3)	
	No	16 (34)	5.37 (0.1-48.8)	.7876
Tobacco usage	Yes	38 (81)	1.98 (0.1-10.3)	
	No	9 (19)	7.52 (0.3-48.8)	.6265
Histologic grade	I	13 (31)	6.72 (0.3-48.8)	
	II	26 (62)	1.58 (0.1-7.7)	
	III	3 (7)	1.96 (1.1-3.6)	**.0830** [d]
	II+III	29 (69)	1.62 (0.1-7.7)	**.0684**
T stage	T1+T2	14 (30)	5.80 (0.3-48.8)	
	T3+T4	33 (70)	1.87 (0.1-10.3)	.3116
N status	Positive	19 (41)	4.45 (0.1-48.8)	
	Negative	27 (59)	2.14 (0.1-13.9)	.4892
Metastasis	Yes	16 (35)	5.88 (0.1-48.8)	
	No	30 (65)	1.54 (0.1-10.3)	.1737
Recurrence	Yes	17 (36)	2.14 (0.3-13.9)	
	No	30 (64)	3.54 (0.1-48.8)	.5281
Perineural invasion	Yes	10 (36)	1.22 (0.4-3.4)	
	No	18 (64)	2.30 (0.1-10.3)	.8667
Angiolymphatic permeation	Yes	0 (0)	–	
	No	27 (100)	1.70 (0.1-10.3)	–
Death due to cancer	No [b]	13 (35)	1.05 (0.1-2.3)	
	Yes	24 (65)	1.85 (0.3-7.7)	.3992
Familial history of cancer	Yes	24 (51)	2.16 (0.1-10.3)	
	No	23 (49)	3.95 (0.1-48.8)	.3888

NOTE. Marginally significant *P* values are in bold.

[b] Alive without recurrence + alive without disease.

[a] Mean gene expression.

[c] Mann-Whitney *U* test.

[d] Kruskal-Wallis test.

each case. The second slides were 25 sections deeper than the first, resulting in at least 250 μm of distance between the 2 sections with different cell samples for each tumor. IHC scoring was blinded to the outcome and clinical aspects of each tumor specimen. Each slide was scanned in low-power field to determine the most stained area. The presence of a clearly visible dark brown precipitation was considered immunopositivity. The pattern of staining for this antibody is cytoplasmatic; the intensity score represented the estimated staining intensity (0, no visible reaction; 1, weak; 2, moderate; and 3, strong immunostaining intensity) and extension (1, ≤1/3; 2, 1/3-2/3; and 3, >2/3 of the total area). The reactions were further analyzed according to the location of the immunodeposits in both the membrane and cytoplasm.

2.4. Statistical analysis

Associations between variables and risk factors were verified at a 5% of significance level using χ^2 and

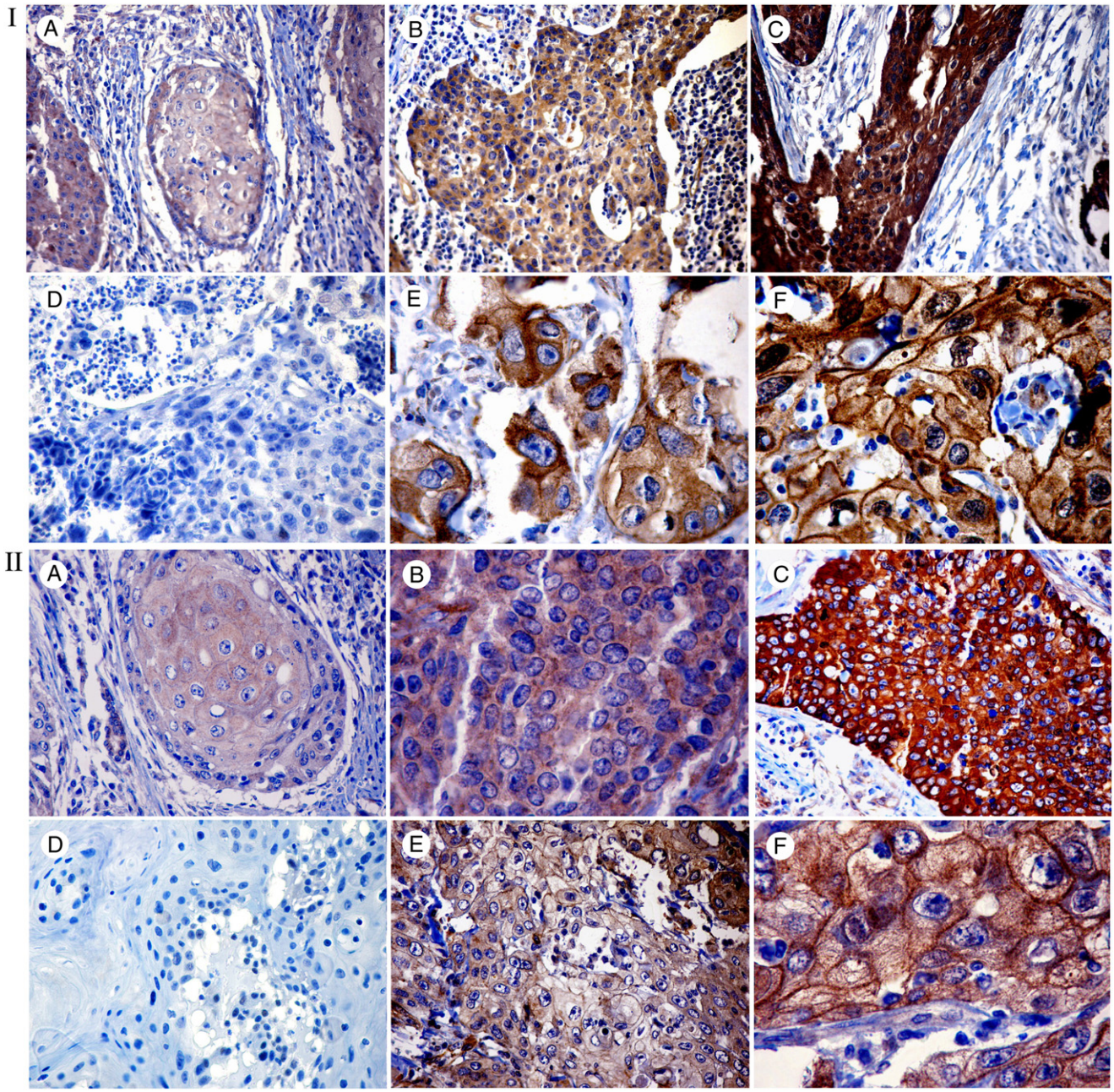

Fig. 1 I: IHC analysis of cortactin at the deep invasive front. A, Score 1 + 3 (intensity + extension) 200×; B, score 2 + 3 400×; C, score 3 + 3 400×; D, negative expression 400x; E, score 2 + 3 with membrane expression 1000×; F, score 3 + 3 with membrane expression and koilocyte presence 1000×. II: IHC analysis of cortactin at the tumor surface. A, Score 1 + 3 400×; B, score 2 + 3 400×; C, score 3 + 3 1000×; D, negative expression 400×; E, score 1 + 3 with membrane expression 400×; F, score 2 + 3 with membrane expression and koilocyte presence 1000×.

Table 2 Comparison between cortactin expression in paired samples (deep front and superficial area) and clinical and histopathologic data

Variables	Categories	CTTN/F, no.		*P*	CTTN/T, no.		*P*
		0-1	2-3		0-1	2-3	
Histologic grade	G_{1-2}	13	55	1.000	10	58	1.000
	G_{3-4}	0	4		0	4	
T stage	T_{1-2}	3	7	.380	1	9	1.000
	T_{3-4}	10	50		9	51	
N status	Positive	2	26	.039	2	26	.186
	Negative	12	33		9	36	
Second primary tumor	Yes	4	7	.203	3	8	.352
	No	10	53		8	55	
Metastasis	Yes	2	14	.718	2	14	1.000
	No	11	43		8	46	
Perineural invasion	Yes	0	25	.003	2	23	.477
	No	13	33		8	38	
Angiolymphatic permeation	Yes	0	4	1.000	1	3	.468
	No	13	53		9	57	
Recurrence	Yes	4	18	1.000	6	16	.074
	No	10	42		5	47	

NOTE. Bonferroni correction ($P < .006$) was applied for histopathologic data comparisons. Abbreviations: CTTN/F, cortactin expression in deep front; CTTN/T, cortactin expression in superficial area.

Fisher exact tests. Kruskal-Wallis or Mann-Whitney *U* tests were applied to compare the *CTTN* gene expression and the clinicopathologic data. The samples were also grouped according to deep invasive front type (infiltrative, expansive, and mixed) and membrane staining (positive and negative) for comparisons with clinicopathologic data and CTTN expression. The Bonferroni correction for multiple comparisons was applied to adjust the *P* value. OS was defined as the interval between the beginning of treatment (surgery) and the date of death or the last information for censored observations. The disease-free interval was measured from the date of treatment to the date when recurrence was diagnosed. Survival probabilities were estimated by the Kaplan-Meier method, and the log-rank test was applied to assess the significance of differences among actuarial survival curves with a confidence interval

Table 3 Distribution of clinical and histopatological data according to perineural invasion

		All cases			Paired cases, front area			Paired cases, superficial area		
		Perineural invasion		*P*	Perineural invasion		*P*	Perineural invasion		*P*
		+	−		+	−		+	−	
Total no. of cases		50	86		25	33		23	38	
Histologic grade	G_{1-2}	46	79	1.000	24	30	.6267	22	35	1.000
	G_{3-4}	4	7		1	3		1	3	
T stage	T_{1-2}	6	14	.4795	2	4	.6862	2	6	.6971
	T_{3-4}	41	66		22	27		20	30	
	NA	3	6		1	2		1	2	
N status	Positive	29	29	.0058	16	10	.0106	16	10	.0009
	Negative	21	57		9	23		7	28	
Second primary tumor	Yes	10	14	.5831	5	2	.2206	4	4	.4610
	No	40	72		20	31		19	34	
Metastasis	Yes	13	12	.0822	7	7	.4722	7	7	.2176
	No	35	70		16	25		14	30	
	NA	2	4		2	1		2	1	
Recurrence	Yes	20	15	.0037	12	6	.0151	11	5	.0029
	No	30	71		13	27		12	33	
Angiolymphatic permeation	Yes	2	5	.6444	1	3	.6267	1	2	.6579
	No	48	81		24	30		22	36	

Abbreviation: NA, not applicable.

of 95%. All analyses were performed using the statistical software package GraphPad Prism 2.01 (GraphPad Software Inc, La Jolla, CA) and SPSS Statistics 17.0.2 (SPSS Inc, Chicago, IL).

3. Results

CTTN overexpression was detected in 14 (30%) of 47 LSCCs (RQ, 2.2-48.8) evaluated by qRT-PCR. No statistical difference was observed in *CTTN* expression levels between tumoral and normal samples (P = .8995; data not shown). Although no significant association was detected between clinicopathologic features and *CTTN* expression levels, 44% of patients with metastasis showed *CTTN* overexpression (7/16 cases) in comparison with patients without metastasis (6/24 cases) (Table 1).

The CTTN protein expression evaluated in deep invasive front and superficial tumor tissues revealed that most samples had CTTN intensity varying from moderate to strong (80% from deep invasive front and 85% from superficial tissue) and extension greater than one third of the total area. Representative immunostaining of both invasive front and tumor surface is presented in Fig. 1.

Although cortactin is a cytoplasmatic protein, 27% of deep invasive front and 20% of superficial area cases had cellular membrane immunostaining. Of these, 9 cases had simultaneous immunostaining in both areas. Considering CTTN-positive membrane immunostaining samples, 67% were from younger patients compared with 47% among negative membrane cases (P = .0284; data not shown).

Eighty-eight samples were classified according to the deep invasive front subtypes as infiltrative (51%), expansive (6%), and mixed (43%). No association was detected when different subtypes were compared for CTTN intensity, membrane immunostaining, and histopathologic variables (data not shown).

The protein immunostaining pattern was compared with clinical and histopathologic data, but no significant associations were observed according to age, sex, alcohol consumption, tobacco use, histologic grade, T stage, N status, metastasis, recurrence, second primary tumor presence, angiolymphatic permeation, or membrane staining (Supplemental Table 2). In deep invasive front cells, the presence of perineural invasion was correlated to cortactin expression (P = .004) after Bonferroni correction (Supplemental Table 2). Furthermore, when only paired deep invasive front and superficial samples (74 cases) were considered, CTTN expression was again significantly associated with perineural invasion (P = .003; Table 2) and marginally associated with lymph node involvement (P = .039; Table 2) in the deep invasive front area. Histologic evidence of perineural invasion (50/136 cases) was significantly associated with N stage and recurrence (P = .0058 and P = .0037, respectively) but not with OS and disease-free survival (data not shown). In the paired cases, comparison with moderate to strong CTTN expression and the presence or absence of perineural invasion, N stage, and recurrence were also significantly correlated with the deep invasive front area (P = .0106 and P = .0151, respectively) and with superficial tissue (P = .0009 and P = .0029, respectively) (Table 3).

The CTTN expression pattern observed differed significantly between deep invasive front and superficial area samples (P = .004; data not shown). Interestingly, 8 samples showed differential expression according to the region: moderate to strong expression in the superficial region and absent to weak expression in the deep invasive front region (11%, group 1; Fig. 2A). However, most samples had moderate to strong expression in the superficial and deep invasive front regions from the same sample (74%, group 2; Fig. 2A). Clinical data of worse prognosis, including the presence of metastasis, positive perineural invasion, lymph node involvement, and positive membrane immunostaining, were more prevalent in group 2 than in group 1 (Fig. 2B).

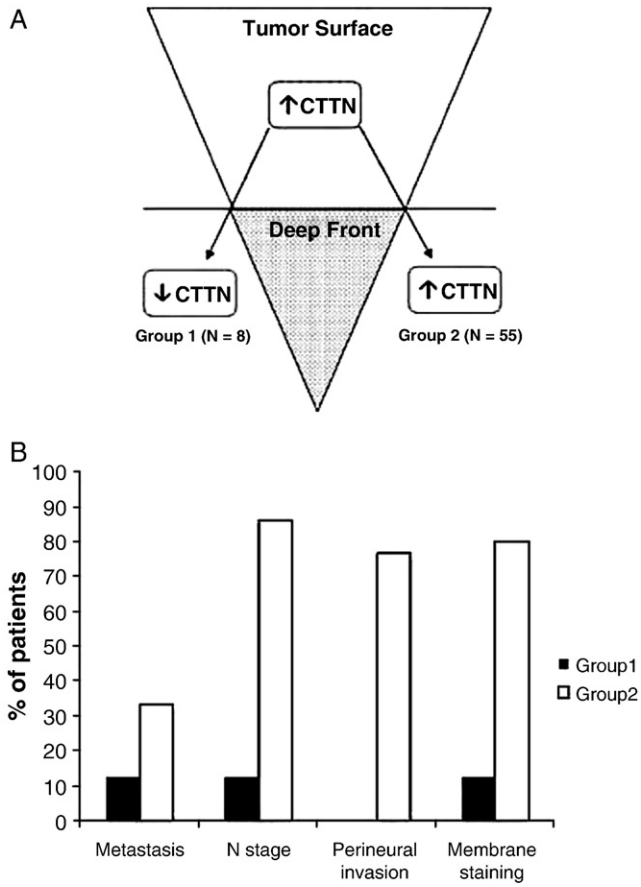

Fig. 2 A, Representation of tumor surface and deep invasive front. Group 1 represents patients with moderate to strong cortactin expression at the tumor surface and absent to weak in the deep invasive front; group 2 indicates patients with moderate to strong cortactin expression in both areas. B, The clinical parameters included in each group are related to CTTN function and worst outcome, including the presence of metastasis, lymph node involvement, perineural invasion, and positive membrane immunostaining.

The Kaplan-Meier method was performed to evaluate OS and disease-free survival. CTTN intensity and extension, presence of cellular membrane immunopositivity, and perineural invasion were considered for the analysis, but no significant values were observed (data not shown).

4. Discussion

It has been suggested that cortactin affects the overall aggressiveness of head and neck carcinomas, regardless of 11q13 amplification status, and promotes tumor growth, survival, vascularization, and invasion. In the present study, analysis of the results demonstrated that the *CTTN* gene and its protein were up-regulated and associated with parameters of worse prognosis in a subset of laryngeal carcinomas.

CTTN mRNA overexpression was observed in 30% of cases. In laryngeal and pharyngeal carcinomas, Rodrigo et al [8] verified *CTTN* up-regulation in 57% of cases. Although no statistical correlation involving metastasis development and transcript overexpression was determined in the present study, a significant number of patients with metastasis showed *CTTN* overexpression (44%). Unfortunately, the small number of patients with metastasis and the relative short-term follow-up in several cases evaluated by qRT-PCR do not permit more elaborate comparisons or conclusions. In xenograft tumor studies, cortactin was reported to enhance metastasis of breast, esophageal squamous cell, and hepatocellular carcinomas to bone, lung, and intrahepatic metastasis, respectively [5,6,25]. The role of overexpressed cortactin in relation to the increase in cell migration and metastatic potential is not only due to its location and function in nontransformed cells; the increase in cell motility is probably due to increased Arp2/3 (actin related protein 2/3 complex, subunit 2, 34kDa) activation and lamellipodia extension [26].

Using IHC, analysis revealed that cortactin was up-regulated in approximately 80% of laryngeal carcinomas. CTTN immunostaining was preferentially detected in the cytoplasm of tumor cells, although some cases showed protein enrichment in the cell membrane, corroborating observations reported by Rodrigo et al [8]. Location and biochemical studies indicate that cortactin plays an important role in regulating cortical actin assembly and organization. In most cell types, cortactin is observed in cytoplasmic punctate structures of unknown composition, concentrated in the perinuclear region, and with F-actin at sites of dynamic peripheral membrane activity. Cortactin translocates from the cytoplasm to the periphery in response to many of the same stimuli that induce its tyrosine phosphorylation, including growth factor treatment, integrin activation, and bacterial entry [27].

CTTN protein expression was compared with the clinicopathologic data, and a significant correlation was found between overexpression and perineural invasion in deep invasive front cells. In addition, perineural invasion was correlated with N stage and recurrence, independent of cortactin expression, because the expression was exclusively correlated with perineural invasion in deep invasive front cells. To our knowledge, no prior study has considered the deep invasive front in laryngeal carcinomas. The first study to develop a simple malignancy grading system restricted to the deep invasive front area was conducted by Bryne et al [22] on oral squamous cell carcinomas. This grading system proved to have additional prognostic value over the established factors [28,29]. Likewise, the data analyzed in this study suggest that the deep invasive front is relevant as a supplementary prognostic tool in laryngeal carcinomas.

Perineural invasion was associated with nodal metastasis and recurrence. Perineural space is known to be a route of extension in squamous cell carcinoma, and it has been widely accepted as an important risk factor for locoregional recurrence and for the decreased survival rate in patients with head and neck carcinoma [30]. Once tumor cells exit the main tumor mass and encroach on a nerve, they may spread both proximally and distally. Involvement often begins near the terminal nerves and spreads proximally and progressively to involve larger trunks. Some genes have already been reported regarding their involvement with perineural invasion in head and neck cancer, such as neural cell adhesion molecules [30]. Furthermore, the CD44v6 isoform, which seems to promote tumor progression, had significantly high expression in head and neck carcinomas showing perineural invasion [31]. In a study using ovarian cells, the authors showed that the binding of hyaluronan to SK-OV-3.ipl cells promotes c-Src kinase recruitment to CD44 and stimulates c-Src kinase activity, which, in turn, increases tyrosine phosphorylation of CTTN. In addition, transfection of SK-OV-3.ipl cells with a dominant active form of c-Src promotes CD44 and c-Src association with CTTN in membrane projections and stimulates ovarian tumor cell migration [32]. Cortactin tyrosine phosphorylation [4] has been correlated with cell migration; however, its effects in the intercellular signaling pathways for cell dynamics control and other functions are still not understood. Based on studies using siRNA [33], it has been hypothesized that hyperphosphorylated cortactin may promote cell migration in cancer cells, indicating the relevance of perineural invasion evaluation when considering CTTN expression in head and neck tumors.

Interesting data in the present study included the verification of 2 clusters according to CTTN expression in the superficial and deep invasive front. In the first group, CTTN expression was moderate to strong in the superficial area and absent to weak in the deep invasive front. This group of patients showed better outcomes in comparison with the second group, in which CTTN expression was moderate to strong in both areas. These data suggest differential cortactin expression in superficial and deep invasive front cells that may be associated with prognosis. In studies using siRNA against cortactin, its diminished expression was correlated with the inhibition of cell motility

[4,34,35]. In addition, several reports suggest that cortactin up-regulation is associated with metastasis [5,6,25] and cell migration [36,37] but not with tumor initiation [35]. A similar study conducted on oral carcinomas reported that E-cadherin transcript/protein expression was heterogeneous between superficial and deep invasive front cells. The authors also reported a correlation between E-cadherin expression at the deep invasive front and invasive front grading score, tumor size, tumor thickness, and 5-year survival [38], confirming the relevance of this region in cancer prognosis. However, improved understanding of the biologic mechanisms that cause the differences in CTTN expression between deep invasive front and superficial areas from the same sample is necessary to elucidate how it affects prognosis in these tumors. Moreover, not only is the CTTN expression level important for the cellular motility process but also its phosphorylation/dephosphorylation regulation, which is necessary for invasive growth and which eventually culminates in distant metastases [39].

In conclusion, the data demonstrated that cortactin is overexpressed in a large number of laryngeal carcinomas, in both deep invasive front and superficial cells. Although not significant, *CTTN* overexpression was verified in patients with metastasis. The association between cortactin up-regulation in deep invasive front cells and perineural infiltration suggests that both this region and the putative marker could be considered a grading system in laryngeal carcinomas.

Supplementary materials related to this article can be found online at doi:10.1016/j.humpath.2010.05.030.

Acknowledgments

We express our gratitude to Dr Sandra A. Drigo for her many helpful suggestions throughout the article's preparation and to Francine Blumental de Abreu for her assistance with the statistical analysis.

References

[1] Rodrigo JP, García LA, Ramos S, Lazo PS, Suárez C. EMS1 gene amplification correlates with poor prognosis in squamous cell carcinomas of the head and neck. Clin Cancer Res 2000;6:3177-82.

[2] Schuuring E, Verhoeven E, Mooi WJ, Michalides RJ. Identification and cloning of two overexpressed genes, U21B31/PRAD1 and EMS1, within the amplified chromosome 11q13 region in human carcinomas. Oncogene 1992;7:355-61.

[3] van Rossum AG, de Graaf JH, Schuuring-Scholtes E, et al. Alternative splicing of the actin binding domain of human cortactin affects cell migration. J Biol Chem 2003;278:45672-9.

[4] Rothschild BL, Shim AH, Ammer AG, et al. Cortactin overexpression regulates actin-related protein 2/3 complex activity, motility, and invasion in carcinomas with chromosome 11q13 amplification. Cancer Res 2006;66:8017-25.

[5] Li Y, Tondravi M, Liu J, et al. Cortactin potentiates bone metastasis of breast cancer cells. Cancer Res 2001;61:6906-11.

[6] Luo ML, Shen XM, Zhang Y, et al. Amplification and overexpression of CTTN (EMS1) contribute to the metastasis of esophageal squamous cell carcinoma by promoting cell migration and anoikis resistance. Cancer Res 2006;66:11690-9.

[7] Gibcus JH, Mastik MF, Menkema L, et al. Cortactin expression predicts poor survival in laryngeal carcinoma. Br J Cancer 2008;98:950-5.

[8] Rodrigo JP, Garcia-Carracedo D, Garcia LA, et al. Distinctive clinicopathological associations of amplification of the cortactin gene at 11q13 in head and neck squamous cell carcinomas. J Pathol 2009;217:516-23.

[9] van Rossum AG, Gibcus J, van der Wal J, Schuuring E. Cortactin overexpression results in sustained epidermal growth factor receptor signaling by preventing ligand-induced receptor degradation in human carcinoma cells. Breast Cancer Res 2005;7:235-7.

[10] Timpson P, Wilson AS, Lehrbach GM, Sutherland RL, Musgrove EA, Daly RJ. Aberrant expression of cortactin in head and neck squamous cell carcinoma cells is associated with enhanced cell proliferation and resistance to the epidermal growth factor receptor inhibitor gefitinib. Cancer Res 2007;67:9304-14.

[11] Clark ES, Brown B, Whigham AS, Kochaishvili A, Yarbrough WG, Weaver AM. Aggressiveness of HNSCC tumors depends on expression levels of cortactin, a gene in the 11q13 amplicon. Oncogene 2009;28:431-44.

[12] Bryne M. Prognostic value of various molecular and cellular features in oral squamous cell carcinoma. J Oral Pathol Med 1991;20:413-20.

[13] Bryne M, Koppang HS, Lilleng R, Kjaerheim A. Malignancy grading of the deep invasive margins of oral squamous cell carcinomas has high prognostic value. J Pathol 1992;166:375-81.

[14] Tumluri V, Thomas GA, Fraser IS. Analysis of Ki-67 antigen at the invasive tumor front of human oral squamous cell carcinoma. J Oral Pathol Med 2003;31:598-604.

[15] Bànkfalvi A, Piffkò J. Prognostic and predictive factors in oral cancer: the role of the invasive tumour front. J Oral Pathol Med 2000;29:291-8.

[16] Graflund M, Sorbe B, Bryne M, Karlsson M. The prognostic value of a histologic grading system, DNA profile, and MIB-1 expression in early stages of cervical squamous cell carcinomas. Int J Gynecol Cancer 2002;12:149-57.

[17] Noguchi M, Kinjyo H, Kohama GI, Nakamori K. Invasive front in oral squamous cell carcinoma: image and flow cytometric analysis with clinicopathologic correlation. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2002;93:682-7.

[18] Po Wing Yuen A, Lam KY, Lam LK, et al. Prognostic factors of clinically stage I and II oral tongue carcinoma—a comparative study of stage, thickness, shape, growth pattern, invasive front malignancy grading, Martinez-Gimeno score, and pathologic features. Head Neck 2002;24:513-20.

[19] Sawair FA, Irwin CR, Gordon DJ, Leonard AG, Stephenson M, Napier SS. Invasive front grading: reliability and usefulness in the management of oral squamous cell carcinoma. J Oral Pathol Med 2003;32:1-9.

[20] Kurokawa H, Zhang M, Matsumoto S, et al. The relationship of the histologic grade at the deep invasive front and the expression of Ki-67 antigen and p53 protein in oral squamous cell carcinoma. J Oral Pathol Med 2005;34:602-7.

[21] Jakobsson PA. Histologic grading of malignancy and prognosis in glottic carcinoma of the larynx. In: Alberti PW, Bryce DP, editors. Workshops from the Centennial Conference on Laryngeal Cancer. New York: Appleton-Century-Crofts; 1976. p. 847-54.

[22] Bryne M, Jenssen N, Boyesen M. Histological grading in the deep invasive front of T1 and T2 glottic squamous cell carcinomas has high prognostic value. Virchows Arch 1995;427:277-81.

[23] Rosa FE, Caldeira JR, Felipes J, et al. Evaluation of estrogen receptor alpha and beta and progesterone receptor expression and correlation with clinicopathologic factors and proliferative marker Ki-67 in breast cancers. Hum Pathol 2008;39:720-30.

[24] Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 2001;29:e45.
[25] Chuma M, Sakamoto M, Yasuda J, et al. Overexpression of cortactin is involved in motility and metastasis of hepatocellular carcinoma. J Hepatol 2004;41:629-36.
[26] Schuuring E, Verhoeven E, Litvinov S, Michalides RJ. The product of the EMS1 gene, amplified and overexpressed in human carcinomas, is homologous to a v-src substrate and is located in cell-substratum contact sites. Mol Cell Biol 1993;13:2891-8.
[27] Weed SA, Parsons JT. Cortactin: coupling membrane dynamics to cortical actin assembly. Oncogene 2001;20:6418-34.
[28] Ondruschka C, Buhtz P, Motsch C, et al. Prognostic value of MMP-2,-9 and TIMP-1,-2 immunoreactive protein at the invasive front in advanced head and neck squamous cell carcinomas. Pathol Res Pract 2002;198:509-15.
[29] Sterz CM, Kulle C, Dakic B, et al. A basal-cell-like compartment in head and neck squamous cell carcinomas represents the invasive front of the tumor and is expressing MMP-9. Oral Oncol 2010;46:116-22.
[30] Chirilă M, Bolboacă SD, Cosgarea M, Tomescu E, Mureşan M. Perineural invasion of the major and minor nerves in laryngeal and hypopharyngeal cancer. Otolaryngol Head Neck Surg 2009;140:165-9.
[31] Wang SJ, Wong G, de Heer AM, Xia W, Bourguignon LY. CD44 variant isoforms in head and neck squamous cell carcinoma progression. Laryngoscope 2009;119:1518-30.
[32] Bourguignon LY, Zhu H, Shao L, Chen YW. CD44 interaction with c-Src kinase promotes cortactin-mediated cytoskeleton function and hyaluronic acid-dependent ovarian tumor cell migration. J Biol Chem 2001;276:7327-36.
[33] Jia L, Uekita T, Sakai R. Hyperphosphorylated cortactin in cancer cells plays an inhibitory role in cell motility. Mol Cancer Res 2008;6:654-62.
[34] Hill A, McFarlane S, Mulligan K, et al. Cortactin underpins CD44-promoted invasion and adhesion of breast cancer cells to bone marrow endothelial cells. Oncogene 2006;25:6079-91.
[35] van Rossum AG, Moolenaar WH, Schuuring E. Cortactin affects cell migration by regulating intercellular adhesion and cell spreading. Exp Cell Res 2006;312:1658-70.
[36] Patel AS, Schechter GL, Wasilenko WJ, et al. Overexpression of EMS1/cortactin in NIH3T3 fibroblasts causes increased cell motility and invasion in vitro. Oncogene 1998;16:3227-32.
[37] Bryce NS, Clark ES, Leysath JL, Currie JD, Webb DJ, Weaver AM. Cortactin promotes cell motility by enhancing lamellipodial persistence. Curr Biol 2005;15:1276-85.
[38] Wang X, Zhang J, Fan M, et al. The expression of E-cadherin at the invasive tumor front of oral squamous cell carcinoma: immunohistochemical and RT-PCR analysis with clinicopathological correlation. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2009;107:547-54.
[39] García-Castillo J, Pedersen K, Angelini PD, et al. HER2 carboxyl-terminal fragments regulate cell migration and cortactin phosphorylation. J Biol Chem 2009;284:25302-13.

OPEN ACCESS Freely available online

Genomic Signatures Predict Poor Outcome in Undifferentiated Pleomorphic Sarcomas and Leiomyosarcomas

Sara Martoreli Silveira[1], Rolando Andre Rios Villacis[1], Fabio Albuquerque Marchi[2], Mateus de Camargo Barros Filho[1], Sandra Aparecida Drigo[3], Cristovam Scapulatempo Neto[4], Ademar Lopes[5], Isabela Werneck da Cunha[6], Silvia Regina Rogatto[1,3]*

1 Neogene Laboratory, A. C. Camargo Cancer Center, São Paulo, São Paulo, Brazil, **2** Institute of Mathematics and Statistics, Inter-Institutional Program on Bioinformatics, USP, São Paulo, São Paulo, Brazil, **3** Department of Urology, Faculty of Medicine, UNESP, Botucatu, São Paulo, Brazil, **4** Department of Pathology, Barretos Cancer Hospital (Pio XII Foundation), Barretos, São Paulo, Brazil, **5** Department of Pelvic Surgery, A. C. Camargo Cancer Center, São Paulo, São Paulo, Brazil, **6** Department of Pathology, A. C. Camargo Cancer Center, São Paulo, São Paulo, Brazil

Abstract

Undifferentiated high-grade pleomorphic sarcomas (UPSs) display aggressive clinical behavior and frequently develop local recurrence and distant metastasis. Because these sarcomas often share similar morphological patterns with other tumors, particularly leiomyosarcomas (LMSs), classification by exclusion is frequently used. In this study, array-based comparative genomic hybridization (array CGH) was used to analyze 20 UPS and 17 LMS samples from untreated patients. The LMS samples presented a lower frequency of genomic alterations compared with the UPS samples. The most frequently altered UPS regions involved gains at 20q13.33 and 7q22.1 and losses at 3p26.3. Gains at 8q24.3 and 19q13.12 and losses at 9p21.3 were frequently detected in the LMS samples. Of these regions, gains at 1q21.3, 11q12.2-q12.3, 16p11.2, and 19q13.12 were significantly associated with reduced overall survival times in LMS patients. A multivariate analysis revealed that gains at 1q21.3 were an independent prognostic marker of shorter survival times in LMS patients (HR = 13.76; $P = 0.019$). Although the copy number profiles of the UPS and LMS samples could not be distinguished using unsupervised hierarchical clustering analysis, one of the three clusters presented cases associated with poor prognostic outcome ($P = 0.022$). A relative copy number analysis for the *ARNT*, *SLC27A3,* and *PBXIP1* genes was performed using quantitative real-time PCR in 11 LMS and 16 UPS samples. Gains at 1q21-q22 were observed in both tumor types, particularly in the UPS samples. These findings provide strong evidence for the existence of a genomic signature to predict poor outcome in a subset of UPS and LMS patients.

Citation: Silveira SM, Villacis RAR, Marchi FA, Barros Filho MdC, Drigo SA, et al. (2013) Genomic Signatures Predict Poor Outcome in Undifferentiated Pleomorphic Sarcomas and Leiomyosarcomas. PLoS ONE 8(6): e67643. doi:10.1371/journal.pone.0067643

Editor: Gerard Roel Rutteman, Utrecht University, Netherlands

Received November 22, 2012; **Accepted** May 22, 2013; **Published** June 25, 2013

Funding: This work was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP (1998/14335-2 and 2008/55693-2) and Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: silvia.rogatto@cipe.accamargo.org.br

Introduction

Sarcomas are a heterogeneous group of mesenchymal tumors that represent approximately 1% of cancers diagnosed in adults and 15% of childhood tumors [1]. Soft-tissue sarcomas (STSs) are classified into two categories. The first group includes tumors with non-pleomorphic morphologies, which are usually associated with genomic translocations and certain specific mutations, and tumors with pleomorphic morphologies, which are associated with complex chromosomal alterations and genomic instability [2]. Leiomyosarcomas (LMSs) and undifferentiated high-grade pleomorphic sarcomas (UPSs) belong to the second STS group.

UPSs, which have been previously known referred to as malignant fibrous histiocytomas (MFHs), represent 5% of STSs diagnosed in adults [3]. Clinically, these aggressive tumors frequently show local recurrence and can metastasize to distant sites [4]. The absence of the lineage with specific differentiation observed in UPS reflects the difficulty of histopathological classification and the reproducibility of sarcoma diagnosis [5]. However, a number of important signaling pathways required for the maintenance of mesenchymal stem cells (MSCs) have been associated with UPS cell tumorigenicity [6,7].

Most UPSs share similar morphologies with undifferentiated and pleomorphic tumor subtypes, particularly LMSs, liposarcomas, and rhabdomyosarcomas [4,8]. LMSs represent more than 20% of STSs. Similar to UPSs, LMSs also display pleomorphic characteristics and often follow an aggressive course [9]. Several studies have evaluated gene-expression profiles from large STS cohorts, and they were unable to distinguish UPSs from LMSs based on hierarchical clustering analysis. However, in some cases, it was possible to identify minor UPS and LMS subgroups with similar gene-expression and/or genomic profiles [10,11,12,13,14].

DNA copy number profiles derived from UPS samples have revealed recurrent genomic alterations that are correlated with

morphological subtypes and patient outcome. These genomic imbalances commonly include gains at the 17q locus, which have been associated with longer disease-free survival times and a lower risk of distant metastasis [15]. In addition, losses of 4q31 and 18q22 have been associated with an increased risk of metastasis and favorable prognosis in UPS and LMS, respectively [16]. Gains at 1p33-p32.3 and 1p21.3 in UPS have been recently associated with increased patient survival times [17]. Unfortunately, DNA copy number studies have evaluated small sample sizes. In addition, the majority of these reports have not described whether the evaluated UPS and LMS samples were obtained from treated or untreated patients. Importantly, accurate diagnoses are essential for these cancer types because distinct diagnostic entities may require different treatment strategies [10].

This study was designed to determine the potential of chromosomal imbalance profiles detected with array CGH methods to reveal biomarkers for diagnosis and/or prognosis. Additionally, the study aimed to identify novel putative molecular targets in untreated patients prior to surgery to improve therapies to treat UPS and LMS.

Patients and Methods

Patients

Thirty seven fresh frozen tissue samples (20 UPS and 17 LMS) were obtained from 36 patients who were followed prospectively at either A.C. Camargo Hospital (São Paulo, Brazil) or Barretos Cancer Hospital (Barretos, São Paulo, Brazil) between 2000 and 2010. The procedures were described to all of the patients, after which time they provided written informed consent. This study was approved by the Ethical Committee in Research of the Antonio Prudente Foundation at A.C. Camargo Hospital (Protocol 1105/08) and by the Ethical Committee in Research of the Pius XII Foundation at Barretos Cancer Hospital (Protocol 302/2010). The medical records of all of the patients were examined to obtain detailed demographic and clinicopathologic data (Table 1), and all of the cases were evaluated by an expert sarcoma pathologist (IWC). The diagnostic criteria were based on World Health Organization (WHO) recommendations and included both the morphology and expression of specific proteins detected using immunohistochemistry [18]. Histological grades were defined according to the recommendations of the Federation Nationale des Centres de Lutte Contre le Cancer (FNCLCC), which considers the mitotic index, tumor necrosis, and cell differentiation [19].

Twenty six out of 37 tumor samples (20 UPS and 17 LMS) were derived from primary tumors, eight from locally recurrent tumors and three from remnant tumors (derived of the surgical margins expansion from different patients). Two primary tumors (UPS8 and UPS18) were derived from the same patient (patient #8, Table 1). None of the patients had received chemotherapy or radiotherapy treatment prior to sample collection. One patient was diagnosed as a Li-Fraumeni Syndrome carrier (patient #34). The average patient age was 59.3 years (ranging from 4–90 years). The anatomical sites commonly affected were lower extremities (14 cases), retroperitoneum (13 cases), trunk (4 cases), upper extremities (4 cases) and head and neck (2 cases). According to the FNCLCC guidelines, the majority of the cases were classified as high histological grade (G2 or G3), and two of the LMS cases (LMS15 and LMS21) were classified as G1. Fourteen patients received only surgical treatment, six underwent neoadjuvant chemotherapy followed by surgery and 15 patients received adjuvant therapy after the surgery (including patient #8). One patient received only chemotherapy without surgery (patient #34). The mean follow-up time was 29.8 months (ranging from 1–109 months). In three patients distant metastases were detected at diagnosis (patients #8, #23 and #32).

A diverse panel of antibodies was used for the immunohistochemical characterization, including SMA (Cell Marque, clone 1A4), Desmin (Ventana, clone DER11), HHF35 (Cell Marque, clone HHF 35), Caldesmon (Dako, clone h-CD), CD34 (Ventana, clone QBEnd 10), CD31 (Ventana, clone JC70), CD99 (Ventana, clone 0.13), S100 (Ventana, clone PAB), NSE (Ventana, clone E27), AE1/AE3 (Ventana, pool), EMA (Ventana, clone E29), HMB45 (Ventana, clone HMB45), MART-1/MelanA (Ventana, clone A-103), CD45 (LCA) (Ventana, clone RP2/18), CD63 (Cell Marque, clone NK1/C3) and MDM-2 (Neomarkers, polyclonal). These markers were used to confirm or exclude other diagnoses, such as melanoma, lymphoma, or undifferentiated tumors. Cases were considered to be pleomorphic sarcomas when tumors showed pleomorphic morphology and were negative for all of the tested markers or when they presented focal expression of the muscle markers SMA, HHF35, Desmin, and/or Caldesmon. Tumors with spindle cell morphology and diffuse expression of muscle markers (SMA, HHF35, Desmin, or Caldesmon) were considered to be LMSs. Cases showing both spindle cell morphology and pleomorphic morphology in addition to the strong or diffuse expression of muscle markers (SMA, HHF35, Desmin, or Caldesmon) were also considered to be LMSs. The retroperitoneal UPS were carefully revised by two pathologists (IWC and CSN) with the aim to exclude occult cases of undifferentiated liposarcoma. Negative staining for MDM2 at immunohistochemistry and absence of any region with differentiated liposacoma features kept the diagnosis at UPS upon morphological grounds. In addition, FISH analysis was performed to investigate *MDM2* amplification in a subgroup of retroperitoneal sarcomas (3 UPS and 3 LMS) (Text S1). None of these cases had detectable *MDM2* amplification (Kreatech MDM2/CEN12, Amsterdam, ND) (Figure A and Table A in Text S1).

Array-based Comparative Genomic Hybridization (array CGH)

Genomic DNA was extracted using a standard phenol/chloroform-based method. Genomic DNA samples from tumors and normal tissue (Promega, Madison, WI, USA) were differentially labeled using the Genomic DNA Enzymatic Labeling Kit (Agilent Technologies, Santa Clara, CA, USA). The hybridizations were performed on Agilent Human CGH 44 K Oligo Microarrays according to the manufacturer's recommendations. The array images were acquired with a DNA microarray scanner using SureScan High-Resolution Technology and the Scan Control (version 8.1) software program (Agilent Technologies, Santa Clara, CA, USA). The data were analyzed using the Nexus Copy Number (version 6.0, Biodiscovery Inc., El Segundo, CA, USA) software program [20]. The Fast Adaptive States Segmentation Technique 2 (FASST2) algorithm and the Significance Testing for Aberrant Copy number (STAC) statistical method were used to identify non-random genomic copy number alterations [21]. Based on these algorithms, DNA copy number alterations were defined as instances that exceeded a significance threshold of 1×10^{-5} and that contained at least five consecutive altered probes per segment. These parameters were used to define the following: copy number gain ($\geq$0.2), high copy number gain ($\geq$0.6), copy number loss ($\leq$0.2), and homozygous loss (≤-1.0). Genomic data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE45573 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc = GSE45573).

Table 1. Clinical and histopathological data from patients (20 UPS and 17 LMS).

Patient	Sample	Age	Sex	Sample origin	Location	TNM	Grade	Local Recurrence	Distance Metastasis	Treatment (QT or RT)	Follow-up (months)[c]
1	UPS1[d]*	51	F	Primary tumor	Lower extremity	T1aN0M0	III	Presence	–	0	NED (108)
2	UPS2*	52	F	Recurrence	Retroperitoneum	T2bN0M0	III	Presence	–	2	DD (33)
3	UPS3[d]*	49	M	Primary tumor	Retroperitoneum	T2bN0M0	III	Presence	–	2	DD (22)
4	UPS4*	90	M	Primary tumor	Lower extremity	T2aN0M0	III	–	–	2	DD (8)
5	UPS5[d]	50	M	Primary tumor	Retroperitoneum	T2bN0M0	III	Presence	–	2	DD (8)
6	UPS6	56	M	Primary tumor	Lower extremity	T2bN0M0	III	–	–	2	NED (44)
7	UPS7*	58	F	Recurrence	Head and Neck	T2aN0M0	III	Presence	–	2	LF (29)
8	UPS8[a,*]	63	M	Primary tumor	Head and Neck	T1aN0M1	III	–	–	2	DD (15)
	UPS18[a]*			Primary tumor	Lower extremity	T2bN0M1	III	–	Lung and adrenal (MD)	2	DD (7)
9	UPS9*	72	M	Primary tumor	Lower extremity	T2bN0M0	III	–	–	1	NED (30)
10	UPS13*	63	M	Recurrence	Lower extremity	T2bN0M0	III	Presence	Lung	1	DD (18)
11	UPS14*	32	F	Primary tumor	Trunk	T1aN0M0	III	–	–	0	NED (30)
12	UPS15*	80	M	Recurrence	Retroperitoneum	T2bN0M0	III	Presence	–	0	LF (11)
13	UPS16*	60	M	Primary tumor	Trunk	T2aN0M0	III	–	–	1	NED (1)
14	UPS17	56	M	Primary tumor	Lower extremity	T2bN0M0	III	–	–	1	NED (20)
15	UPS19	77	M	Primary tumor	Upper extremity	T2bN0M0	III	–	–	2	LF (33)
16	UPS20*	41	M	Primary tumor	Retroperitoneum	T2bN0M0	III	–	–	2	NED (6)
17	UPS21*	82	F	Primary tumor	Lower extremity	T2bN0M0	III	–	–	0	DD (3)
18	UPS22*	78	M	Primary tumor	Upper extremity	T2bN0M0	III	Presence	Lung	0	AD (11)
19	UPS23*	60	M	Recurrence	Lower extremity	T2aN0M0	III	Presence	Lung and bones	2	AD (5)
20	LMS3	60	F	Recurrence	Upper extremity	T1aN0M0	II	Presence	-	0	NED (64)
21	LMS4*	89	F	Recurrence	Trunk	T2bN0M0	II	Presence	Lung	0	DD (12)
22	LMS5	77	F	Recurrence	Trunk	T2bN0M0	III	–	Pelvis	0	DD (54)
23	LMS6	54	F	Primary tumor	Retroperitoneum	T2bN0M1	III	Presence	Liver (MD)	2	DD (15)
24	LMS7*	37	M	Primary tumor	Lower extremity	T1aN0M0	III	–	–	1	NED (45)
25	LMS8	61	F	Primary tumor	Retroperitoneum	T2bN0M0	III	–	Liver	0	DD (28)
26	LMS9	62	M	Primary tumor	Lower extremity	T2bN0M0	II	–	–	2	NED (99)
27	LMS15*	49	M	Primary tumor	Lower extremity	T2bN0M0	I	–	–	0	NED (65)
28	LMS16*	50	M	Recurrence	Retroperitoneum	T2bN0M0	III	Presence	–	0	NED (109)
29	LMS17	48	M	Primary tumor	Retroperitoneum	T2bN0M0	III	–	–	0	LF (4)
30	LMS18*	61	M	Primary tumor	Retroperitoneum	T2bN0M0	III	Presence	–	2	AD (36)
31	LMS19[b,*]	4	M	Primary tumor	Lower extremity	T2bN0M0	III	–	–	2	NED (25)
32	LMS20*	81	F	Primary tumor	Lower extremity	T2bN0M1	III	–	Lung, liver and abdominal wall (MD)	1	DD (10)
33	LMS21*	52	F	Primary tumor	Retroperitoneum	T2bN0M0	I	–	–	0	NED (23)
34	LMS22[b,*]	74	F	Primary tumor	Retroperitoneum	T2bN0M0	II	–	–	3	LF (2)

Table 1. Cont.

Patient	Sample	Age	Sex	Sample origin	Location	TNM	Grade	Local Recurrence	Distance Metastasis	Treatment (QT or RT)	Follow-up (months)[c]
35	LMS23*	45	F	Primary tumor	Retroperitoneum	T2bN0M0	III	Presence	–	0	NED (21)
36	LMS24*	58	F	Primary tumor	Upper extremity	T2aN0M0	III	–	–	2	NED (48)

Abbreviations - F: Female, M: Male, DD: Death by disease, NED: No evidence of disease, AD: Alive with disease, LF: Loss of follow-up; MD: metastasis at diagnosis.
Treatment – QT: Chemotherapy; RT: Radiotherapy; 0: Surgery; 1: Neoadjuvant therapy; 2: Adjuvant therapy; 3: Chemotherapy without surgery.
*Selected for qPCR validation.
[a]Samples from the same patient.
[b]Patients with Li-Fraumeni Syndrome.
[c]Time to last follow-up from diagnosis.
[d]Samples of different patients obtained from expansion of primary tumor surgical (remnant of primary tumor).
doi:10.1371/journal.pone.0067643.t001

Quantitative Real-Time PCR (qPCR)

The genomic DNA sequences of candidate regions were obtained from the Ensembl Genome Browser website (GRCh37/hg19 Human Reference Assembly; February 2009). Primer sequences were designed using the Primer-Blast online software tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). Eight primer pairs (*ARNT*-P1, *ARNT*-P2, *ARNT*-P3, *PBXIP1*-P1, *PBXIP1*-P2, *SLC27A3*-P1, *CCND1*-P1, and *CCND1*-P2) were designed to amplify the altered regions detected using array CGH, including the 60-nucleotide probe present on the Agilent platform (Table S1). Standard curves generated to ensure optimal amplification efficiency (90–100%) were created using five template concentrations from four-fold serial dilutions (ranging from 80–0.31 ng). The reactions were carried out by automated pipetting using the QIAgility system (Qiagen, Courtaboeuf, France) in a total volume of 12.5 µl. Each reaction contained Power SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA), 20 ng of DNA, and 200 nM of each primer. The reactions were performed in duplicate, and the following PCR cycling conditions were used: an initial hold at 95°C for 10 min and 40 cycles of 95°C for 15 s and 58–59°C for 1 min. A dissociation curve was performed after the amplification cycle using the 7500 Real-Time PCR System (Applied Biosystems, Foster City, CA). The specificity of the amplified products was verified by analyzing the dissociation curves and the variation between replicates. Any instances in which the cycle quantification (Ct) values were greater than 0.5 were reassessed.

By qPCR analyses, DNA samples from 10 healthy individuals (reference controls) were compared with 11 LMS and 16 UPS samples (previously evaluated by array CGH). The relative copy numbers were calculated according to the delta-delta Ct model [22] using *GAPDH* as the reference gene. The relative copy number was calculated based on the target gene/*GAPDH* ratio, and this value was defined as a loss when the ratio was <0.55 and as a gain when the ratio was >1.35 (based on reference intervals).

Statistical Analyses

Comparisons between groups with clinicopathological and molecular alterations were performed using the Fisher exact test and Student's t-test. Overall survival (OS) probabilities were calculated using the Kaplan-Meier method and the Log Rank test for significance. The end-point for the OS analysis was restricted to deaths due to cancer. A multivariate analysis was performed using Cox proportional hazards with a model that included significant chromosomal alterations in LMS, tumor size, topography, tumor depth, local recurrence, and treatment (i.e., chemotherapy and/or radiotherapy). Statistical analyses were carried out using the software programs Nexus Copy Number (version 6.0; Biodiscovery Inc., El Segundo, CA, USA), Graphpad Prism 5 (Graphpad Software Inc., La Jolla, CA, USA), and SPSS version 17.0 (SPSS, Chicago, Illinois, USA) for Windows.

Results

Genome-wide Profiling of UPS and LMS

Several genomic changes were detected in the UPS and LMS samples, with UPS showing more complex genomic alterations. Changes in DNA copy number were identified in more than 20% of the UPS and LMS cases ($P<0.05$) as shown in Table 2.

The most frequently observed significant alterations in the UPS cases involved gains at 20q13.33 (75% of cases); 1q21.3-q23.1 (60%); 7q22.1 (60%); 9q34.11 and 20p11.21 (45%); and 1q21.1-q21.2, 8p11.21, 11q13.1 and 16p13.3 (40%) (Table 2). Genomic amplification ($\log_2$ ratio >0.6) was observed at 1q21.1-q21.2

Table 2. Genomic imbalances more frequently detected in UPS and LMS.

Cytoband location	Start (bp)	Stop (bp)	Size (Mb)	Event	Genes	miRNAs	Frequency (%)	P-Value
UPS								
1q21.1-q21.2	147,458,669	149,236,666	1.8	Gain	64	0	40.0	0.004
1q21.3-q23.1	151,345,357	155,272,136	3.9	Gain	186	5	60.0	0.004
2q11.1-q11.2	95,562,577	98,202,102	2.6	Gain	53	0	45.0	0.013
3p26.3	0	726,469	0.7	Loss	4	0	60.0	0.015
3p12.1-p11.2	85,966,634	87,627,650	1.7	Gain	7	0	30.0	0.009
7q22.1	99,861,211	100,667,677	0.8	Gain	61	0	60.0	0.042
8p11.21	41,627,121	43,175,310	1.5	Gain	25	1	40.0	0.032
9q34.11	129,506,829	130,229,037	0.7	Gain	49	2	45.0	0.008
11p15.5	0	2,197,662	2.2	Gain	118	3	45.0	0.020
11q13.1	65,171,847	66,296,450	1.1	Gain	60	0	40.0	0.021
11q13.1-q13.2	66,882,158	67,549,110	0.7	Gain	36	0	30.0	0.021
16p13.3	1,062,920	1,341,725	0.3	Gain	14	0	40.0	0.036
16q24.3	87,762,155	88,283,196	0.5	Gain	22	0	20.0	0.020
18p11.32	0	846,102	0.8	Gain	12	0	35.0	0.034
20p11.21	22,798,105	23,302,271	0.5	Gain	10	0	45.0	0.007
20q13.33	60,236,430	61,684,607	1.4	Gain	64	3	75.0	0.004
LMS								
1q21.3	151,256,551	151,856,750	0.6	Gain	27	0	23.5	0.018
1q21.3-q22	152,467,522	154,387,590	1.9	Gain	88	2	23.5	0.018
6p21.32	31,923,769	32,132,073	0.2	Gain	24	1	23.5	0.008
7q22.1	99,647,068	100,857,537	1.2	Gain	78	0	29.4	0.006
8q24.3	143,523,381	146,274,826	2.8	Gain	127	4	47.0	0.045
9p21.3	21,199,776	22,226,425	1.0	Loss	23	1	41.2	0.023
11p15.5	0	1,567,792	0.9	Gain	87	1	23.5	0.011
11q12.2-q12.3	61,352,779	62,229,449	0.9	Gain	37	0	29.4	0.047
11q13.1-q13.2	64,322,265	67,979,861	3.7	Gain	181	3	29.4	0.047
11q13.2-q13.3	68,744,800	70,077,870	1.3	Gain	19	1	29.4	0.047
14q11.2	21,580,291	22,230,297	0.7	Gain	44	0	23.5	0.007
16p11.2	34,059,589	34,361,998	0.3	Loss	1	0	23.5	0.031
17q25.1	68,836,129	70,494,896	1.7	Gain	40	0	41.2	0.050
19q13.12	40,964,925	41,016,382	0.1	Gain	5	0	52.9	0.008
19q13.43	63,347,573	63,811,651	0.5	Loss	36	0	23.5	0.043

Legend: UPS - Undifferentiated Pleomorphic Sarcomas; LMS - Leiomyosarcomas.
doi:10.1371/journal.pone.0067643.t002

(UPS7), 1q21.3-q23.1 (UPS7 and UPS8), 3p12.1-p11.2 (UPS2, UPS9, UPS15, UPS19, and UPS22), 7q22.1 (UPS4), 8p11.21 (UPS22), 11q13.1 (UPS3 and UPS8), 16p13.3 (UPS2), 18p11.32 (UPS13), and 20p11.21 (UPS19 and UPS22). Losses at 3p26.3 were observed in 60% of the cases (Table 2). No homozygous deletions ($\log_2$ ratio<-1) were consistently observed in the UPS cases. None of these UPS alterations were significantly associated with clinical variables.

The LMS samples showed fewer genomic alterations compared with the UPS samples. Fifteen significant genomic imbalances were more frequently observed in the LMS samples (Table 2), the most notable being gains at 19q13.12 (53%) and 8q24.3 (47%). Losses at 9p21.3 ($P=0.023$) and gains at 17q25.1 ($P=0.050$) were observed in 41.2% of the samples, including two cases (LMS19 and LMS23) with homozygous deletions (log2 ratio<-1) at 9p21.3.

Gains at 1q21.3, 11q12.2-q12.3, and 19q13.12 were significantly associated with death caused by LMS (Table 3). Furthermore, reduced overall survival time was significantly associated with gains at 1q21.3 ($P=0.002$), 11q12.2-q12.3 ($P=0.005$), 16p11.2 ($P=0.033$), and 19q13.12 ($P=0.027$) (Figures 1A–D; Table 3). A multivariate analysis indicated that gains at 1q21.3 are an independent prognostic marker for shorter overall survival time ($P=0.019$; HR = 13.76; $CI_{95\%}=1.534$ to 123.427).

Although the majority of the UPS and LMS tumors were found in the retroperitoneum and lower extremities, two LMS samples preferentially localized to the trunk were characterized by significant recurrent alterations, including gains at 6p21.32 ($P=0.044$), 14q11.2 ($P=0.044$), 17q25.1 ($P=0.029$), and 19q13.43 ($P=0.007$).

An unsupervised hierarchical clustering analysis could not distinguish between the UPS and LMS samples, nor could it segregate the samples according to anatomical origin; however,

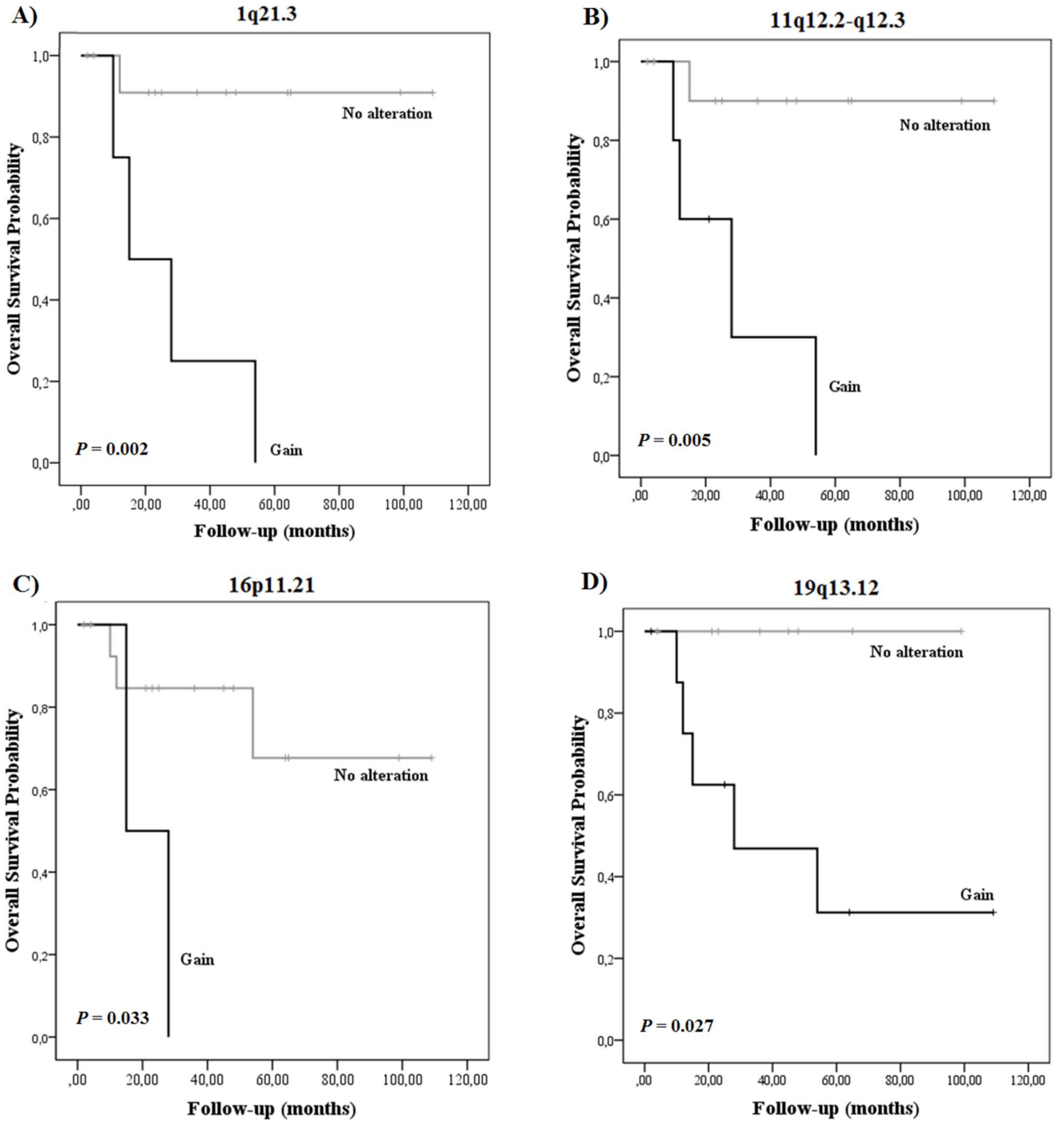

Figure 1. Overall survival curves from LMS patients with specific genomic alterations. Gains at (A) 1q21.3, (B) 11q12.2-q12.3, (C) 16p11.2, and (D) 19q13.12 were associated with shorter survival times. P-values were determined using the Log-rank test.
doi:10.1371/journal.pone.0067643.g001

three different clusters (1–3) were observed (Figures 2A and B). Similar numbers of genomic alterations were observed in clusters 1 (4 UPS and 7 LMS) and 3 (6 UPS and 4 LMS), whereas cluster 2 (10 UPS and 6 LMS) exhibited a more complex genomic profile (Figure 2B). Furthermore, the chromosomal changes in these three clusters were correlated with the clinicopathological features of each patient (Table 1). The presence of genomic alterations in cluster 2 was significantly correlated with female patients ($P = 0.020$) and with death from UPS or LMS ($P = 0.022$). Furthermore, cluster 2 included 70% of the cases (5/8 UPS and 4/5 LMS) in which patients died from the disease and 62% (3/5 UPS and 2/3 LMS) of the cases in which patients developed metastases during the follow-up period. Cluster 3 was primarily composed of male patients ($P = 0.023$). Interestingly, gains at 1q21.3 were more frequently detected in cluster 2 (62.5%) compared with clusters 3 (30%) and 1 (0%) ($P = 0.0035$).

Table 3. Genomic alterations associated with clinical outcome in LMS patients.

Chromosome region	Alteration	Prognosis*	Survival**
1q21.3	Gain	Increased risk of death (P = 0.002)	Decreased overall survival (P = 0.002)
11q12.2-q12.3	Gain	Increased risk of death (P = 0.0099)	Decreased overall survival (P = 0.005)
16p11.2	Gain	Increased risk of death (P = 0.073)	Decreased overall survival (P = 0.033)
19q13.12	Gain	Increased risk of death (P = 0.003)	Decreased overall survival (P = 0.027)

*P values obtained by Fisher's exact test.
**P values obtained by log-rank test.
doi:10.1371/journal.pone.0067643.t003

Quantitative Analysis of Copy Number Alterations in UPS and LMS

A subset of the cases (11 LMS and 16 UPS; Table 1) was evaluated by qPCR to confirm the gains at 1q21.1-q21.2 (*ARNT*), 1q21.3 (*PBXIP1* and *SCL27A3*), and 11q13.2-q13.3 (*CCND1*). Eight primer pairs were designed to cover the candidate regions. Three primer pairs covered the same probe sequence included in the Agilent 4×44 K platform (*ARNT*-P1, *PBXIP1*-P2 and *CCND1*-P2), whereas two primer pairs flanked the specific probes to determine the extent of the alteration (*SLC27A3*-P1 and *ARNT*-P2). Additionally, other three primers pairs were also designed in regions of the exon-intron junctions (*ARNT*-P3, *PBXIP1*-P1, and *CCND1*-P1). The ten reference samples isolated from healthy individuals displayed normal copy numbers for each of the sequences evaluated by qPCR.

For *ARNT* region, relative DNA copy number gains were observed in both UPS and LMS, but these genomic alterations were most frequently observed in UPS (10/16 samples) compared with LMS (4/11 samples). Three UPS (UPS7, UPS8, and UPS9) and two LMS (LMS4 and LMS7) samples showed gains over a large region (17 kb) surrounding the *ARNT* gene (for all primer pairs) (Figure 3A, Table S2). In addition, seven UPS and two LMS samples showed gains at two primer sets (Table S2). Similarly, gains at *PBXIP1* were more frequently detected in UPS (6/16) than in LMS (1/11) samples. Two UPS (UPS7 and UPS9) showed gains in both primer sets flanking *PBXIP1*, which span a region covering approximately 2 kb (Table S2). For *SLC27A3* gene, 6 UPS and two LMS cases exhibited increased copy numbers (Figure 3B, Table S2).

The *ARNT*, *PBXIP1* and *SCL27A3* genes, which are mapped to 1q21, cover a chromosomal region of approximately 4.12 Mb. Copy number gains involving these three genes were detected in

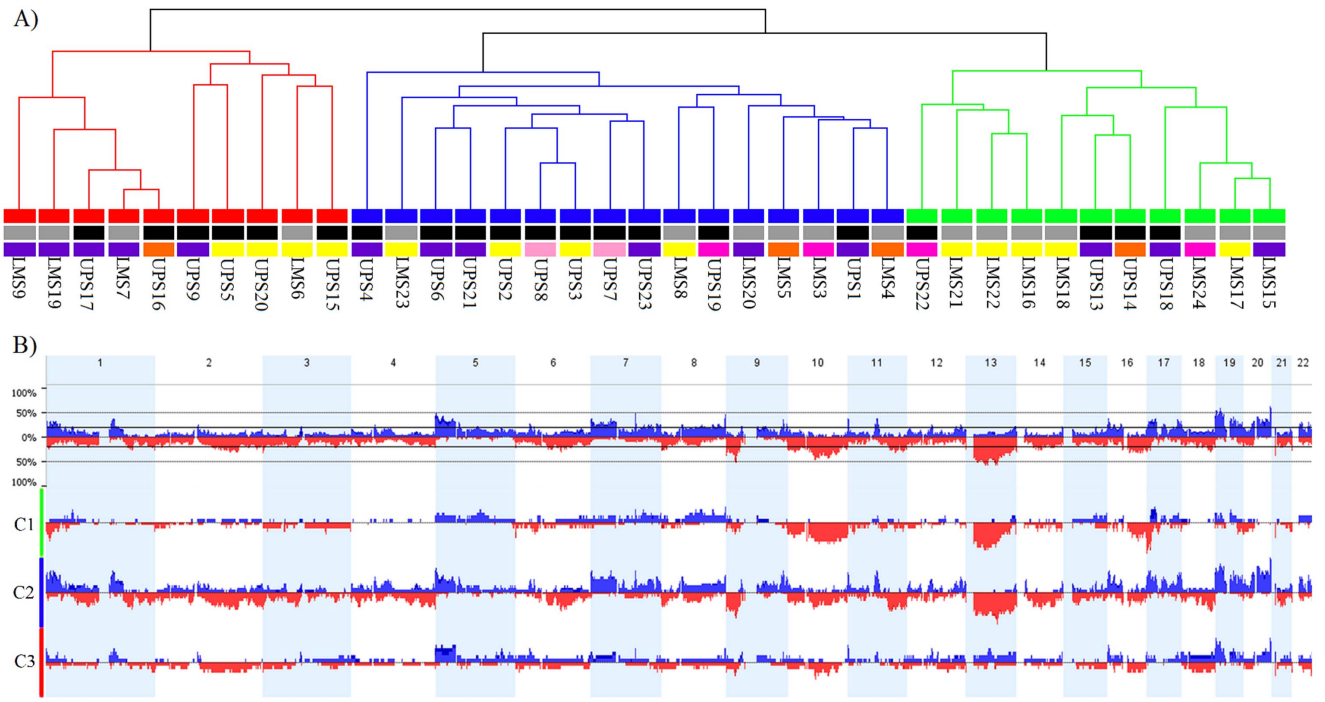

Figure 2. Unsupervised hierarchical clustering of 20 undifferentiated pleomorphic sarcomas (UPSs) and 17 leiomyosarcomas (LMSs). (A) In the dendrogram, cluster 1 is shown in green, cluster 2 is shown in blue, and cluster 3 is shown in red. Clusters related to the sites of anatomical origin were not observed for these tumors; origin sites include the following regions: upper extremity (pink), lower extremity (purple), trunk (orange), retroperitoneum (yellow), and head and neck (rose). (B) Genomic alterations were detected in clusters 1 (C1; 11 cases), 2 (C2; 16 cases), and 3 (C3; 10 cases). The top bars (blue) indicate genetic gains, whereas the lower bars (red) indicate genetic losses. The images shown were adapted from the output of the Nexus 6.0 software program.
doi:10.1371/journal.pone.0067643.g002

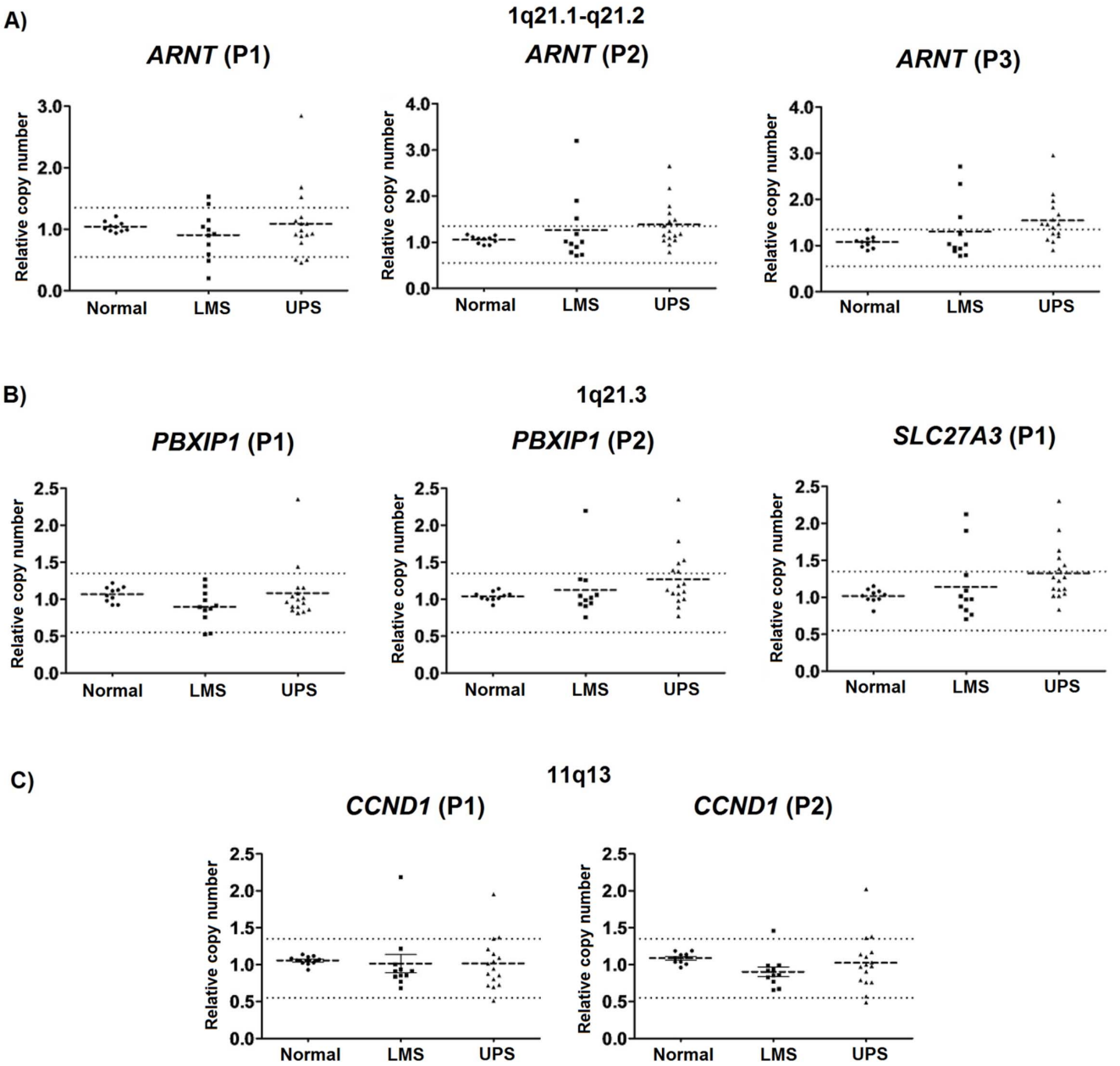

Figure 3. Quantification of DNA copy number alterations using qPCR for the *ARNT, PBXIP1, SLC27A3,* and *CCND1* genes. Eight primer pairs were designed, including (A) three for *ARNT* (ARNT-P1, ARNT-P2, and ARNT-P3); (B) two for *PBXIP1* (PBXIP1-P1 and PBXIP1-P2) and one for *SLC27A3* (SLC27A3-P1); and (C) two for *CCND1* (CCND1-P1 and CCND1-P2).
doi:10.1371/journal.pone.0067643.g003

two UPS cases (UPS7 and UPS9, Table S2), one of which displayed amplification (UPS9), whereas the other (UPS7) exhibited complex morphology.

Interestingly, the proportion of cases with concordant results of copy number gains in UPS samples detected by both array CGH and qPCR was 83% (5/6) for 1q21.1-q21.2 (*ARNT*) region and 63% (7/11) for 1q21.3 9 (*PBXIP1* and S*CL27A3*). For *CCND1* gene (11q13), copy number gains in both primer sets were detected in three UPS samples (UPS2, UPS20, and UPS22), suggesting that at least 5 kb of genomic sequence was altered in these samples. Two LMS samples showed increased copy number for one or more of the sequences amplified by each primer pair (P1 and P2) (Figure 3C, Table S2).

Discussion

In general, UPSs and LMSs display similar profiles of recurrent chromosomal imbalances even when compared with other sarcoma subtypes [9,12,14,16,17,23,24]. Some studies have suggested that the similarities shared by UPSs and LMSs may indicate a common origin [16,25,26,27]. However, the majority of these studies did not report whether the tested samples were collected from patients who had received either chemotherapy or radiotherapy prior to surgery. In this study, none of the patients had received treatment prior to sample collection, thereby excluding the possibility of treatment-induced genetic changes.

UPS and LMS frequently show complex karyotypes, pleomorphic histology and undifferenciated molecular profiles, thus hindering the correct diagnosis and the development of therapeutic options for these patients. Consequently, studies addressing the identification of molecular events driving oncogenesis in these tumors can ultimately be translated into meaningful biomarkers [18]. To address this challenge, we performed a comparative analysis of genomic copy number profiles for UPS and LMS samples carefully diagnosed by combining histology evaluation, immunohistochemical characterization using a panel of antibodies as well as FISH analysis for *MDM2*. Although an unsupervised hierarchical clustering analysis could not distinguish between the tumors subtypes, three patient clusters were identified based on patterns of copy number alterations. Of these three clusters, cluster 2 (16 UPS and 10 LMS) was strongly associated with poor prognostic outcome. More specifically, 105 genomic alterations were exclusively observed in cluster-2 cases (>50% of cases), including gains at 1q21.2, 1q21.3, 9q34.11, 11p15.5, 11q13.1, 16p13.3, and 20q13.33. Notably, gains at 1q21.3 in samples from cluster 2 were significantly associated with poor prognostic outcome.

The genetic similarities observed between UPS and LMS may indicate a common origin for these tumor subtypes. This hypothesis has also been raised by other comparative studies that assessed genomic patterns (using chromosomal CGH and array CGH) and gene-expression profiles (using cDNA microarrays and RT-qPCR) in UPS and LMS [16,17,23,24,25]. For example, Kresse et al. [17] compared genomic gains and losses in 33 UPS and 44 LMS samples using array CGH analysis (BAC and PAC arrays of 4.6K). The authors reported seven chromosomal regions (1p36.32-p35.2, 1p21.3-p21.1, 1q32.1-q42.13, 2q14.1-q22.2, 4q33-q34.3, 6p25.1-p21.32, and 7p22.3-p13) that differed significantly between the UPS and LMS tumor subtypes; three of these regions (1p35.1-p36.32, 1q32.1-q42.13, and 7p22.3-p13) were also identified in this study.

In another study using integrative analyses, similar patterns of genomic alterations were observed in 18 UPS and 31 LMS samples taken from tumors of the extremities [16]. The gene-expression analysis revealed nine differentially expressed genes (*TAGLN3, D4S234E, KIAA1729, PDLIM5, TEAD3, TPM2, ALDH1B1, TRDMT1,* and *DHODH*) but failed to differentiate between tumor subtypes. In the current study, these genes were not observed among the recurrent genomic regions that were differentially altered. Larramendy et al. [27] analyzed 102 untreated primary UPS samples and 82 LMS samples using chromosomal CGH, and the authors reported similar profiles of genetic alterations between these two tumor subtypes. Although a clustering analysis could not differentiate between the UPS and LMS samples, the authors reported one cluster (2 LMS and 10 UPS) that was characterized by high-level amplifications of the 1p33-p34.3, 17q22-q23, 17q25-qter, 19p, 22p, and 22q loci. Similarly, we observed genetic amplification at 17q25, 19p, and 22q; however, these alterations were not restricted to one specific cluster.

Large genomic regions (up to 4 Mb) displaying changes in DNA copy number were detected in both UPS and LMS samples. To better characterize these alterations, four genes (*ARNT, PBXIP1, SLC27A3,* and *CCND1*) were selected and evaluated using qPCR in a subset of cases. In general, the alterations detected by array CGH in the UPS and LMS groups were confirmed by qPCR. Although *ARNT* copy number gains were observed in both tumor subtypes, these alterations were predominantly observed in the UPS samples (63%). Several studies have demonstrated that amplification of the 1q21-q22 locus occurs in UPS and a variety of other STSs, including liposarcomas and osteosarcomas [15,17,28,29,30,31,32,33]; however, no specific candidate genes from this region had been studied in the context of UPS. Aryl hydrocarbon receptor nuclear translocator (*ARNT*), which is also known as hypoxia-induced factor-1 beta (HIF-1beta), is constitutively expressed in all normal human tissues with increased expression in the ovary, lung, spleen, testis, and pancreas [34]. *ARNT* overexpression has been reported in breast cancer, hepatocellular carcinoma, and colon carcinoma cell lines [35,36]. *PBXIP1* and *SLC27A3* copy number gains were also observed in UPS samples (38% of the cases for each). The *SLC27A3* gene encodes Acetyl-CoA synthetase, which is important for fatty-acid metabolism, particularly in neoplastic cells [37]. Although the contribution of lipid-metabolic pathways to tumor development is poorly understood, it is known that a high rate of lipid synthesis is necessary for the biogenesis of plasma membranes, which is required for tumor growth [38]. Lipids also play important roles as second messengers, which can be misregulated in tumor cells. Indeed, increases in the levels of specific messenger lipids are often associated with malignant phenotypes [39]. Functional studies have shown *SLC27A3* to be an effective therapeutic target in gliomas because it maintains the oncogenic properties of glioma cell lines through the regulation of the AKT protein [37].

Sixteen recurrent genomic changes were identified in UPSs. Although none of these changes were significantly associated with clinical variables, high-level amplification of the 3p12.1-p11.2 locus was observed in five cases. This amplified region spans the seven following genes: *BC040985, BC050344, CADM2, CHMP2B, MIR4795, POU1F1,* and *VGLL3*. Consistent with our findings, Hallor et al. [40] demonstrated that amplification of the 3p11-12 region was associated with *CHMP2B* and *VGLL3* overexpression in UPS cases with prominent inflammation. In another study, Carneiro et al. [16] identified recurrent amplifications at 3p11-p12 in a subset of UPS and LMS tumors using genomic and transcriptomic analyses. These findings suggest that the 3p11-p12 region, which includes the *CHMP2B* and *VGLL3* genes, may play an important role in UPS and LMS tumors.

Among the 15 minimally recurrent altered regions identified in the LMS samples, four regions showing gains were significantly associated with reduced overall survival time (1q21.3, 11q12.2-q12.3, 16p11.2, and 19q13.12). To our knowledge, this signature of poor prognosis has not been previously reported for LMS. In our study, copy number gains at 11q12.2-q12.3 and 19q13.12 were associated with death from LMS. The 11q12.2-q12.3 locus contains 37 genes, including genes related to the processes of chromosome segregation (*INCENP*), chromatin remodeling and histone deacetylation (*MTA2*), transcriptional regulation (*EEF1G*), and RNA processing (*TUT1*). To date, these genes have not been linked to LMS, and they represent potential targets for further validation. For example, metastatic tumor antigen 2 (MTA2) is a member of the MTA family and is closely associated with tumor progression and metastasis. MTA2 overexpression has been correlated with advanced TNM stages, tumor size, and lymph-node metastasis in non-small-cell lung cancer [41]. Gains at 19q13.12 have been described in sporadic cases using CGH approaches without clinical association [25,42,43]. We report an association between cases with poor prognosis and genomic gains at 19q13.12, indicating that this region may be a useful marker for LMS outcome. Furthermore, this amplified region has been linked to several cancer types, including pancreatic carcinoma [44], ovarian carcinoma [45], and breast cancer [46].

Genomic gains at 1q21.3 were associated with reduced overall survival time in LMS. A multivariate analysis revealed that gains

at the 1q21.3 locus were an independent prognostic marker of shorter survival times for LMS patients. These data suggest that gains at 1q21.3 confer an increased risk of death from the disease compared with other genomic changes and known prognostic factors in STSs. Although increased copy numbers and high-level amplification of 1q21.3 have been frequently observed in chromosomal and array CGH studies of LMS samples [16,28,42,47,48], none have reported an association with poor prognosis. This region covers 27 genes, including *MUC1, SPRR1B, SPRR2A, SPRR3, RAB25*, and 13 members of the S100 family. Of these genes, low amplification levels of *MUC1, SPRR1B, SPRR2A, SPRR3*, and *S100A6* (mapped to 1q21-q22) have been observed in LMS samples using FISH analysis [28]. In addition to *RAB25* amplification, rearrangements and increased expression of the *S100A4* gene have been correlated with advanced disease stages and poor survival in other malignancies, such as ovarian osteosarcoma metastasis and ovarian carcinoma [49]. Further analysis of gene-expression profiles and functional data should be conducted to determine whether these genes play a similar role in LMS.

In conclusion, we describe a large number of genomic changes observed in UPS and LMS patients who were not previously treated with chemotherapy or radiotherapy. Importantly, a subset of patients with poor prognosis displayed recurrent gains at the 1q21.2, 1q21.3, 9q33.3-q34.11, 11p15.5, 11q13.1, 16p13.3, and 20q13.33 loci. These loci may be useful as diagnostic markers to distinguish between patient outcomes. Three novel candidate genes associated with the amplified 1q21 region, including the *ARNT* gene, were identified in UPS patients. Gains at 1q21.3 were shown to be an independent prognostic marker for shorter survival times in LMS patients, suggesting that genes mapped to this region may be involved in the aggressiveness of LMS tumors. Therefore, this study describes several novel molecular markers that may be used to identify leiomyosarcoma patients with poor outcomes.

Supporting Information

Table S1 Primer sets used by quantitative real time PCR for confirmation of DNA genomic imbalances in UPS and LMS cases. (DOC)

Table S2 DNA copy number alterations detected by qPCR at *ARNT, PBXIP1, SLC27A3* and *CCND1* genes in 16 UPS and 11 LMS samples. Legends: DNA copy number alterations are shown in filled boxes, including gains (light gray) and high copy gain (dark gray). Empty boxes represent absence of alteration. (DOC)

Text S1 Fluorescence in situ hybridization for *MDM2/CEN12*. (DOCX)

Acknowledgments

The authors acknowledged Dr Ariane Fidelis Busso for their many helpful suggestions throughout the preparation of the experiments and A. C. Camargo Hospital Nucleic Acid Bank.

Author Contributions

Conceived and designed the experiments: SMS IWC SRR. Performed the experiments: SMS RARV. Analyzed the data: SMS FAM MCBF SAD. Contributed reagents/materials/analysis tools: RARV AL. Wrote the paper: SMS SRR SAD. Pathological revision: IWC CSN.

References

1. Osuna D, de Alava E (2009) Molecular pathology of sarcomas. Rev Recent Clin Trials 4: 12–26.
2. Wardelmann E, Schildhaus HU, Merkelbach-Bruse S, Hartmann W, Reichardt P, et al. (2010) Soft tissue sarcoma: from molecular diagnosis to selection of treatment. Pathological diagnosis of soft tissue sarcoma amid molecular biology and targeted therapies. Ann Oncol 21: 265–269.
3. Fletcher CD (2006) The evolving classification of soft tissue tumours: an update based on the new WHO classification. Histopathology 48: 3–12.
4. Nascimento AF, Raut CP (2008) Diagnosis and management of pleomorphic sarcomas (So-Called "MFH") in adults. J Surg Oncol 97: 330–339.
5. Erlandson RA, Antonescu CR (2004) The rise and fall of malignant fibrous histiocytoma. Ultrastruct Pathol 28: 283–289.
6. Matushansky I, Hernando E, Socci ND, Mills JE, Matos TA, et al. (2007) Derivation of sarcomas from mesenchymal stem cells via inactivation of the Wnt pathway. J Clin Invest 117: 3248–3257.
7. Wang CY, Wei Q, Han I, Sato S, Ghanbari-Azarnier R, et al. (2012) Hedgehog and Notch signaling regulate self-renewal of undifferentiated pleomorphic sarcomas. Cancer Res 72: 1013–1022.
8. Rubin BP, Nishijo K, Chen HI, Yi X, Schuetze DP, et al. (2011) Evidence for an unanticipated relationship between undifferentiated pleomorphic sarcoma and embryonal rhabdomyosarcoma. Cancer Cell 19: 177–191.
9. Mills AM, Beck AH, Montgomery KD, Zhu SX, Espinosa I, et al. (2011) Expression of subtype-specific group 1 leiomyosarcoma markers in a wide variety of sarcomas by gene expression analysis and immunohistochemistry. Am J Surg Pathol 35: 583–589.
10. Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, et al. (2002) Molecular characterisation of soft tissue tumours: a gene expression study. Lancet 359: 1301–1307.
11. Baird K, Davis S, Antonescu CR, Harper UL, Walker RL, et al. (2005) Gene expression profiling of human sarcomas: insights into sarcoma biology. Cancer Res 65: 9226–9235.
12. Francis P, Namlos HM, Müller C, Edén P, Fernebro J, et al. (2007) Diagnostic and prognostic gene expression signatures in 177 soft tissue sarcomas: hypoxia-induced transcription profile signifies metastatic potential. BMC Genomics 8: 73.
13. Nakayama R, Nemoto T, Takahashi H, Ohta T, Kawai A, et al. (2007) Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. Mod Pathol 20: 749–759.
14. Gibault L, Pérot G, Chibon F, Bonnin S, Lagarde P, et al. (2010) New insights in sarcoma oncogenesis: a comprehensive analysis of a large series of 160 soft tissue sarcomas with complex genomics. J Pathol 223: 64–71.
15. Weng WH, Ahlén J, Lui WO, Brosjö O, Pang ST, et al. (2003) Gain of 17q in malignant fibrous histiocytoma is associated with a longer disease-free survival and a low risk of developing distant metastasis. Br J Cancer 89: 720–726.
16. Carneiro A, Francis P, Bendahl PO, Fernebro J, Akerman M, et al. (2009) Indistinguishable genomic profiles and shared prognostic markers in undifferentiated pleomorphic sarcoma and leiomyosarcoma: different sides of a single coin? Lab Invest 89: 668–675.
17. Kresse SH, Ohnstad HO, Bjerkehagen B, Myklebost O, Meza-Zepeda LA (2010) DNA copy number changes in human malignant fibrous histiocytomas by array comparative genomic hybridisation. PLoS One 5: e15378.
18. Fletcher CD, Unni KK, Mertens F (2002) World Health Organization classification of tumors. Pathology and genetics of tumours of soft tissue and bone. Lyon: IARC Press.
19. Guillou L, Coindre JM, Bonichon F, Nguyen BB, Terrier P, et al. (1997) Comparative study of the National Cancer Institute and French Federation of Cancer Centers Sarcoma Group grading systems in a population of 410 adult patients with soft tissue sarcoma. J Clin Oncol 15: 350–362.
20. Abyzov A, Mariani J, Palejev D, Zhang Y, Haney MS, et al. (2012) Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. Nature. 20: 438–442.
21. Diskin SJ, Eck T, Greshock J, Mosse YP, Naylor T, et al. (2006) STAC: A method for testing the significance of DNA copy number aberrations across multiple array- experiments. Genome Res 16: 1149–1158.
22. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402–408.
23. Lee YF, John M, Edwards S, Clark J, Flohr P, et al. (2003) Molecular classification of synovial sarcomas, leiomyosarcomas and malignant fibrous histiocytomas by gene expression profiling. Br J Cancer 88: 510–515.
24. Chibon F, Lagarde P, Salas S, Pérot G, Brouste V, et al. (2010) Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. Nat Med 16: 781–787.
25. Derré J, Lagacé R, Nicolas A, Mairal A, Chibon F, et al. (2001) Leiomyosarcomas and most malignant fibrous histiocytomas share very similar comparative genomic hybridization imbalances: an analysis of a series of 27 leiomyosarcomas. Lab Invest 81: 211–215.

26. Chibon F, Mariani O, Mairal A, Derré J, Coindre JM, et al. (2003) The use of clustering software for the classification of comparative genomic hybridization data. an analysis of 109 malignant fibrous histiocytomas. Cancer Genet Cytogenet 141: 75–78.
27. Larramendy ML, Gentile M, Soloneski S, Knuutila S, Böhling T (2008) Does comparative genomic hybridization reveal distinct differences in DNA copy number sequence patterns between leiomyosarcoma and malignant fibrous histiocytoma? Cancer Genet Cytogenet 187: 1–11.
28. Forus A, Weghuis DO, Smeets D, Fodstad O, Myklebost O, et al. (1995) Comparative genomic hybridization analysis of human sarcomas: I. Occurrence of genomic imbalances and identification of a novel major amplicon at 1q21-q22 in soft tissue sarcomas. Genes Chromosomes Cancer 14: 8–14.
29. Larramendy ML, Tarkkanen M, Blomqvist C, Virolainen M, Wiklund T, et al. (1997) Comparative genomic hybridization of malignant fibrous histiocytoma reveals a novel prognostic marker. Am J Pathol 151: 1153–1161.
30. Szymanska J, Virolainen M, Tarkkanen M, Wiklund T, Asko-Seljavaara S, et al. (1997) Overrepresentation of 1q21–23 and 12q13–21 in lipoma-like liposarcomas but not in benign lipomas: a comparative genomic hybridization study. Cancer Genet Cytogenet. 99: 14–18.
31. Mertens F, Fletcher CD, Dal Cin P, De Wever I, Mandahl N, et al. (1998) Cytogenetic analysis of 46 pleomorphic soft tissue sarcomas and correlation with morphologic and clinical features: a report of the CHAMP Study Group. Chromos Morphlogy Genes Chromos Can 22: 16–25.
32. Mairal A, Terrier P, Chibon F, Sastre X, Lecesne A, et al. (1999) Loss of chromosome 13 is the most frequent genomic imbalance in malignant fibrous histiocytomas. A comparative genomic hybridization analysis of a series of 30 cases. Cancer Genet Cytogenet 111: 134–138.
33. Louis-Brennetot C, Coindre JM, Ferreira C, Pérot G, Terrier P, et al. (2011) The CDKN2A/CDKN2B/CDK4/CCND1 pathway is pivotal in well-differentiated and dedifferentiated liposarcoma oncogenesis: an analysis of 104 tumors. Genes Chromosomes Cancer 50: 896–907.
34. Yamamoto J, Ihara K, Nakayama H, Hikino S, Satoh K, et al. (2004) Characteristic expression of aryl hydrocarbon receptor repressor gene in human tissues: organ-specific distribution and variable induction patterns in mononuclear cells. Life Sci 74: 1039–1049.
35. Yamakuchi M, Lotterman CD, Bao C, Hruban RH, Karim B, et al. (2010) P53-induced microRNA-107 inhibits HIF-1 and tumor angiogenesis. Proc Natl Acad Sci U S A 107: 6334–6339.
36. Oda Y, Nakajima M, Mohri T, Takamiya M, Aoki Y, et al. (2012) Aryl hydrocarbon receptor nuclear translocator in human liver is regulated by miR-24. Toxicol Appl Pharmacol 1: 222–231.
37. Pei Z, Sun P, Huang P, Lal B, Laterra J, et al. (2009) Acyl-CoA synthetase VL3 knockdown inhibits human glioma cell proliferation and tumorigenicity. Cancer Res 69: 9175–9182.
38. Patra SK (2008) Dissecting lipid raft facilitated cell signaling pathways in cancer. Biochim Biophys Acta 1785: 182–206.
39. Pei Z, Fraisl P, Berger J, Jia Z, Forss-Petter S, et al. (2004) Mouse very long-chain Acyl-CoA synthetase 3/fatty acid transport protein 3 catalyzes fatty acid activation but not fatty acid transport in MA-10 cells. J Biol Chem 279: 54454–54462.
40. Hallor KH, Sciot R, Staaf J, Heidenblad M, Rydholm A, et al. (2009) Two genetic pathways, t(1;10) and amplification of 3p11–12, in myxoinflammatory fibroblastic sarcoma, haemosiderotic fibrolipomatous tumour, and morphologically similar lesions. J Pathol 217: 716–727.
41. Liu SL, Han Y, Zhang Y, Xie CY, Wang EH, et al. (2012) Expression of metastasis-associated protein 2 (MTA2) might predict proliferation in non-small cell lung cancer. Target Oncol 7: 135–143.
42. Hu J, Rao UN, Jasani S, Khanna V, Yaw K, et al. (2005) Loss of DNA copy number of 10q is associated with aggressive behavior of leiomyosarcomas: a comparative genomic hybridization study. Cancer Genet Cytogenet 161: 20–27.
43. Wang R, Lu YJ, Fisher C, Bridge JA, Shipley J (2001) Characterization of chromosome aberrations associated with soft-tissue leiomyosarcomas by twenty-four-color karyotyping and comparative genomic hybridization analysis. Genes Chromosomes Cancer 31: 54–64.
44. Curtis LJ, Li Y, Gerbault-Seureau M, Kuick R, Dutrillaux AM, et al. (1998) Amplification of DNA sequences from chromosome 19q13.1 in human pancreatic cell lines. Genomics 53: 42–55.
45. Thompson FH, Nelson MA, Trent JM, Guan XY, Liu Y, et al. (1996) Amplification of 19q13.1-q13.2 sequences in ovarian cancer. G-band, FISH, and molecular studies. Cancer Genet Cytogenet 87: 55–62.
46. Muleris M, Almeida A, Gerbault-Seureau M, Malfoy B, Dutrillaux B (1995) Identification of amplified DNA sequences in breast cancer and their organization within homogeneously staining regions. Genes Chromosomes Cancer 14: 155–163.
47. Yang J, Du X, Chen K, Ylipää A, Lazar AJ, et al. (2009) Genetic aberrations in soft tissue leiomyosarcoma. Cancer Lett 275: 1–8.
48. Levy B, Mukherjee T, Hirschhorn K (2000) Molecular cytogenetic analysis of uterine leiomyoma and leiomyosarcoma by comparative genomic hybridization. Cancer Genet Cytogenet 121: 1–8.
49. Fujiwara M, Kashima TG, Kunita A, Kii I, Komura D, et al. (2011) Stable knockdown of S100A4 suppresses cell migration and metastasis of osteosarcoma. Tumour Biol 32: 611–622.

OPEN ACCESS Freely available online

PLOS ONE

Gene Expression Profiling in Leiomyosarcomas and Undifferentiated Pleomorphic Sarcomas: SRC as a New Diagnostic Marker

Rolando A. R. Villacis[1], Sara M. Silveira[1], Mateus C. Barros-Filho[1], Fabio A. Marchi[2], Maria A. C. Domingues[3], Cristovam Scapulatempo-Neto[4], Samuel Aguiar Jr[5], Ademar Lopes[5], Isabela W. Cunha[6], Silvia R. Rogatto[1,7]*

1 Neogene Laboratory, Research Center (CIPE), A. C. Camargo Cancer Center, São Paulo, Brazil, **2** Inter-Institutional Grad Program on Bioinformatics, Mathematics and Statistics Institute, USP - University of São Paulo, São Paulo, Brazil, **3** Department of Pathology, School of Medicine, UNESP - São Paulo State University, Botucatu, São Paulo, Brazil, **4** Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, São Paulo, Brazil, **5** Department of Pelvic Surgery, A. C. Camargo Cancer Center, São Paulo, Brazil, **6** Department of Pathology, A. C. Camargo Cancer Center, São Paulo, Brazil, **7** Department of Urology, School of Medicine, UNESP - São Paulo State University, Botucatu, São Paulo, Brazil

Abstract

Background: Undifferentiated Pleomorphic Sarcoma (UPS) and high-grade Leiomyosarcoma (LMS) are soft tissue tumors with an aggressive clinical behavior, frequently developing local recurrence and distant metastases. Despite several gene expression studies involving soft tissue sarcomas, the potential to identify molecular markers has been limited, mostly due to small sample size, in-group heterogeneity and absence of detailed clinical data.

Materials and Methods: Gene expression profiling was performed for 22 LMS and 22 UPS obtained from untreated patients. To assess the relevance of the gene signature, a meta-analysis was performed using five published studies. Four genes (*BAD*, *MYOCD*, *SRF* and *SRC*) selected from the gene signature, meta-analysis and functional *in silico* analysis were further validated by quantitative PCR. In addition, protein-protein interaction analysis was applied to validate the data. SRC protein immunolabeling was assessed in 38 UPS and 52 LMS.

Results: We identified 587 differentially expressed genes between LMS and UPS, of which 193 corroborated with other studies. Cluster analysis of the data failed to discriminate LMS from UPS, although it did reveal a distinct molecular profile for retroperitoneal LMS, which was characterized by the over-expression of smooth muscle-specific genes. Significantly higher levels of expression for *BAD*, *SRC*, *SRF*, and *MYOCD* were confirmed in LMS when compared with UPS. *SRC* was the most value discriminator to distinguish both sarcomas and presented the highest number of interaction in the *in silico* protein-protein analysis. SRC protein labeling showed high specificity and a positive predictive value therefore making it a candidate for use as a diagnostic marker in LMS.

Conclusions: Retroperitoneal LMS presented a unique gene signature. SRC is a putative diagnostic marker to differentiate LMS from UPS.

Citation: Villacis RAR, Silveira SM, Barros-Filho MC, Marchi FA, Domingues MAC, et al. (2014) Gene Expression Profiling in Leiomyosarcomas and Undifferentiated Pleomorphic Sarcomas: SRC as a New Diagnostic Marker. PLoS ONE 9(7): e102281. doi:10.1371/journal.pone.0102281

Editor: Adriano Angelucci, University of L'Aquila, Italy

Received February 14, 2014; **Accepted** June 17, 2014; **Published** July 16, 2014

Funding: This work was supported by grants from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Conselho Nacional de Pesquisa (CNPq) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: silvia.rogatto@cipe.accamargo.org.br

Introduction

Soft Tissue Sarcomas (STS) comprise a heterogeneous group of mesenchymal tumors that represent around 1% of all neoplasms [1,2].

Diagnosis of these tumors poses a challenge to the pathologist due to their rarity, pleomorphic nature and histologic overlap with numerous sarcoma subtypes [2]. The majority of undifferentiated pleomorphic sarcomas (UPS) demonstrate similar morphology to undifferentiated and pleomorphic tumor subtypes, particularly leiomyosarcoma (LMS) and liposarcoma [1–3]. UPS, previously known as malignant fibrous histiocytoma (MFH), represents 17% of STS and has an extremely aggressive pattern of behavior [4,5]. Leiomyosarcoma accounts for 24% of all STS [4]. In general, Immunohistochemical (IHC) assays for well-differentiated LMS show positivity for actin, desmin, h-caldesmon, transgelin and sirtuin [6–8], however, none of these markers specifically differentiate smooth muscle. High-grade LMS shows histological similarities to UPS, which can cause difficulties in the distinction between these neoplasms [6].

The morphological and histopathological similarities between high-grade LMS and UPS are also observed at molecular level. Although copy number alterations have been described as similar in both tumors [9,10], gains of 1q21.3 have been reported as an independent prognostic marker for shorter survival in patients with LMS [11]. Hierarchical clustering analysis of transcriptomic data for a large series of STS failed to discriminate LMS and UPS. However, tumor subgroups with similar gene expression profiles have been reported [12–14]. Studies that combine gene-expression and DNA copy number alterations have also failed to differentiate between these neoplasms [15], however, three novel molecular subtypes of LMS have been described [6], [16], in addition to a genetic signature capable of predicting metastasis in UPS and LMS [17]. More recently, retroperitoneal LMS (LMS-R) has been described as different from LMS of extremities at the molecular level, demonstrating a distinct clinical outcome [18].

In this study, gene expression profiles were evaluated in 44 STS samples obtained from untreated patients prior to surgery, aiming to identify molecular biomarkers.

Materials and Methods

Patients

Forty-four fresh frozen tissue samples (22 UPS and 22 LMS) were obtained prior to chemotherapy or radiotherapy from 43 patients, at the A.C. Camargo Cancer Center (São Paulo, Brazil) and Barretos Cancer Hospital (Barretos, São Paulo, Brazil), between 2002 and 2010. Two primary tumors (UPS8 and UPS18) were obtained from the same patient. Ninety formalin-fixed paraffin-embedded (FFPE) samples (52 LMS and 38 UPS, which included 12 LMS and 7 UPS evaluated by oligoarrays) were analyzed using IHC assays. Study design is shown in Figure 1A.

This study was approved by the Human Research Ethics Committees of both institutions (A. C. Camargo Cancer Center Protocol 1240/09 and Barretos Cancer Hospital Protocol 302/2010). All patients were advised of the procedures and provided written informed consent. The pathological information of tumor samples ($N = 44$) is summarized in Table 1. Diagnostic criteria were based on the recommendations of the World Health Organization (WHO) [19]. Tumor grade was defined according to the recommendations of the *Federation Nationale des Centres de Lutte Contre le Cancer* (FNCLCC) [20]. An antibody panel for IHC was used to confirm or refute the diagnosis of LMS and UPS (Text S1).

The median age of patients at diagnosis was 58.8 years (ranging from 4–90 years). Twenty-four out of 43 cases were male. The majority of the patients was treated by surgery (18) or surgery combined with chemotherapy and radiotherapy (10). Distant metastases were identified in 12 patients. The mean follow-up period was 42.1 years (ranging from 1–209 months). In the last follow-up, 28 patients were alive and 15 were dead of disease.

Transcriptional profiling (oligoarrays)

Total RNA was extracted using Trizol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The RNA quality was assessed using the RNA 6000 Nano Kit on the Agilent 2100 Bioanalyzer platform (Agilent Technologies, Palo Alto, CA, USA). Only samples with RIN (RNA integrity number) >7 were considered. The Two-Color Human GE 4X44K Microarrays platform (Agilent Technologies) was used [21]. A combination of equal amounts of total RNA obtained from 15 different cell lines was used as reference [22]. Hierarchical clustering analysis (HCL) of the most variant probes (standard deviation, SD >1.5) was used to visualize cluster samples. Significance Analysis of Microarray (SAM) was applied to

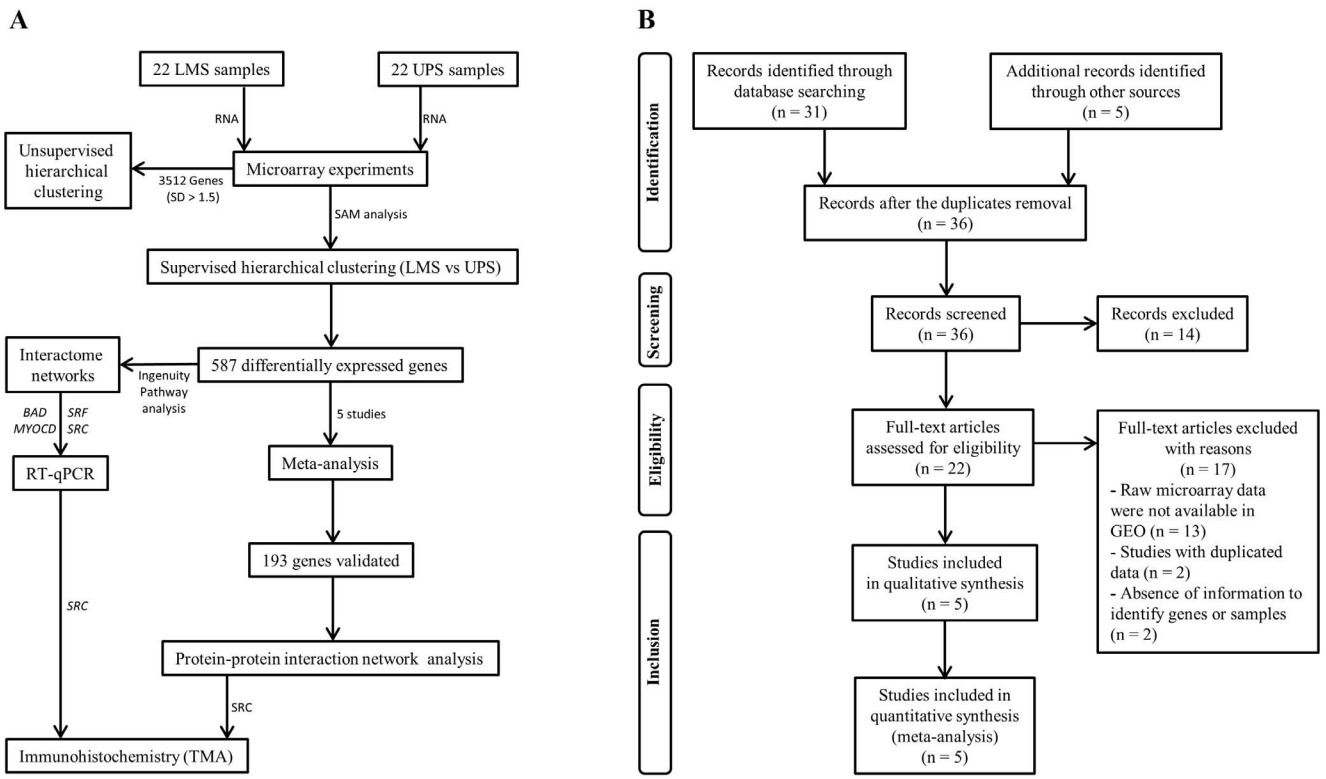

Figure 1. Flow diagrams showing the study design (A) and meta-analysis (B).
doi:10.1371/journal.pone.0102281.g001

Table 1. Pathological features of 22 LMS and 22 UPS samples evaluated by large-scale expression analysis.

Variable	LMS (*N* = 22)	UPS (*N* = 22)
Sample source		
Primary tumour	17 (77%)	17 (77%)
Recurrence	5 (23%)	5 (23%)
Topology		
Extremity	9 (41%)	11 (50%)
Retroperitoneum	10 (45%)	6 (27%)
Trunk	2 (9%)	2 (9%)
Head and neck	0	2 (9%)
Pelvis	1 (5%)	1 (5%)
Tumor size		
<5 cm	5 (23%)	3 (14%)
>5 cm	17 (77%)	19 (86%)
Histological grade		
I	4 (18%)	0
II	5 (23%)	0
III	13 (59%)	22 (100%)
Local recurrence*		
Yes	8 (36%)	10 (45%)
No	14 (64%)	12 (55%)

*not considered surgery extension of surgical margins.

doi:10.1371/journal.pone.0102281.t001

compare LMS and UPS [23]. All experiments were analyzed using the MeV 4.8 software (available at: http://www.tm4.org/mev/). Ingenuity Pathway Analysis (IPA, Ingenuity System Inc, Redwood City, CA, USA) was used to construct interactome networks with the differentially expressed genes between LMS and UPS. The microarray data have been deposited in NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE49941.

Meta-analysis of microarray studies

The meta-analysis was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [24]. The literature search was performed in the PubMed database (http://www.ncbi.nlm.nih.gov/pubmed/) using the following terms: "soft tissue sarcoma", "leiomyosarcoma", "undifferentiated pleomorphic sarcoma", "malignant fibrous histiocytoma" and "gene expression array". Additional search was identified by manually cross-referencing abstracts and articles published up to November 2013 independently of language, publication year or other limits (Figure 1B). Titles and abstracts were evaluated to identify relevant information and the full texts were archived for analysis. The criteria for inclusion were: (1) The diagnostic of UPS and LMS had to be confirmed and described in the same study; (2) studies with reference standard for the diagnosis of UPS and LMS; (3) studies with raw microarray data available in GEO. Exclusion criteria were: (1) studies with duplicate data reported in other studies; (2) absence of clinical and histological information; (3) studies that were letters, editorials, case reports or case series. Five microarray datasets, containing 309 tumors (134 LMS and 175 UPS), were selected for the final analysis [12,13,17,25,26]. The ratio of LMS - UPS was calculated individually for each dataset and expression direction (over-expression or down-expression) of each gene was checked and recorded. Differentially expressed genes were submitted to *in silico* protein-protein interaction analysis [27,28].

Transcript expression levels using real time quantitative reverse transcription PCR (RT-qPCR)

Expression levels of *BAD*, *SRC*, *SRF* and *MYOCD* were assessed for 13 UPS and 12 LMS (previously analyzed by microarray), which were selected according to the availability of RNA. Primer sequences are provided in Table S1. The reactions were carried out in duplicate as previously described [21]. Only replicates with low variability (ΔCycle quantification <0.5) were considered for analyses. *HMBS*, *HPRT* and *GAPDH* were selected as reference transcripts [29]. The relative gene expression was calculated according to the Pfaffl method [30].

Protein expression analysis

A tissue microarray (TMA) was constructed with the Tissue Microarrayer (Beecher Instruments, Silver Springs, USA). Each sample was represented in quadruplicates. IHC was carried out using rabbit polyclonal anti-SRC (Abcam: ab47405, Cambridge, UK) (dilution 1:200). Positive breast tumor tissue and two negative controls were assessed by IHC. The negative controls were created via omission of the primary antibody and incubation of the slides in PBS, followed by replacement of the primary antibody with normal rabbit serum. The final scores (median of the four scores) were obtained according to staining intensity of the cytoplasm or membrane and were described as negative/weak (score 0–1) or positive (score 2–3). The samples were scored blind with respect to clinical data of the patient.

Data analyses

Fisher's exact and chi-square tests were used to determine the association between the categorical variables. Overall survival (OS) probability was calculated using the Kaplan-Meier method and the Log Rank test for significance. The Mann Whitney test was used to compare RT-qPCR results. Receiver Operating Characteristic (ROC) curves and classification models designed with Fisher discriminant analysis evaluated the classification properties of markers assessed by RT-qPCR. Statistical analyses were performed using the SPSS 17.0 software (SPSS, Chicago, IL, USA) and GraphPad Prism 5.0 (GraphPad Software Inc., La Jolla, CA).

Results

LMS-R as a single entity in gene expression profiling analysis

Unsupervised hierarchical clustering (UHCL) analysis using 3512 genes (SD >1.5) revealed two clusters (I and II) including 43 samples, with one case (UPS17) clustering separately (Figure 2A). Cluster I was composed of LMS samples only (13) ($P<0.001$). Interestingly, all LMS-R samples ($N=10$) clustered together ($P<0.001$). One LMS of trunk (LMS5) and two LMS of extremities (LMS14 and LMS15) were also grouped in Cluster I. Cluster II contained the remaining nine LMS from different topologies and 21/22 UPS samples ($P<0.001$). Tumors with histological grade I were detected exclusively in cluster I ($P=0.004$). No significant difference was found in the comparison between the clusters and patient age ($P=0.504$), gender ($P=0.186$), tumor size ($P=1.0$), presence of metastasis ($P=0.460$), local recurrence ($P=0.332$) and overall survival ($P=0.392$) (data not shown).

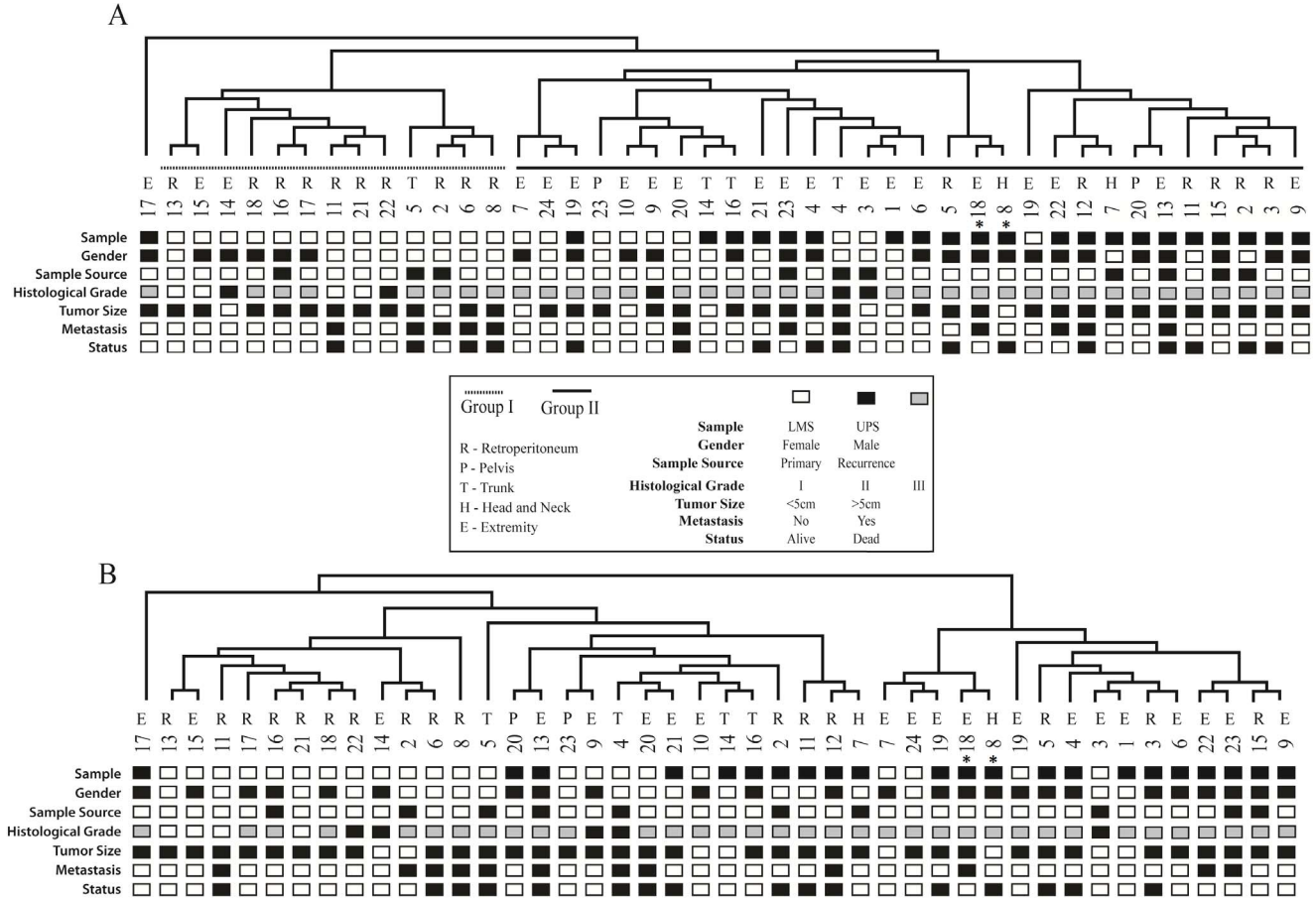

Figure 2. Hierarchical clustering of microarray data. (A) Unsupervised hierarchical clustering (UHCL) for 22 LMS and 22 UPS revealed two global clusters (I and II) and 3,512 genes differentially expressed (standard deviation >1.5). Cluster I included all retroperitoneal LMS (LMS-R), while cluster II contained the majority of the sarcomas of extremities. (B) Supervised hierarchical clustering (SHCL) using the 587 differentially expressed genes from the two-class unpaired Significance Analysis of Microarray (SAM) with FDR <1% between UPS and LMS. LMS-R samples demonstrated the most distinct molecular profile. Samples from the same patient are indicated with an asterisk. Images adapted from the output of the MeV 4.8 software.
doi:10.1371/journal.pone.0102281.g002

Supervised hierarchical clustering (SHCL) using SAM analysis revealed 587 differentially expressed genes in LMS compared to UPS, being 580 over-expressed and 7 down-expressed, but failed to discriminate between the neoplasms. However, over again the retroperitoneal LMS cases demonstrated similar expression profiles (Figure 2B). Only two LMS of the extremities (LMS14 and LMS15) clustered with the LMS-R samples. The IPA network analysis generated three main interactions networks (Table S2). The first network was associated with skeletal and muscular system development and function, tissue morphology, and cellular assembly and organization, including *ACTA2*, *MYLK*, *MYOCD* and *SRF* (Figure S1A). The second was related to cellular movement, cell morphology and cellular assembly and organization, and encompassed *ILK*, *CXCL1*, *LAMA5* and *SRC* (Figure S1B). The third network harbored genes associated with cell death, DNA replication, recombination, repair and gene expression, including *FLNA*, *MEF2C*, *CHEK1* and *BAD* (Figure S1C).

Meta-analysis and protein-protein interaction network analysis

Aiming to validate the signature obtained in the comparison between LMS and UPS from the SAM analysis, a meta-analysis including five studies was conducted (Table 2) and revealed 316 of 587 differentially expressed genes obtained by comparison between LMS and UPS. A total of 193 differentially expressed genes (191 over-expressed and 2 down-expressed) in LMS compared with UPS exhibited the same expression pattern in all studies, including several genes related to muscular function (Table S3). Protein-protein interaction analysis involving the 193 validated meta-analysis genes revealed a network of 59 proteins with more than 10 interaction partners (Figure 3). SRC protein presented the highest number of interactions, in particular with ACTN1, AR, FLNA and MUC1.

Real-time quantitative RT-PCR of *BAD, SRC, SRF*, and *MYOCD*

Based on the gene signature, gene network analysis, gene function and meta-analysis, *BAD*, *SRC*, *SRF* and *MYOCD* were selected to be further investigated by RT-qPCR. *SRF*, *SRC* and *BAD* were detected as central node genes in the main networks generated by IPA software (Figure S1), supporting their potential involvement in LMS development.

Similarly to the gene expression microarray, all genes showed significant over-expression ($P<0.01$) in LMS when compared to UPS (Figure 4A). Although not significant, *MYOCD* presented over-expression in LMS-R compared to non-retroperitoneal LMS

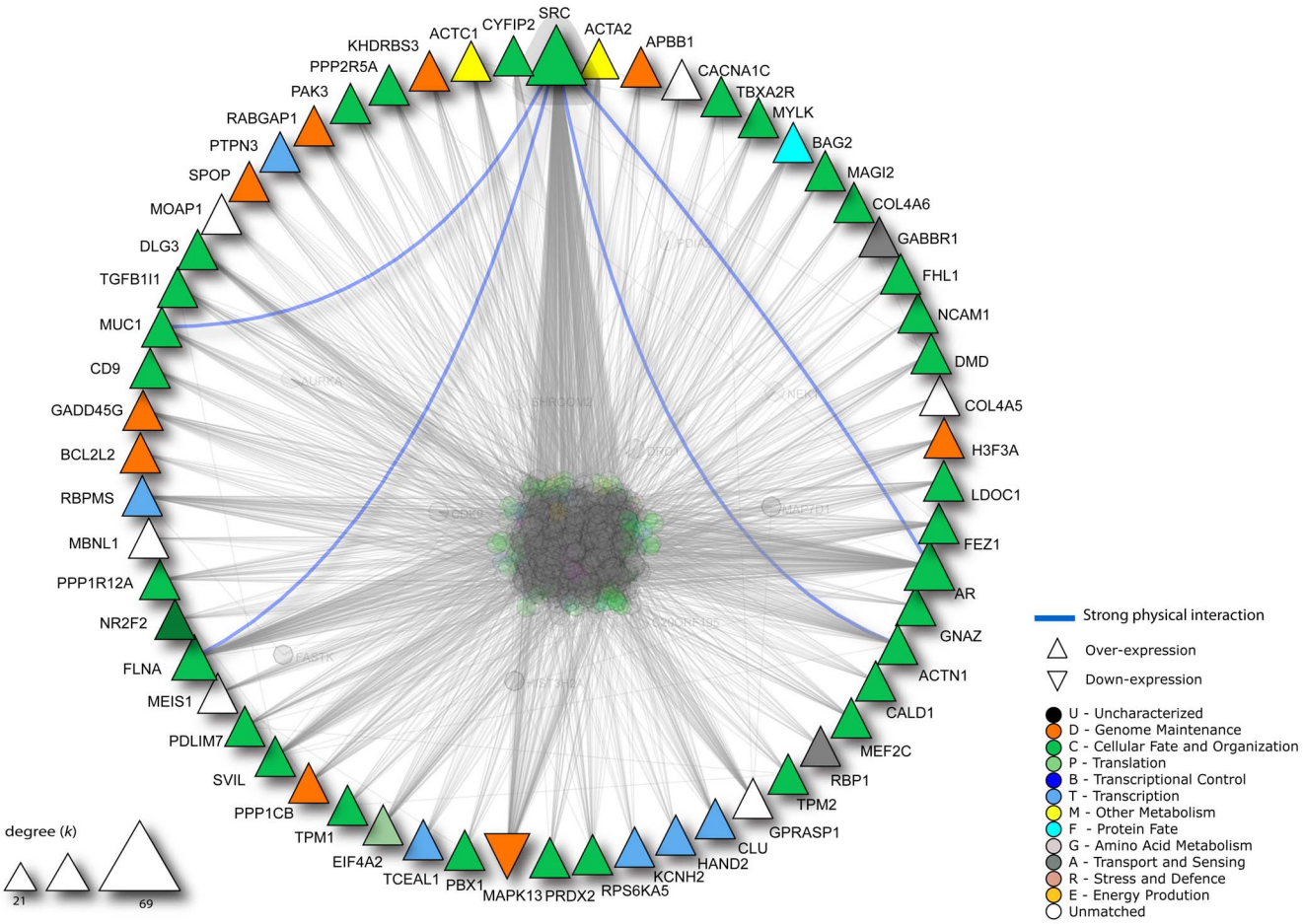

Figure 3. Protein-protein interaction network of 59 genes derived from the meta-analysis. Only proteins with more than 10 interactions partners were considered. The triangle size is proportional to the number of interactions for each protein. SRC presented the largest number of interactions, including four strong physical interactions found in the network. The interaction partners for each protein were obtained from the Interologous Interaction Database (I2D) 2.0 and the network was visualized and analyzed with the NAViGaTOR 2.3 software.
doi:10.1371/journal.pone.0102281.g003

(LMS-NR) ($P = 0.123$). Of all genes tested, *SRC* was revealed to be the most valuable discriminator for distinguishing between LMS and UPS (area under the ROC curve, AUC = 0.897 and accuracy = 92.0%) (Table S4). Two samples were classified incorrectly (LMS24 and UPS8) when using SRC as the discriminator.

SRC protein expression by immunohistochemistry

SRC protein revealed negative/weak immunostaining (score 0–1) in 30 LMS and 36 UPS while positivity (scores 2–3) was detected in 22 LMS and two UPS ($P<0.001$) (Figure 4B). Although most LMS-R presented positivity for SRC (5/7), no significant difference was observed between LMS-R and LMS-NR (Table 3). SRC positivity exhibited high specificity (94.7%) and a positive predictive value of 91.7% in discriminating LMS from UPS.

Table 2. Soft-tissue sarcoma profiling studies included in the meta-analysis.

GEO accession	Study	Platform	Samples	Tumors included in the current study
GSE3443	Nielsen et al. [12]	Custom cDNA array	46	11 LMS vs. 6 UPS (MFH)
GSE2719	Detwiller et al. [25]	Affymetrix HG-U133A	54	6 LMS vs. 9 UPS (MFH)
GSE6481	Nakayama et al. [13]	Affymetrix HG-U133A	105	6 LMS vs. 21 UPS (MFH)
GSE21122	Barretina et al. [26]	Affymetrix HG-U133A	158	26 LMS vs. 3 UPS (pleomorphic MFH)
GSE21050	Chibon et al. [17]	Affymetrix HG-U133A Plus 2	310	85 LMS vs.136 UPS

GEO: Gene Expression Omnibus; LMS: Leiomyosarcoma; MFH: malignant fibrous histiocytoma; UPS: undifferentiated pleomorphic sarcoma.
doi:10.1371/journal.pone.0102281.t002

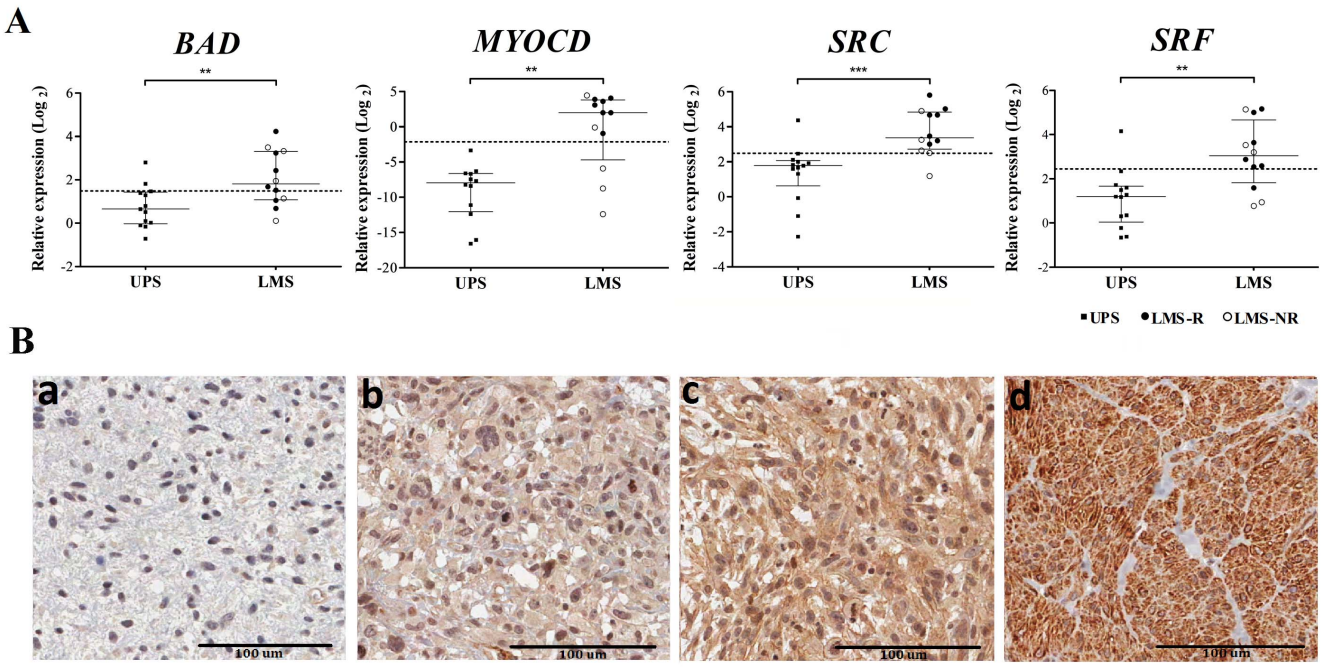

Figure 4. RT-qPCR of four genes and SRC protein expression by immunohistochemistry. (A) Dotplot showing the relative expression of the genes evaluated by RT-qPCR in 13 UPS and 12 LMS cases. For MYOCD gene 12 UPS and 12 LMS were evaluated. The dashed line represents an optimized threshold from ROC curves to discriminate LMS from UPS (**P<0.01; *** P<0.001) (Mann-Whitney U test). Error bars represent the median and interquartile range. (B) Examples of SRC immunostaining: negative (a); weak (b); moderate (c) and intense (d). Scores 0 and 1 were defined as SRC negative expression and scores 2 and 3 were considered SRC positive expression. Scale bars at 100 μm. Images captured with Scanscope XT Scanner System (Aperio Technologies, Inc., Vista, CA, EUA).
doi:10.1371/journal.pone.0102281.g004

Discussion

Large-scale gene expression studies in STS have many limitations, including evaluation of only a small number of fresh-frozen samples, a large number of histological subtypes included within the same report, a small number with the same histological subtype and the inclusion of cases treated with chemotherapy or radiotherapy prior to sample collection [3,31]. Overall, these limitations reduce the probability of identifying recurrent genetic alterations with clinical significance. In this study, 22 UPS and 22 LMS samples were collected prior to treatment with chemotherapy or radiotherapy, with their diagnosis confirmed using the markers recommended in the literature.

Although the UHCL analysis was not able to accurately discriminate LMS from UPS, a distinct molecular profile for LMS-R was established, independent of histological grade (cluster I). Cluster II was divided in two subgroups and included the high-grade LMS-NR and UPS from different anatomical sites. It is important to highlight that one patient developed two UPS (UPS8 and UPS18), which clustered together and demonstrated a similar gene expression profile, suggesting that these samples were not two primary tumors as previously diagnosed. Comparison of the clusters with the clinicopathological findings and overall survival revealed no significant difference, probably due to the fact that most of the tumors presented with a high histological grade and similar prognosis.

Several reports have shown a strong similarity between UPS and LMS, suggesting that they share common oncogenic pathways and may correspond to different stages of the same tumor entity [9,14,15]. However, this study has demonstrated that LMS-R clusters separately. Both UHCL and SHCL analysis suggested that high-grade LMS-NR demonstrates a gene expression profile with a closer proximity to UPS than to LMS-R. Almost all genes that distinguished LMS-R from other sarcomas were over-expressed (580 of 587 genes).

The function and interaction network analysis revealed a large number of genes associated with muscle structure and function, including *ACTG2*, *CALD1*, *DMD*, *MYOCD*, *MYLK*, *SRF* and *TAGLN*. A meta-analysis with datasets from five studies supported these findings. In all studies, 193 of the 316 genes evaluated were over-expressed in LMS compared to UPS. A considerable number of genes related to muscle function were found to be over-expressed in all studies, including *ACTC1*, *ACTA2*, *ACTN1*, *CALD1*, *CNN1*, *MYLK*, *MYL9*, *DMD*, *SGCA*, *TAGLN*, *TPM1*, *TPM2* and *SMTN*.

In LMS, over-expression of genes associated with muscle structure has been previously reported in the literature [12,14,16,32]. Recently, Italiano et al. [18] used genomic and transcriptomic analysis for 73 LMS and described a distinct clinical and molecular profile for LMS-R compared to LMS of the extremities. This finding was based mainly on over-expression of genes associated to muscle differentiation. MYOCD, a transcriptional cofactor of SRF, was found to be associated with muscular differentiation in well-differentiated retroperitoneal LMS [14,18,32].

In this study, overexpression of *MYOCD, SRF, BAD* and *SRC* in LMS was found, predominantly in LMS-R. These genes were tested for their potential as diagnostic markers to distinguish LMS from UPS. Combined accuracy, sensitivity and specificity were calculated for these genes, and *SRC* was identified as the most valuable discriminator. In addition, a higher level of SRC protein expression was detected in both LMS-R and LMS-NR when compared with UPS. Despite the use of SRC as a unique marker not being highly sensitive (<50%), its high specificity and positive

Table 3. SRC protein expression levels evaluated by immunohistochemistry in 52 LMS and 38 UPS.

		Immunostaining						
Cases	**Total**	**Negative Expression**		**Positive Expression**			***P***	***P***
		Score 0	**Score 1**	**Score 2**	**Score 3**		**(UPS vs. LMS)***	**(LMS-R vs. LMS-NR)#**
UPS-NR	32 (100%)	25 (78.1%)	5 (15.7%)	1 (3.1%)	1 (3.1%)			
UPS-R	6 (100%)	6 (100%)	0 (0%)	0 (0%)	0 (0%)		<0.001	0.216
LMS-NR	45 (100%)	22 (48.9%)	6 (13.3%)	3 (6.7%)	14 (31.1%)			
LMS-R	7 (100%)	2 (28.6%)	0 (0%)	0 (0%)	5 (71.4%)			

UPS: undifferentiated pleomorphic sarcoma; LMS: leiomyosarcoma; UPS-R: retroperitoneal UPS; UPS-NR: non-retroperitoneal UPS (including head and neck, trunk and extremity); LMS-R: retroperitoneal LMS; LMS-NR: non-retroperitoneal LMS (including head and neck, trunk and extremity);
*Chi-square test;
#Fisher exact test.
doi:10.1371/journal.pone.0102281.t003

predictive value (>90%) supports its potential use when in combination with other conventional muscular markers. It is important to highlight that *in silico* analysis of protein-protein interaction using the genes obtained in the meta-analysis showed a large number of interactions involving SRC, therefore suggesting its important role in the pathogenesis of LMS.

Conclusions

A distinct molecular profile was demonstrated for LMS-R, mediated by over-expression of genes involved in muscular development and function. Both well-differentiated LMS-R [14,20] and high grade LMS-R presented over-expression of *MYOCD*. Meta-analysis confirmed the involvement of 193 differentially expressed genes in the comparison between LMS and UPS. Our findings suggested that LMS-R is a new molecular entity of sarcomas whose alterations could be useful to stratify patients on specific therapeutic protocols. In addition, *in-silico* analysis revealed *SRC* as a central gene associated with LMS. Furthermore, SRC protein expression could be useful as a diagnostic marker to differentiate between LMS and UPS, especially in cases where muscular markers are negative.

Supporting Information

Figure S1 Graphic representation of the three interaction networks with the genes over-expressed (red) and down-expressed (green) in LMS compared to UPS. Genes were associated with skeletal and muscular system development and function, tissue morphology, cellular assembly and organization in first network (A); related to cellular movement, cell morphology and cellular assembly and organization in second network (B); and associated with cell death, DNA replication, recombination, repair and gene expression in third network (C). The red and green colors tones are proportional to intensity of expression for each gene. The genes selected for validation are indicated in blue circles. Image adapted from Ingenuity Pathway Analysis (IPA) software.
(TIF)

Table S1 Primer sequences and properties of the transcripts evaluated by RT-qPCR.
(XLSX)

Table S2 Description of differentially expressed genes in LMS and UPS categorized in three main interaction networks according to IPA software.
(XLSX)

Table S3 Meta-analysis of 316 genes from the supervised analysis (LMS vs. UPS) present in five array datasets.
(XLSX)

Table S4 Accuracy, specificity, sensitivity and confidence interval analysis used to evaluate the potential of the genes tested, individually or in association, as diagnostic markers in LMS and UPS.
(XLSX)

Text S1 Diagnostic criteria and IHC antibody panel used to define LMS and UPS.
(DOCX)

Acknowledgments

We thank Priscila Daniele Ramos Cirilo for her help during the research, and the Biobank of the A.C. Camargo Cancer Center and Barretos Cancer

Hospital (São Paulo, Brazil). Our special thanks to Dr. Sandra A. Drigo for her helpful suggestions during this study.

Author Contributions

Conceived and designed the experiments: SRR IWC RARV. Performed the experiments: RARV SMS. Analyzed the data: RARV SMS MCBF FAM MACD CSN IWC. Contributed reagents/materials/analysis tools: SA AL FAM. Wrote the paper: RARV SSR. Performed the histopathological analysis: MACD CSN IWC. Collected the tissue samples and clinical information: SA AL. Read and approved the final version of the manuscript: RARV SMS MCBF FAM MACD CSN SA AL IWC SRR.

References

1. Guillou L, Aurias A (2010) Soft tissue sarcomas with complex genomic profiles. Virchows Arch 456: 201–217. doi: 10.1007/s00428-009-0853-4.
2. Nielsen TO, West RB (2010) Translating gene expression into clinical care: sarcomas as a paradigm. J Clin Oncol 28: 1796–1805. doi: 10.1200/JCO.2009.26.1917.
3. West RB (2010) Expression profiling in soft tissue sarcomas with emphasis on synovial sarcoma, gastrointestinal stromal tumor, and leiomyosarcoma. Adv Anat Pathol 17: 366–373. doi: 10.1097/PAP.0b013e3181ec7428.
4. Toro JR, Travis LB, Wu HJ, Zhu K, Fletcher CD, et al. (2006) Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology and end results program, 1978–2001: An analysis of 26,758 cases. Int J Cancer 119: 2922–2930. doi: 10.1002/ijc.22239.
5. Fletcher CD (2014) The evolving classification of soft tissue tumours: an update based on the new 2013 WHO classification. Histopathology 64: 2–11. doi: 10.1111/his.12267.
6. Mills AN, Beck AH, Montgomery KD, Zhu SX, Espinosa I, et al. (2011) Expression of subtype-specific group 1 leiomyosarcomas markers in a wide variety of sarcomas by gene expression analysis and immunohistochemistry. Am J Surg Pathol 35: 583–589. doi: 10.1097/PAS.0b013e318211abd6.
7. Robin YM, Penel N, Pérot G, Neuville A, Vélasco V, et al. (2012) Transgelin is a novel marker of smooth muscle differentiation that improves diagnostic accuracy of leiomyosarcomas: a comparative immunohistochemical reappraisal of myogenic markers in 900 soft tissue tumors. Mod Pathol 26: 502–510. doi: 10.1038/modpathol.2012.192.
8. Dickson BC, Riddle ND, Brooks JS, Pasha TL, Zhang PJ (2013) Sirtuin 1 (SIRT1): a potential immnunohistochemical marker and therapeutic target in soft tissue neoplasms with myoid differentiation. Hum Pathol 44: 1125–1130. doi: 10.1016/j.humpath.2012.10.001.
9. Larramendy ML, Gentile M, Soloneski S, Knuutila S, Böhling T (2008) Does comparative genomic hybridization reveal distinct differences in DNA copy number sequence patterns between leiomyosarcoma and malignant fibrous histiocytoma? Cancer Genet Cytogenet 187: 1–11. doi: 10.1016/j.cancergencyto.2008.06.005.
10. Kresse SH, Ohnstad HO, Bjerkehagen B, Myklebost O, Meza-Cepeda LA (2010) DNA copy number changes in human malignant fibrous histiocytomas by array comparative genomic hybridisation. PLoS One 5: e15378. doi: 10.1371/journal.pone.0015378.
11. Silveira SM, Villacis RAR, Marchi FA, Barros Filho MdeC, Drigo SA, et al. (2013) Genomic signatures predict poor outcome in Undifferentiated Pleomorphic Sarcomas and Leiomyosarcomas. PLoS One 8: e67643. doi: 10.1371/journal.pone.0067643.
12. Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, et al. (2002) Molecular characterisation of soft tissue tumours: a gene expression study. Lancet 359: 1301–1307. doi: 10.1016/S0140-6736(02)08270-3.
13. Nakayama R, Nemoto T, Takahashi H, Ohta T, Kawai A, et al. (2007) Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. Mod Pathol 20: 749–759. doi: 10.1038/modpathol.3800794.
14. Gibault L, Pérot G, Chibon F, Bonnin S, Lagarde P, et al. (2011) New insights in sarcoma oncogenesis: a comprehensive analysis of a large series of 160 soft tissue sarcomas with complex genomics. J Pathol 223: 64–71. doi: 10.1002/path.2787.
15. Carneiro A, Francis P, Bendahl PO, Fenebro J, Akerman M, et al. (2009) Indistinguishable genomic profiles and shared prognostic markers in undifferentiated pleomorphic sarcoma and leiomyosarcoma: different sides of a single coin? Lab Invest 89: 668–675. doi: 10.1038/labinvest.2009.18.
16. Beck AH, Lee CH, Witten DM, Gleason BC, Edris B, et al. (2010) Discovery of molecular subtypes in leiomyosarcoma through integrative molecular profiling. Oncogene 29: 845–854. doi: 10.1038/onc.2009.381.
17. Chibon F, Lagarde P, Salas S, Pérot G, Brouste V, et al. (2010) Validated prediction of clinical outcome in sarcomas and multiple types of cancer on the basis of a gene expression signature related to genome complexity. Nat Med 16: 781–787. doi: 10.1038/nm.2174.
18. Italiano A, Lagarde P, Brulard C, Terrier P, Laë M, et al. (2013) Genetic profiling identifies two classes of soft-tissue leiomyosarcomas with distinct clinical characteristics. Clin Cancer Res 19: 1190–1196. doi: 10.1158/1078-0432.CCR-12-2970.
19. Fletcher CDM, Unni KK, Mertens F (2002) World Health Organization classification of tumours: pathology and genetics of tumours of soft tissue and bone. Lyon: IARC Press. 427 p.
20. Guillou L, Coindre JM, Bonichon F, Nguyen BB, Terrier P, et al. (1997) Comparative study of the National Cancer Institute and French Federation of Cancer Centers Sarcoma Group grading systems in a population of 410 adult patients with soft tissue sarcoma. J Clin Oncol 15: 350–362. doi: 10.1111/j.1550-7408.1980.tb04667.x.
21. Cirilo PD, Marchi FA, Barros Filho MdeC, Rocha RM, Domingues MA, et al. (2013) An integrative genomic and transcriptomic analysis reveals potential targets associated with cell proliferation in uterine leiomyomas. PLoS One 8: e57901. doi: 10.1371/journal.pone.0057901.
22. Boccardo E, Manzini Baldi CV, Carvalho AF, Rabachini T, Torres C, et al. (2010) Expression of human papillomavirus type 16 E7 oncoprotein alters keratinocytes expression profile in response to tumor necrosis factor-alpha. Carcinogenesis 31: 521–531. doi: 10.1093/carcin/bgp333.
23. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98: 5116–5121. doi: 10.1073/pnas.091062498.
24. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6:e1000097. doi: 10.1371/journal.pmed.1000097.
25. Detwiller KY, Fernando NT, Segal NH, Ryeom SW, D'Amore PA, et al. (2005) Analysis of hypoxia-related gene expression in sarcomas and effect of hypoxia on RNA interference of vascular endothelial cell growth factor A. Cancer Res 65: 5881–5889. doi: 10.1158/0008-5472.CAN-04-4078.
26. Barretina J, Taylor BS, Banerji S, Ramos AH, Lagos-Quintana M, et al. (2010) Subtype-specific genomic alterations define new targets for soft-tissue sarcoma therapy. Nat Genet 42: 715–721. doi: 10.1038/ng.619.
27. Brown KR, Jurisica I (2005) Online predicted human interaction database. Bioinformatics 21: 2076–2082. doi: 10.1093/bioinformatics/bti273.
28. Brown KR, Otasek D, Ali M, McGuffin MJ, Xie W et al. (2009) NAViGaTOR: Network Analysis, Visualization and Graphing Toronto. Bioinformatics 25: 3327–3329. doi: 10.1093/bioinformatics/btp595.
29. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3: research0034.1. doi: 10.1186/gb-2002-3-7-research0034.
30. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45. doi: 10.1093/nar/29.9.e45.
31. Weigelt B, Baehner FL, Reis-Filho JS (2010) The contribution of gene expression profiling to breast cancer classification, prognostication and prediction: a retrospective of the last decade. J Pathol 220: 263–280. doi: 10.1002/path.2648.
32. Pérot G, Derré J, Coindre JM, Tirode F, Lucchesi C, et al. (2009) Strong smooth muscle differentiation is dependent on myocardin gene amplification in most human retroperitoneal leiomyosarcomas. Cancer Res 69: 2269–2278. doi: 10.1158/0008-5472.CAN-08-1443.

Annals of Oncology

original articles

Annals of Oncology 25: 69–75, 2014
doi:10.1093/annonc/mdt421
Published online 26 November 2013

ATM down-regulation is associated with poor prognosis in sporadic breast carcinomas

R. C. Bueno[1,2], R. A. Canevari[3], R. A. R. Villacis[1,2], M. A. C. Domingues[4], J. R. F. Caldeira[5], R. M. Rocha[6], S. A. Drigo[1,2*,†] & S. R. Rogatto[1,2,†]

[1]NeoGene Laboratory, Department of Urology, São Paulo State University, Botucatu; [2]A. C. Camargo Cancer Center, São Paulo; [3]Laboratory of Biomedical Vibrational Spectroscopy (LEVB), Institute of Research and Development, IP&D, Paraíba Valley University, São José dos Campos; [4]Department of Pathology, São Paulo State University, Botucatu; [5]Department of Senology, Amaral Carvalho Hospital, Jaú; [6]Department of Pathology, A. C. Camargo Cancer Center, São Paulo, Brazil

Received 19 February 2013; revised 16 July 2013 and 28 August 2013; accepted 2 September 2013

Background: *Ataxia telangiectasia-mutated* (*ATM*) gene downexpression has been reported in sporadic breast carcinomas (BC); however, the prognostic value and mechanisms of *ATM* deregulation remain unclear.

Patients and methods: *ATM* and miRNAs (miR-26a, miR-26b, miR-203, miR-421, miR-664, miR-576-5p and miR-18a) expression levels were evaluated by quantitative real-time PCR (RT-qPCR) in 52 BC and 3 normal breast samples. ATM protein expression was assessed by immunohistochemistry in 968 BC and 35 adjacent normal breast tissues. *ATM* copy number alteration was detected by array comparative genomic hybridization (aCGH) in 42 tumours.

Results: Low ATM levels were associated with tumour grade. Absence of ATM protein expression was associated with distant metastasis ($P < 0.001$), reduced disease-free survival (DFS, $P < 0.001$) and cancer-specific survival (CSS, $P < 0.001$). Multivariate analysis indicated ATM protein expression as an independent prognostic marker for DFS ($P = 0.001$, HR = 0.579) and CSS ($P = 0.001$, HR = 0.554). *ATM* copy number loss was detected in 12% of tumours and associated with lower mRNA levels. miR-421 over-expression was detected in 36.5% of cases which exhibit lower *ATM* transcript levels ($P = 0.075$, $r = -0.249$).

Conclusions: The data suggest that ATM protein expression is an independent prognostic marker in sporadic BC. Gene copy number loss and miR-421 over-expression may be involved in *ATM* deregulation in BC.

Key words: *ataxia-telangiectasia-mutated* (*ATM*) gene, breast cancer, copy number alteration, microRNA, prognostic marker

introduction

Ataxia telangiectasia-mutated (*ATM*) is a tumour suppressor gene that encodes a serine/threonine kinase involved in DNA repair. ATM activates DNA damage response pathways, mainly due to double-strand breaks. These harmful DNA lesions lead to genomic rearrangements and chromosomal instability contributing to tumorigenesis [1].

ATM mutations cause Ataxia Telangiectasia (AT), an autosomal-recessive neurodegenerative disorder characterized by cerebellar ataxia associated with immunodeficiency, cancer predisposition, radiosensitivity, insulin-resistant diabetes and premature ageing. In addition to its pivotal role in DNA damage response, ATM is involved in cell cycle control, apoptosis, gene regulation, oxidative stress and telomere maintenance and is deregulated in several malignancies, including BC [2, 3]. Higher risk of developing BC and other tumours was first reported in mothers of patients with AT. Subsequently, *ATM* mutation was associated with moderate risk to BC development [4, 5]. Although several *ATM* mutations have been described, they are rare and frequently associated with hereditary BC [6].

In sporadic BC, ATM transcript and protein down-regulation has been reported [7–9], though the mechanisms involved in *ATM* deregulation are still unknown. Allelic loss has been proposed as one of the mechanisms for *ATM* gene inactivation in sporadic BC cases [10], but the second allele inactivation mechanism is still unclear [4]. *ATM* epigenetic silencing mediated by CpG island methylation has been proposed, but contradictory results have been reported. *ATM* hypermethylation was found in 18 of 33 (78%) locally advanced BC [7]. Nonetheless, recent reports have indicated that *ATM* promoter hypermethylation is not associated with *ATM* under-expression in BC [11–13], indicating the existence of an alternative *ATM* regulatory mechanism.

A post-transcriptional *ATM* regulation mechanism mediated by microRNAs has been reported in neuroblastoma [14], glioma [15] and BC [16]. MicroRNAs are small non-coding RNAs that

*Correspondence to: Dr Sandra A. Drigo, NeoGene Laboratory, Department of Urology, São Paulo State University – UNESP, Distrito de Rubião Junior, Botucatu, SP 18618-970, Brazil. Tel: +55-14-38116436; Fax: +55-14-38116271; Email: sandradrigo@gmail.com

†Both authors contributed equally.

regulate gene expression through degradation or inhibition of mRNA translation. MicroRNAs have been associated with cell cycle regulation, growth, apoptosis, differentiation and stress response. It is well known that microRNA deregulation plays a role in tumour development and progression by modulating oncogenic or tumour-suppressor pathways [17].

Clinical significance of *ATM* deregulation in prognosis of BC patients is limited and controversial. This study aimed to assess ATM gene and protein expression in sporadic BC and their association with clinical outcome. In addition, we sought to investigate *ATM* gene regulatory mechanisms, focussing on DNA copy number alterations and miRNA expression.

material and methods

patients

Ductal invasive BC samples ($N = 978$) were obtained from Amaral Carvalho Hospital and AC Camargo Hospital. Clinical and histopathological data are summarized in supplementary Table S1, available at *Annals of Oncology* online.

Fifty-two macro-dissected fresh tumour tissues (>80% of tumour cells) were obtained for RNA and DNA extraction. Transcript analyses were performed in 52 samples and array-CGH (aCGH) in 42 of 52 samples.

Protein analysis included 42 formalin fixed paraffin-embedded (FFPE) BC (also evaluated for mRNA analysis) in conventional slides and 926 tumours arranged in four tissue microarrays (TMAs). In 35 of 42 tumours, adjacent normal breast tissues were also evaluated (conventional slides).

Six normal breast tissues from healthy individuals subjected to breast reduction surgery were obtained as controls. Three macro-dissected fresh tissues (>80% of mammary epithelial cells) and three FFPE samples were used for transcript and Immunohistochemistry (IHC) analyses, respectively.

Patients were followed prospectively with a mean follow-up of 81.9 ± 35.1 months. All patients were untreated before sample collection. Of the 978 patients, 793 have received surgery as the primary treatment; 185 and 160 received neoadjuvant chemotherapy and radiotherapy, respectively; 339 were treated by adjuvant hormonal therapy, 585 by radiotherapy and 404 by chemotherapy.

The patients were advised of the procedures and provided written informed consent. The Human Research Ethics Committee of both institutions approved the study.

ATM gene expression by reverse transcription-quantitative PCR

Total RNA was extracted from using Trizol reagent (Invitrogen Life Technologies, Inc., Carlsbad, CA). cDNA synthesis and amplification are described in supplementary material, available at *Annals of Oncology* online. Primer sequences are provided in supplementary Table S2, available at *Annals of Oncology* online.

ATM and p53 protein expression by immunohistochemistry

IHC for ATM and p53 was performed as detailed in supplementary material, available at *Annals of Oncology* online.

ATM gene copy number alteration by aCGH

Gene copy number alteration was assessed by aCGH using Agilent Human CGH 180K Oligo Microarrays according to the manufacturer's recommendations (detailed in supplementary material, available at *Annals of Oncology* online).

microRNA expression by RT-qPCR

Five algorithms were used to select the top ranked candidates for *ATM* regulation. The selected microRNAs were: hsa-miR-421, hsa-miR-26a, hsa-miR-26b, hsa-miR-203, hsa-miR-664, hsa-miR-576-5p and hsa-miR-18a. Target and reference primer sequences are provided in supplementary Table S3 and S4, available at *Annals of Oncology* online, respectively. Reverse transcription-quantitative PCR (RT-qPCR) for microRNA analysis is given in supplementary material, available at *Annals of Oncology* online.

data analysis

Kruskal–Wallis or Mann–Whitney tests was applied to compare *ATM* transcript levels and clinicopathological characteristics or *ATM* gene copy number alteration. Chi-square test and Fisher's exact test were used to determine the association between the categorical variables. Bonferroni correction for multiple comparisons was applied to adjust the *P*-values. Spearman's test was performed for correlation analyses between *ATM* and microRNA levels. Disease-free survival (DFS) and cancer-specific survival (CSS) were calculated using the Kaplan–Meier method. The end-point for CSS and DFS analysis was restricted to distant metastasis development and death due to BC, respectively. Patients with stage IV ($N = 72$), treated with neoadjuvant therapy ($N = 170$) and missing follow-up information ($N = 47$), were excluded from these analyses. Significant variables detected in univariate analyses were included in multivariate Cox proportional hazard regression model. GraphPad Prism 5 (GraphPad Software, Inc., La Jolla, CA) and SPSS version 17.0 (SPSS; Chicago, IL) were used for statistical analysis.

results

ATM expression and association with clinical and histopathological variables

Decreased *ATM* transcript levels were detected in tumours compared with controls ($P = 0.064$, Figure 1A). Lower *ATM* mRNA levels were observed according to increased tumour grade ($P = 0.005$, Figure 1B). No association was found between *ATM* levels and other clinical and histopathological parameters. Although not significant, *ATM* transcript down-expression was associated with decreased DFS and CSS (data not shown).

ATM protein patterns are shown in Figure 1. Negative ATM expression was associated with tumours compared with adjacent normal breast tissues ($P < 0.0001$; supplementary Table S5, available at *Annals of Oncology* online).

ATM negativity was associated with higher grade, but it is not significant after Bonferroni correction (Table 1). Negative ATM expression was significantly associated with distant metastasis even after Bonferroni correction ($P < 0.001$, Table 1).

ATM-negative tumours were associated with reduced DFS ($P < 0.001$, Figure 1G) and CSS ($P < 0.001$, Figure 1H) in a large cohort of BC patients (supplementary Table S6, available at *Annals of Oncology* online). Significant associations were also observed between prognostic factors (T, N, clinical stage, grade, ER, PR and HER-2 status) and clinical outcome (supplementary Table S6, available at *Annals of Oncology* online).

Multivariate analyses showed that ATM protein status is an independent prognostic factor for DFS and CSS (Table 2). ATM-positive tumours presented lower risk of developing metastasis (HR = 0.579, $P = 0.001$) and cancer-specific death (HR = 0.554, $P = 0.001$). Tumour size and regional lymph nodes involvement were detected as independent prognostic factors for DFS; and tumour grade, tumour size and involvement of regional lymph

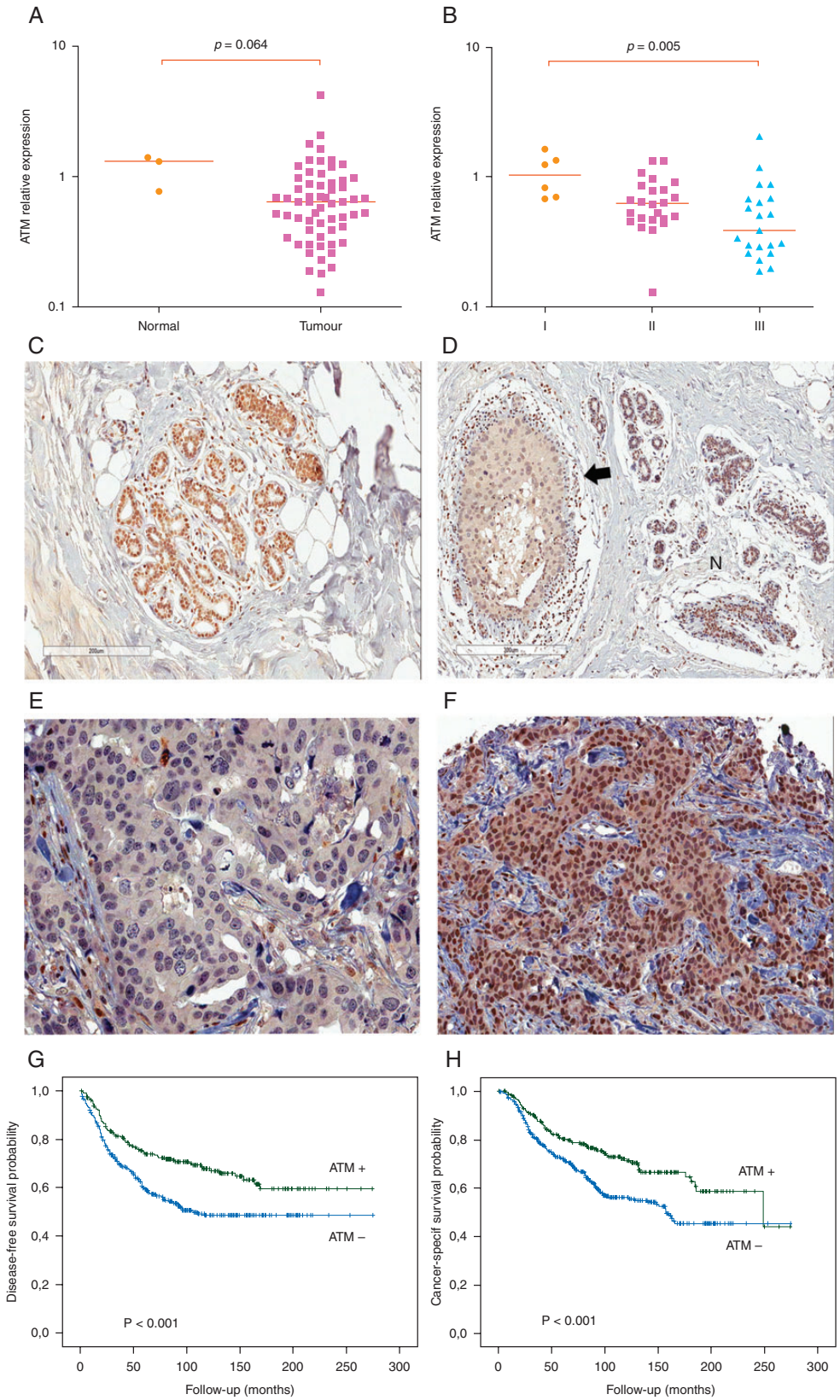

Figure 1. (**A**) Relative expression of *ATM* mRNA in breast tumour samples and normal breast tissues. (**B**) *ATM* mRNA relative expression according to the tumour grade. The transcript expression values are shown in log scale and the bars indicate the median value in each group. (**C–F**) ATM protein expression detected by immunohistochemistry in breast tissues. Positive ATM immunostaining in normal breast tissue (C). Positive ATM adjacent normal epithelial cells (N) and ATM-negative tumour cells (arrow) (D). Negative (E) and positive (F) ATM immunostaining in breast tumour cells. Disease-free survival (**G**) and cancer-specific survival (**H**) curves according to ATM protein expression in breast cancer patients. *P*-values were determined by Log-rank test.

Table 1. ATM protein expression detected by immunohistochemistry (IHC) in breast tumour samples according to the clinical and histopathological variables ($N = 968$)

Variables	N	ATM, N (%)		P
		Negative	Positive	
Age (years)				
<50	372	220 (59)	152 (41)	0.457
≥50	591	364 (62)	227 (38)	
NA	5			
T				
T1	90	51 (57)	39 (43)	0.307
T2	412	249 (60)	163 (40)	
T3	126	86 (68)	40 (32)	
T4	295	178 (60)	117 (40)	
NA	45			
N				
N0	304	174 (57)	130 (43)	0.310
N1	272	168 (62)	104 (38)	
N2	224	136 (61)	88 (39)	
N3	140	93 (66)	47 (34)	
NA	28			
M				
Negative	872	530 (61)	342 (39)	1.000
Positive	72	44 (61)	28 (39)	
NA	24			
Tumour grade				
G1	135	68 (50)	67 (50)	0.025
G2	541	337 (62)	204 (38)	
G3	286	181 (63)	105 (37)	
NA	6			
ER				
Negative	300	186 (62)	114 (38)	0.614
Positive	613	369 (60)	244 (40)	
NA	55			
PR				
Negative	478	300 (63)	178 (37)	0.113
Positive	407	234 (57)	173 (43)	
NA	83			
HER2				
Negative	699	429 (61)	270 (39)	0.099
Positive	113	60 (53)	53 (47)	
NA	156			
Distant metastasis[a]				
Negative	478	262 (55)	216 (45)	**<0.001**
Positive	379	260 (69)	119 (31)	
NA	111			

[a]Patients with metastasis at diagnosis were excluded in this analysis.
T, tumour size; N, lymph nodes involvement; M, metastasis at diagnosis; ER, oestrogen receptor alpha, PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; NA, information not available; significant data after Bonferroni correction for multiple variables is indicated in bold.

nodes for CSS (Table 2). We also evaluated if ATM stratified by p53 protein status was associated with DFS or CSS in a subset of cases ($N = 221$) and no association was detected (supplementary Figure S1, available at *Annals of Oncology* online).

Table 2. Multivariate model for disease-free survival (DFS) and cancer-specific survival (CSS) analysis in breast cancer patients ($N = 679$)

Variables	P	HR (95% CI)
DFS		
ATM (positive versus negative)	0.001	0.579 (0.421–0.797)
T (T4 versus T3 versus T2 versus T1)	<0.001	1.613 (1.366–1.903)
N (N3 versus *N2* versus N1)	<0.001	1.524 (1.323–1.754)
CSS		
ATM (positive versus negative)	0.001	0.554 (0.391–0.784)
Tumour Grade (G3 versus G2 versus G1)	0.013	1.361 (1.066–1.736)
T (T4 versus T3 versus T2 versus T1)	<0.001	1.659 (1.395–1.973)
N (N3 versus *N2* versus N1)	<0.001	1.591 (1.369–1.849)

HR, hazard ratio; 95% CI, 95% confidence interval; *P*-value obtained by Log-rank test.

evaluation of ATM gene regulatory mechanisms in BC

ATM transcript levels were compared with *ATM* copy number dosage obtained from another study conducted by our group using aCGH. Eleven probes span the *ATM* gene consisting of 66 exons. *ATM* locus was examined in 42 cases also evaluated for mRNA expression. Gene copy number loss was found in five cases (12%) (Figure 2A).

Tumours with *ATM* losses showed transcript down-regulation ($P = 0.013$, Figure 2B). ATM protein expression was assessed in three of five cases, and two of them were negative (data not shown).

To evaluate whether *ATM* may be regulated by miRNAs, expression levels of predicted miRNAs (miR-26a, miR-26b, miR-203, miR-421, miR-576-5p, miR-664 and miR-18a) were detected in 52 tumours and controls also evaluated for *ATM* levels. Although not significant, a negative correlation was observed between the *ATM* and miR-421 transcript levels ($r = -0.249$, $P = 0.075$, Figure 2C, *P*-value was not Bonferroni-corrected). Overexpression of miR-421 was detected in 36.5% of tumours (data not shown). No significant correlation was detected between the others miRs and *ATM* levels (data not shown).

discussion

ATM gene deregulation has been described in sporadic BC [4], but the clinical relevance of ATM remains to be established. *ATM* transcript and protein down-regulation were detected in BC compared with normal breast samples, corroborating other studies [8, 9, 18–20]. We found lower levels of ATM transcript and protein expression in high-grade tumours. Ding et al. [21] showed that the combined abnormalities of *ATM*, *BRCA1* and *TP53* genes were associated with poorly differentiated BC. Additionally, ATM aberrant protein expression was associated with grade II–III tumours [9]. These findings suggest the association of ATM in more aggressive disease.

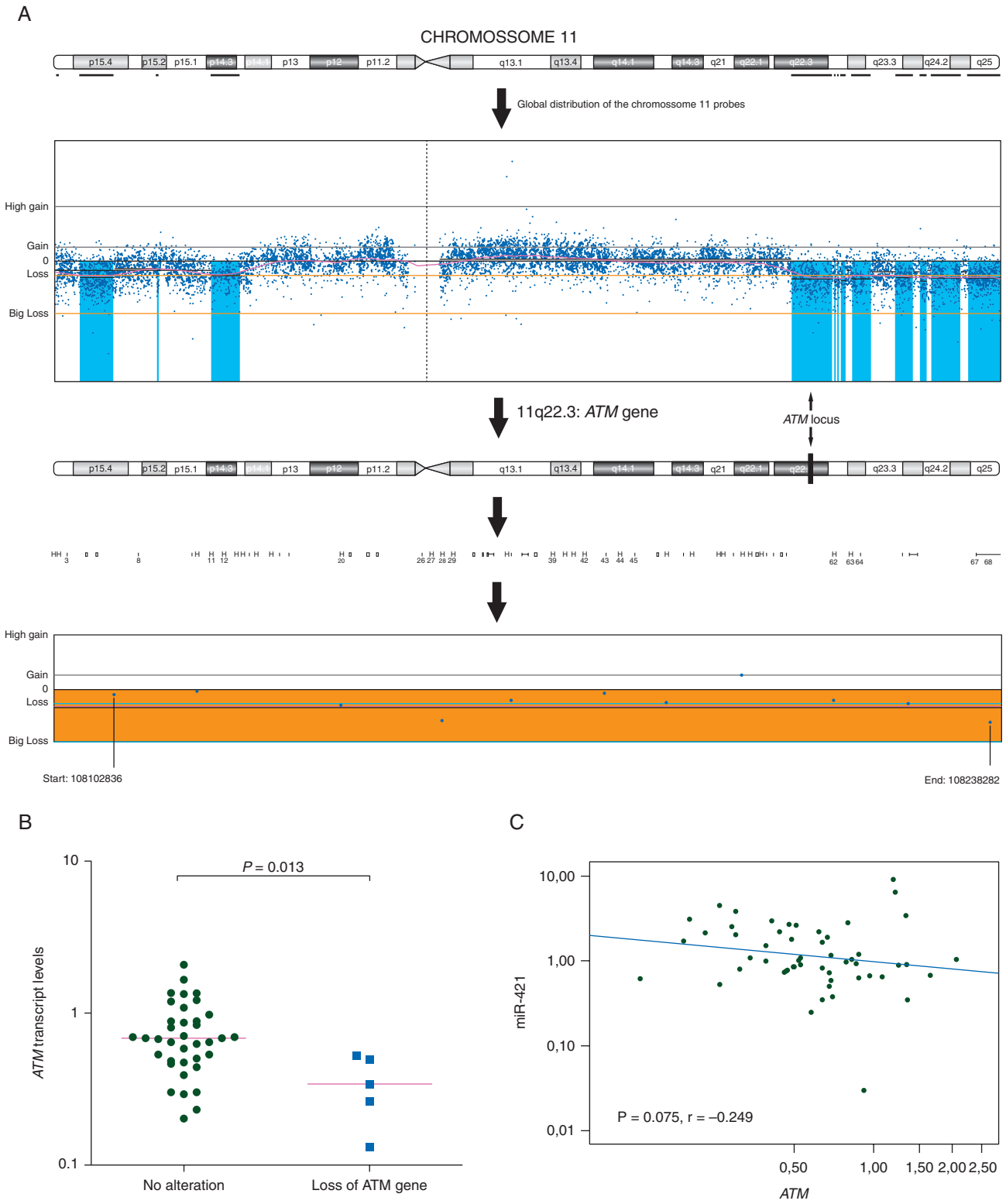

Figure 2. (**A**) Graphic representation of the chromosome 11 showing the global distribution of the array probes, the location of all copy number variations (CNVs) and more specifically the *ATM* locus with the probes covering this region. The image was adapted from Nexus 6.0 software. (**B**) *ATM* mRNA relative expression in tumour samples according to *ATM* copy number loss detected by aCGH in paired samples. The transcript expression values are shown in log scale and the bars indicate the median value in each group. (**C**) Correlation analysis between transcript levels of *ATM* and miR-421. Spearman's test.

By evaluating a large cohort of patients with a long-term follow-up, an association between absence of ATM protein expression and distant metastasis was detected, suggesting that loss of ATM expression may be involved in aggressive tumour behaviour and, consequently, a worse outcome in BC patients. Survival analysis revealed that patients with ATM-negative

tumours had shorter DFS and CSS. Moreover, multivariate analysis showed that ATM protein status was an independent prognostic marker for both DFS and CSS. ATM-positive cases showed lower risk of metastasis and death due to BC. Our data suggested that ATM is an independent prognostic factor, together with tumour size and lymph nodes involvement. Ye et al. [8] evaluated *ATM* mRNA levels in breast tumour samples ($N = 471$) and showed that lower *ATM* transcript levels were associated with worse outcome (DFS and CSS). However, the authors did not demonstrate that *ATM* gene expression is as an independent prognostic marker in BC. Other reports have shown association between ATM down-regulation and poor survival in BC patients with p53 wild-type tumours treated with DNA-damaging chemotherapy [22, 23]. Although with limited specificity and sensibility, *TP53* missense mutations have been reported as associated with increased protein stability. Here, ATM combined with p53 status was not associated with clinical outcome in BC. Further studies should be conducted to better understand the prognostic value of ATM in the context of *TP53* mutation. Thus, our data are in agreement with other studies showing that low ATM levels are associated with poor outcome in BC. However, by multivariate analysis, we found that ATM is a clinically meaningful prognostic marker in BC.

The mechanisms involved in ATM deregulation in sporadic BC are poorly understood. LOH in *ATM* region has been reported in sporadic BC [10, 19]. Copy number loss for the entire *ATM* locus detected by MLPA-analysis was recently reported in 13.6% of BC samples [23]. Similarly, we showed that *ATM* gene copy number loss detected in 12% of tumours was associated with transcript down-regulation. This finding suggests that *ATM* loss is one of the mechanisms involved in gene deregulation in BC. A recent study integrated genomic, transcriptomic and proteomic analyses on a large series of BC and identified four main tumour subtypes associated with different subsets of genetic and epigenetic abnormalities [24]. In this report, besides the disruption of p53 pathway, other pathway-inactivating events including *ATM* loss and *MDM2* amplification were detected in BC, specifically in more aggressive luminal B subtypes [24]. Taken altogether, these data indicate that *ATM* gene loss and the related-pathways are important mechanisms in breast carcinogenesis.

Post-transcriptional *ATM* regulation can also be altered in BC. Accordingly, recent studies have indicated that *ATM* gene expression can be regulated by miRNAs [14–16]. ATM down-expression was detected in 38% and 61% of tumours by transcript and protein analyses, respectively. In addition, only 8 of 24 cases had both *ATM* mRNA and protein down-expression in paired BC samples, suggesting that post-transcriptional mechanisms such as miRNAs could be involved in gene deregulation. To address this issue, putative microRNAs regulators of *ATM* were investigated. Although not significant, overexpression of miR-421 was correlated with lower *ATM* transcript levels. It has been postulated that microRNAs can also be responsible for fine regulation of gene expression through modestly repressing a large number of mRNAs and consequently protein output [25]. Additionally, cooperation between different miRNAs in regulating target genes has been proposed [26]. Experimental validation is needed to prove the regulation of *ATM* by miR-421 in BC cell lines.

The involvement of miRNA in ATM expression was firstly described in HeLa and neuroblastoma cell lines, showing that miR-421 suppresses *ATM* expression by targeting the 3′-UTR of *ATM* [14]. However, *ATM* regulation by miRs in BC has been fairly explored. Recently, Song et al. [16] evaluated the miR-18a as a putative ATM regulatory miRNA in BC. Up-regulation of miR-18a was detected by the comparison between nine tumour cell lines and one normal epithelial cell lineage, and between 10 BC and adjacent normal tissues. ATM protein down-regulation was observed by transfection of miR-18a in BC cell lines. Conversely, ATM up-regulation was detected in these cells after miR-18a suppression. Here, miR-18a showed no correlation with *ATM* transcript and protein expression. Different events isolated or in cooperation can coordinate the regulation of *ATM*, including gene deletions, mutations, other miRNAs and promoter hypermethylation.

Taken altogether, the present study demonstrated that decreased ATM expression is associated with worse prognosis in BC. As ATM protein expression is an independent prognostic marker in BC, we suggest that ATM may be a clinically applicable marker of disease outcome. Our results also suggest that gene copy number loss and miRNA regulation may be potentially involved in *ATM* deregulation in sporadic BC.

acknowledgements

The authors thank Dr Robson Francisco Carvalho for his help with the miRNA *in silico* analysis, Dr Patrícia Pintor dos Reis for the manuscript revision and Dr Márcia Hatakeyama and Department of Pathology staff (A. C. Camargo Cancer Center, São Paulo, Brazil) for their assistance with protein analyses.

funding

This study was supported by grants from the National Institute of Science and Technology in Oncogenomics (INCITO—São Paulo Research Foundation—FAPESP 2008/57887-9 and Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq 573589/08-9) and FAPESP 2007/52632-0.

disclosure

The authors have declared no conflicts of interest.

references

1. Bensimon A, Aebersold R, Shiloh Y. Beyond ATM: the protein kinase landscape of the DNA damage response. FEBS Lett 2011; 585: 1625–1639.
2. Lavin MF. Ataxia-telangiectasia: from a rare disorder to a paradigm for cell signalling and cancer. Nat Rev Mol Cell Biol 2008; 9: 759–769.
3. Morris LG, Veeriah S, Chan TA. Genetic determinants at the interface of cancer and neurodegenerative disease. Oncogene 2010; 29: 3453–3464.
4. Prokopcova J, Kleibl Z, Banwell CM et al. The role of ATM in breast cancer development. Breast Cancer Res Treat 2007; 104: 121–128.
5. Fanale D, Amodeo V, Corsini LR et al. Breast cancer genome-wide association studies: there is strength in numbers. Oncogene 2012; 31: 2121–2128.
6. Goldgar DE, Healey S, Dowty JG et al. Rare variants in the ATM gene and risk of breast cancer. Breast Cancer Res 2011; 13: R73.
7. Vo QN, Kim WJ, Cvitanovic L et al. The ATM gene is a target for epigenetic silencing in locally advanced breast cancer. Oncogene 2004; 23: 9432–9437.

8. Ye C, Cai Q, Dai Q et al. Expression patterns of the ATM gene in mammary tissues and their associations with breast cancer survival. Cancer 2007; 109: 1729–1735.
9. Tommiska J, Bartkova J, Heinonen M et al. The DNA damage signalling kinase ATM is aberrantly reduced or lost in BRCA1/BRCA2-deficient and ER/PR/ERBB2-triple-negative breast cancer. Oncogene 2008; 27: 2501–2506.
10. Meng ZH, Ben Y, Li Z et al. Aberrations of breast cancer susceptibility genes occur early in sporadic breast tumors and in acquisition of breast epithelial immortalization. Genes Chromosomes Cancer 2004; 41: 214–222.
11. Brandes JC, Carraway H, Herman JG. Optimal primer design using the novel primer design program: MSP primer provides accurate methylation analysis of the ATM promoter. Oncogene 2007; 26: 6229–6237.
12. Treilleux I, Chapot B, Goddard S et al. The molecular causes of low ATM protein expression in breast carcinoma; promoter methylation and levels of the catalytic subunit of DNA-dependent protein kinase. Histopathology 2007; 51: 63–69.
13. Pal R, Srivastava N, Chopra R et al. Investigation of DNA damage response and apoptotic gene methylation pattern in sporadic breast tumors using high throughput quantitative DNA methylation analysis technology. Mol Cancer 2010; 9: 303.
14. Hu H, Du L, Nagabayashi G et al. ATM is down-regulated by N-Myc-regulated microRNA-421. Proc Natl Acad Sci USA 2010; 107: 1506–1511.
15. Ng WL, Yan D, Zhang X et al. Over-expression of miR-100 is responsible for the low-expression of ATM in the human glioma cell line: M059J. DNA Repair (Amst) 2010; 9: 1170–1175.
16. Song L, Lin C, Wu Z et al. miR-18a impairs DNA damage response through downregulation of ataxia telangiectasia mutated (ATM) kinase. PLoS One 2011; 6: e25454.
17. Di Leva G, Croce CM. Roles of small RNAs in tumor formation. Trends Mol Med 2010; 16: 257–267.
18. Salimi M, Mozdarani H, Majidzadeh K. Expression pattern of ATM and cyclin D1 in ductal carcinoma, normal adjacent and normal breast tissues of Iranian breast cancer patients. Med Oncol 2012; 29: 1502–1509.
19. Waha A, Sturne C, Kessler A et al. Expression of the ATM gene is significantly reduced in sporadic breast carcinomas. Int J Cancer 1998; 78: 306–309.
20. Bhattacharya N, Mukherjee N, Singh RK et al. Frequent alterations of MCPH1 and ATM are associated with primary breast carcinoma: clinical and prognostic implications. Ann Surg Oncol 2012; doi: 10.1245/s10434-012-2715-6.
21. Ding SL, Sheu LF, Yu JC et al. Abnormality of the DNA double-strand-break checkpoint/repair genes, ATM, BRCA1 and TP53, in breast cancer is related to tumour grade. Br J Cancer 2004; 90: 1995–2001.
22. Jiang H, Reinhardt HC, Bartkova J et al. The combined status of ATM and p53 link tumor development with therapeutic response. Genes Dev 2009; 23: 1895–1909.
23. Knappskog S, Chrisanthar R, Lokkevik E et al. Low expression levels of ATM may substitute for CHEK2/TP53 mutations predicting resistance towards anthracycline and mitomycin chemotherapy in breast cancer. Breast Cancer Res 2012; 14: R47.
24. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. Nature 2012; 490: 61–70.
25. Baek D, Villen J, Shin C et al. The impact of microRNAs on protein output. Nature 2008; 455: 64–71.
26. Ivanovska I, Cleary MA. Combinatorial microRNAs: working together to make a difference. Cell Cycle 2008; 7: 3137–3142.

Annals of Oncology 25: 75–80, 2014
doi:10.1093/annonc/mdt505

Final results from the prospective phase III WSG-ARA trial: impact of adjuvant darbepoetin alfa on event-free survival in early breast cancer†

U. Nitz[1,2*], O. Gluz[1,2], I. Zuna[2], C. Oberhoff[3,4], T. Reimer[5], C. Schumacher[6], J. Hackmann[7], M. Warm[8,9], C. Uleer[10], V. Runde[11], J. Dünnebacke[12], N. Belzl[13], D. Augustin[14], R. E. Kates[2], N. Harbeck[2,15] & on behalf of the West German Study Group

[1]Breast Center Niederrhein, Evangelic Hospital Bethesda, Moenchengladbach; [2]West German Study Group, Moenchengladbach; [3]Women Clinics, University Clinics Essen, Essen; [4]Clinics for Gynaecology, Clinics Bremen, Bremen; [5]Women Clinics, University Clinics Rostock, Rostock; [6]Breast Center, St. Elisabeth Hospital, Köln-Hohenlind; [7]Breast Center, Marienhospital Witten, Witten; [8]Women Clinics, University Clinics Cologne, Cologne; [9]Breast Center, Clinics of Cologne Holweide, Cologne; [10]Praitice for Gynaecology and Oncology, Hildesheim; [11]Clinics for Inner Medicine, Wilhelm-Anton-Hospital, Goch; [12]Gynaecology, Hospital Marienhof, Koblenz; [13]Radiation Clinics, General Hospital Hagen, Hagen; [14]Mamma Center Ostbayern, Clinics Deggendorf, Deggendorf; [15]Breast Center, Women Clinics and CCCLMU of the University Munich, Munich, Germany

Received 14 March 2013; revised 5 September 2013; accepted 13 September 2013

Background: WSG-ARA plus trial evaluated the effect of adjuvant darbepoetin alfa (DA) on outcome in node positive primary breast cancer (BC).

*Correspondence to: Prof. Ulrike Nitz, West German Study Group, Ludwig-Weber-Straße 15b, D-41061 Moenchengladbach, Germany. Tel: +49-2161-566-2313; Fax: +49-2161-566-2319; E-mail: ulrike.nitz@wsg-online.com

†Presented in part at the 40th Annual Meeting of the American Society of Clinical Oncology (ASCO) 2004, New Orleans, at the 43rd Annual Meeting of the American Society of Clinical Oncology (ASCO) 2007, Chicago and at the 34th Annual San Antonio Breast Cancer Symposium (SABCS) 2011.

Cancer Genetics 208 (2015) 341–344

Cancer Genetics

BRIEF COMMUNICATION

Genomic profile of a Li-Fraumeni-like syndrome patient with a 45,X/46,XX karyotype, presenting neither mutations in *TP53* nor clinical stigmata of Turner syndrome

Tatiane R. Basso [a,1], Rolando A.R. Villacis [a,1], Luisa M. Canto [a], Vinicius M.F. Alves [a], Rainer M.L. Lapa [b], Amanda F. Nóbrega [c], Maria I. Achatz [c], Silvia R. Rogatto [a,b,*]

[a] *International Research Center (CIPE), A.C. Camargo Cancer Center, São Paulo, Brazil;* [b] *Department of Urology, Faculty of Medicine, São Paulo State University (UNESP), Botucatu, Brazil;* [c] *Department of Oncogenetics, A.C. Camargo Cancer Center, São Paulo, Brazil*

Li–Fraumeni syndrome (LFS) is a hereditary disorder that predisposes patients to several types of cancer and is associated with *TP53* germline mutations. Turner syndrome (TS) is one of the most common aneuploidies in women. Patients with TS have a higher risk of developing cancer, although multiple malignant tumors are extremely rare. Herein, we describe a patient with a 45,X/46,XX karyotype with no classic phenotype of TS. She presented with a clinical diagnosis of Li-Fraumeni-like syndrome (LFL), showing papillary thyroid carcinoma and fibrosarcoma of the left flank, and had no *TP53* germline mutations. Genome-wide analysis of copy number variations (CNVs) was assessed in DNA from peripheral blood cells and saliva. A total of 109 rare CNVs in the blood cells, including mosaic loss of the X chromosome (76% of cells), were identified. In saliva, three rare CNVs were detected, all of them were also detected in the blood cells: loss of 8q24.11 (*EXT1*), gain of 16q24.3 (*PRDM7* and *GAS8*), and the mosaic loss of the X chromosome (50% of cells). Results of conventional G-banding confirmed the 45,X/46,XX karyotype. Surprisingly, the patient presented with an apparently normal phenotype. The *PRDM* and *GAS8* genes are potential candidates to be associated with the risk of developing cancer in this LFL/TS patient.

Keywords Li-Fraumeni-like syndrome, Turner syndrome, genomic alterations, mosaicism

Li–Fraumeni syndrome (LFS) and its variant Li-Fraumeni-like syndrome (LFL) are rare autosomal dominant disorders associated with *TP53* germline mutations, which increase the risk of developing a wide spectrum of malignant tumors (1). It has been estimated that 20% of LFS and 60% of LFL patients have no *TP53* mutations (2). Patients with LFS and mutations in *TP53* have a lifetime risk of developing cancer of 73% in men and almost 100% in women (3). Moreover, the relative risk for a second primary tumor in these patients was estimated to be 5.3% (4).

Received February 9, 2015; received in revised form March 4, 2015; accepted March 5, 2015.

* Corresponding author.

E-mail address: silvia.rogatto@cipe.accamargo.org.br

[1] Both authors contributed equally to this study, and both are considered first author.

Turner syndrome (TS) is associated with the complete or partial loss of one X chromosome (5). Short stature, lymphedema, gonadal dysgenesis, webbed neck, and cubitus valgus are some of the clinical features of TS (6). The diagnosis may occur at various stages in life, with up to 10% of cases detected in adult women (7). The severity of TS usually correlates with the number of 45,X cells (8).

The chromosomal and hormonal abnormalities found in TS patients have been associated with an increased risk of developing cancer, including gonadoblastoma in cases with Y chromosome material, melanoma, cervical tumors, bladder tumors, and childhood brain tumors (9,10). Few studies have reported multiple tumors in TS patients (11,12).

In this study, we describe, for the first time, one patient that presented with both LFL without *TP53* pathogenic mutations and a 45,X/46,XX karyotype with no typical TS

2210-7762/$ - see front matter © 2015 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cancergen.2015.03.004

phenotype. In addition, we present a comprehensive analysis of the genomic alterations found in this patient.

Patient and methods

Case report

A 56-year-old female (76 kg and 159 cm), the fourth child of nonconsanguineous parents, is described. Physical examination showed no skeletal abnormalities, with the exception of very mild cubitus valgus. Menarche occurred at age 11 and full pubertal development ensued. The patient showed a papillary thyroid carcinoma (29 y), treated by thyroidectomy, which evolved with cervical lymph nodes metastasis. At age 35 years, she was diagnosed with fibrosarcoma of the left flank with pulmonary metastases. Subsequently, she was diagnosed with uterine leiomyoma (46 y) and angiomyolipoma of the kidney (51 y). Second- and third-degree maternal relatives (grandfather with gastric cancer at 71 y, aunt with endometrial cancer at 64 y, and a cousin with breast cancer at 42 y) had history of cancer (Figure 1). The patient bore two children; both pregnancies and labors were uneventful. The patient was diagnosed with LFL according to the Eeles 1 criteria (13). Sanger sequencing showed an absence of *TP53* mutations. This study was approved by the Human Research Ethics Committee (CEP 1726112), and the patient provided written informed consent and received genetic counseling.

Cytogenetic and molecular cytogenetic analyses

Genomic DNA from peripheral blood and saliva was extracted using the standard Gentra Puregene blood kit (Qiagen, Venlo, The Netherlands) and DNA Genotek prepIT-L2P kit protocols (OraSure Technologies Inc., Bethlehem, PA), respectively. Molecular karyotyping was performed using the CytoScan HD array (Affymetrix, Santa Clara, CA) following the manufacturer's recommendations. Copy number variations (CNVs) and copy neutral loss of heterozygosity (cnLOH) analyses were performed with the Chromosome Analysis Suite (ChAS) software v.2.0.1 (Affymetrix). The criteria used for the analyses considered at least 50 markers for gains, 25 markers for losses, and cnLOHs with a minimum length of 5 Mb.

Genomic alterations were compared with the Database of Genomic Variants (DGV, http://dgv.tcag.ca/dgv/app/home) and with the Affymetrix Database of Variants (aDGV), which consists of array results obtained in 2,421 phenotypically normal individuals evaluated with the same platform used in this study. We excluded polymorphic variations and focused on only rare CNVs, which are described in less than 1% of the reference population. Conventional cytogenetic analysis, using the GTG-banding technique, was performed.

Results

DNA from blood cells revealed 109 rare CNVs, including 47 gains with an average of 62 Kb in 15 chromosomes and 62 losses with an average of 129 Kb in 19 chromosomes. No cnLOHs were detected with the parameters used. Six alterations were found to be mosaic (Supplementary Table 1), including loss of the entire X chromosome, which was detected in 76% of cells.

Interestingly, in the DNA obtained from the saliva, only three rare CNVs previously identified in the blood cells were detected: deletion of 8q24.11, encompassing the *EXT1* gene [arr 8q24.1 (119,107,433–119,122,050) × 1 hg19]; gain of 16q24.3, covering the genes *PRDM7*, *LOC100130015*, *C16orf3* and *GAS8* [arr 16q24.3 (90,089,010 90,155,062) × 3 hg19]; and the mosaic loss of the X chromosome (45,X/46,XX), which was detected in 50% of cells from saliva (Figure 2A). Cytogenetic analysis of peripheral blood using GTG-banding revealed a karyotype mos45,X[75]/46,XX[27] (Figures 2B and 2C). These findings revealed that this LFL patient also has TS.

Discussion

The presence of multiple tumors and a family history of cancer in addition to a diagnosis of LFL (Eeles 1 criteria) and the

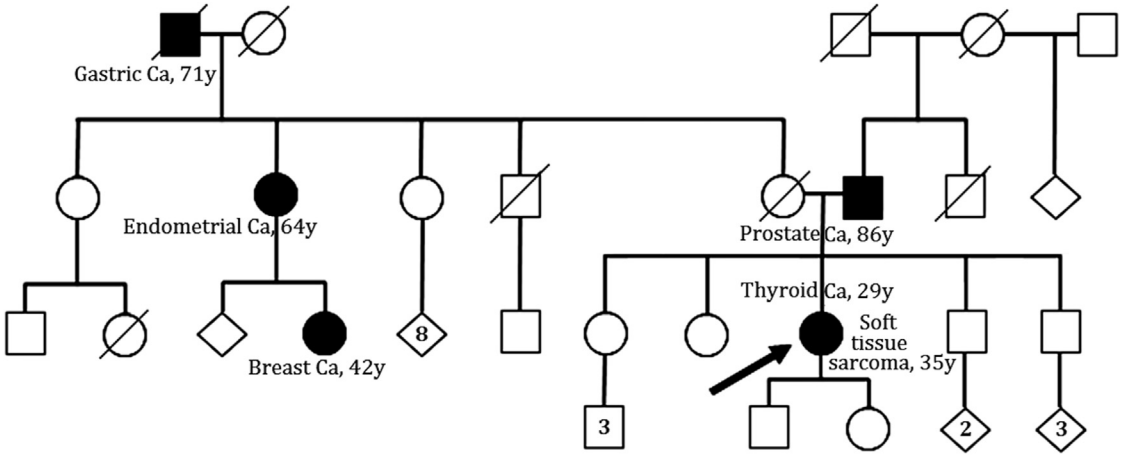

Figure 1 A pedigree chart of the patient's family, consistent with LFL according to Eeles 1 criteria. The arrow indicates the proband who developed two malignant tumors (papillary thyroid carcinoma at age 29 y and soft tissue sarcoma at age 35 y). The patient had four relatives with tumors. *Abbreviation*: Ca, cancer.

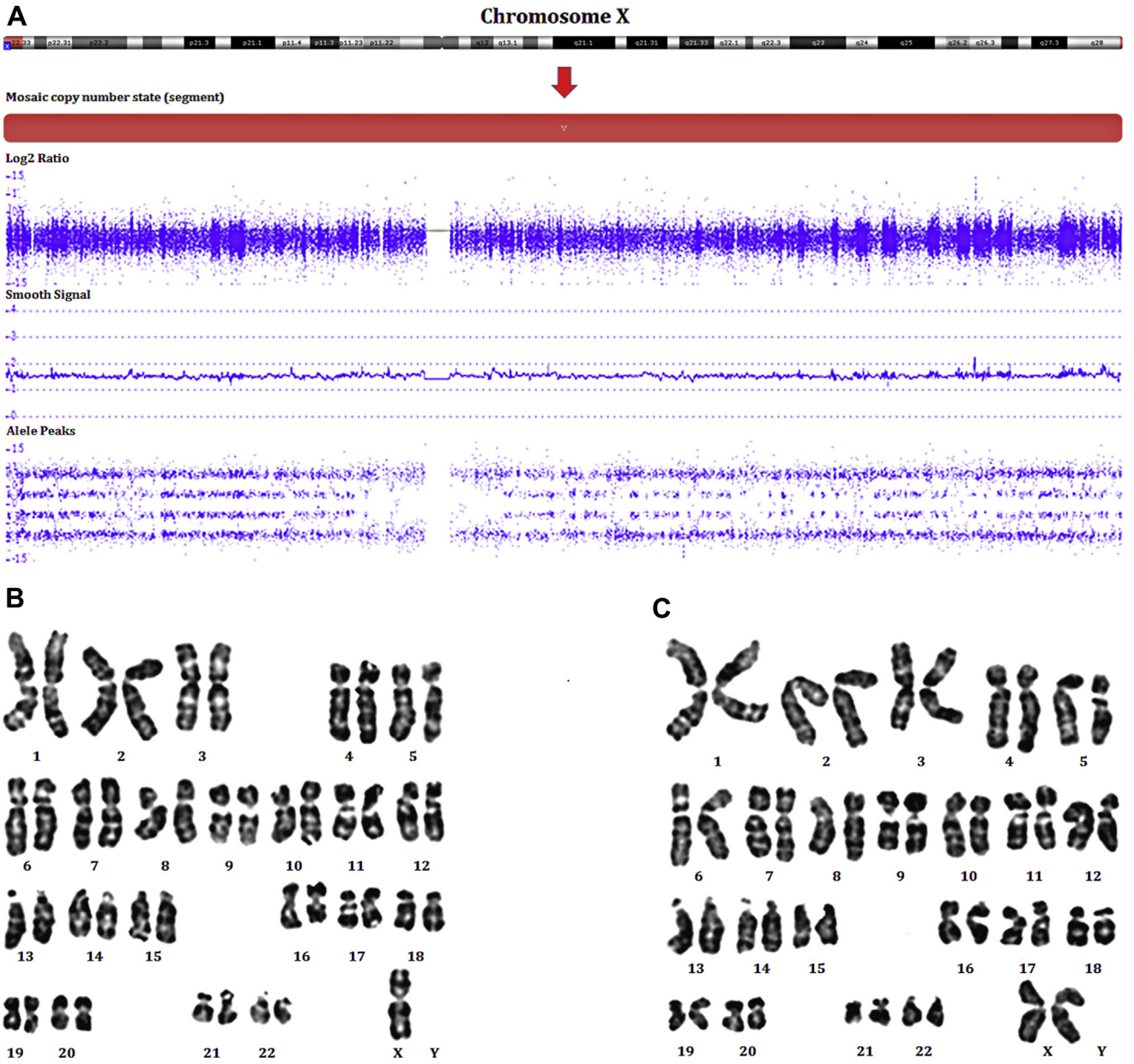

Figure 2 Mosaic loss of the X chromosome detected using the CytoScan HD array (Affymetrix, Santa Clara, CA). (A) The parameters (log_2 ratio, smooth signal, and allele peaks) show the mosaic status of the X chromosome. (B,C) G-banding showing (B) 45,X and (C) 46,XX karyotypes.

absence of *TP53* mutations warranted the investigation of possible genomic alterations associated with cancer predisposition. A recent study of 70 Brazilian LFS/LFL patients with no mutations in *TP53* suggested that multiple loci might be involved in cancer risk (14). None of the CNVs described by the authors coincided with the rare CNVs detected in this study.

Herein, microarray analysis of leukocyte DNA and G-banding analysis revealed the mosaic loss of the X chromosome in similar proportions of cells. Remarkably, the patient had no clinical phenotype of TS. She was taller (159 cm) than the median stature of Brazilian women of age 55 to 64 (156.6 cm) (15). Women with the 45,X karyotype and a normal phenotype have been reported to experience fertility problems (16) or to not undergo puberty (17); our patient went through puberty and experienced no fertility problems. The wide range of TS phenotypes may be the result of different levels of mosaicism within tissues, as well as X-chromosome imprinting, incomplete X inactivation, and gene-dosage effects (18,19).

The high percentage of mosaicism detected in our patient was observed in two different germ layers of derived tissues, ectoderm (50% of oral epithelial cells) and mesoderm (76% of lymphocytes), which strongly suggests that the aneuploidy occurred in a very early stage of embryonic development. A plausible explanation for the reproductive ability and the absence of the classical TS phenotype found in the patient is the presence of a higher number of 46,XX cells in other organs. The increased level of the diploid cell line could be the result of a reduplication of the X chromosome or uniparental disomy (20).

The comparison of genomic profiles in lymphocytic DNA and epithelial cells revealed that only three rare CNVs were shared, with leukocytes demonstrating a large number of alterations. These differences may be related to the proportion of T-cells present in the sample from which the DNA was extracted (21). The loss of 8q24.1 comprised only an intronic region of *EXT1* and appears to be benign. Regarding the genes mapped on the 16q24.3 gain (*PRDM7*, *LOC100130015*, *C16orf3* and *GAS8*), several members of the *PRDM* family may act as tumor suppressors or oncogenes, whereas *GAS8* has been suggested as a potential tumor suppressor (22,23). Nonetheless, although this gain encompass genes associated with cancer, this rare CNV could be benign and not related to cancer risk development. Alternatively, a cryptic germline gene point mutation could explain the LFL syndrome phenotype (24).

Cases of multiple malignant tumors associated with TS are extremely rare (11,12). To our knowledge, only two patients have been reported presenting with TS and LFS, both of them with *TP53* mutations (25,26). The risk of cancer in TS patients increases with the percentage of 45,X cells (9). Our patient presented with a high frequency of cells with 45,X. The X chromosome contains several genes associated with cancer; therefore, it has been suggested that monosomy X is related to the emergence of neoplasms (9).

Overall, we described a rare case of a woman with a TS genotype and a clinical diagnosis of LFL with wild type *TP53*. We demonstrated that genome-wide molecular screening has the potential to uncover unexpected genotypes. New genomic alterations were found in this patient with LFL/TS, including two genes mapped on 16q24.3 with the potential to contribute to cancer-risk development.

Acknowledgments

The authors thank the patients for participating in this study, and Laura Uribe for her support in case analyses. This study was supported by grants from the National Institute of Science and Technology in Oncogenomics (INCITO - São Paulo Research Foundation, FAPESP 2008/57887-9 and Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq 573589/08-9).

Supplementary data

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.cancergen.2015.03.004.

References

1. Malkin D. Li-fraumeni syndrome. Genes Cancer 2011;2:475–484.
2. Bougeard G, Sesboüé R, Baert-Desurmont S, et al. French LFS working group. Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families. J Med Genet 2008;45:535–538.
3. Chompret A, Brugieres L, Ronsin M, et al. P53 germline mutations in childhood cancers and cancer risk for carrier individuals. Br J Cancer 2000;82:1932–1937.
4. Hisada M, Garber JE, Fung CY, et al. Multiple primary cancers in families with Li-Fraumeni syndrome. J Natl Cancer Inst 1998;90:606–611.
5. Stochholm K, Juul S, Juel K, et al. Prevalence, incidence, diagnostic delay, and mortality in Turner syndrome. J Clin Endocrinol Metab 2006;91:3897–3902.
6. Hook EB, Warburton D. Turner syndrome revisited: review of new data supports the hypothesis that all viable 45,X cases are cryptic mosaics with a rescue cell line, implying an origin by mitotic loss. Hum Genet 2014;133:417–424.
7. Elsheikh M, Dunger DB, Conway GS, et al. Turner's syndrome in adulthood. Endocr Rev 2002;23:120–140.
8. El-Mansoury M, Barrenäs ML, Bryman I, et al. Chromosomal mosaicism mitigates stigmata and cardiovascular risk factors in Turner syndrome. Clin Endocrinol (Oxf) 2007;66:744–751.
9. Schoemaker MJ, Swerdlow AJ, Higgins CD, et al. Cancer incidence in women with Turner syndrome in Great Britain: a national cohort study. Lancet Oncol 2008;9:239–246.
10. Brant WO, Rajimwale A, Lovell MA, et al. Gonadoblastoma and Turner syndrome. J Urol 2006;175:1858–1860.
11. Ochi H, Takeuchi J, Sandberg AA. Multiple cancers in a Turner's syndrome with 45,X/46,XXp-/46,XX/47,XXX karyotype. Cancer Genet Cytogenet 1985;16:335–339.
12. Caballé EL, Solé F, Besses C, et al. A new case of Turner syndrome associated with multiple myeloma. Cancer Genet Cytogenet 2000;117:80–81.
13. Eeles RA. Germline mutations in the TP53 gene. Cancer Surv 1995;25:101–124.
14. Silva AG, Krepischi AC, Pearson PL, et al. The profile and contribution of rare germline copy number variants to cancer risk in Li-Fraumeni patients negative for TP53 mutations. Orphanet J Rare Dis 2014;9:63.
15. Instituto Brasileiro de Geografia e Estatística [Brazilian Institute of Geography and Statistics]. Estimativas de população [Estimates of population]. Available at: http://www.ibge.gov.br/home/estatistica/populacao/. Accessed January 21, 2015.
16. Sugawara N, Kimura Y, Araki Y. Case report: a successful pregnancy outcome in a patient with non-mosaic Turner syndrome (45,X) via in vitro fertilization. Hum Cell 2013;26:41–43.
17. Ko JM, Kim JM, Kim GH, et al. Influence of parental origin of the X chromosome on physical phenotypes and GH responsiveness of patients with Turner syndrome. Clin Endocrinol (Oxf) 2010;73:66–71.
18. Nazarenko SA, Timoshevsky VA, Sukhanova NN. High frequency of tissue-specific mosaicism in Turner syndrome patients. Clin Genet 1999;56:59–65.
19. Levitsky LL, Luria AH, Hayes FJ, et al. Turner syndrome: update on biology and management across the life span. Curr Opin Endocrinol Diabetes Obes 2015;22:65–72.
20. Denes AM, Landin-Wilhelmsen K, Wettergren Y, et al. The proportion of diploid 46,XX cells increases with time in women with Turner syndrome-a 10-year follow-up study. Genet Test Mol Biomarkers 2015;19:82–87.
21. Schwienbacher C, De Grandi A, Fuchsberger C, et al. Copy number variation and association over T-cell receptor genes—influence of DNA source. Immunogenetics 2010;62:561–567.
22. Fog CK, Galli GG, Lund AH. PRDM proteins: important players in differentiation and disease. Bioessays 2012;34:50–60.
23. Whitmore SA, Settasatian C, Crawford J, et al. Characterization and screening for mutations of the growth arrest-specific 11 (GAS11) and C16orf3 genes at 16q24.3 in breast cancer. Genomics 1998;52:325–331.
24. Hehir-Kwa JY, Pfundt R, Veltman JA, et al. Pathogenic or not? Assessing the clinical relevance of copy number variants. Clin Genet 2013;84:415–421.
25. Pivnick EK, Furman WL, Velagaleti GV, et al. Simultaneous adrenocortical carcinoma and ganglioneuroblastoma in a child with Turner syndrome and germline p53 mutation. J Med Genet 1998;35:328–332.
26. Krutilkova V, Trkova M, Fleitz J, et al. Identification of five new families strengthens the link between childhood choroid plexus carcinoma and germline TP53 mutations. Eur J Cancer 2005;41:1597–1603.

PLOS | ONE

RESEARCH ARTICLE

Molecular Expression Profile Reveals Potential Biomarkers and Therapeutic Targets in Canine Endometrial Lesions

Fabiana Azevedo Voorwald[1], Fabio Albuquerque Marchi[2], Rolando Andre Rios Villacis[2], Carlos Eduardo Fonseca Alves[3], Gilson Hélio Toniollo[1], Renee Laufer Amorim[3], Sandra Aparecida Drigo[4☯], Silvia Regina Rogatto[2,4☯*]

1 Veterinary Clinic and Department of Surgery, São Paulo State University (UNESP), Jaboticabal, São Paulo, Brazil, **2** International Research Center (CIPE), A. C. Camargo Cancer Center, São Paulo, Brazil, **3** Veterinary Clinic Department, São Paulo State University (UNESP), Botucatu, São Paulo, Brazil, **4** Department of Urology, Faculty of Medicine, São Paulo State University (UNESP), Botucatu, São Paulo, Brazil

☯ These authors contributed equally to this work.
* rogatto@fmb.unesp.br

Citation: Voorwald FA, Marchi FA, Villacis RAR, Alves CEF, Toniollo GH, Amorim RL, et al. (2015) Molecular Expression Profile Reveals Potential Biomarkers and Therapeutic Targets in Canine Endometrial Lesions. PLoS ONE 10(7): e0133894. doi:10.1371/journal.pone.0133894

Editor: Francesco Staffieri, University of Bari, ITALY

Received: February 3, 2015

Accepted: July 2, 2015

Published: July 29, 2015

Data Availability Statement: The microarray gene expression data used in this work was submitted to Gene Expression Omnibus (GEO) database repository (http://www.ncbi.nlm.nih.gov/geo/), accession number GSE69481.

Funding: This study was supported by FAPESP (Fundação de Amparo a Pesquisa do Estado de São Paulo). SRR received research grants from FAPESP (2011/51976-2). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Cystic endometrial hyperplasia (CEH), mucometra, and pyometra are common uterine diseases in intact dogs, with pyometra being a life threatening disease. This study aimed to determine the gene expression profile of these lesions and potential biomarkers for closed-cervix pyometra, the most severe condition. Total RNA was extracted from 69 fresh endometrium samples collected from 21 healthy female dogs during diestrus, 16 CEH, 15 mucometra and 17 pyometra (eight open and nine closed-cervixes). Global gene expression was detected using the Affymetrix Canine Gene 1.0 ST Array. Unsupervised analysis revealed two clusters, one mainly composed of diestrus and CEH samples and the other by 12/15 mucometra and all pyometra samples. When comparing pyometra with other groups, 189 differentially expressed genes were detected. *SLPI*, *PTGS2/COX2*, *MMP1*, *S100A8*, *S100A9* and *IL8* were among the top up-regulated genes detected in pyometra, further confirmed by external expression data. Notably, a particular molecular profile in pyometra from animals previously treated with exogenous progesterone compounds was observed in comparison with pyometra from untreated dogs as well as with other groups irrespective of exogenous hormone treatment status. In addition to *S100A8* and *S100A9* genes, overexpression of the inflammatory cytokines *IL1B*, *TNF* and *IL6* as well as *LTF* were detected in the pyometra from treated animals. Interestingly, closed pyometra was more frequently detected in treated dogs (64% *versus* 33%), with *IL1B*, *TNF*, *LBP* and *CXCL10* among the most relevant overexpressed genes. This molecular signature associated with potential biomarkers and therapeutic targets, such as *CXCL10* and *COX2*, should guide future clinical studies. Based on the gene expression profile we suggested that pyometra from progesterone treated dogs is a distinct molecular entity.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Cystic endometrial hyperplasia (CEH), mucometra, and pyometra are very common uterine diseases in countries where spaying healthy dogs is not a routine practice [1]. Among these diseases, pyometra is of particular importance in veterinary medicine due to its association with septicemia and toxemia [2]. Pyometra or chronic purulent endometritis affects approximately 25% of intact female dogs before the age of 10 and is characterized by bacterial infection and inflammation within the uterine cavity [3], [4]. In open cervix pyometra, inflammation leads to a distended uterus with a purulent vaginal discharge. Conversely, in closed pyometra an intrauterine purulent exudate accumulates leading to an increased risk of sepsis and subsequent death. Ovariohysterectomy is the most effective treatment in the prevention of overwhelming sepsis and disease recurrence [5].

Despite presenting distinct types of uterine fluid and degree of mucin hydration, mucometra and pyometra are considered similar diseases, both causing an accumulation of intrauterine fluid [6]. While pyometra presents an infected purulent fluid, mucometra is characterized by the presence of a sterile seromucous fluid within the uterine cavity [1]. Mucometra is thought to occur with CEH, leading to decreased fertility in breeding animals and having the potential risk of progression to pyometra [1], [2], [4].

The pathogenesis of pyometra is not completely understood. An association between pyometra and diestrus has been reported, confirmed by the absence of the development of pyometra in dogs who have undergone a bilateral oophorectomy [4]. CEH is the most common uterine disease in canines and has been proposed as a lesion that predisposes female dogs to pyometra [6]. CEH is a progressive proliferative process that is mediated by progesterone and potentially exacerbated by estrogen [2], [4]. In the pathogenesis of pyometra, progesterone has been reported by as being involved in endometrial gland secretion stimulation, suppression of the immune response and induction of cervical closure, providing a favorable environment for bacterial growth [2], [4]. In contrast, there is some evidence to suggest that uterine infection or endometrial irritation by foreign bodies may lead to an excessive endometrial hypertrophy and hyperplasia resulting in pronounced endometrial glandular proliferation. In addition, luminal epithelial cell secretion can initiate the development of pyometra or mucometra, depending on its origin, bacterial or not [2], [4], [7].

Although CEH, mucometra and pyometra are very common uterine diseases affecting intact female dogs, there have been a limited number of molecular studies performed on these disorders. To the best of our knowledge, only one study, using a limited number of cases, has reported the gene expression profile in pyometra samples in comparison with normal endometrium [8]. Thus, this study aimed to determine the gene expression profile of pyometra, comparing it with CEH and mucometra, as well as with endometrium samples obtained of healthy female dogs during diestrus. In addition, we sought to identify the molecular signature of closed pyometra, the most life-threatening form of the disease. These data have the potential to identify the biological mechanisms that contribute to uterine endometrial dysfunctions and reveal potential biomarkers that could be useful in clinical practice.

Materials and Methods

Animals

This study was performed according to the National and International Recommendations for the Care and Use of Animals. All procedures were performed under the approval of the Ethics Committee for Animal Experimentation from FCAV-UNESP (Permit Number: 008105/11).

We obtained written owner consent before including any subject in the experiment. The animals were treated according to the norms of the Veterinary Hospital.

Endometrial samples were collected from female dogs, admitted to the Veterinary Hospital for elective or therapeutic ovariohysterectomy (OHE). The sample was composed of 16 cystic endometrial hyperplasia (CEH), 15 mucometra and 17 pyometra specimens (8 open and 9 closed-cervix pyometra samples). Two uterine samples with similar morphological features (approximately 1 cm^2 each) per case were collected from mid-portion of the right and left horn for molecular and histological analyses, respectively. Luminal exudate was excluded from the endometrial samples. Normal endometrium, CEH, mucometra and pyometra cases were diagnosed based on clinical information, physical examination, radiographic and/or ultrasonographic images, and further histologically confirmed by two experienced pathologists (FAV and RLA). The determination of estrus cycle phase was performed through anamnesis, clinical control, vaginal cytology and serum progesterone levels [9,10]. Diestrus phase was also subdivided into early, mid- and late diestrus (S1 Table). Fifty-five animals were in diestrus and 14 in anestrus. Normal endometrial samples were obtained from 21 healthy female dogs during early diestrus. Seventeen of 69 dogs (1 from diestrus, 4 from CEH, 2 from mucometra and 10 from pyometra groups) were previously treated with exogenous progesterone-like compounds to suppress estrous signs. Clinical features of all animals are shown in Table 1. Statistical analysis included Chi-square or Fisher exact test to determine the association between the categorical variables.

Blood Samples and Analysis

Before surgery, blood samples for hematological and progesterone (P4) serum levels were collected from the jugular vein using the vacuum system EDTA (BD Vacutainer Blood Collection Tube–BD, Franklin Lakes, New Jersey, USA) and EDTA-free (BD Vacutainer Serum Tube–BD, Franklin Lakes, New Jersey, EUA) respectively. Biochemical and hematological analysis were performed following routine protocol. Hematological parameters included hematocrit (HCT), white blood cell count (WBC) and platelet count (PLT). Serum P4 levels were measured by chemiluminescence immunoassay according to Tahir et al. [11]. Kruskal-Wallis or Mann-Whitney tests were applied to compare hematological data and P4 levels between the groups. Statistical analysis was carried out using SPSS version 17.0 (SPSS) and the GraphPad Prism 5 (GraphPad Software Inc.) software.

Tissue samples and Histopathological Examination

Uterine tissues were collected immediately after OHE. Samples were both formalin-fixed for histopathology and snap-frozen in liquid nitrogen, the latter being immediately stored at -80°C for mRNA extraction. Formalin-fixed material was paraffin embedded, with hematoxylin and eosin (HE) slides prepared for histological diagnosis (CEH, mucometra or pyometra). Representative photomicrographs of endometrium tissue sections from diestrus, CEH, mucometra and pyometra groups are shown in S1 Fig.

RNA extraction

Fresh frozen tissue samples were macrodissected using sterile scalpel blades, based on areas of endometrium identified following HE evaluation. Tissue samples were submitted to cleavage using lysing tubes in Precellys R tissue homogenizer equipment (BioAmerica Inc, Florida, USA). Total RNA was extracted using TRIzol reagent (Invitrogen Life Technologies Inc., Carlsbad, CA, USA) and the mRNA purified using the RNeasy MiniKit (Qiagen), according to the manufacturer's recommendations. RNA samples were quantified on a Nano-Drop ND-8000

Table 1. Clinical parameters in female dogs from diestrus, cystic endometrial hyperplasia (CEH), mucometra and pyometra groups.

Features	Diestrus	CEH	Mucometra	Pyometra
	N = 21 (%)	N = 21 (%)	N = 15 (%)	N = 17 (%)
Age (years)				
Median (range)	3 (1–9)	10 (3–16)	8 (3–13)	9 (6–14)
Estrus cycle stage				
Proestrus	-	-	-	-
Estrus	-	-	-	-
Diestrus	21 (100)	10 (62)	9 (60)	15 (88)
Early diestrus	21	9	3	9
Mid-diestrus	-	1	3	4
Late diestrus	-		3	2
Anestrus	-	6 (38)	6 (40)	2 (12)
Breed				
American Pit Bull Terrier	-	-	-	2 (12)
American Staff Terrier	-	1 (6)	-	-
Beagle	1 (5)	-	-	-
Boxer	-	1 (6)	1 (7)	-
Brazilian Mastiff	-	-	-	2 (12)
Brazilian Terrier	-	1 (6)	-	-
Cocker Spaniel	-	-	1 (7)	-
German Shepherd	-	2 (12.5)	1 (7)	-
Labrador Retriever	1 (5)	-	-	1 (5.5)
Pinscher	-	1 (6)	-	2 (12)
Poodle	-	2 (12.5)	-	2 (12)
Teckel	-	1 (6)	2 (13)	1 (5.5)
Mixed-breed	19 (90)	7 (44)	10 (67)	7 (41)
Serum progesterone levels (P4) ng/mL				
Mean ± SD error	20.13 ± 3.537	6.656± 2.034	3.576 ± 0.9891	9.893 ± 3.206
WBC (x10^3 μL)				
Mean ± SD error	7996 ± 1109	13070 ± 4327	9617 ± 1781	20789 ± 4535*
HCT (%)				
Mean ± SD error	40.38 ± 3.067	42.19 ± 4.657	37.58 ± 3.505	35.17 ± 2.994
PLT (x10^3 μL)				
Mean ± SD error	251.2 ± 19.350	357.3 ± 91.79	353.9 ± 64.3	232.7 ± 47.46
Previous treatment with exogenous progesterone**				
Yes	1 (5)	4 (25)	2 (13)	10 (59)
No	20 (95)	12 (75)	13 (87)	7 (41)

WBC, White Blood Count; HCT, Hematocrit; PLT, Platelet Count.

* Higher WBC was detected in pyometra compared with mucometra (P = 0.035), CEH (P = 0.0048) and diestrus (P<0.0001, Mann-Whitney test).

** Previous treatment with exogenous progesterone compounds to suppress estrous signs. A significant association of previous progesterone treatment and pyometra was detected (P = 0.001, Chi-square test).

doi:10.1371/journal.pone.0133894.t001

spectrophotometer (Thermo Scientific, Wilmington, NC, USA) and analyzed using Agilent 2100 Bioanalyzer 6000 Nanochip (Agilent Technologies Inc., Waldbronn, BW, Germany). Only samples with a RIN (RNA integrity number) higher than 7.0 were considered for use in the gene expression experiments.

Gene expression profiling and pathway analysis

Gene expression profiling of 69 endometrial samples was performed using Affymetrix Canine Gene 1.0 ST Array (Affymetrix, Santa Clara, CA, USA). Data were extracted using the Affymetrix Genotyping Console (Affymetrix) and normalized using quantile normalization and robust multi-array analysis (RMA) background correction. Filtering characteristics of fold-change -2.0 to 2.0 and a FDR at $P<0.05$ were used to identify the differentially expressed genes. For normalization, processing and statistical analysis the limma v3.22.1 package was used (http://www.bioconductor.org/packages/release/bioc/html/limma.html). The Prism GraphPad software was used for Student's t-test and Kruskal-Wallis test. The molecular processes, functions and molecular networks were further evaluated by analyzing differentially expressed genes using Ingenuity Pathways Analysis (IPA) (Ingenuity Systems, www.ingenuity.com). Protein-protein interaction (PPI) networks were annotated, visualized and analyzed using NAViGaTOR v2.03 (http://ophid.utoronto.ca/navigator/). Enrichment analysis by the IPA and PPI network analyses were performed based on human data. The microarray data are available on the Gene Expression Omnibus repository (GEO, http://www.ncbi.nlm.nih.gov/geo/), accession number GSE69481.

Results

Clinical data

Eleven different breeds, with the majority being mixed-breed dog, were represented in all groups (Table 1).

Higher serum levels of progesterone were observed in diestrus group compared with CEH, mucometra and pyometra ($P<0.0001$; $P<0.0001$; $P = 0.006$, respectively; S1 Fig and Table 1). Although pyometra samples showed a higher P4 levels in comparison with other uterine disorders, it was not statistically significant. Interestingly, significantly higher serum levels of P4 were detected in dogs with closed pyometra when compared with open pyometra ($P = 0.0006$, S2 Fig).

As expected, higher WBC were detected in pyometra group when compared with mucometra ($P = 0.035$), CEH ($P = 0.0048$) and diestrus ($P<0.0001$) (Table 1). No difference was observed for the other hematological parameters.

A significant association of previous exogenous progesterone treatment and pyometra ($P = 0.001$) was detected. The highest frequency of treated animals was observed in pyometra (59%, 10/17), followed by CEH (25%, 4/16), mucometra (13%, 2/15) and diestrus cases (5%, 1/21) (Table 1).

Molecular analysis

Unsupervised hierarchical clustering analysis revealed two main clusters, with mucometra and pyometra samples showing a strong trend to cluster together, and diestrus and CEH samples mainly grouped in the other cluster (Fig 1). Pyometra showed 189 differentially expressed genes in comparison with other groups (diestrus, CEH and mucometra), with 169 overexpressed and 20 underexpressed (S2 Table). The *SLPI* (*secretory leukocyte peptidase inhibitor*) gene was detected as having the highest fold change, being 30 times more expressed in pyometra compared with other endometrial tissues (S2 Table). Among the highest upregulated genes in pyometra, three metalloproteinase genes (*MMP13*, *MMP1* and *MMP12*) and three S100 family members (*S100A12*, *S100A8* and *S100A9*) were observed. Conversely, the *EPHA7* (ephrin receptor A7) gene was detected as having the lowest fold change in pyometra.

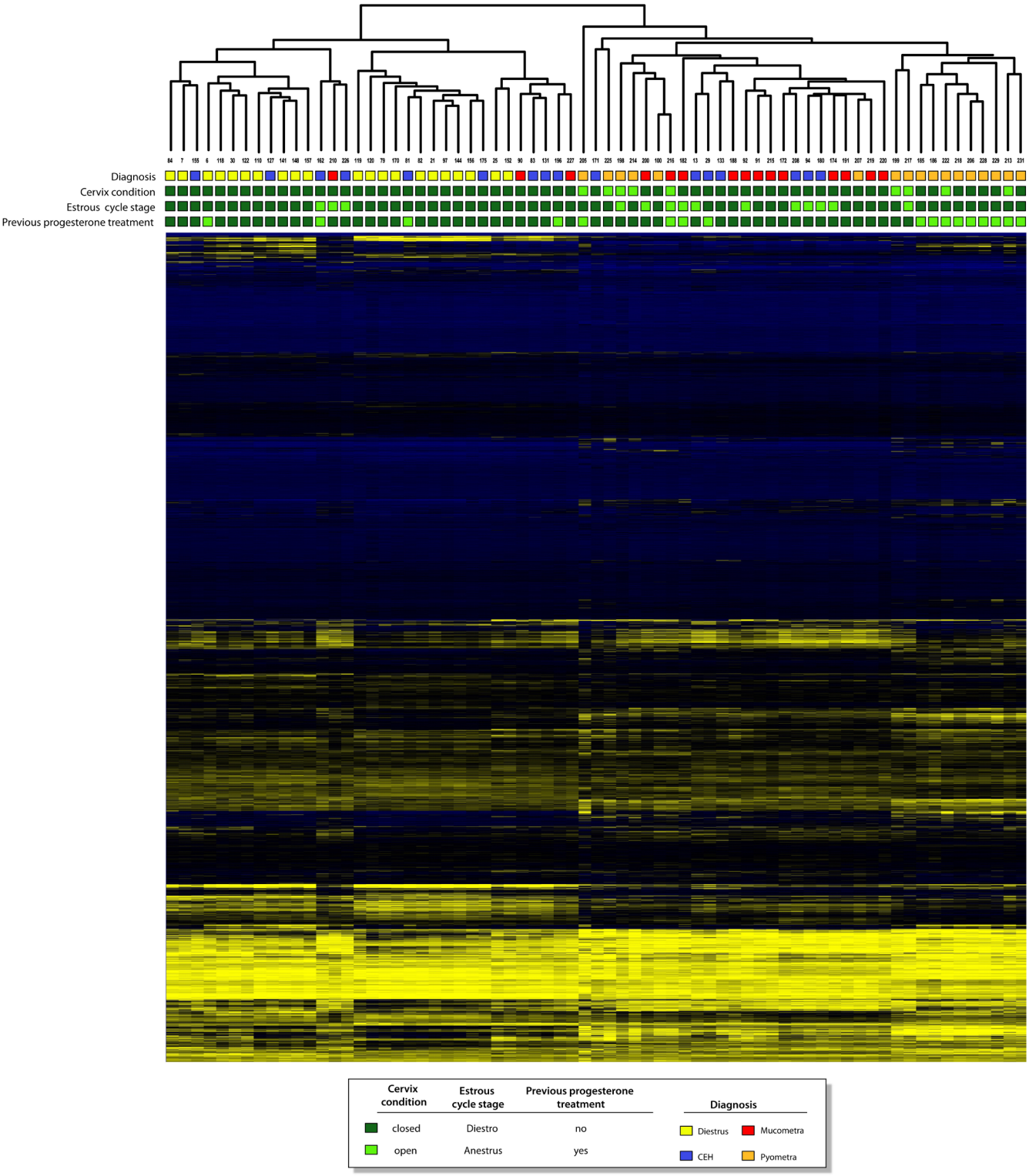

Fig 1. Hierarchical clustering analysis. Dendogram resulted from unsupervised analysis including diestrus, cystic endometrial hyperplasia (CEH), mucometra and pyometra samples. Cervix condition, estrus cycle phase and previous exogenous progesterone treatment status for each sample are shown.

doi:10.1371/journal.pone.0133894.g001

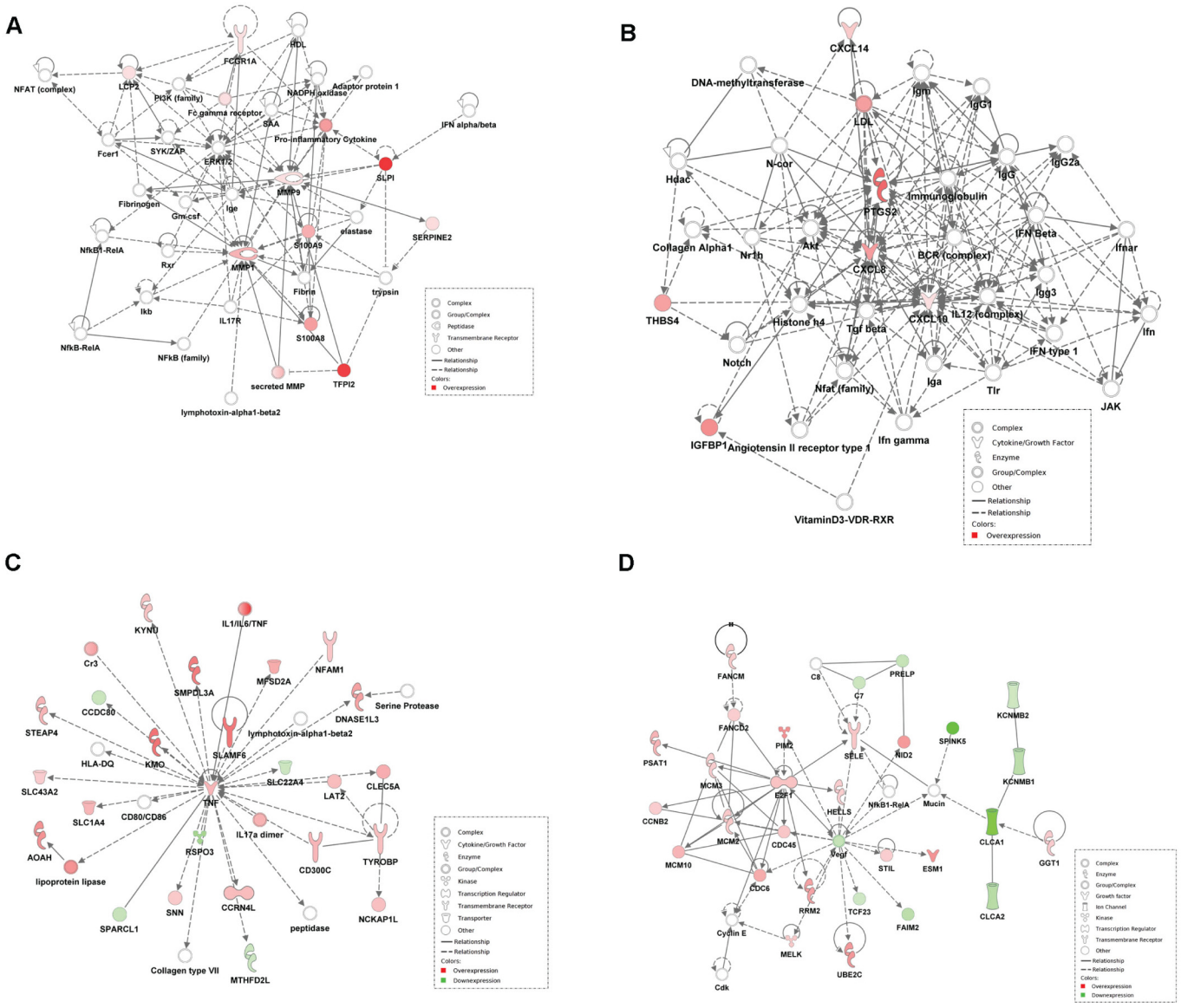

Fig 2. IPA network analysis. (A-D) Top networks identified with IPA software for differentially expressed genes in different group's comparisons: A-B: Pyometra compared with diestrus, CEH and mucometra. A: Family members of the MMP and *S100* genes were detected as central nodes in pyometra and connected with pro-inflammatory cytokines. The *SLP1* gene (highest fold change in pyometra) is also depicted. B: Multiple interactions between the *CXCL8/IL8*, a pro-inflammatory gene, and other genes are shown, including the *PTGS2/COX2* gene. C-D: Pyometra of animals previously treated with progesterone compounds compared to pyometra of untreated dogs. C: The proinflammatory cytokine TNF was detected as central with multiple connections with different genes. D: *E2F1* (overexpression) and *VEGF* (underexpression) products in the treated group. The lines between genes represent known interactions, with solid lines and dashed lines representing direct and indirect interactions, respectively. Different node shapes represent the functional class of the gene product. Red and green nodes represent overexpressed and underexpressed genes in each comparison.

doi:10.1371/journal.pone.0133894.g002

Canonical pathway and network analysis by IPA were performed for the differentially expressed genes in pyometra. Network analysis showed multiple interactions between the *MMP* genes and *S100* family genes detected with the highest fold change (Fig 2A). In addition, important interactions between overexpressed *CXCL8* gene, a chemokine that is one of the major mediators of the inflammatory response, and other genes were highlighted, including the *PTGS2/COX2* gene (Fig 2B).

A comparison between gene expression data in pyometra *versus* normal endometrium using the Hagman et al. (2009) [8] findings and our data (pyometra *vs*. diestrus) was also performed. Concordance of 58 differentially expressed genes was identified, with 33 being underexpressed and 25 overexpressed in pyometra (Table 2). The *SLPI* (*secretory leukocyte peptidase inhibitor*) gene was detected as being the top up-regulated gene in both studies. In addition, overexpression of the *MMP* and *S100* family genes was confirmed. To identify specific altered genes in pyometra, the list of differentially expressed genes detected in pyometra versus other groups (diestrus, CEH and mucometra) was compared with the list of 58 common genes detected by Hagman et al [8] data's comparison. As a result, only 29 (50%) differentially expressed genes remained on the list, with 22 being overexpressed and 7 underexpressed (Table 2, highlighted in bold).

The unsupervised clustering analysis (Fig 1) also revealed two distinct groups of pyometra, one of them being closer to diestrus and CEH samples, with the other clustering with CEH and mucometra samples, yet still isolated from them. Interestingly, the latter comprised a group of pyometra specimens obtained from bitches previously treated with exogenous progesterone compounds. In contrast, the pyometra samples that clustered near the diestrus and CEH samples were mostly obtained from untreated dogs (Fig 1). These findings indicated different molecular alterations in pyometra as a result of previous progesterone treatment. Estrous cycle phase and cervix conditions (open or closed) in each sample are also shown and were randomly distributed over the groups (Fig 1). In addition, the molecular profile was not correlated with histopathological features in each group (data not shown).

Further supervised clustering analysis was performed in CEH, mucometra and pyometra samples according to exogenous progesterone treatment status (Fig 3A). In accordance with previous unsupervised analysis results, two major clusters were observed, being one composed by all pyometra samples from treated dogs and the other cluster comprised by the pyometra samples from untreated animals as well as all by the other groups irrespective of treatment status (Fig 3A). These data suggest that pyometra from hormone-treated dogs present a distinct molecular signature. Therefore, the molecular profile of pyometra according to previous hormone treatment was investigated. Supervised clustering analysis revealed 194 differentially expressed genes, in treated (N = 10) versus untreated (N = 7) dogs, with 57 being underexpressed and 134 overexpressed (Fig 3B). The top 20 differentially expressed genes detected in this analysis are shown in Table 3. Interestingly, three members of the S100 family (*S100A8*, *S100A9* and *S100A12*) presented the highest fold changes among the up-regulated genes in pyometra from the hormone-treated group (Table 3). In addition, pro-inflammatory cytokines and chemokines (*TNF*, *IL1B*, *IL6*, and *CCL3*, among others) were also upregulated in the treated animals. Network analysis showed *TNF* as a central gene with multiple connections with other genes (Fig 2C) and an overexpression of the transcript factor *E2F1* gene interacting with different genes (Fig 2D).

In order to identify a molecular signature for closed pyometra, a life-threatening condition, the expression profile of closed pyometra compared with open pyometra was investigated (Fig 3C). Eighty-two differentially expressed genes were detected, but no significant difference was observed after Bonferroni correction. Characterization of exclusive molecular alterations in closed and open pyometra was also sought, with the aim of identifying putative biomarkers. Firstly, two lists of significant genes exclusively expressed in open pyometra *versus* diestrus and closed pyometra *versus* diestrus were generated and further compared to reveal exclusively altered genes in each group. Interestingly, closed pyometra revealed 70 exclusively altered genes, while open pyometra had 34 (S3 and S4 Tables, respectively). The top five-upregulated genes in closed pyometra were *LBP*, *CCL3*, *IL1B*, *CXCL10* and *ITGAM*; while in open pyometra were *FABP3*, *IL7*, *TNC*, *SDC1* and *CLDN2*. With the aim of revealing potential biomarkers and

Table 2. Comparative analysis of the differentially expressed genes in pyometra *versus* normal endometrium as described by Hagman et al (2009) [8] and our data.

Genes	Fold change	
	(Pyometra *versus* Normal*)	(Pyometra *versus* Normal [8])
SLPI	50.26	344.8
TFPI2	47.10	48.1
PTGS2	36.59	88.7
SMPDL3A	27.78	194.4
IGFBP1	27.29	104.4
IL8	23.68	242.9
S100A8	23.02	56.6
C6	22.93	48.5
THBS4	22.82	31.2
S100A9	19.43	160.7
SPP1	18.54	27.7
MMP1	18.21	44.7
SRGN	12.96	53.2
CXCL14	12.74	47.0
CCL2	12.54	44.8
SELL	9.94	28.3
MS4A7	9.52	22.7
SERPINE2	8.33	25.6
AOAH	8.05	37.0
MMP9	7.31	29.4
CASP4	6.92	28.1
CXCL10	6.91	31.1
FCGR1A	6.67	42.5
LCP2	6.11	23.4
C5AR1	5.65	43.4
EPHA7	-22.05	-26.0
SULT1D1	-18.49	-30.6
FEZ1	-6.85	-10.7
MSX2	-6.15	-7.9
LEF1	-5.01	-7.4
MSX1	-4.99	-6.3
TFCP2L1	-4.89	-15.2
ANK3	-4.24	-8.1
CMTM8	-4.22	-5.7
DEPDC7	-4.09	-6.6
RASGRP1	-4.04	-5.4
SLC30A2	-3.99	-7.4
RGS22	-3.88	-8.9
DLX5	-3.79	-7.4
GCLC	-3.72	-7.0
PPAPDC1A	-3.68	-6.9
CTH	-3.58	-6.1
RHPN2	-3.40	-8.0
ALDH1A1	-3.19	-7.1
TFF2	-3.14	-9.2

(*Continued*)

Table 2. (*Continued*)

Genes	Fold change	
	(Pyometra *versus* Normal*)	(Pyometra *versus* Normal [8])
SNCAIP	-3.08	-10.9
EPHX2	-2.98	-8.3
GRIP1	-2.96	-5.9
WIF1	-2.86	-5.5
SH3BGRL2	-2.85	-6.0
EFHC2	-2.72	-6.1
PPP1R1B	-2.65	-6.6
HYI	-2.55	-7.9
ENPP6	-2.40	-7.9
FOXA2	-2.38	-7.8
STRBP	-2.38	-6.2
NAALAD2	-2.29	-6.4
NDP	-2.26	-8.3

* Normal endometrium samples were obtained from dogs in early diestrus.

In bold are indicated those genes that were identified as differentially expressed when pyometra group was compared with the other groups (diestrus, CEH and mucometra).

doi:10.1371/journal.pone.0133894.t002

therapeutic targets for closed pyometra, a gene set enrichment analysis by IPA was performed, revealing 21 genes exclusively expressed in closed pyometra (Table 4). The *IL1B* gene, which encodes a proinflammatory cytokine, was detected as the highest upregulated gene (Fold change = 9.29, Table 4). Overexpression of the *CXCL10*, *NNMT*, *MMP8*, *F3* and *TNF* genes was also identified in closed pyometra (Table 4). Thereafter, PPI networks were constructed using NAViGaTOR based on genes with altered expression detected exclusively in closed and open pyometra, in order to highlight potential biomarkers and/or therapeutic targets to therapy for each condition (Fig 4). In closed pyometra, four up-regulated genes were revealed as potential biomarkers and therapeutic targets, including *CXCL10*, *IL1B*, *KDR*, and *TNF*. The *LBP* gene, detected with the highest fold change, was indicated as a potential diagnostic marker in closed pyometra. In open pyometra, four overexpressed genes (*ITGAV*, *FGFR3*, *SRC* and *PTGS1*) were indicated as potential biomarkers and drug targets.

Discussion

In this study, the global expression profile of pyometra was described in comparison to diestrus, CEH and mucometra, with a significant number of cases in each group. A distinct molecular profile for pyometra in female dogs previously treated with exogenous progesterone compounds compared with untreated dogs was described, indicating that they are, in fact, distinct molecular entities. This study also explored the molecular profile of closed pyometra, a condition that remains both a diagnostic and therapeutic challenge, in order to identify putative biomarkers and/or molecular therapeutic targets.

Unsupervised analysis of transcriptomic profiles allowed the identification of two major groups, with one of them having essentially diestrus and CEH samples, and the other composed of 7/16 CEH samples, 12/15 mucometra and all pyometra specimens. These data corroborate the current knowledge that CEH may predispose to the development of mucometra and

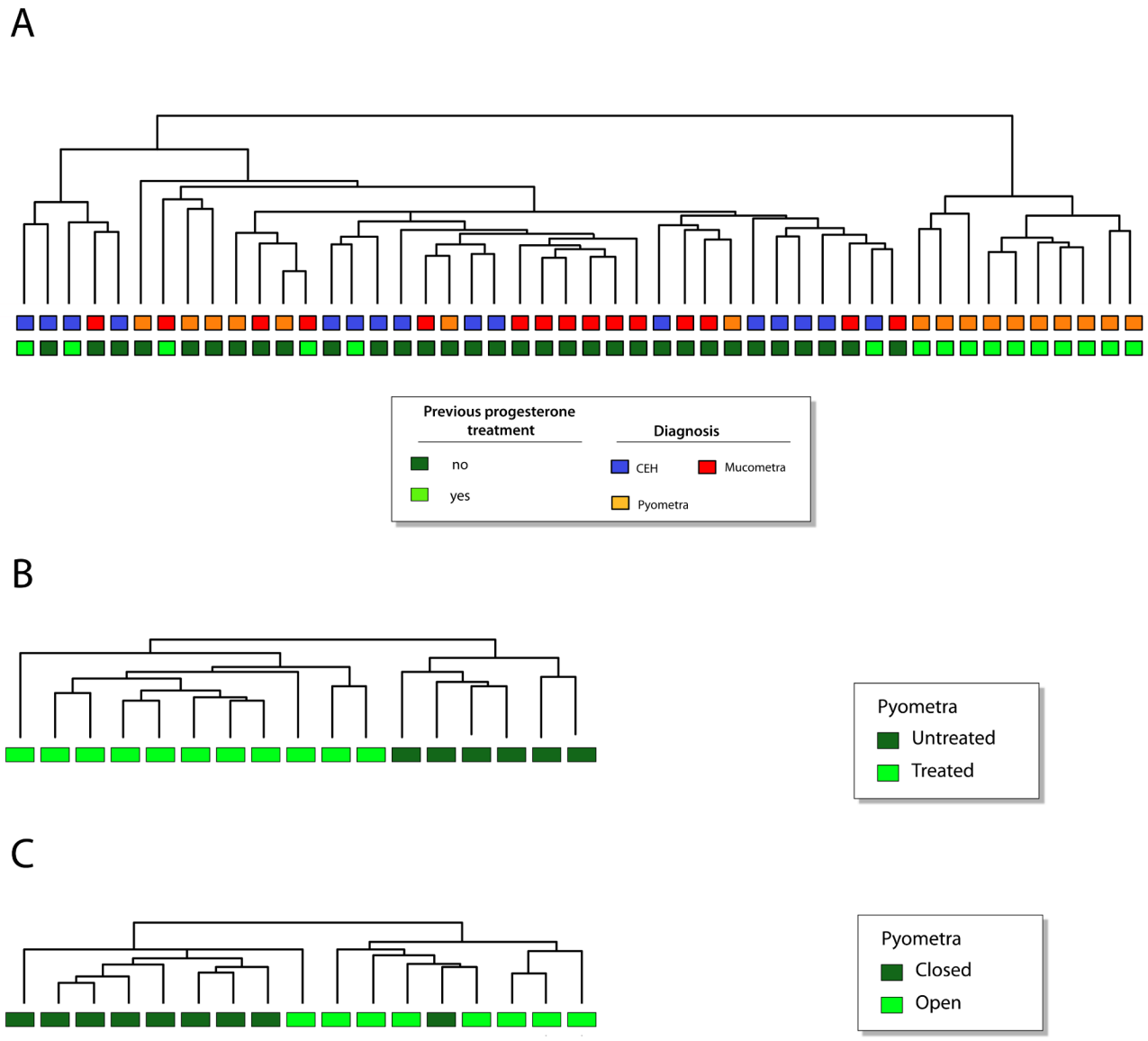

Fig 3. Supervised hierarchical clustering analysis of global gene expression data. Supervised analysis according to: **A:** absence or presence of previous exogenous progesterone treatment in female dogs with CEH, mucometra and pyometra; **B**. absence or presence of previous exogenous progesterone treatment in female dogs with pyometra. **C:** cervix condition in pyometra (open versus closed).

doi:10.1371/journal.pone.0133894.g003

pyometra. CEH can lead to endometrial thickening with subsequent accumulation of fluids within uterine cavity, increasing the risk of pyometra or mucometra development [1].

A cross study validation was performed via the comparison of our data (pyometra *versus* diestrus) with the Hagman's data [8] revealing 58 genes in common in both studies, thus confirming our findings. However, 50% of these genes (29/58) remained as differentially expressed in pyometra after the comparison between pyometra versus other groups (diestrus, CEH and mucometra). This analysis, allowed the identification of altered genes exclusive to pyometra. The *SLPI* gene identified as the top overexpressed gene in pyometra encodes an antimicrobial peptide secreted by epithelial tissues. In addition, SPLI modulates infection and inflammation by neutralizing lipopolysaccharide (LPS) from Gram-negative bacteria impairing innate immune activation by toll-like receptors (TLRs) [12]. *SLPI* was previously reported as

Table 3. Top 20 differentially expressed genes in pyometra according to previous exogenous progesterone treatment.

Genes	Average treated group	Average untreated group	Fold change	*P* value	Adjusted *P* value
S100A12	10.76	7.18	11.98	0.0001	0.0028
S100A9	11.47	8.18	9.79	0.0000	0.0010
IL1B	7.89	4.64	9.46	0.0000	0.0010
S100A8	10.20	7.06	8.82	0.0000	0.0013
PI3	11.12	8.17	7.68	0.0001	0.0024
EMR1	9.29	6.52	6.83	0.0000	0.0022
LTF	11.40	8.68	6.59	0.0001	0.0036
IL6	8.05	5.38	6.36	0.0000	0.0016
SLC7A11	8.76	6.10	6.36	0.0000	0.0002
CCNB1	5.92	3.28	6.21	0.0001	0.0029
PLEK	9.79	7.20	6.02	0.0000	0.0010
KMO	7.81	5.25	5.91	0.0002	0.0044
CLEC4E	8.28	5.73	5.83	0.0000	0.0012
DGAT2	9.25	6.72	5.76	0.0000	0.0008
CCL3	9.54	7.02	5.74	0.0000	0.0011
CHL1	5.56	9.61	-16.57	0.0000	0.0001
KRT5	5.01	8.98	-15.67	0.0000	0.0009
MAMDC2	6.54	9.81	-9.64	0.0000	0.0000
SPINK5	4.48	7.62	-8.83	0.0004	0.0066
ALDH1A1	5.40	8.30	-7.47	0.0000	0.0012

* Differentially expressed genes were defined by a significant Bonferroni correction ($P < 0.05$).

doi:10.1371/journal.pone.0133894.t003

expressed in woman reproductive tract and induced by progesterone [13]. Up-regulation of the *SPLI* in pyometra may be endometrium-protective, both against microorganisms and from immune-mediated tissue damage.

Notably, both supervised analyses according to hormone treatment status evaluating all uterine disease samples (CEH, mucometra and pyometra) and only pyometra group confirmed a molecular signature for pyometra in treated dogs. These findings highlight the molecular alterations underlying the well-known association of pyometra with progesterone [2], [4]. Overexpression of *S100A8*, *S100A9* and *S100A12* genes were detected in pyometra, particularly in the hormone-treated animals. Several S100 proteins, including S100A8, S100A9 and S100A12, have been identified as endogenous danger-associated molecular patterns (DAMPs). DAMPs are intracellular molecules, which are released following cell death and can be recognized by the innate immune system, signaling tissue damage [14], [15]. S100A8 and S100A9 stimulate the production of the pro-inflammatory cytokines TNF, IL-6, IL-1B and IL-8 [14]. Accordingly, the *IL1B* and *IL8* genes were detected as overexpressed in pyometra group, being *IL1B*, *IL6* and *TNF* overexpressed particularly in pyometra of hormone-treated animals. Drugs targeting the S100A8/S100A9 complex leading to modulation of inflammatory response have been proposed in the treatment of cardiovascular disease [14]. Thus, high levels of *S100A8* and *S100A9* in pyometra may be a result of tissue damage leading to amplification and/or perpetuation of the local inflammation, being an attractive therapeutic target.

Overexpression of *PTGS2/COX2* was detected in pyometra, confirming previous report [16]. COX-2, also detected as differentially expressed in pyometra of hormone-treated dogs, is the more important source of prostaglandins and thromboxane A2 in inflammation. Furthermore, COX-2 is downstream of TLR signaling after activation by endogenous S100 proteins

Table 4. Gene set enrichment analysis in closed pyometra detected by IPA analysis.

Gene	Location	Type(s)	Fold change
IL1B	Extracellular Space	Cytokine	9.29
CXCL10	Extracellular Space	Cytokine	6.22
NNMT	Cytoplasm	Enzyme	4.8
MMP8	Extracellular Space	Peptidase	3.83
F3	Plasma Membrane	Transmembrane receptor	3.37
TNF	Extracellular Space	Cytokine	3.07
MMP3	Extracellular Space	Peptidase	2.84
KDR	Plasma Membrane	Kinase	2.49
RARB	Nucleus	Ligand-dependent nuclear receptor	2.35
PDGFRB	Plasma Membrane	Kinase	2.34
ANGPT1	Extracellular Space	Growth factor	2.29
ADORA2A	Plasma Membrane	G-protein coupled receptor	2.27
PDGFRA	Plasma Membrane	Kinase	2.19
IL2RB	Plasma Membrane	Transmembrane receptor	2.1
FN1	Extracellular Space	Enzyme	2.1
TEK	Plasma Membrane	Kinase	2.07
FOLH1	Plasma Membrane	Peptidase	-2.02
ALDH2	Cytoplasm	Enzyme	-2.07
ESR1	Nucleus	Ligand-dependent nuclear receptor	-2.4
EPCAM	Plasma Membrane	Other	-2.36
CHRNA7	Plasma Membrane	Transmembrane receptor	-2.2

doi:10.1371/journal.pone.0133894.t004

and other stimulus like LPS [15]. Thus, in pyometra a continuous activation of TLR by both endogenous and exogenous stimulus can lead to an exacerbated inflammatory response. COX-2 interacts directly with CXCL8/IL-8 and CXCL14 (Fig 2B), chemokines that were also overexpressed in pyometra. High serum levels of IL-8 were observed in dogs with pyometra, particularly in those that developed Systemic Inflammatory Response Syndrome (SIRS), suggesting that IL-8 may contribute to the development of a systemic disease in dogs with pyometra [17]. Overexpression of *CXCL14* can contribute to inflammatory cell infiltration in pyometra due to its chemoattractive action in immune cells, like monocytes and natural killer cells [18]. Both *IL8* and *CXCL14* are induced by *COX2*, thus selective COX-2 inhibitors could be an interesting treatment option in pyometra in terms of control the inflammatory response.

Overexpression of the *LTF* gene was detected in pyometra of hormone-treated dogs. Lactoferrin gene and protein overexpression were previously reported in pyometra [19]. Lactoferrin interacts with LPS on the bacterial surface and activates TLR-4 on the surface of phagocytes and epithelial cells. Interestingly, LTF can also be considered an inflammatory regulator by impairing LPS ligation to TLR and subsequent activation of the inflammatory response [20]. Thus, LTF might be induced in pyometra to counterbalance the inflammatory response. In contrast, the pro-inflammatory *TNF* gene was identified as a central gene in pyometra of hormone-treated dogs (Fig 3C). Considering that TNF has a pivotal role in inflammation, therapies targeting TNF could be considered in pyometra. In conclusion, a more pronounced inflammatory gene signature for pyometra of treated dogs was revealed, suggesting that repetitive progesterone exposure may contribute to bacterial colonization of the endometrium and an inflammatory response. This inflammatory environment enables the exposition and recognition of DAMPs by innate immune receptors, such as TLRs and production of pro-

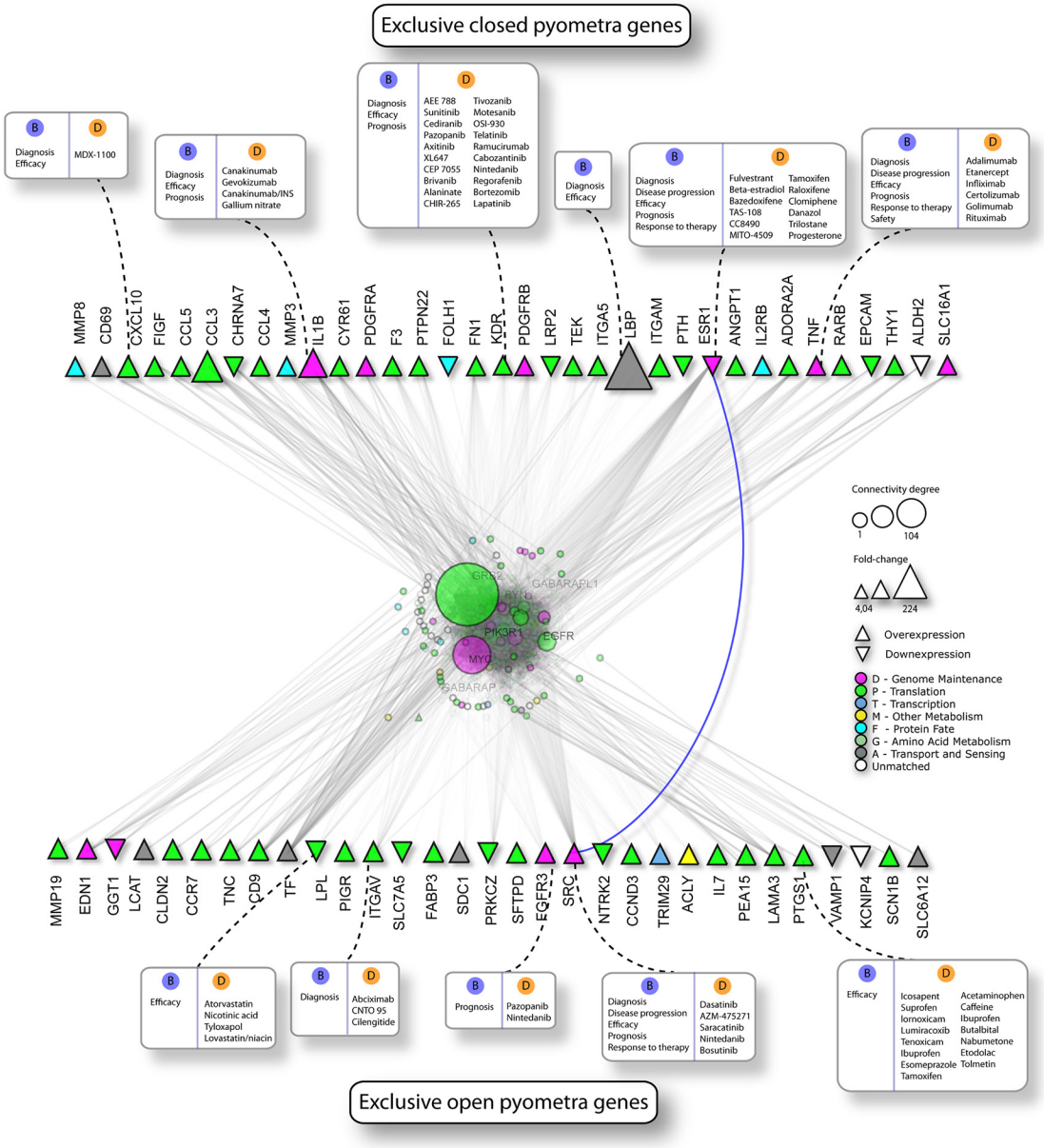

Fig 4. Protein-protein interaction (PPI) networks in closed and open pyometra. PPI networks based on altered genes exclusively detected in closed pyometra and open pyometra and their interactive partners built and visualized with Navigator v.2.3. Potential candidates for biomarkers (B blue circles) and targets for therapy (Drugs, D orange circles) in closed pyometra and open pyometra are highlighted. Triangles represent the genes with altered expression in each group and are color-coded according to Gene Ontology (GO). Upright and inverted triangles indicate overexpressed and underexpressed genes, respectively.

doi:10.1371/journal.pone.0133894.g004

inflammatory cytokines and other inflammatory molecules, such as COX-2, aggravating the inflammatory process.

With regard to closed pyometra, the inflammatory process can progress to sepsis and SIRS, present either with or without infection [21]. Interestingly, a high frequency of closed pyometra was observed among hormone-treated dogs (7/10 cases compared to untreated dogs (2/7). Further studies are needed to confirm if repeated exogenous progesterone treatment predisposes to development of closed pyometra. Treatment with progesterone compounds to suppress heat

cycles in female dogs is common in countries where spaying dogs is not a routine practice. Thus, these results are clinically relevant and neutering strategies in intact animals should be considered since treatment with progesterone compounds might ultimately produce undesired results. However, care should be taken considering that ovariohysterectomy during diestrus can lead to subtle progesterone decline and hyperprolactinemia with subsequent pseudopregnancy [22].

In this study, detection of exclusively altered genes in closed and open pyometra combined with enrichment analysis by IPA and PPI analysis were described, suggesting potential biomarkers and molecular targets for therapy. In closed pyometra, overexpression of *IL1B* and *TNF* were suggested as therapeutic targets, specifically anti-IL1 and anti-TNF antibodies. In addition, *CXCL10* overexpressed in closed pyometra is important in host immunity; but its excessive activation may lead to a detrimental immune response. The CXCL10 protein and its receptor (CXCR3) have been proposed as therapeutic targets in cancer and in immune-mediated diseases. Ongoing phase II clinical trials MDX-1100-anti-CXCL10, indicated here by PPI analysis in closed pyometra, have been conducted in inflammatory bowel disease and rheumatoid arthritis [23]. High serum levels of CXCL10, IL-6, IL-10, TNF and IL-8 were reported in a canine sepsis model [24]. Thus, CXCL10 in combination with other cytokines such as TNF are potential biomarkers for sepsis as a consequence of pyometra progression and also as targets in anti-inflammatory therapy in pyometra. *LBP* gene was also suggested as a diagnostic marker in closed pyometra. LBP is an acute-phase protein synthesized in the liver, but also by epithelial cells of other tissues, including the reproductive tract [25]. The LBP binds to LPS and other microorganism components leading to activation of the inflammatory response [25]. LBP serum levels have been proposed as a useful diagnostic marker in both urinary infections in children and atherosclerosis, as well as a prognostic marker in acute appendicitis [26]-[28]. Future studies are warranted to confirm LBP as meaningful diagnostic marker in closed pyometra. In open pyometra, *PTGS1/COX1* overexpression was detected as an exclusively altered gene. COX-1 is expressed constitutively in most cells and is also responsible for prostaglandin production during inflammation, which has the potential for being targeted by numerous anti-inflammatory drugs, including ibuprofen.

Conclusions

To the best of our knowledge, this is the first report that described the expression profile of pyometra compared with other common uterine diseases, such as CEH and mucometra, in intact female dogs. In addition, a molecular signature for closed pyometra was described, which should be further explored to reveal prognostic and predictive biomarkers. Finally, distinct pyometra expression patterns from progesterone treated and untreated dogs were observed, with a more pronounced inflammatory signature in treated dogs, which suggests that previous progesterone exposure may contribute to an exacerbated inflammation in pyometra. A high frequency of closed pyometra in the hormone-treated group was observed and future studies are warranted to verify the impact of progesterone treatment on the development of closed pyometra.

Supporting Information

S1 Fig. Representative photomicrographs of endometrial specimens. HE stained endometrial sections obtained from diestrus, CEH, mucometra and pyometra groups (x5 and x10 magnification).
(TIF)

S2 Fig. Serum levels of progesterone (P4) detected in diestrus, CEH, mucometra and pyometra (open and closed) groups. Mann-Whitney test.
(TIF)

S1 Table. Diestrus subphases diagnostic criteria.
(DOCX)

S2 Table. Top 20 differentially expressed genes in pyometra samples compared to other groups (diestrus, CEH and mucometra samples).
(DOCX)

S3 Table. Exclusively altered genes in closed pyometra.
(DOCX)

S4 Table. Exclusively altered genes in open pyometra.
(DOCX)

Acknowledgments

The authors would like to thank the staff of the Pathology Department and Veterinary Clinic Department, UNESP–Botucatu, SP.

Author Contributions

Conceived and designed the experiments: SRR FAV. Performed the experiments: FAV RARV. Analyzed the data: FAV FAM CEFA RLA SAD SRR. Contributed reagents/materials/analysis tools: SRR GHT. Wrote the paper: CEFA RLA SAD SRR.

References

1. Hagman R. Diagnostic and prognostic markers for uterine diseases in dogs. Reprod Domest Anim. 2014; 49 Suppl 2: 16–20. doi: 10.1111/rda.12331 PMID: 24947856
2. Pretzer SD. Clinical presentation of canine pyometra and mucometra: a review. Theriogenology 2008; 70: 359–363. doi: 10.1016/j.theriogenology.2008.04.028 PMID: 18513791
3. Egenvall A, Hagman R, Bonnett BN, Hedhammar A, Olson P, Lagerstedt AS. Breed risk of pyometra in insured dogs in Sweden. J Vet Intern Med. 2001; 15: 530–538. PMID: 11817057
4. Verstegen J, Dhaliwal G, Verstegen-Onclin K. Mucometra, cystic endometrial hyperplasia, and pyometra in the bitch: advances in treatment and assessment of future reproductive success. Theriogenology 2008; 70: 364–374. doi: 10.1016/j.theriogenology.2008.04.036 PMID: 18501959
5. Fieni F, Topie E, Gogny A. Medical treatment for pyometra in dogs. Reprod Domest Anim. 2014; 49 Suppl 2:28–32. doi: 10.1111/rda.12302 PMID: 24947858
6. De Bosschere H, Ducatelle R, Vermeirsch H, Van Den Broeck W, Coryn M. Cystic endometrial hyperplasia-pyometra complex in the bitch: should the two entities be disconnected? Theriogenology 2001; 55: 1509–1519. PMID: 11354710
7. Smith FO. Canine pyometra. Theriogenology 2006; 66: 610–612. PMID: 16828152
8. Hagman R, Rönnberg E, Pejler G. Canine uterine bacterial infection induces upregulation of proteolysis-related genes and downregulation of homeobox and zinc finger factors. PLoS One 2009; 4:e8039. doi: 10.1371/journal.pone.0008039 PMID: 19956711
9. Sorribas CE. Atlas de reprodução canina. São Caetano do Sul: Interbook; 2006.
10. Johnston SD, Kustritz MV, Olson PN. Disorders of the canine uterus and uterine tubes (oviducts). In: Johnston SD, Kustritz MV, Olson PN, editors. Canine and feline theriogenology. Philadelphia: WB. Saunders Company; 2001. pp. 206–224.
11. Tahir MZ, Thoumire S, Raffaelli M, Grimard B, Reynaud K, Chastant-Maillard S. Effect of blood handling conditions on progesterone assay results obtained by chemiluminescence in the bitch. Domest Anim Endocrinol. 2013; 45:141–144. doi: 10.1016/j.domaniend.2013.07.002 PMID: 23988180
12. Scott A, Weldon S, Taggart CC. SLPI and elafin: multifunctional antiproteases of the WFDC family. Biochem Soc Trans. 2011; 39: 1437–1440. doi: 10.1042/BST0391437 PMID: 21936829

13. Horne AW, Stock SJ, King AE. Innate immunity and disorders of the female reproductive tract. Reproduction 2008; 135: 739–749. doi: 10.1530/REP-07-0564 PMID: 18502890
14. Schiopu A, Cotoi OS. S100A8 and S100A9: DAMPs at the crossroads between innate immunity, traditional risk factors, and cardiovascular disease. Mediators Inflamm. 2013; 2013: 828354. doi: 10.1155/2013/828354 PMID: 24453429
15. Sangiuliano B, Pérez NM, Moreira DF, Belizário JE. Cell death-associated molecular-pattern molecules: inflammatory signaling and control. Mediators Inflamm. 2014; 2014: 821043. doi: 10.1155/2014/821043 PMID: 25140116
16. Silva E, Leitão S, Henriques S, Kowalewski MP, Hoffmann B, Ferreira-Dias G, et al. Gene transcription of TLR2, TLR4, LPS ligands and prostaglandin synthesis enzymes are up-regulated in canine uteri with cystic endometrial hyperplasia-pyometra complex. J Reprod Immunol. 2010; 84: 66–74. doi: 10.1016/j.jri.2009.10.004 PMID: 19945173
17. Karlsson I, Hagman R, Johannisson A, Wang L, Karlstam E, Wernersson S. Cytokines as immunological markers for systemic inflammation in dogs with pyometra. Reprod Domest Anim. 2012; 47 Suppl 6:337–341. doi: 10.1111/rda.12034 PMID: 23279533
18. Hara T, Tanegashima K. Pleiotropic functions of the CXC-type chemokine CXCL14 in mammals. 2012; 151: 469–476. doi: 10.1093/jb/mvs030 PMID: 22437940
19. Kida K, Baba E, Torii R, Kawate N, Hatoya S, Wijewardana V, et al. Lactoferrin expression in the canine uterus during the estrous cycle and with pyometra. Theriogenology 2006; 66: 1325–1333. PMID: 16730786
20. Drago-Serrano ME, de la Garza-Amaya M, Luna JS, Campos-Rodríguez R. Lactoferrin-lipopolysaccharide (LPS) binding as key to antibacterial and antiendotoxic effects. Int Immunopharmacol. 2012; 12: 1–9. doi: 10.1016/j.intimp.2011.11.002 PMID: 22101278
21. Hagman R. Clinical and molecular characteristics of pyometra in female dogs. Reprod Domest Anim. 2012; 47 Suppl 6: 323–325. doi: 10.1111/rda.12031 PMID: 23279529
22. Harvey MJ, Cauvin A, Dale M, Lindley S, Ballabio RJ. Effect and mechanisms of the anti-prolactin drug cabergoline on pseudopregnancy in the bitch. J Small Anim Pract. 1997; 38: 336–339. PMID: 9282338
23. Van Raemdonck K, Van den Steen PE, Liekens S, Van Damme J, Struyf S. CXCR3 ligands in disease and therapy. Cytokine Growth Factor Rev. 2015; 26: 311–321. doi: 10.1016/j.cytogfr.2014.11.009 PMID: 25498524
24. Floras AN, Holowaychuk MK, Bienzle D, Bersenas AM, Sharif S, Harvey T, et al. N-terminal pro-C-natriuretic peptide and cytokine kinetics in dogs with endotoxemia. J Vet Intern Med. 2014; 28: 1447–1453. doi: 10.1111/jvim.12409 PMID: 25056958
25. Schumann RR. Old and new findings on lipopolysaccharide-binding protein: a soluble pattern-recognition molecule. Biochem Soc Trans. 2011; 39: 989–993. doi: 10.1042/BST0390989 PMID: 21787335
26. Tsalkidou EA, Roilides E, Gardikis S, Trypsianis G, Kortsaris A, Chatzimichael A, et al. Lipopolysaccharide-binding protein: a potential marker of febrile urinary tract infection in childhood. Pediatr Nephrol. 2013; 28: 1091–1097. doi: 10.1007/s00467-013-2432-9 PMID: 23463341
27. Serrano M, Moreno-Navarrete JM, Puig J, Moreno M, Guerra E, Ortega F, et al. Serum lipopolysaccharide-binding protein as a marker of atherosclerosis. Atherosclerosis 2013; 230: 223–227. doi: 10.1016/j.atherosclerosis.2013.07.004 PMID: 24075748
28. Brănescu C, Şerban D, Şavlovschi C, Dascălu AM, Kraft A. Lipopolysaccharide binding protein (L.B.P.)—an inflammatory marker of prognosis in the acute appendicitis. J Med Life. 2012; 5: 342–347. PMID: 23125878