

**IDENTIFICAÇÃO DE MUTAÇÕES NOS  
GENES *APC* E *MUTYH* EM FAMÍLIAS COM  
POLIPOSE ADENOMATOSA INTESTINAL**

**GIOVANA TARDIN TORREZAN**

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

**Área de Concentração: Oncologia**

**Orientadora: Dra. Dirce Maria Carraro**

**Co-Orientador: Dr. Benedito Mauro Rossi**

**São Paulo**

**2014**

## FICHA CATALOGRÁFICA

Preparada pela Biblioteca da Fundação Antônio Prudente

Torrezan, Giovana Tardin

**Identificação de mutações nos genes *APC* e *MUTYH* em famílias com polipose adenomatosa intestinal** / Giovana Tardin Torrezan – São Paulo, 2014. 117p.

Tese (Doutorado)-Fundação Antônio Prudente.

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

Orientadora: Dirce Maria Carraro

Descritores: 1. POLIPOSE INTESTINAL 2. SÍNDROMES NEOPLÁSICAS HEREDITÁRIAS. 3. NEOPLASIAS COLORRETAIS. 4. GENES APC. 5. TÉCNICAS GENÉTICAS.

*"O jovem que quer ser cientista – e à ciência dedicar todo o seu tempo e amor – tem pelo menos três certezas: a de que morrerá um dia (como todo mundo), a de que não ficará rico (como quase todo mundo) e a de que se divertirá muito (como pouca gente)"*

*Newton Freire-Maia*

## AGRADECIMENTOS

À Dra. Dirce Maria Carraro, minha orientadora, obrigada pela confiança e pelo incentivo em todos os momentos, e também pela oportunidade de crescimento que tive ao participar de outros projetos do grupo.

Ao Dr. Benedito Mauro Rossi, meu co-orientador, obrigado pela dedicação e auxílio em todas as horas, que mesmo à distância sempre se fez tão presente.

À Dra. Erika Maria Monteiro Santos e à Dra. Ana Cristina Victorino Krepischi, pela participação fundamental neste trabalho.

À minha banca de acompanhamento, Dr. Fábio de Oliveira Ferreira e Dr. Raphael Parmigiani, obrigado pela disponibilidade, pelas sugestões e correções ao longo do desenvolvimento deste projeto.

A todas as famílias participantes e ao corpo clínico envolvidos nesta pesquisa, sobretudo à Dra. Maria Isabel Achatz e ao Dr. Samuel Aguiar Junior, sem os quais este trabalho não poderia ser realizado.

A todos do Laboratório de Genômica e Biologia Molecular, Adriana, Bruna, Daiana, Elisa, Felipe, Mabel, Márcia, Rafael, Roberto, Tatiana, Vanina, e às meninas do Banco de Tumores, Ana Paula, Eloisa e Louise, pelo companheirismo, convivência e amizade.

Em especial ao Felipe, pela sua contribuição neste trabalho e à Bruna, Elisa e Adriana pelo esforço conjunto em nosso “projeto B”.

A todos os alunos, funcionários e pesquisadores do CIPE por tudo que aprendi e pelos momentos de confraternização.

À equipe da Pós-Graduação e Biblioteca, Ana Maria, Luciane, Vanuza e Suely pelo apoio para a realização desta tese.

À CAPES e à FAPESP pelo apoio financeiro (Processo FAPESP: 2010/15326-0).

À Cristiane e sua família, por me fazerem sentir parte dela e tornarem minha adaptação em São Paulo bem mais fácil.

Ao meu irmão Hugo e às minhas amigas irmãs de coração, que passaram quatro anos tentando entender meu trabalho (e provavelmente passarão mais 40), agradeço pela força em todos os momentos que precisei e pela amizade sincera.

Aos meus pais, Irineu e Iracema Torrezan, pelo apoio e amor em todas as horas, e por serem sempre meus exemplos. Vocês são os maiores responsáveis pelas minhas conquistas.

Ao Márcio, pela compreensão, paciência e carinho, e por estar sempre ao meu lado, seja na ida para Houston, na vinda para São Paulo, ou em nosso cotidiano.

## RESUMO

Torrezan GT. **Identificação de mutações nos genes *APC* e *MUTYH* em famílias com polipose adenomatosa intestinal**. São Paulo; 2014. [Tese de Doutorado-Fundação Antônio Prudente].

Atualmente, pacientes com múltiplos adenomas colorretais são avaliados para mutações germinativas em dois genes, *APC* e *MUTYH*. Pacientes com mutações em *APC* apresentam Polipose Adenomatosa Familiar Clássica ou Atenuada (FAP/AFAP), enquanto que pacientes portadores de mutações bialélicas em *MUTYH* apresentam Polipose Associada ao *MUTYH* (MAP). O espectro das mutações em *APC* e *MUTYH*, assim como as correlações genótipo-fenótipo nestas síndromes, apresentam importante impacto clínico e podem ser distintas em cada população, tornando necessária a obtenção de dados genéticos e clínicos de diferentes populações. Além disso, cerca de 10-15% dos pacientes com polipose não apresentam mutações nesses genes, o que sugere a existência de outros genes de predisposição ainda desconhecidos. Assim, os objetivos deste estudo foram caracterizar mutações germinativas nos genes *APC* e *MUTYH* em pacientes Brasileiros com polipose, além de identificar novos genes associados com a síndrome através de sequenciamento de exoma dos pacientes negativos. No total, 23 pacientes não relacionados foram avaliados para mutações pontuais na região codificante dos genes *APC* e *MUTYH* através de sequenciamento capilar, e para rearranjos genômicos nos mesmos genes por meio de MLPA (*Multiplex Ligation-Dependent Probe Amplification*), arrays de hibridação genômica comparativa (CGH-array), e PCR duplex quantitativo. Este último método de avaliação do número de cópias genômicas foi desenvolvido e validado no presente estudo. Foram identificados 21 pacientes mutados nesta coorte (91%) - 6 pacientes apresentaram mutações patogênicas em *MUTYH*, 14 apresentaram mutações patogênicas em *APC* e um paciente foi portador de uma nova variante *missense* de significado clínico desconhecido em *APC* (p.Val1789Leu); seis mutações foram descritas pela primeira vez neste trabalho. Em um destes pacientes identificamos a primeira grande deleção

genômica descrita no gene *MUTYH*. Correlações genótipo-fenótipo dos dados clínicos dos probandos e seus familiares (totalizando 113 indivíduos afetados) revelaram resultados divergentes em comparação com os descritos em outros estudos, especialmente em relação à ocorrência de tumores desmóides em famílias com mutações antes códon 1444 (6/8 famílias com desmóides). Dois pacientes não apresentaram mutações em nenhum dos genes avaliados e foram investigados para mutações em outros genes através do sequenciamento completo do exoma na plataforma SOLiD 5500XL. Foi identificada em um desses pacientes uma alteração nova em um gene da família das polimerases, cuja principal função é o reparo de DNA, não presente em bancos de dados e nem em população controle Brasileira. Análises de segregação e experimentos funcionais serão realizados para verificar a associação deste e de outros candidatos selecionados com o fenótipo de polipose. Esta primeira investigação abrangente do espectro mutacional dos genes *APC* e *MUTYH* em pacientes com polipose brasileiros mostrou uma alta taxa de detecção e identificou novas mutações patogênicas. Notavelmente, um número significativo de famílias com mutação em *APC* não foram consistentes com as correlações genótipo-fenótipo previstos de outras populações.

## SUMMARY

Torrezan GT. [Mutational screening of *APC* and *MUTYH* genes in families with adenomatous polyposis]. São Paulo; 2014. [Tese de Doutorado-Fundação Antônio Prudente].

Patients with multiple colorectal adenomas are currently screened for germline mutations in two genes, *APC* and *MUTYH*. *APC*-mutated patients present classic or attenuated familial adenomatous polyposis (FAP/ AFAP), while patients carrying biallelic *MUTYH* mutations exhibit *MUTYH*-associated polyposis (MAP). The spectrum of *APC* and *MUTYH* mutations as well as the genotype-phenotype correlations in polyposis syndromes present clinical impact and can be population specific, making important to obtain genetic and clinical data from different populations. Furthermore, up to 10-15% of polyposis patients do not harbor mutations in these genes, suggesting that other yet unknown polyposis-predisposing genes could exist. Thus, the aim of this study was to characterize germline mutations in *APC* and *MUTYH* genes in Brazilian polyposis and to investigate novel susceptibility genes by exome sequencing of negative patients. At total, 23 unrelated polyposis patients were screened for *APC/MUTYH* point mutations through DNA capillary sequencing, and for *APC* and *MUTYH* genomic rearrangements through MLPA (multiplex ligation-dependent probe amplification), array-comparative genomic hybridization, and duplex quantitative PCR. This last gene dosage method was developed and validated in this study. We identified 21 mutated patients in this cohort (91%) – 6 patients carried *MUTYH* pathogenic mutations, 14 carried *APC* pathogenic mutations and one carried a novel *APC* missense variant of unknown clinical significance (p.Val1789Leu); six mutations were described for the first time in this series. One of these patients harbored the first large genomic deletion identified in *MUTYH* gene. Genotype-phenotype correlations of clinical data from probands and their affected relatives (113 individuals in total) revealed divergent results compared with those described in other studies, particularly regarding the occurrence of desmoid tumors in families with mutations before codon 1444 (6/8



families with desmoid) . Two patients were negative for mutations in the evaluated genes and were screened for mutations in other genes through exome sequencing at SOLiD 5500xl platform. One of these patients was found to harbor a novel alteration in a polymerase gene, which main function is DNA repair, and was not described in any database nor present in a Brazilian control population. The association of this and other selected candidates with the polyposis phenotype will be further investigated through segregation analysis and functional assays. This first comprehensive investigation of the *APC* and *MUTYH* mutation spectrum in Brazilian polyposis patients showed a high detection rate and identified novel pathogenic mutations. Notably, a significant number of *APC*-positive families were not consistent with the predicted genotype-phenotype correlations from other populations. Furthermore, we expect that exome sequencing can reveal novel genes associated with the polyposis phenotype.

## LISTA DE FIGURAS

<b>Figura 1</b>	Distribuição dos domínios proteicos e das mutações patogênicas mais frequentes de <i>APC</i> .....	8
<b>Figura 2</b>	Distribuição dos domínios protéicos e das mutações patogênicas mais frequentes de <i>MUTYH</i> .....	10
<b>Figura 3</b>	Delineamento experimental.....	18
<b>Figura 4</b>	Heredograma da família ID20.....	96
<b>Figura 5</b>	Heredograma da família ID22.....	96
<b>Figura 6</b>	Localização da variante de <i>POLQ</i> p.Lys2155Asn identificada no paciente ID22.....	101
<b>Figura 7</b>	Localização da variante de <i>APC</i> p.Val1789Leu identificada no paciente ID13.....	108

## LISTA DE TABELAS E QUADRO

<b>Tabela 1</b>	Cobertura do sequenciamento de exoma.....	97
<b>Tabela 2</b>	Número e tipo das variantes identificadas por sequenciamento de exoma.....	98
<b>Tabela 3</b>	Variantes indels identificadas nos pacientes ID20 e ID22 por sequenciamento de exoma.....	103
<b>Tabela 4</b>	Variantes <i>nonsense</i> e <i>splice site</i> identificadas nos pacientes ID20 e ID22 por sequenciamento de exoma.....	104
<b>Quadro 1</b>	Genes relacionados ao câncer colorretal hereditário segundo dados de literatura.....	99

## LISTA DE ABREVIACÕES

<b>8-oxoG</b>	8-dihydro-8-oxoguanina
<b>AFAP</b>	Polipose Adenomatosa Familiar Atenuada
<b>APC</b>	<i>Adenomatose Polipose Coli</i>
<b>BER</b>	Reparo por exisão de base
<b>CCR</b>	Câncer Colorretal
<b>CGH-array</b>	Arrays de hibridação genômica comparativa
<b>CHRPE</b>	Hipertrofia congénita do epitélio pigmentado da retina
<b>DHPLC</b>	Cromatografia líquida desnaturante de alta performance
<b>DNA</b>	Ácido desoxirribonucléico
<b>FAP</b>	Polipose Adenomatosa Familiar
<b>Indels</b>	Alteração de inserção ou deleção
<b>LOVD</b>	<i>Leiden Open Variation Database</i>
<b>MAP</b>	Polipose Associada ao <i>MUTYH</i>
<b>MCR</b>	Região de cluster de mutação
<b>MLPA</b>	Amplificação de múltiplas sondas dependente de ligação
<b><i>MUTYH</i></b>	<i>MutY Homolog</i>
<b>NGS</b>	Sequenciamento de nova geração
<b>PCR</b>	Reação em cadeia da polimerase
<b>PTT</b>	Teste da proteína truncada
<b>RCCH</b>	Registro de Câncer Colorretal Hereditário
<b>SNVs</b>	Variações de nucleotídeos únicos
<b>SEER</b>	The Surveillance, Epidemiology, and end Results

# ÍNDICE

<b>1</b>	<b>REVISÃO DE LITERATURA .....</b>	<b>1</b>
1.1	Aspectos clínicos e genéticos das síndromes de polipose.....	2
1.1.1	O Gene <i>APC</i> .....	5
1.1.2	O Gene <i>MUTYH</i> .....	9
1.2	Correlação genótipo-fenótipo .....	10
1.3	Técnicas Moleculares.....	13
<b>2</b>	<b>OBJETIVOS.....</b>	<b>15</b>
2.1	Objetivo principal .....	15
2.2	Objetivos específicos .....	15
<b>3</b>	<b>COMENTÁRIOS GERAIS.....</b>	<b>16</b>
<b>4</b>	<b>ARTIGO 1: TORREZAN ET AL. 2011 – BMC MEDICAL GENETICS 12:128 .....</b>	<b>20</b>
<b>5</b>	<b>ARTIGO 2: TORREZAN ET AL. 2012 BMC MEDICAL GENETICS 13:55 .....</b>	<b>39</b>
<b>6</b>	<b>ARTIGO 3: TORREZAN ET AL. 2013 - ORPHANET JOURNAL OF RARE DISEASES, 8:54.....</b>	<b>60</b>
<b>7</b>	<b>ARTIGO 4: IDENTIFICAÇÃO DE NOVOS GENES DE SUSCEPTIBILIDADE À POLIPOSE INTESTINAL POR SEQUENCIAMENTO DE EXOMA (EM PREPARAÇÃO).....</b>	<b>96</b>
<b>8</b>	<b>COMENTÁRIOS FINAIS .....</b>	<b>105</b>
<b>9</b>	<b>CONCLUSÕES .....</b>	<b>110</b>
<b>10</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>111</b>

## 1 REVISÃO DE LITERATURA

O Câncer Colorretal (CCR) representa um importante problema médico e de saúde pública, sendo o terceiro câncer mais comum e a segunda causa de morte por câncer nos países desenvolvidos. Nos Estado Unidos, foram estimados 142 mil novos casos e 50 mil mortes em 2013 decorrentes de CCR (The Surveillance, Epidemiology, and End Results-SEER 2013). No Brasil, o câncer de cólon e reto é o segundo mais frequente em mulheres e o terceiro em homens, excluindo os tumores de pele não melanoma, e o número de casos novos de câncer colorretal estimado para o ano de 2014 será de 15.440 casos em homens e de 17.240 em mulheres (Ministério da Saúde 2014).

Atualmente, estima-se que 20-30% de todos os tumores colorretais apresentem algum componente hereditário, incluindo-se aqueles devido a genes de baixa penetrância e com herança multifatorial (MARKOWITZ e BERTAGNOLLI 2009). Dentre os casos de câncer colorretal familiar em que se identificam mutações de alta penetrância em único gene (aproximadamente 5% a 10% de todos os casos de CCR), existem duas principais síndromes: a Polipose Adenomatosa Familiar (FAP) e a Síndrome de Lynch. O entendimento da base genética dessas doenças, mesmo não estando completo, melhorou as estratégias de rastreamento e prevenção, e poderá também melhorar as opções de tratamento. Além disso, estas síndromes representam modelos únicos para estudar a tumorigênese colorretal, e, portanto podem prover informações sobre a etiologia e acompanhamento de outros tumores hereditários ou esporádicos de cólon (STRATE e SYNGAL 2005).

## 1.1 ASPECTOS CLÍNICOS E GENÉTICOS DAS SÍNDROMES DE POLIPOSE

A Polipose Adenomatosa Familiar (FAP) é uma condição rara, com incidência em torno de 2 por milhão e prevalência de 40 por milhão, afetando igualmente homens e mulheres (BÜLOW 2003). Apenas uma pequena fração (<1%) de todos os CCR é devida a FAP, e esta porção aparentemente está diminuindo devido a diagnóstico precoce e medidas preventivas, como cirurgias profiláticas que removem o intestino antes da malignização dos pólipos (STRATE e SYNGAL 2005; ROZEN e MACRAE 2006).

FAP é uma doença hereditária marcada pelo surgimento de centenas a milhares de pólipos adenomatosos colorretais durante a segunda ou terceira década de vida. Na maioria dos casos, é herdada como uma doença autossômica de alta penetrância e grandes genealogias são descritas nas quais o fenótipo está presente em um parental afetado e em aproximadamente metade de seus filhos por muitas gerações. Entretanto, cerca de 30% dos pacientes com FAP não apresentam história familiar, e apresentam o fenótipo devido à existência de alterações genéticas *de novo* (VARESCO 2004).

Pacientes com FAP tem uma forte predisposição para início precoce de CCR, assim como outras malignidades (STRATE e SYNGAL 2005). A forma clássica de FAP é caracterizada pela presença de 100 a 5.000 pólipos adenomatosos no cólon e reto, que geralmente são diagnosticados entre os 20-30 anos de idade e inevitavelmente levam ao desenvolvimento de carcinoma colorretal aos 40-50 anos (VASEN et al. 2008). Os pólipos podem ser assintomáticos e indistinguíveis de

adenomas esporádicos, porém a incidência dos adenomas é muito maior em pacientes com FAP. O número de pólipos aumenta significativamente o risco de carcinogênese colorretal, comparado a um único adenoma encontrado nos casos esporádicos. O risco de desenvolvimento do câncer em pacientes FAP não tratados é de 100% (BEGGS e HODGSON 2008).

Em 1987, estudos de ligação com famílias de indivíduos afetados por FAP demonstraram que o gene responsável pela doença estava localizado no braço longo do cromossomo 5 (LEPPERT et al. 1987). Após esta descoberta, estudos de clonagem realizados em 1991 identificaram o gene *Adenomatose Polipose Coli* (*APC*) como responsável pela síndrome, e muitas mutações patogênicas em sua região codificadora foram encontradas em diversos pacientes com FAP (KINZLER et al. 1991; NISHISHO et al. 1991).

Além do câncer colorretal, várias manifestações extracolônicas também podem ser observadas em pacientes FAP, principalmente tumores desmóides, osteomas e pólipos na porção superior do trato gastrointestinal (KNUDSEN et al. 2003; NIELSEN et al. 2007) e, em geral, há uma correlação entre o local da mutação e a severidade das manifestações clínicas (ROZEN e MACRAE 2006).

Mutações germinativas no gene *APC* podem ser identificadas em 80-90% das famílias com FAP (POWELL et al. 1993). Entretanto, aproximadamente um terço dos casos de FAP apresenta uma mutação germinativa *de novo* em *APC*. A ausência de história familiar nesses casos pode atrasar o diagnóstico, que às vezes só irá ocorrer após o desenvolvimento do câncer (STRATE e SYNGAL 2005). Mutações no gene *APC* também têm papel no desenvolvimento do CCR esporádico e são encontradas em até 80% desses tumores (TSAO e SHIBATA 1994).



Existem algumas variantes da síndrome, como a Síndrome de Gardner, Turcot, e a Polipose Adenomatosa Atenuada (AFAP). Para todas estas síndromes a colectomia profilática representa uma excelente estratégia para prevenir o câncer colorretal, mas prevenir as demais malignidades extracolônicas continua um desafio (STRATE e SYNGAL 2005). Nos últimos anos tem se dado muita atenção para a forma atenuada da doença (AFAP) que está associada a um menor número de pólipos, evolução menos agressiva, início mais tardio e menor incidência de manifestações extracolônicas (KNUDSEN et al. 2003; NIELSEN et al. 2007). AFAP geralmente é causada por mutações nas extremidades 3', 5' e em uma porção do éxon 9 do gene *APC* que sofre *splicing* alternativo e está ausente em algumas isoformas da proteína.

Mutações em um segundo gene, denominado *MUTYH* (homólogo humano do MutY), ocasionam um fenótipo semelhante a AFAP (BEGGS e HODGSON 2008). A presença de mutações germinativas em *MUTYH* associadas ao câncer colorretal foi primeiramente descrita na “Família N” por AL-TASSAN et al. (2002). Nesta família, três de sete irmãos eram afetados por um fenótipo AFAP com carcinoma colorretal e pólipos, porém na ausência de mutações germinativas em *APC*. Estes pacientes, entretanto, apresentavam em seus tumores mutações somáticas em *APC* ocasionadas por transversões G:C para T:A, sugerindo que existia uma mutação germinativa ocasionando um defeito na via de reparo de mutações relacionadas a 8-oxo-guanina (AL-TASSAN et al. 2002). A identificação dessas mutações levou ao surgimento do conceito de polipose associada à *MUTYH* (MAP) (SIEBER et al. 2003).

A MAP difere do fenótipo clínico da polipose clássica, pois apesar de o risco de desenvolver CCR ser em torno de 100%, os pacientes tendem a desenvolver

polipose mais tardiamente (30-50 anos de idade) e desenvolvem um menor número de pólipos (SAMPSON et al. 2005). Além disso, diferentemente do que ocorre com mutações no gene *APC*, a herança por mutação em *MUTYH* é autossômica recessiva, ou seja, para que o fenótipo se manifeste duas cópias alteradas de *MUTYH* devem ser herdadas (mutações bialélicas) (BEGGS e HODGSON 2008).

Quanto menos agressivo é o fenótipo da polipose, mais difícil se torna o diagnóstico, visto que pacientes com poucos pólipos podem ser portadores de AFAP, MAP, Lynch ou outras poliposes menos comuns. Estudos recentes em pacientes com fenótipo de polipose atenuada (3 a 100 adenomas), nos quais não foram encontradas mutações germinativas em *APC*, demonstraram que aproximadamente 25% dos afetados eram portadores de mutações bialélicas em *MUTYH* (SIEBER et al. 2003; ISIDRO et al. 2004; VARESCO 2004).

### 1.1.1 O Gene APC

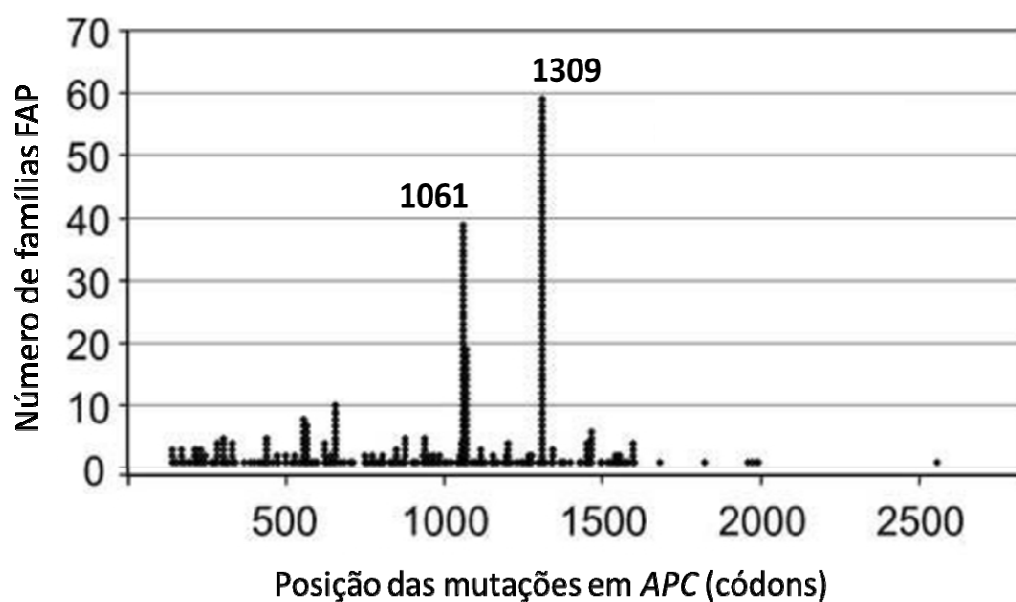
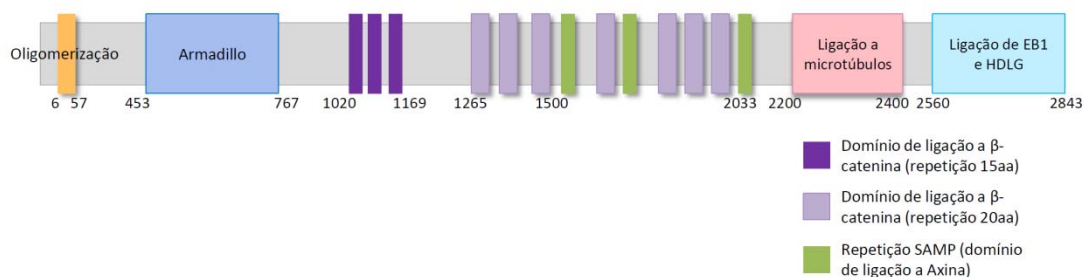
O gene *APC* é um supressor de tumor localizado no braço longo do cromossomo 5 (5q21). Mais de 1100 mutações germinativas diferentes já foram descritas em famílias com FAP e estão depositadas no banco de dados LOVD (*Leiden Open Variation Database*). Essas mutações são semelhantes às que ocorrem em adenomas esporádicos e a maioria (>90%) são mutações *nonsense* (mutações de uma única base gerando de códon de parada prematuro) e *frameshift* (pequenas inserções/deleções que ocasionam mudança de quadro de leitura e levam à formação de códons de parada prematuros) (STRATE e SYNGAL 2005). O resultado destas alterações é a ausência da proteína ou uma proteína APC truncada, resultando em um nível insuficiente de atividade da proteína nas células do epitélio colônico, levando a

um descontrole da proliferação celular (VARESCO 2004). Mutações em sítios de *splice* e de troca de aminoácido (*missense*) são encontradas em baixa frequência e podem apresentar uma significância desconhecida, representando um desafio para o diagnóstico e aconselhamento genético (HEINIMANN et al. 2001; AZZOPARDI et al. 2008). Nenhuma mutação *missense* em *APC* foi claramente associada com o fenótipo de FAP, mas algumas variantes, como I1307K e E1317Q, já foram descritas como possíveis alelos de suscetibilidade para formação de adenomas e CCR (VARESCO 2004).

O transcrito mais abundante de *APC* tem 8532pb e é dividido em 15 éxons, sendo que o éxon 15 corresponde a 77% da região codificadora. O gene *APC* codifica uma proteína multifuncional com domínios distintos (FODDE et al. 2001) que regula diversos processos celulares, incluindo transcrição, controle do ciclo celular, migração, diferenciação e apoptose. A proteína APC desempenha um papel importante na via de sinalização Wnt, visto que ela se liga e regula negativamente a  $\beta$ -catenina (MORIN et al. 1997). Em células indiferenciadas, a glicoproteína Wnt se liga ao receptor Frizzled, gerando um sinal de transdução que permite que a  $\beta$ -catenina seja translocada e se acumule no núcleo, onde ela irá agir como transativadora do fator de transcrição Tcf4. Os genes *MYC* e *CCND1* (Ciclina D1) foram identificados como os principais alvos dessa via e são responsáveis por estimular a proliferação celular (GOSS e GRODEN 2000). Em células que apresentam perda da função do gene *APC*, como algumas células do CCR, o acúmulo de  $\beta$ -catenina e a ativação da transcrição de seus alvos ocorre constitutivamente, mesmo na ausência de proteínas Wnt. A interação APC/ $\beta$ -catenina também

influencia as interações célula-célula por interagir com a E-caderina (KLAUS e BIRCHMEIER 2008; PINEDA et al. 2009).

Atualmente, métodos baseados em PCR permitem a detecção de mutações germinativas em *APC* em uma grande proporção de famílias com FAP. Em mais de 70-80% dos pacientes com FAP clássica pode ser identificada alguma mutação em *APC* que ative constitutivamente a via Wnt. Em pacientes AFAP, entretanto, a porcentagem de mutações identificadas em *APC* é bem menor, em torno de 25% (NIELSEN et al. 2007). Mutações patogênicas são encontradas em toda a região codificadora do gene, mas existem dois *hotspots* mutacionais (aminoácidos 1061 e 1309) que contam por aproximadamente 11% e 17% de todas as mutações germinativas em descritas *APC*, respectivamente (Figura 1) (PINEDA et al. 2009).



Fonte: Modificado de ARETZ et al. (2004) e BARROS (2012)

**Figura 1** - Distribuição dos domínios proteicos e das mutações patogênicas mais frequentes de *APC*. O painel superior apresenta uma representação esquemática dos diversos domínios da proteína APC. O gráfico inferior representa a posição das mutações mais frequentemente encontradas em famílias FAP, destacando-se as mutações nos códons 1061 e 1309 que representam *hotspots* no gene *APC*.

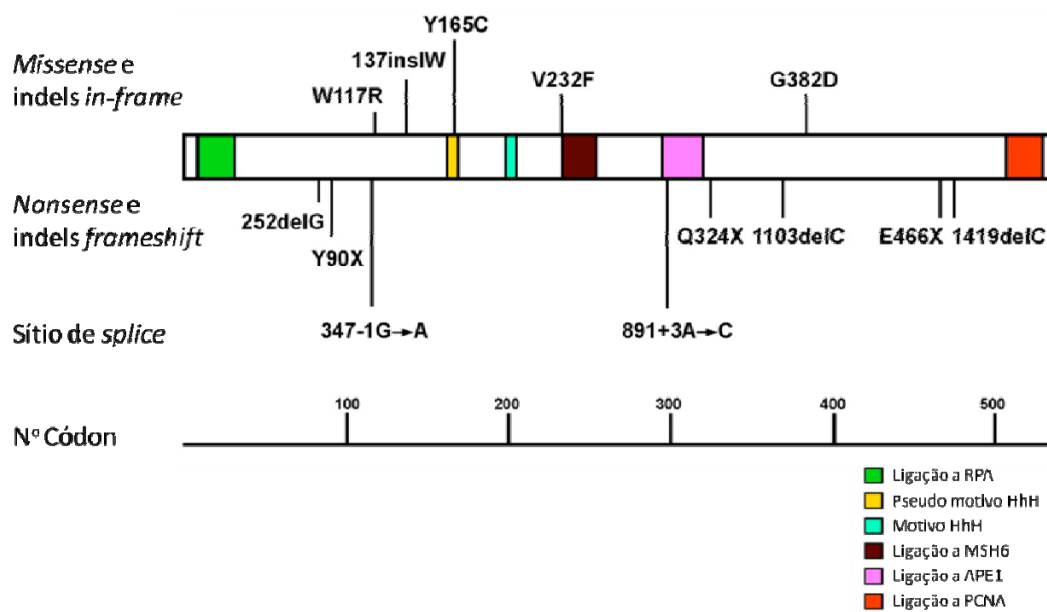
### 1.1.2 O Gene *MUTYH*

O gene *MUTYH* está localizado em 1p32-34, apresenta 16 éxons e codifica uma proteína de 535 aminoácidos que atua no reparo de DNA por excisão de base (SAMPSON et al. 2005). O produto do gene *MUTYH* é uma DNA glicosilase que juntamente com os produtos dos genes *OGG1* e *MTH1* realizam o reparo de excisão de bases danificadas do DNA (BER - Base Excision Repair). No processo normal de metabolismo aeróbico, as formas 8-dihydro-8-oxoguanina (8-oxoG) geradas por danos oxidativos ao DNA pareiam-se rapidamente com adenina. A proteína *MUTYH* age removendo da fita de DNA a base adenina mal pareada a 8-oxo-guanina. Este pareamento errôneo, caso não seja reparado, levará à formação de mutações por transversões de G:C para T:A. Por este motivo, tumores de pacientes MAP apresentam uma elevada taxa dessas transversões em genes comumente mutados em CCR, como *APC* e *KRAS* (AL-TASSAN et al. 2002; BEGGS e HODGSON 2008).

Foram identificados dois *hotspots* no gene *MUTYH*, p.Tyr179Cys e p.Gly396Asp (anteriormente p.Tyr165Cys e p.Gly382Asp) (Figura 2). Estas variantes são as mutações mais frequentes em caucasianos de origem do norte da Europa e afetam resíduos de aminoácidos altamente conservados, diminuindo a atividade enzimática da proteína (VARESCO 2004; CHEADLE e SAMPSON 2007). Além dessas variantes, aproximadamente 300 alterações foram descritas em *MUTYH*, sendo que cerca de 80 delas são comprovadamente mutações de susceptibilidade a adenomas e carcinomas colorretais (LOVD Database).

Mutações somáticas ou adquiridas em *MUTYH* aparentemente são incomuns, assim a expressão fenotípica da MAP requer duas mutações germinativas e apresenta um padrão de herança autossômico recessivo. Por este motivo, a maioria dos casos

de MAP é visto em pacientes esporádicos, famílias onde apenas os irmãos são afetados e famílias com relatos de casamentos consanguíneos (STRATE e SYNGAL 2005; VARESCO 2004).



Fonte: Modificado de CHEADLE e SAMPSON (2003).

**Figura 2** - Distribuição dos domínios proteicos e das mutações patogênicas mais frequentes de *MUTYH*. As mutações p.Tyr165Cys e p.Gly382Asp (atualmente p.Tyr179Cys e p.Gly396Asp) representam *hotspots* no gene *MUTYH*. APE1: endonuclease apurínica 1; HhH: *helix-hairpin-helix*; PCNA: antígeno nuclear de proliferação celular; RPA: proteína de replicação A.

## 1.2 CORRELAÇÃO GENÓTIPO-FENÓTIPO

Vários estudos descreveram uma associação entre o sítio da mutação em *APC* e o fenótipo dos pacientes com FAP. Características como o número de pólipos, a idade de acometimento e a ocorrência de manifestações extracolônicas parecem estar relacionadas com sítios mutacionais específicos. Estes resultados abrem novas perspectivas para a medicina personalizada, pois o conhecimento genético pode

direcionar a análise de mutações e provavelmente auxiliar na tomada de decisões terapêuticas (NIEUWENHUIS e VASEN 2007). Por exemplo, um paciente que apresente polipose extensa na infância pode ser rastreado diretamente para mutações no códon 1309 e, no caso de ser portador, a vigilância nos familiares deste paciente deveria ser iniciada na infância, devido ao início excepcionalmente precoce dos sintomas em pacientes com mutação neste sítio (GEBERT et al. 1999).

NIEUWENHUIS e VASEN em 2007 revisaram diversos estudos de correlação genótipo-fenótipo e buscaram categorizar os fenótipos de acordo com a severidade da polipose e o sítio de mutação em *APC*. O fenótipo atenuado (<100 pólipos) é restrito a mutações localizadas antes do códon 157, após o códon 1595 e na região do éxons 9 que sofre *splicing* alternativo. Um número abundante de pólipos (milhares) está associado com mutações dos códons 1250 a 1464. Um número intermediário de pólipos (100-1000) está relacionado com mutações no restante do gene. Em relação às manifestações extracolônicas, a maioria pode ocorrer associada a mutações em qualquer local do gene *APC*, exceto a ocorrência de CHRPE (hipertrofia congênita do epitélio pigmentado da retina), que parece estar restrita a mutações nos códons 311-1444, e tumores desmóides, que geralmente estão relacionados a mutações após o códon 1444. Tumores do trato gastrointestinal superior frequentemente estão associados com mutações após o códon 1395, mas também foram descritos para alterações em outras regiões do gene (NIEUWENHUIS e VASEN 2007).

Embora vários modelos para explicar a alteração da função de *APC* e o fenótipo correspondente tenham sido propostos, nenhum modelo único explica todos os dados disponíveis (SIEBER et al. 2000). O seguinte modelo é frequentemente



utilizado para explicar a variabilidade de fenótipos de FAP: os pacientes com FAP herdam uma mutação germinativa e de acordo com a hipótese dos dois eventos de Knudson, um segundo hit é necessário para que os adenomas se desenvolvam. O tipo de mutação germinativa parece determinar a natureza do segundo evento em *APC*. Pacientes com mutações próximas ao códon 1300 tendem a ter o segundo *hit* por perda alélica, enquanto pacientes com mutações distantes deste códon apresentam mutações que produzem proteínas truncadas. No caso da perda alélica, não resta nenhum produto de *APC* e os pacientes desenvolvem polipose severa. Entretanto, mutações nas porções 5' e 3' do gene podem causar uma alteração leve na funcionalidade da proteína e manter uma grande porção do gene intacta, resultando em fenótipos mais atenuados (LAMLUM et al. 1999). O *splicing* alternativo de mutações no éxon 9 podem moderar o efeito das mutações retirando a parte da informação genética que contém a mutação (YOUNG et al. 1998).

Deve-se considerar que o local da mutação não irá prever exatamente o produto de um gene e um fenótipo subsequente, e que embora correlações genótipo-fenótipo gerais tenham sido estabelecidas, inconsistências e contradições também foram reportadas e elas enfatizam as limitações dessas correlações. Provavelmente fatores genéticos adicionais e fatores ambientais apresentam papéis importantes na determinação das manifestações colônicas e extracolônicas (BURT et al. 2004). A utilização da correlação genótipo-fenótipo na prática clínica deve estar sempre combinada à utilização dos dados clínicos. Investigações futuras são necessárias, em particular em relação a tumores desmóides e adenomas no duodeno, visto que estes são comuns em pacientes com FAP e são responsáveis por grande parte da mortalidade desta síndrome (NIEUWENHUIS e VASEN 2007).

### 1.3 TESTES GENÉTICOS

Anteriormente, o diagnóstico de molecular de FAP era feito pelo rastreamento de mutações no gene *APC* utilizando o teste da proteína truncada (PTT), seguido de sequenciamento. Alternativamente, utilizava-se a metodologia de DHPLC (cromatografia líquida desnaturante de alta performance) seguida de sequenciamento dos *amplicons* que apresentassem um padrão aberrante. Entretanto, ambas as tecnologias apresentavam uma baixa sensibilidade, podendo ocasionar resultados falso-negativos (VAN DER LUIJT et al. 1997). Hoje em dia, o padrão ouro para avaliação de mutações nos genes *APC* e *MUTYH* é o sequenciamento de todos os éxons codificantes e limites íntron-éxon, que é capaz de identificar com alta sensibilidade alterações de ponto e pequenas inserções e deleções.

Além da realização do sequenciamento para a avaliação de alterações pontuais, também são necessárias técnicas que avaliem o número de cópias genômicas, para poder identificar grandes deleções ou amplificações de éxons únicos ou múltiplos. Para esta finalidade, as técnicas mais utilizadas na atualidade são o MLPA (amplificação de múltiplas sondas dependente de ligação) e o CGH-array (*arrays* de hibridação genômica comparativa). Com o uso conjunto destas metodologias, a taxa de detecção de mutações em pacientes com polipose chega a cerca de 70-90% (NIELSEN et al. 2007; GROVER et al. 2012).

Atualmente, com o advento do sequenciamento de nova geração (NGS), a velocidade de sequenciamento aumentou substancialmente a um custo inversamente proporcional. Estas plataformas tornaram factível a aplicação de metodologias de rastreamento de painéis de genes ou até mesmo do exoma completo (região

correspondente a todos os éxons do genoma). Para isso, o sequenciamento paralelo massivo é combinado com métodos de captura de ou amplificação de DNA para focar a análise em regiões específicas.

O sequenciamento paralelo de múltiplos genes nas plataformas de NGS tem sido cada vez mais utilizado em centros de diagnóstico molecular, visto que por um custo menor que do sequenciamento capilar é possível sequenciar todos os genes conhecidos como associados à determinada síndrome em um único experimento. Esta estratégia já foi utilizada com sucesso para o rastreamento de genes associados ao câncer colorretal hereditário (PRITCHARD et al. 2012). Entretanto, esta metodologia permite avaliar apenas aqueles genes previamente identificados por outros estudos.

Por outro lado, ao avaliar todos os genes conhecidos através da utilização do sequenciamento de exoma se torna possível identificar novos genes de susceptibilidade ao câncer. Recentemente, três estudos baseados em sequenciamento de exoma identificaram mais de 50 genes como possíveis novos genes de susceptibilidade ao câncer colorretal hereditário, sendo que dois genes (*POLD1* e *POLE*) foram considerados como fortes candidatos após extensas análises de validação (DERYCKE et al. 2013; GYLFE et al. 2013; PALLES et al. 2013). Assim, quanto mais genes forem identificados e um maior espectro de mutações for encontrado, mais eficiente será o rastreamento e aconselhamento genético dos pacientes.

## 2 OBJETIVOS

### 2.1 OBJETIVO PRINCIPAL

O objetivo geral deste trabalho é caracterizar mutações germinativas nos genes *APC* e *MUTYH* em pacientes com múltiplos pólipos adenomatosos intestinais e identificar outros genes possivelmente associados à síndrome.

### 2.2 OBJETIVOS ESPECÍFICOS

- Caracterizar mutações germinativas em *APC* e *MUTYH* em pacientes com polipose pelas técnicas de PCR seguido de sequenciamento direto, CGH-array e MLPA.
- Classificar a significância patogênica das mutações encontradas por análise dos resultados de mutação e conferência nos bancos de dados. Mutações de substituição de aminoácido não previamente descritas serão avaliadas quanto o impacto na função protéica através de programas de predição de patogenicidade.
- Correlacionar as características clínicas de pacientes portadores de diferentes mutações germinativas.
- Identificar novos genes possivelmente associados com a síndrome através de sequenciamento de exoma dos pacientes com polipose e negativos para mutações em *APC* e *MUTYH*.

### 3 COMENTÁRIOS INICIAIS

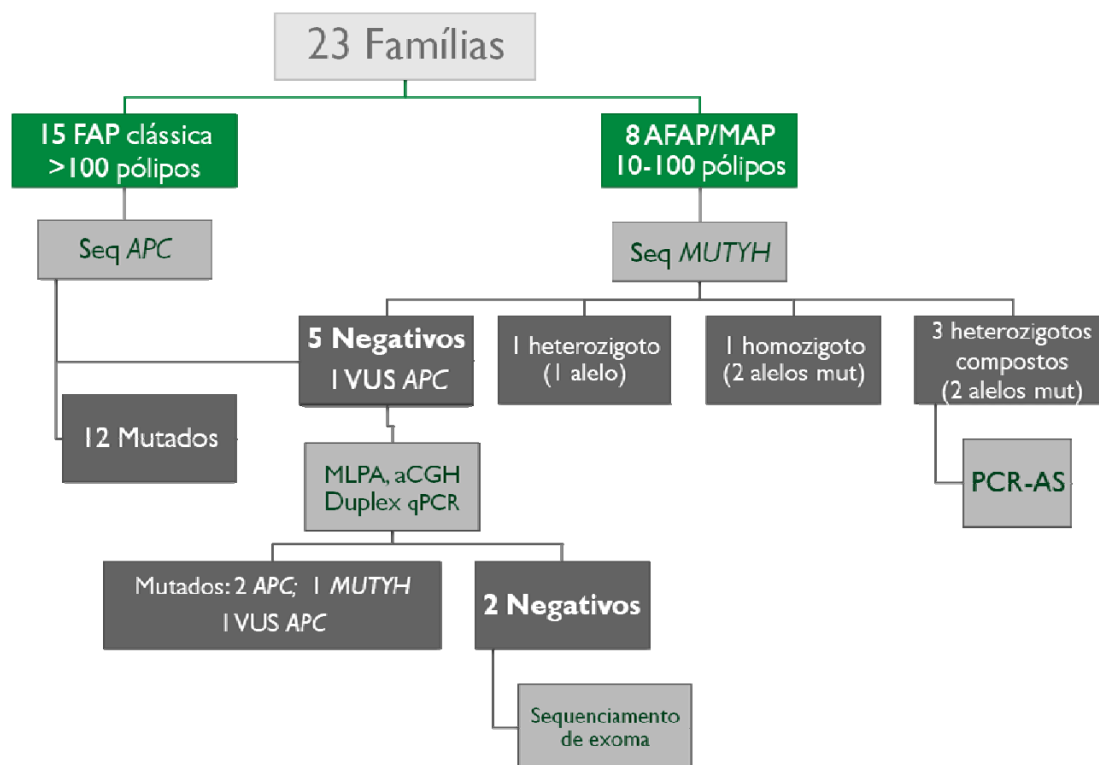
Neste primeiro item de comentários iniciais, é feita uma breve introdução sobre a casuística, o delineamento experimental e um resumo dos resultados finais deste trabalho. Os três capítulos seguintes da tese correspondem aos três manuscritos publicados a partir deste trabalho (TORREZAN et al. 2011, 2012 e 2013). Por fim, um último capítulo de resultados é dedicado às análises dos dados provenientes do sequenciamento de exoma, que ainda estão em andamento. Todas as metodologias utilizadas neste estudo estão apresentadas em detalhe no Anexo 1.

Para a realização deste estudo foram selecionados pacientes com critérios clínicos para FAP Clássica (mais de 100 pólipos adenomatosos) e AFAP/MAP (10-100 pólipos adenomatosos) a partir do banco de dados do Registro de Câncer Colorretal Hereditário do A.C. Camargo Cancer Center. Para o estudo de rastreamento de mutações, foram recrutados 23 probandos afetados não relacionados, 8 familiares afetados e 13 não afetados, totalizando 44 indivíduos.

Além dos dados clínicos dos 44 indivíduos recrutados, informações clínicas também foram obtidas para os demais familiares afetados de cada família através da pesquisa de prontuários ou durante a entrevista do probando com o aconselhador genético para montagem do heredograma, totalizando 113 pacientes afetados por polipose. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa do Centro de Tratamento e Pesquisa do A.C. Camargo Cancer Center, sob o nº 1169/08-B. As informações sobre os objetivos da pesquisa foram apresentadas aos pacientes para a assinatura do Consentimento Livre e Esclarecido (Anexo 2).

O rastreamento de mutações de ponto e rearranjos genômicos nos genes *APC* e *MUTYH* foi realizado inicialmente nos 23 pacientes portadores de polipose. Após a identificação da mutação presente na família, os demais familiares recrutados foram sequenciados pontualmente em busca das alterações já conhecidas.

Em resumo, foram identificadas mutações patogênicas em 20 dos 23 probandos (87%), sendo que dos três pacientes restantes, um apresentou uma nova variante *missense* no gene *APC* de significado clínico desconhecido, enquanto dois não apresentaram nenhum tipo de alteração e foram avaliados através de sequenciamento do exoma. O fluxograma abaixo representa o delineamento do estudo e resume os resultados encontrados com o emprego dos métodos utilizados.



**Figura 3** - Delineamento experimental. Foram selecionadas neste estudo 23 famílias portadoras de polipose, das quais 15 apresentavam uma polipose clássica (>100 pólipos) e foram sequenciadas inicialmente para o gene *APC*, enquanto oito apresentavam o fenótipo atenuado e foram primeiramente sequenciadas para o *MUTYH*. Os pacientes negativos para o primeiro gene analisado foram então sequenciados para o segundo gene. Após o sequenciamento, foram encontrados 12 pacientes com alterações patogênicas em *APC*, e cinco com alterações em *MUTYH*, sendo que destes três eram heterozigotos compostos para duas mutações distintas. Para estes pacientes, a técnica de PCR-alelo específica (PCR-AS) foi utilizada para confirmar a presença de alterações bialélicas. Cinco pacientes não apresentaram mutações pontuais em nenhum dos genes, e um paciente apresentou uma variante de significado clínico desconhecido (do Inglês VUS – *variant of unknown significance*). Para estes seis pacientes, as técnicas MLPA, CGH-array e duplex PCR quantitativa foram empregadas na busca de grandes rearranjos genômicos. Três alterações de número de cópias foram identificadas – duas em *APC* e uma em *MUTYH*. Por fim, dois pacientes foram negativos para mutações nesses genes e foram selecionados para o sequenciamento do exoma.

Os resultados obtidos encontram-se organizados em três artigos científicos provenientes do estudo. A primeira publicação resultante deste trabalho (TORREZAN et al. 2011) refere-se à identificação e caracterização do primeiro grande rearranjo genômico descrito para o gene *MUTYH*. Esta alteração é uma deleção que envolve os éxons 4 a 16 deste gene, e foi identificada em homozigose em um paciente através de PCR multiplex e CGH-array.

O segundo manuscrito originado deste projeto (TORREZAN et al. 2012) descreve as características fenotípicas de um paciente com uma deleção de 5,2Mb que engloba o gene *APC* inteiro e mais 19 genes. Esta alteração foi identificada por CGH-array e validada por um novo método de avaliação do número de cópias genômicas baseado em uma PCR duplex quantitativa (duplex qPCR), que foi desenvolvido neste trabalho e vem sendo utilizado frequentemente por nosso grupo para validar deleções e ampliações de éxons inteiros.

O último artigo publicado (TORREZAN et al 2013) refere-se à descrição detalhada dos dados genotípicos e fenotípicos de toda a casuística. A avaliação das correlações genótipo-fenótipo revelou algumas divergências entre a população deste estudo e outras populações já descritas na literatura.

Por fim, os resultados provenientes do sequenciamento de exoma de dois pacientes com polipose são apresentados no quarto capítulo.



**4 ARTIGO 1: TORREZAN ET AL. 2011 – BMC MEDICAL GENETICS 12:128**

**Breakpoint characterization of a novel large intragenic deletion of *MUTYH* detected in a MAP patient: Case report**

Giovana T Torrezan<sup>1</sup>, Felipe C C da Silva<sup>1</sup>, Ana C V Krepischi<sup>1,2</sup>, Érika M M Santos<sup>1</sup>, Fábio de O Ferreira<sup>3</sup>, Benedito M Rossi<sup>4</sup>, Dirce M Carraro<sup>1,2§</sup>

1- International Center of Research and Training (CIPE), A. C. Camargo Hospital, São Paulo, SP, Brazil.

2- National Institute of Science and Technology in Oncogenomics (INCITO), São Paulo, SP, Brazil.

3- Colorectal Tumors Department, A. C. Camargo Hospital, São Paulo, SP, Brazil.

4- Barretos Cancer Hospital - Pio XII Foundation, Barretos, São Paulo, Brazil.

§Corresponding author

Dirce Maria Carraro

Principal Investigator of Genomic and Molecular Biology Laboratory

CIPE - A. C. Camargo Hospital

Rua Taguá, 440. CEP: 01508-010 São Paulo/SP - BR

Phone number: 55 11 2189-2954; Fax number: 55 11 3272-0495

**Abstract**

**Background:** *MUTYH*-associated polyposis (MAP) is a recessive, hereditary, colorectal cancer-predisposing syndrome caused by biallelic mutations in the *MUTYH* gene. Most *MUTYH* pathogenic variants are missense mutations, and until recently no gross genomic deletions had been described. **Case Presentation:** We have identified a large deletion in the *MUTYH* gene: a >4.2 kb deletion encompassing exons 4-16. This is the second description of this rearrangement, which has been recently described as the first large deletion in this gene. The clinically suspected MAP patient was homozygous for this mutation and presented with no amplification products for 14 exons of *MUTYH* on initial screening. Deletion breakpoints were refined to base pair level through array comparative genomic hybridization (aCGH) analysis followed by sequencing. The identified breakpoints were located within intron 3 and 146 bp downstream of the 3' end of the gene, with the presence of an *Alu*Jr element adjacent to the distal breakpoint. The presence of a 2 bp insertion at the junction suggests the involvement of the non-homologous end joining (NHEJ) repair mechanism, possibly facilitated by rearrangement-promoting elements. Examination of the *MUTYH* locus revealed a high *Alu* density that may make this region prone to rearrangements. **Conclusion:** Large deletions are a possible mechanism for loss of function of the *MUTYH* gene, and investigation of such mutations may be important in identifying causative mutations in MAP patients.

## Background

*MUTYH*-associated polyposis (MAP) (MIM#608456), a recessive inherited syndrome characterized by colorectal adenomatous polyposis and a high risk of colorectal cancer, is a disorder caused by biallelic pathogenic germline variants in the *human mutY homologue (MUTYH)* gene [1]. *MUTYH* spans 11.2 kb on chromosome 1p34.1 [2] and encodes a DNA glycosylase that plays a key role in base excision repair (BER) of 8-oxoG: A mismatches by removing the mismatched adenine [3]. The oxidation product 8-oxoG is the most stable product of oxidative DNA damage [4], which can lead to G:C to T:A transversions if not repaired<sup>5</sup>. Tumorigenesis in MAP patients is thought to be initiated by somatic G:C →T:A transversions in *KRAS* and/or *APC* [5].

The clinical manifestations of MAP resemble familial adenomatous polyposis (FAP; MIM#175100), which is caused by autosomal dominantly inherited pathogenic variants in the *APC* gene. Most biallelic *MUTYH* mutation carriers have between ten and several hundred polyps, usually with later onset compared to FAP patients [1,6]. Also, a number of MAP patients with CRC and no polyps have been reported [5]. Mutations in *MUTYH* account for approximately 40% of patients with 10-100 colorectal adenomas (attenuated FAP patients) and positive familial history, a proportion slightly higher than that of *APC* mutations in these patients (30%) [7].

Nearly 300 different sequence variants have been identified so far in this gene (*MUTYH* Leiden Open Variation Database; [http://chromium.liacs.nl/LOVD2/colon\\_cancer/home.php?select\\_db=MUTYH](http://chromium.liacs.nl/LOVD2/colon_cancer/home.php?select_db=MUTYH)). In contrast to pathogenic *APC* variants, which mostly result in a truncated or absent protein, most pathogenic *MUTYH* variants are missense variants and only a minority

consists of splice site and truncating variants [2]. The two hotspot mutations p.Tyr179Cys (exon 7) and p.Gly396Asp (exon 13) are prevalent in populations of European origin, probably due to a founder effect, and account for approximately 80% of all reported mutant alleles [5,8]. In the Brazilian population, a recently published paper on patients with clinical phenotypes of MAP, FAP and Lynch syndrome identified 4/60 (6.6%) patients with these two hotspot mutations in a biallelic state [9]. Until recently, no gross genomic deletions or duplications had been described in this gene in any population.

Here, we report a case study in which we have characterized a large *MUTYH* deletion in a MAP patient. During the composition of this manuscript, an independent study performed by a French group identified this same rearrangement in one of their polyposis patients [10]. We have refined the breakpoint of this >4.2 kb deletion to the base pair level. Based on the analysis of the sequences at breakpoints, we suggest a possible mechanism of origin for this alteration. Presence of this deletion was also analyzed in familial colorectal patients and a control group.

### **Case Presentation**

A Brazilian female patient (FAP15) from the Hereditary Colorectal Cancer Registry of Hospital AC Camargo (São Paulo, Brazil), who was clinically suspected for MAP, was screened for mutations in the *MUTYH* gene by direct sequencing. The age at onset of the attenuated polyposis (approximately 40 polyps) was 42 years and of the rectal cancer was 44 years. No extracolonic manifestations were observed in this patient. Family history was accessed through personal report. The patient stated that she had unaffected deceased parents, four unscreened siblings (without

colonoscopy) and one affected sister from whom biological material was unavailable due to residence distance (figure 1). The affected sister presented with attenuated polyposis at the age of 44 years. The proband had a Caucasian ancestry, since her great-grandparents were Portuguese.

In addition, the presence of this *MUTYH* deletion was screened through PCR in other 183 Brazilian individuals: three *APC/MUTYH*-mutation negative and 18 *APC/MUTYH*-mutation positive polyposis patients (12 *APC* mutation carriers, 5 *MUTYH* biallelic and one *MUTYH* monoallelic mutation carriers); 51 clinically suspected Lynch syndrome patients (fulfilling Amsterdam II or Bethesda criteria), who were non-carriers of germline mutations in the mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6* and *PMS2*; and 111 healthy controls. Written informed consent was obtained from all patients and controls. This study was approved by the ethics committee of A.C. Camargo Hospital (approval number: 1169/08-B).

### **Identification and characterization of the deletion**

We detected a homozygous deletion encompassing several exons of the *MUTYH* gene in this MAP patient. Genomic DNA was obtained from blood using the Puregene Genomic DNA Isolation Kit (Gentra Systems, USA) according to manufacturer's instructions. Eleven primer pairs were designed to amplify all 16 exons of *MUTYH* gene (NM\_001128425.1), including intron-exon boundaries. Primers used for these analyses and PCR conditions are available upon request. Unexpected results were observed when performing standard PCR for *MUTYH* mutation screening on this patient; exons 3-16 could not be amplified (data not shown). As a positive control for the reaction, multiplex PCR in the presence of primers for control genes (*GAPDH* intron 7; *HPRT1* exon 3) was performed for each

*MUTYH* exon and confirmed the absence of genomic template for this gene. The results of multiplex PCR for the patient and a control DNA are shown in figure 2A.

With the purpose of elucidating the extent of the deletion, comparative genomic hybridization based on microarrays (aCGH) was performed. Briefly, samples were labeled with Cy3- and Cy5-dCTPs by random priming. Experiments were performed in duplicate using a 180K whole-genome platform (OGT Technologies). Purification, hybridization, and washing were performed as recommended by the manufacturer. Data extraction was conducted using the Feature Extraction software (Agilent Technologies). We applied the Genomic Workbench software (Agilent Technologies) for identifying the constitutive genomic imbalances using the statistical algorithm ADM-2, with a sensitivity threshold of 6.7, a threshold  $\log_2$  ratio of 0.4 or 1.1 for duplication or high copy number gain, and -0.4 and -1.1 for deletion and homozygous loss, respectively.

Data analysis detected two sequences mapped within the *MUTYH* gene that exhibited  $\log_2$  ratio values compatible with a homozygous deletion (oligoprobes A\_14\_P107667, mapped at chr1:45794902-45794961, and A\_14\_P122926, at chr1:45797594-45797640; coordinates given according to UCSC Feb. 2009 (GRCh37/hg19) assembly) (figure 2B). Oligoprobes flanking the deletion at the 5' end of *MUTYH* gene sequence and downstream of the 3' end had  $\log_2$  ratios values consistent with two copies (A\_14\_P122926 - chr1:45804348-45804405, and A\_14\_P118536 - chr1:45793780-45793830) (figure 2B). Using an exon 2 forward primer (5'-GGTGGATGAGAGGGAGATAG-3') and reverse primers designed on the position of the A\_14\_P118536 undeleted probe and 600 bp upstream (primers 16R2: 5'-CTTCAGGTCTTCACCAAGTCC-3' and 16R1: 5'-

CTTCTCCTGTGCTTCCTCTC-3', respectively), we successfully amplified the junction fragments for this deletion. After confirming that the primer 16R1 was located at an undeleted region and as it was closest to the breakpoint, further analyses were performed only with this primer. The PCR reaction of the control DNA pool failed to amplify any detectable PCR product (theoretically 5889 bp in length), whereas PCR products of  $\approx 1600$  bp were detected using genomic DNA from the patient (figure 2E). Briefly, the PCR conditions consisted of 25 ng of genomic DNA, 0.3  $\mu$ mol of each primer and 1X Platinum® PCR SuperMix High Fidelity (Invitrogen) in a 20  $\mu$ L reaction. Cycling conditions entailed a preincubation at 95°C for 2 minutes followed by 35 cycles of denaturation at 95°C for 20 seconds, annealing at 62°C and extension at 72°C for 2 minutes. Direct sequencing of the junction fragments identified the exact breakpoints (genomic position: chr1:g.45,794,768\_45,799,052del4285insTA, cDNA position: c.348+33\_\*64+146del4285insTA), the precise size of the deletion interval (4,285 kb) and the affected exons (4-16) (figure 2C and D). Genomic positions were confirmed using MUTALYZER 2.0 software. Previously, exon 3 appeared to be deleted in the PCR analysis because the breakpoint is located upstream of the reverse primer site. No region of the adjacent *HPDL* gene is deleted downstream of the 3' end of *MUTYH*.

As the patient investigated in this study presented a homozygous mutation and reported no inbreeding in her family, we examined the presence of this deletion among 111 healthy controls, 21 polyposis patients and 51 patients with familial CRC negative for mutations on MMR genes. The junction fragment could not be amplified from the DNA of any other patient or control (data not shown). One possible

explanation for the homozygous occurrence of the deletion in this patient, without consanguineous parents, is the inheritance of two alleles of a founder mutation in the population. The proband's great-grandparents were Portuguese, therefore she had Caucasian ancestry. In this sense, very recently this same deletion was found in a European MAP patient in compound heterozygosity with the common *MUTYH* Caucasian mutation p.Gly396Asp [10]. The description of the same rearrangement in patients from different geographic regions is a very exciting discovery and additional studies to uncover the origin and frequency of this mutation would be of great value. Another potential explanation for the homozygosity is the occurrence of isodisomy, a type of uniparental disomy that results in two identical segments from one parental homologue and can occur due to a recombination event in the zygote, , though the existence of an affected sibling does not support this hypothesis.

Analysis of the junction sequence revealed several important features, including an insertion of 2 bp of non-template DNA (figure 2F), several recombination motifs such as GCG and GAS [11] and the presence of an *AluJr* element located 2 bases upstream of the distal breakpoint (figure 2G). No repetitive element is present at the upstream breakpoint. Examination of the *MUTYH* locus using the Repeat-Masker software (<http://www.repeatmasker.org/>) revealed an *Alu* density of 22.2% for an 11,800 kb region (chr1:45,794,343-45,806,142), which includes the entire gene (11,229 bp), and 571 bp downstream, which encompasses the distal breakpoint characterized in this study (figure 3).

This study is particularly important from both a clinical and research perspective because we have identified the first large intragenic deletion in *MUTYH*: a 4.285 kb homozygous deletion in a MAP patient



(c.[348+33\_\*64+146del4285insTA];[348+33\_\*64+146del4285insTA])). This mutation was submitted to the *MUTYH* LOVD Mutation Database and has been publicly available since May 12th: ([http://chromium.liacs.nl/LOVD2/colon\\_cancer/variants.php?select\\_db=MUTYH&action=view&view=1012041%2C0002450%2C10](http://chromium.liacs.nl/LOVD2/colon_cancer/variants.php?select_db=MUTYH&action=view&view=1012041%2C0002450%2C10)). The patient presented a homozygous deletion encompassing exons 4 to 16 detected and mapped through PCR, aCGH and sequencing. Breakpoints were sequenced and analysis of the junction sequence allowed us to postulate that classical nonhomologous end joining (NHEJ) is the most likely mechanism responsible for the deletion, as a 2-bp non-template insertion sequence is observed at the breakpoint junction [12,13]. This repair process involves the double strand breakage of DNA followed by end joining in the absence of extensive sequence homology, and is associated with small insertions at the junction sites [14] and with very short stretches of sequence identity (a few bp) between the two ends of the breakpoint junctions [15]. Analysis of breakpoints surrounding sequences revealed the presence of an *Alu*Jr adjacent to the distal breakpoint and the existence of recombination/deletion promoting motifs [11] such as GAG and GCS trinucleotides. These motifs are known DNA polymerase pause-site core elements [16] and might have facilitated deletion.

*In silico* analysis of the *Alu* content in the *MUTYH* gene region revealed a high ratio of *Alu* sequences per kb, with *Alu* elements occurring once in every 1.8 kb (22.2%) when compared with the average of the human genome where *Alu* elements account for about 10% and occur once every 3 kb [17]. A recent study of more than 20 genes found evidence that a high content of transposable elements causes increased frequency of gene disruption by gross deletions in human disease [18].

Additionally, the local enrichment of *Alus* has been observed in regions with recurrent *Alu*-mediated rearrangements, e.g., the *VHL* locus in von Hippel-Lindau disease patients [19], and in genes implicated in colorectal cancer-predisposing disorders such as Lynch and Peutz–Jeghers syndromes (*EPCAM-MSH2* [20] and *STK11* [21], respectively). However, not every *Alu*-rich gene is prone to this type of deletion formation. Examples include the thymidine kinase or  $\beta$ -tubulin genes [20,22]. In the *MUTYH* gene, a recent study has shown that a polymorphic *AluY* insertion in intron 15 when in a homozygous state, leads to a 2.15 fold increase in the levels of 8-oxoG in leukocyte DNA and is associated with type 2 diabetes mellitus [23]; however, this insertion has not yet been studied in the context of colorectal cancer.

## Conclusions

Currently, there is no information on the percentage of defects in the *MUTYH* gene caused by deletions/duplications of complete exons, and screening for this type of mutation is not a common practice as it is with the *APC* gene. Our finding highlights the importance of performing screening methods, such as qPCR, MLPA or long-range PCR, for large *MUTYH* deletions in polyposis patients. Furthermore, it remains to be addressed whether the high density of *Alu* elements in this gene can cause an elevated frequency of *Alu*-mediated deletions and rearrangements. Nevertheless, the evidence shown here suggests that deletions could represent an important mechanism of loss of function in the *MUTYH* gene.

## Figures and Legends

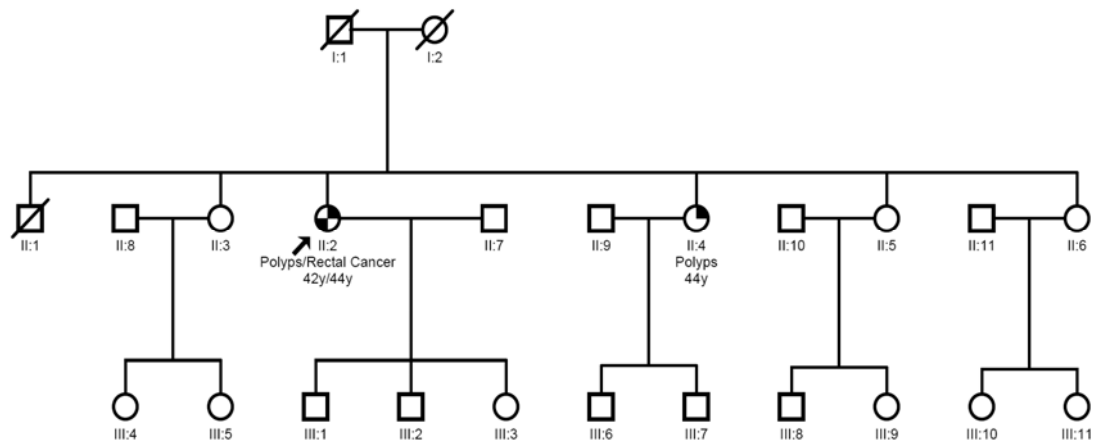


Figure 1: Family tree of patient FAP15. The proband is indicated by the arrow. The proband's age at onset of the polyposis (42 years) and of the rectal cancer (44 years) are indicated. The patient presented unaffected deceased parents, four unscreened (without colonoscopy) siblings and one affected sister. The affected sister presented with polyposis at the age of 44 years. The Caucasian patient stated having Portuguese great-grandparents and no report of inbreeding in her family.

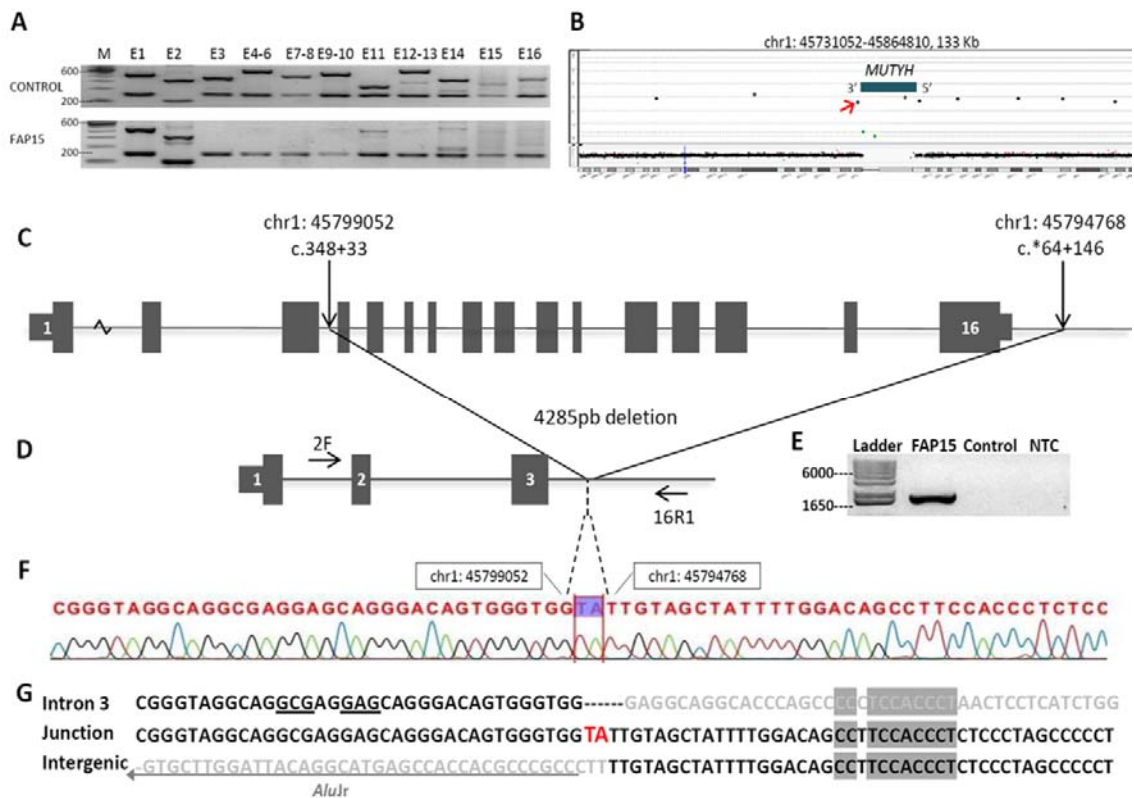


Figure 2: Homozygous deletion and characterization of the breakpoint in patient FAP15. **A:** Agarose gel of multiplex PCR performed for *MUTYH* exons 1 to 16 (E1-E16) and control genes *GAPDH* or *HPRT1*. The patient (FAP15) exhibited absence of amplification for *MUTYH* exons 3 to 16, whereas a healthy individual (Control) presented all expected fragments. **B:** The blue line on the profile of chromosome 1 (bottom panel) indicates the mapping at 1p34.1 of the two intragenic probes with  $\log_2$  ratio values compatible with a homozygous deletion (green circles, upper panel); probes flanking this deletion (at 5' of *MUTYH*, and downstream of its 3' end) exhibited normal values. The red arrow indicates the position of probe A\_14\_P118536 where novel reverse primers were designed. **C:** Schematic drawing of the normal gene with genomic and cDNA (NM\_001128425.1) positions of the breakpoints. **D:** Schematic drawing of the deleted gene and primer locations (2F

primer: chr1: 45800257-45800276; 16R1 primer: chr1: 45794388-45794407). **E:** Agarose gel showing the amplification of the deletion junction in the patient and amplicon absence in the control DNA pool. **F:** Partial chromatogram of the junction sequencing showing in highlight the two inserted nucleotides (TA). **G:** Nucleotide sequence around the deletion breakpoints and the deletion junction. Red larger nucleotides are the non-template insertion (filler DNA). Light gray nucleotides indicate the deleted sequence. Shaded nucleotides represent shared sequences between the breakpoints. Rearrangement-promoting elements GAG/GCS\* are shown underlined. The gray arrow beneath the intergenic deleted sequence shows the location of the *AluJr* element. \*S denotes the ambiguity code symbol for G/C. Coordinates at chromosome 1 are given according UCSC Feb. 2009 (GRCh37/hg19) assembly.

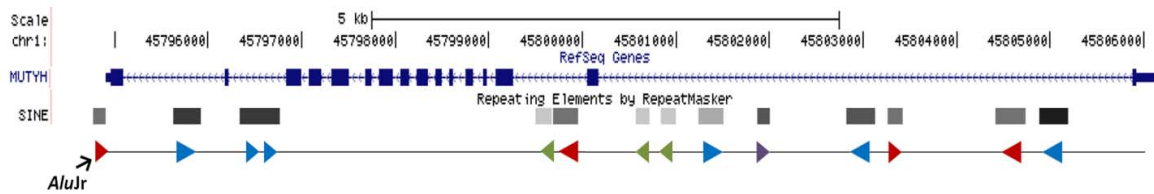


Figure 3: Short interspersed nucleotide element (SINE) content of the *MUTYH* genomic region (chr1:45794343-45806142) (data from UCSC database and RepeatMasker). A region of 11.8 kb is shown, extending 571 bp downstream of the *MUTYH* gene. Vertical bars in the genomic sequence indicate exons. The direction of the arrowheads denotes the orientation of the SINE repeat elements and each family is represented by a different color (Red: *AluJ*, Blue: *AluS*, Green: MIR, Purple: FLAM\_C). The black arrow indicates the *AluJr* element involved in the distal breakpoint; no repetitive element is present on the proximal breakpoint.

**Consent**

Written informed consent was obtained from all patients and controls for publication of this case report and any accompanying images. Copies of the written consents are available for review by the Editor-in-Chief of this journal.

**Competing Interests**

The authors state that they have no competing financial interests.

**Authors' Contributions**

GTT designed and performed the molecular genetic studies and drafted the manuscript. FCCS participated in the experimental design and revised the manuscript critically. ACVK carried out the aCGH analysis and revised the manuscript critically. EMMS participated in the clinical data acquisition and analysis. FOF conceived the study, and participated in its coordination. BMR conceived of the study, and participated in its design. DMC participated in its experimental design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

**Acknowledgments and Funding**

The authors acknowledge the patients for their collaboration, the Nucleic Acid Bank of A. C. Camargo Hospital for sample processing and Bianca C. G. Lisboa for guidance in establishing the sequencing protocols. This work was supported by National Institute of Science and Technology in Oncogenomics (INCITO) – FAPESP: 2008/57887-9.

## References

- 1 Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, Cheadle JP: **Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors.** *Nat Genet* 2002, **30(2)**:227-32.
- 2 Out AA, Tops CMJ, Nielsen ÅM, Weiss MM, Minderhout IJH MV, Fokkema IFAC, Buisine M-pierre, Claes K, Colas C, Fodde R, Fostira F, Franken PF, Gaustadnes M, Heinimann K, Hodgson SV, Hogervorst FBL, Holinski-feder E, Lagerstedt-robinson K, Olschwang S, Ouweland AMWVD, Redeker EJW, Scott RJ, Vankeirsbilck B, Grønlund RV, Wijnen JT, Wikman FP, Aretz S, Sampson JR, Devilee P, Dunnen JTD, Hes FJ: **Leiden Open Variation Database of the MUTYH Gene Human Mutation.** *Hum Mutat* 2010, **31(11)**:1205-1215.
- 3 Nghiem Y, Cabrera M, Cupples CG, Miller JH: **The mutY gene: a mutator locus in Escherichia coli that generates G.C---T.A transversions.** *PNAS* 1988, **85(8)**:2709-13.
- 4 Sampson JR, Jones S, Dolwani S, Cheadle JP: **MutYH ( MYH ) and colorectal cancer.** *Biochem Soc Trans* 2005, **33(4)**:679-683.
- 5 Nielsen M, Morreau H, Vasen HFA, Hes FJ: **MUTYH-associated polyposis (MAP)** *Crit Rev Oncol Hematol* 2010, 1-16.
- 6 Sampson JR, Dolwani S, Jones S, Eccles D, Ellis A, Evans DG, Frayling I, Jordan S, Maher ER, Mak T, Maynard J, Pigatto F, Shaw J, Cheadle JP: **Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH.** *Lancet* 2003, **362(9377)**:39-41.
- 7 Filipe B, Baltazar C, Albuquerque C, Fragoso S, Lage P, Vitoriano I, Mão de Ferro S, Claro I, Rodrigues P, Fidalgo P, Chaves P, Cravo M, Nobre Leitão C: **APC or MUTYH mutations account for the majority of clinically well-characterized families with FAP and AFAP phenotype and patients with more than 30 adenomas.** *Clin Genet* 2009, **76**:242-255.

- 8 Aretz S, Hes FJ: **Clinical utility gene card for : MUTYH-associated polyposis ( MAP ), autosomal recessive colorectal adenomatous polyposis.** *Eur J Hum Genet* 2010, 2-4.
- 9 Pitroski CE, Cossio SL, Koehler-Santos P, Graudenz M, Prolla JC, Ashton-prolla P: **Frequency of the common germline MUTYH mutations p.G396D and p.Y179C in patients diagnosed with colorectal cancer in Southern Brazil.** *Int J Colorectal Dis* 2011, 3:1-16.
- 10 Rouleau E, Zattara H, Lefol C, Noguchi T, Briaux A, Buecher B, Bourdon V, Sobol H, Lidereau R, Olschwang S: **First large rearrangement in the MUTYH gene and attenuated familial adenomatous polyposis syndrome.** *Clin Genet* 2011,80:301-303.
- 11 Abeysinghe SS, Chuzhanova N, Krawczak M, Ball EV, Cooper DN: **Translocation and gross deletion breakpoints in human inherited disease and cancer I: Nucleotide composition and recombination-associated motifs.** *Hum Mutat* 2003, 22(3):229-44.
- 12 Korbel JO, Urban AE, Affourtit JP, Godwin B, Grubert F, Simons JF, Kim PM, Palejev D, Carriero NJ, Du L, Taillon BE, Chen Z, Tanzer A, Saunders ACE, Chi J, Yang F, Carter NP, Hurles ME, Weissman SM, Harkins TT, Gerstein MB, Egholm M, Snyder M: **Paired-End Mapping Reveals Extensive Structural Variation in the Human Genome.** *Science* 2007, 420(318):420-426.
- 13 Perry GH, Ben-dor A, Tsalenko A, Sampas N, Rodriguez-revenga L, Tran CW, Scheffer A, Steinfeld I, Tsang P, Yamada NA, Park HS, Kim J-il, Seo J-sun, Yakhini Z, Laderman S, Bruhn L, Lee C: **The Fine-Scale and Complex Architecture of Human Copy-Number Variation.** *Am J Hum Genet* 2008, 82:685-695.
- 14 Linardopoulou EV, Williams EM, Fan Y, Friedman C, Young JM, Trask BJ: **Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication.** *Nature* 2005, 437:94-100.
- 15 Chan CY, Kiechle M, Manivasakam P, Schiestl RH: **Ionizing radiation and restriction enzymes induce microhomology-mediated illegitimate**



- recombination in *Saccharomyces cerevisiae*. *Nuc Ac Res* 2007, **35(15):5051-9.****
- 16 Weaver DT, DePamphilis ML: **Specific Sequences in Native DNA that Arrest Synthesis by DNA Polymerase  $\alpha$ .** *Biol Chem* 1982, **257(4):2075-2086.**
- 17 Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford a, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan a, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian a, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson a, Deadman R, Deloukas P, Dunham a, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt a, Jones M, Lloyd C, McMurray a, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall a, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra M a, Mardis ER, Fulton L a, Chinwalla a T, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty a, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen a, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs R a, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama a, Hattori M, Yada T, Toyoda a, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal a, Platzer M, Nyakatura G, Taudien S, Rump a, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan a, Qin S, Davis RW, Federspiel N a, Abola a P, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans G a, Athanasiou M, Schultz R, Roe B a, Chen F, Pan H, Ramser J, Lehrach H,

- Reinhardt R, McCombie WR, Bastide M de la, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey J a, Bateman a, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones T a, Kasif S, Kasprzyk a, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght a, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit a F, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams a, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld a, Wetterstrand K a, Patrinos a, Morgan MJ, Jong P de, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409(6822):**860-921.
- 18 Zelm MCV, Geertsema C, Nieuwenhuis N, Ridder DD, Conley ME, Schiff C, Tezcan I, Bernatowska E, Hartwig NG, Sanders EAM, Litzman J, Kondratenko I, Dongen JJMV, Burg MVD: **Gross Deletions Involving IGHM , BTK , or Artemis : A Model for Genomic Lesions Mediated by Transposable Elements.** *J Hum Genet* 2008, **(82):**320-332.
- 19 Franke G, Bausch B, Hoffmann MM, Cybulla M, Wilhelm C, Kohlhase J, Scherer G, Neumann HPH: **Alu-Alu recombination underlies the vast majority of large VHL germline deletions: Molecular characterization and genotype-phenotype correlations in VHL patients.** *Hum Mutat* 2009, **30(5):**776-86.
- 20 Kuiper RP, Vissers LELM, Venkatachalam R, Bodmer D, Hoenselaar E, Goossens M, Haufe A, Kamping E, Niessen RC, Hogervorst FBL, Gille JJP, Redeker B, Tops CMJ, Gijn ME van, Ouweland AMW van den, Rahner N, Steinke V, Kahl P, Holinski-Feder E, Morak M, Kloor M, Stemmler S, Betz B, Hutter P, Bunyan DJ, Syngal S, Culver JO, Graham T, Chan TL, Nagtegaal ID, Krieken JHJM van, Schackert HK, Hoogerbrugge N, Kessel

- AG van, Ligtenberg MJL: **Recurrence and variability of germline EPCAM deletions in Lynch syndrome.** *Hum Mutat* 2011, **32(4)**:407-14.
- 21 Resta N, Giorda R, Bagnulo R, Beri S, Della Mina E, Stella A, Piglionica M, Susca FC, Guanti G, Zuffardi O, Ciccone R: **Breakpoint determination of 15 large deletions in Peutz-Jeghers subjects.** *Hum Genet* 2010, **128(4)**:373-82.
- 22 Batzer MA, Deininger PL: **Alu repeats and human genomic diversity.** *Nat Rev Genet* 2002, **3**:370–379.
- 23 Chen H, Sun C, Guo W, Meng R, Du H, Qi Q, Gu X, Li L, Zhang K, Zhu D, Wang Y: **AluYb8 insertion in the MUTYH gene is related to increased 8-OHdG in genomic DNA and could be a risk factor for type 2 diabetes in a Chinese population.** *Mol Cell Endocrinol* 2011, **332**:301-305.

**5 ARTIGO 2: TORREZAN ET AL. 2012 BMC MEDICAL GENETICS 13:55**

**A novel SYBR-based duplex qPCR for the detection of gene dosage: Detection of an *APC* large deletion in a familial adenomatous polyposis patient with an unusual phenotype.**

Giovana Tardin Torrezan<sup>1</sup>, Felipe Cavalcanti Carneiro da Silva<sup>1</sup>, Ana Cristina Victorino Krepischi<sup>1,2</sup>, Érika Maria Monteiro dos Santos<sup>1</sup>, Benedito Mauro Rossi<sup>3</sup>, Dirce Maria Carraro<sup>1,2</sup>

**AUTHORS AFFILIATIONS:**

1- CIPE - International Center of Research and Training - A. C. Camargo Hospital, São Paulo, SP, Brazil

2- National Institute of Science and Technology in Oncogenomics (INCITO), São Paulo, SP, Brazil

3- Barretos Cancer Hospital - Pio XII Foundation, Barretos, SP, Brazil

**CORRESPONDING AUTHOR:**

Dirce Maria Carraro, PhD - [dirce.carraro@accamargo.org.br](mailto:dirce.carraro@accamargo.org.br)

Head, Genomic and Molecular Biology Laboratory

Coordinator, Nucleic Acid Biobank

International Center of Research and Training - A. C. Camargo Hospital

Rua Taguá, 440. CEP: 01508-010 São Paulo/SP - BR

Phone and Fax number: 55 11 2189.5023

**Abstract**

**Background:** Familial adenomatous polyposis (FAP) is a hereditary colorectal cancer syndrome caused by a loss of function of the *APC* gene. Large deletions in *APC* are a common cause of FAP; despite the existence of a variety of gene dosage detection methodologies, most are labor intensive and time and resource consuming.

**Methods:** We describe a new duplex qPCR method for gene dosage analysis based on the coamplification of a target and a reference gene in a SYBR Green reaction, followed by a comparison of the ratio between the target and the reference peaks of the melting curve for the test (patient) and control samples. The reliability of the described duplex qPCR was validated for several genes (*APC*, *HPRT1*, *ATM*, *PTEN* and *BRCA1*). **Results:** Using this novel gene dosage method, we have identified an *APC* gene deletion in a FAP patient undergoing genetic testing. Comparative genomic hybridization based on microarrays (aCGH) was used to confirm and map the extent of the deletion, revealing a 5.2 MB rearrangement (5q21.3-q22.3) encompassing the entire *APC* and 19 additional genes. **Conclusion:** The novel assay accurately detected losses and gains of one copy of the target sequences, representing a reliable and flexible alternative to other gene dosage techniques. In addition, we described a FAP patient harboring a gross deletion at 5q21.3-q22.3 with an unusual phenotype of the absence of mental impairment and dysmorphic features.

**Keywords:** Gene dosage; quantitative PCR; familial adenomatous polyposis; *APC* whole gene deletion

## Background

Familial adenomatous polyposis (FAP) is a dominantly inherited syndrome characterized by the development of hundreds of adenomatous colorectal polyps and, consequently, colorectal cancer (CRC) [1,2]. FAP is caused by mutations in the tumor suppressor gene *APC* (adenomatous polyposis coli - NM\_000038.5) mapped at 5q21-q22 [3, 4]. Over 1100 different pathogenic *APC* mutations have been reported to date (LOVD Mutation Database - <http://www.lovd.nl/2.0/>), and the great majority are nonsense mutations or small deletions and insertions leading to a truncated protein. Large gene deletions account for approximately 5% of the germline *APC* mutations, but the true prevalence of this type of alteration remains unknown due to the lack of easy screening techniques [5]. Such alterations have been detected by Southern blotting, MLPA®, competitive and differential qPCR, long-range PCR and DNA array-based methods [6]. Despite the great variety of techniques that exist to determine gene dosage, most of them are labor intensive and time and resource consuming.

The aim of this study was to develop a novel gene dosage method using SYBR Green to confirm a genomic alteration in a FAP patient. The technique is based on the coamplification, by duplex PCR, of the target and reference genes and the comparison of the peaks of the melting curve (PMc) ratio between patient and controls. This method was successfully validated and applied to identify a large genomic deletion encompassing the *APC* and 19 additional genes. Thus, this paper presents the detailed description of this novel gene dosage method, as well as evaluates the genotype-phenotype correlation of the *APC*-deleted patient.

## **Methods**

### *Samples*

The index patient (FAP02) was a Brazilian male with clinically suspected FAP identified in the Hereditary Colorectal Cancer Registry of AC Camargo Hospital (São Paulo, Brazil). His family history was accessed through the index patient report: his paternal grandmother, his father and one paternal uncle were affected with polyps and CRC at unknown ages, and one sister presented with polyps/CRC at the age of 44 that progressed to liver metastasis. The affected relatives were deceased; therefore, no biological material was available for mutation screening. One unaffected sister and one unaffected niece were available for genetic testing (Additional file 1, supplementary figure 1). The index patient was diagnosed with colorectal cancer (T1N0M0) at the age of 40, and harbored more than 100 synchronous adenomatous colorectal polyps, including duodenal and gastric polyps; no signs of other extracolonic manifestations commonly associated with FAP were observed. The patient was referred for total proctocolectomy with ileoanal pouch anastomosis. Examinations performed by a gastroenterologist and a genetic counselor did not identify any dysmorphic features or severe mental impairment. The patient completed his high school education at a regular school, and then has been working and living independently.

The validation of the gene dosage assay was performed on samples from 24 healthy controls and three patients with previously detected genomic alterations. In addition, the novel missense variant identified in the FAP patient was screened for in 95 healthy individuals.

This study was executed in compliance with the Helsinki Declaration and was approved by the ethics committee of AC Camargo Hospital (approval number: 1169/08-B). Written informed consent was obtained from all patients and controls. Genomic DNA from patients and controls was obtained from blood samples using the Puregene Genomic DNA Isolation Kit (Gentra Systems - Minneapolis, MN, USA) according to the manufacturer's instructions. The DNA concentration was verified with a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific – Waltham, MA, USA) and DNA samples were diluted to a concentration of 25 ng/ $\mu$ L. Immediately before the qPCR assay, the concentration of DNA samples was reassessed and they were diluted to a 3.6 ng/ $\mu$ L solution, from which 5  $\mu$ L was used in each qPCR reaction (a total of 18 ng).

#### *PCR and sequence analysis*

Sanger sequencing of all *APC* exons (NM\_000038.5), including intron-exon boundaries, was performed in the FAP patient. The amplified fragments were directly sequenced in both directions using the ABI3130xl sequencer (Applied Biosystems - Foster City, USA). The output results were aligned to the reference sequence using CLC Bio Main Workbench Software (Muehltal, Germany). In addition, the presence of the novel missense variant was screened for in 95 control samples by PCR followed by sequencing. The primer sequences and PCR conditions are available upon request.



### *SYBR-based duplex qPCR*

Two main principles were followed for designing the duplex qPCR primers, as follows.

1) The total length of the amplicon (including the primers) should be between 90 and 220 bases.

2) Amplicons for the target and reference genes should have a minimum of 5 °C of difference between their melting temperatures ( $T_m$ ), such that a clear individualization of target and reference gene melting peaks can be achieved.

In all duplex qPCR reactions, the target and reference genes were coamplified with 0.15  $\mu$ mol of each primer, 18 ng of DNA and 1X SYBR<sup>®</sup> Green PCR MasterMix (Applied Biosystems - Foster City, CA, USA) in a 20  $\mu$ L reaction. qPCR was performed using an ABI Prism 7500 detection system (Applied Biosystems - Foster City, CA, USA). The amplification conditions were as follows: a 10 min preincubation at 95°C followed by 25 cycles of 15 sec at 95°C and one minute at 60°C. The reduced number of cycles (25) maintains the reaction in the exponential phase, ensuring that the final amount of amplified product (demonstrated by the PMc height) is proportional to the starting amount of template DNA.

After coamplification, the PCR products were subjected to a linear temperature transition from 60°C to 95°C at 0.1°C/s, and melting curves of the decrease of SYBR Green fluorescence were generated using the ABI Prism 7500 software (Applied Biosystems - Foster City, USA). The PMc were identified by plotting the negative first derivate of the change in fluorescence ( $-dF/dT$ , the rate of change of fluorescence) vs. temperature, allowing discrimination of the two PMcs, one from the reference gene and the other from the target gene.

The ratio between the target and reference PMc height reflects the relative concentration of the target gene in the sample. Normalizing the PMc target/reference ratio of the patient sample with the PMc ratio of the control sample generates a value that corresponds to the DNA dosage, revealing whether the target sequence is deleted, duplicated or normal. Ratios between 0.9 and 1.1 were considered normal, ratios between 0.6 and 0.7 indicated deletions of one allele, and ratios above 1.4 indicated amplifications; ratios outside these intervals were considered inconclusive and repeated. Values were determined empirically based on the X chromosome dosage assay, target genes validation tests and sensibility assay described below. All the assays included three healthy controls and were performed in duplicate for patients and controls.

Validation of this novel gene dosage assay was accomplished through the assessment of X chromosome dosage in 24 healthy controls (12 males and 12 females) by determining the ratio between *HPRT1* exon 3 (Xq26.1), as the target gene, and *GAPDH* intron 7 (12p13), as the reference gene. A control sample from a female was used to normalize the ratios. Further validation was performed for different target genes (*ATM* exon 4, *PTEN* exon 2 and *BRCA1* exon 24) in samples from patients harboring germline gene deletions and duplications previously detected by aCGH in other studies of our group.

For the FAP patient, the duplex qPCR was performed for *APC* exons 2 and 15 using *GAPDH* as the reference gene. After confirming the *APC* deletion in this patient, the detection sensitivity of this method was assessed by mixing five different DNA proportions of the *APC*-deleted patient (P) and a healthy control (C). We used 18 ng of DNA in each reaction as follows: 1 P : 0 C (18 ng P : 0 ng C); 3 P : 1 C

(13.5 ng P : 4.5 ng C); 1 P : 1 C (9 ng P : 9 ng C); 1 P : 3 C (4.5 ng P : 13.5 ng C); and 0 P : 1 C (0 ng P : 18 ng C). All mixtures were submitted to a duplex qPCR for the *APC* exon 15 as the target gene and *GAPDH* as the reference gene. Experiments were performed in duplicate and data obtained from the sample containing only DNA from the wild-type control was used to normalize the PMc ratios.

#### *aCGH experiments and data analysis*

Comparative genomic hybridization based on microarrays (aCGH) was performed in duplicate for patient FAP02 using an 180K whole-genome platform (OGT – Oxford, UK). Briefly, the samples were labeled with Cy3- and Cy5-dCTPs by random priming. Purification, hybridization, and washing were performed as recommended by the manufacturer. Data extraction was conducted using the Feature Extraction software (Agilent Technologies - Santa Clara, USA). Genomic Workbench software (Agilent Technologies - Santa Clara, USA) was used to identify the constitutive genomic imbalances using the statistical algorithm ADM-2 with a sensitivity threshold of 6.7 and threshold log<sub>2</sub> ratios of 0.4 and 1.1 for duplication and high copy number gains, and -0.4 and -1.1 for deletion and homozygous loss, respectively.

## **Results**

A novel homozygous missense variant (c.6965A>G p.Gln2322Arg – figure 1A) at exon 15 was identified in one FAP Brazilian patient undergoing genetic testing of the *APC* gene (consulted databases: LOVD -<http://www.lovd.nl/2.0/>, dbSNP -<http://www.ncbi.nlm.nih.gov/projects/SNP/>, and 1000 Genomes -

<http://www.1000genomes.org/>). Because FAP is generally characterized by loss of function mutations in only one allele, the presence of an undescribed homozygous missense mutation led us to suspect the existence of a large *APC* deletion in the second allele. To test this hypothesis, we developed and validated a new SYBR® Green-based duplex qPCR method for gene dosage assessment.

To determine the ability of our technique to dose chromosomal copy alterations, an X chromosome dosage assessment was carried out. Male and female control samples were evaluated to determine the ratio of the *HPRT1* (Xq26.1) and a reference gene (*GAPDH*). The normalized PMc ratios from the male samples ranged from 0.63 to 0.70, and the ratios from the female samples ranged from 0.90 to 1.09, indicating the presence of one X chromosome in the male samples and two in the female samples (figure 2A). Visual inspection revealed a clear reduction in the *HPRT1* melting peak in the male samples relative to that in the female samples (figure 2B).

To further validate this approach for different target genes and samples, we used genomic DNA previously submitted to aCGH in which chromosomal alteration had been detected (data not shown). The duplex qPCR was performed in two samples harboring gene deletions and one sample carrying gene duplication. All genomic alterations were confirmed by this method: the deleted samples showed normalized PMc ratios of 0.63 and 0.64 (*ATM* and *PTEN* genes, respectively), and the *BRCA1* duplicated sample displayed a normalized PMc ratio of 1.5. No overlap between the ratios of any of the wild-type controls and the ratios of the deleted/duplicated patients was detected.

For the *APC* gene, the duplex qPCR was performed for exons 2 and 15, and gave normalized PMc ratios of 0.67 and 0.61, respectively (figure 1B), revealing a heterozygous deletion for these *APC* exons in the FAP patient. These results were confirmed using standard differential qPCR with  $2^{-\Delta\Delta Cq}$  calculations [7] (see Additional file 1, Methods), which revealed values of 0.56 and 0.49 for *APC* exons 2 and 15, respectively. With the purpose of elucidating the extent of the patient *APC* deletion, aCGH was executed using an OGT 180K whole-genome platform; we detected a 5.2 Mb deletion at 5q21.3-q22.3 – chr5:g.(107,755,923\_107,818,559)\_(113,079,145\_113,113,875)del (UCSC Feb. 2009 – GRCh37/hg19) – that encompasses the entire *APC* gene and 19 additional genes (figure 1C).

Finally, the sensitivity of this method was assessed by mixing different proportions of DNA from the deleted FAP patient and a healthy control (figure 3A and B). The results demonstrated that this approach could detect gene dosage alterations even when only 25% of the DNA carried the heterozygous gene deletion (proportion 4), making this technology capable of detecting mosaic gene deletions.

## Discussion

We have described a simple and inexpensive duplex qPCR method employing SYBR Green for the assessment of DNA copy number, thereby revealing genomic duplications and deletions. We also identified a FAP patient carrying a 5.2 MB deletion in 5q21.3-q22.3 in addition to a novel missense *APC* variant (p.Gln2322Arg). We examined the presence of this undescribed missense variant in 95 healthy controls, and the alteration was not detected in any control individual.

Because FAP is a dominant condition that is usually caused by truncating mutations or large deletions in the *APC* gene, we strongly believe that the causative mutation of the polyposis observed in this family is the large deletion and that the missense variant represents a rare neutral variant in the index patient.

Data regarding clinical characteristics of individuals carrying whole-gene *APC* deletions are limited. However, reported cases have consistently exhibited a degree of polyposis typical of classical FAP (between 100 and 1000 adenomatous polyps) [8,9,10]. Our patient was diagnosed with colorectal cancer at the age of 40 and harbored more than 100 adenomatous colorectal polyps, including duodenal and gastric polyps, but no other extracolonic manifestations commonly associated with FAP (desmoids, osteomas, cutaneous soft-tissue tumors, dental abnormalities and CHRPE).

Severe to mild mental impairment is an additional common phenotype described for FAP individuals with chromosomal deletions at 5q15-q22, 5q21-q22 and 5q22.1-q31.1, as revised by Readle et al. [9] and described in the Decipher Database (<http://decipher.sanger.ac.uk/>). Furthermore, those patients have generally been characterized by dysmorphic facial features, including macrognathia, hypertelorism, a high forehead, a well-demarcated philtrum, a high arched palate, and “carp mouth” [10]. The similarities between these patients raise the possibility that the various phenotypes represent a contiguous gene syndrome resulting from partial or complete loss of adjacent genes at 5q, usually occurring as *de novo* mutations, with only a few deletions being described as inherited within families [9]. In these patients, distal deletions (5q22 to 5q31.1) appear to be associated with severe intellectual disability, while deletions encompassing the 5q15 to 5q22 region

result in a milder phenotype in terms of both intellectual disability and physical features [8,9]. In our study, the identified gross deletion (5q21.3-q22.3) encompasses the entire *APC* gene and 19 additional genes and is likely to have been present in this family for at least three generations, with a quite unusual phenotype of absence of mental impairment and dysmorphic features. However, mild mental impairment cannot be ruled out because no test was applied to measure the cognitive ability of the index patient.

To date, the most frequently used techniques to detect gene copy number alterations include aCGH, which is expensive; MLPA®, which depends on the availability of the gene of interest; and quantification using real-time PCR, which can be performed with intercalating dyes (such as SYBR® Green) or TaqMan® assays. The latter are more financially onerous because specific primer and probes sets for each target gene must be ordered commercially.

Regular SYBR® Green qPCR can be performed either in a single tube with an internal control fragment (competitive PCR) or using separate reactions for the target and control genes (differential PCR). However, competitive qPCR is a time-consuming process that is limited to sets of primers available from one supplier [11], and the reproducibility of differential qPCR is inevitably compromised by the variable efficiency of the PCR itself because small variations in reaction components can greatly influence the final yield of the amplified product [12,13]. In this sense, a duplex PCR, as performed in this study, avoids the variations in template starting amounts that can occur in independent PCR reactions, e.g., due to operator loading errors.

It is important to notice that this method requires that target and reference gene amplicons display a minimum difference between the melting temperatures, such that a clear individualization of the target and the reference gene melting peaks can be achieved. In our experience, this minimum difference was  $\geq 5^{\circ}\text{C}$ . Another important point to consider is that, similar to other amplicon-based gene dosage approaches, the duplex qPCR only assesses the copy number of the amplified region (90-220 bp), meaning that primers for each exon should be designed to screen the entire coding-sequence of a gene.

The novel assay presented here accurately detected losses and gains of one copy of the target sequence in all analyses (including mosaics of down to 25% of heterozygous mutated cells). Therefore, this methodology represents a reliable and flexible alternative both for screening gene dosage changes and as a validation assay for previously detected alterations. Nevertheless, validation of this technique for diagnostic purposes demands additional analytical studies in larger cohorts and for different genes.

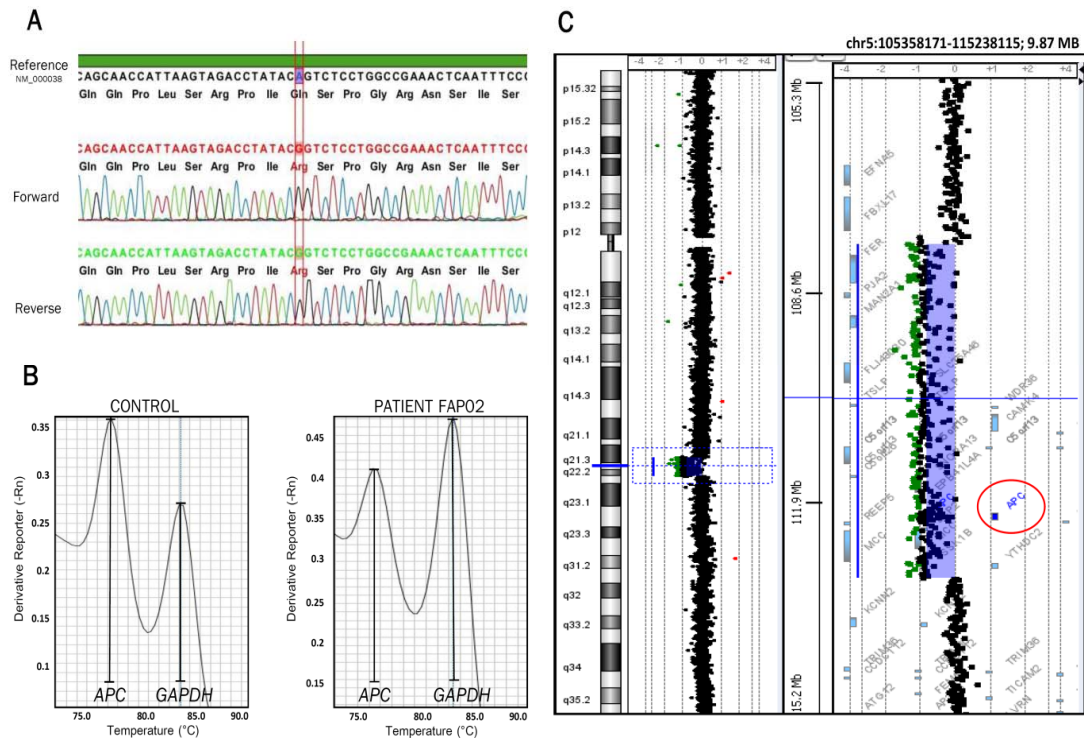
## **Conclusions**

We have described a simple and inexpensive duplex qPCR method employing SYBR Green for the assessment of DNA copy number, thereby revealing genomic duplications and deletions. We also reported a FAP patient carrying a 5.2 MB deletion at 5q21.3-q22.3 in addition to a novel missense *APC* variant (p.Gln2322Arg). This gross deletion that encompasses the entire *APC* gene and 19 additional genes is likely to have been present in this family for at least three



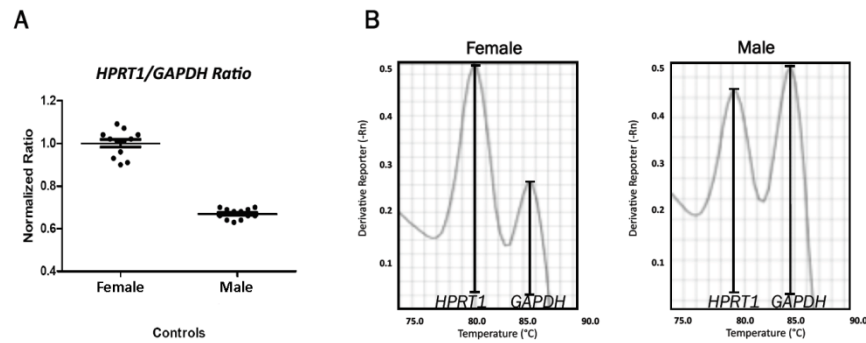
generations, with a quite unusual phenotype of absence of mental impairment and dysmorphic features.

## Figures and Legends

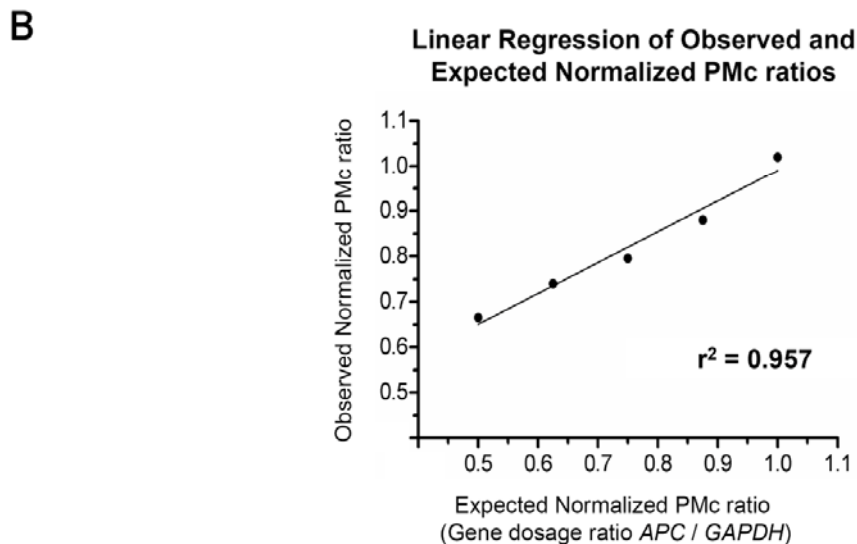
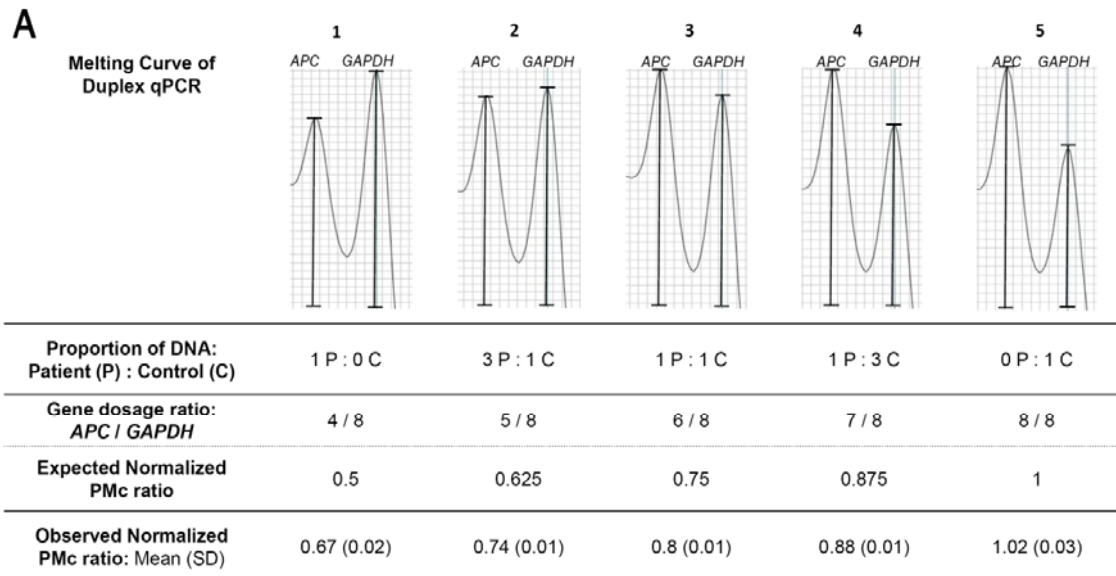


**Figure 1: Missense mutation and deletion of the APC gene in one FAP patient.**

**A:** Chromatogram of the novel missense variant c.6965A>G p.Gln2322Arg in APC exon 15 apparently in homozygosity. **B:** Melt curve of duplex qPCR of APC exon 15 and GAPDH intron 7 (reference gene) – PMc ratio of APC/GAPDH in the control sample was 1.4 and in the FAP patient it was 0.86, leading to a normalized PMc ratio of 0.61 for the FAP patient, indicating the presence of deletion. **C:** aCGH chromosome 5 profile of the FAP patient showing a 5.2 Mb deletion at 5q21.3-q22.3 that encompasses the entire APC sequence (red circle) and 19 additional genes.



**Figure 2: X chromosome dosage assay for validation of the SYBR Green Duplex qPCR. A:** Distribution of the normalized *HPRT1* / *GAPDH* ratio in 24 control samples (12 females and 12 males); the normalized ratio of the 12 female samples ranged from 0.90 to 1.09, while in the 12 male samples the range was from 0.63 to 0.70, when using a female sample as the control. **B:** Visual comparison of the melt curve revealing a reduction of the *HPRT1* peak when comparing a male to a female sample.



**Figure 3: Sensitivity of the Duplex qPCR.** The sensitivity of this methodology was investigated by mixing different DNA proportions of the deleted patient and a healthy control. **A:** First panel presents the melting curve of the five different DNA proportions used (described in the first line of the table); second and third line show the gene dosage ratio between *APC* and *GAPDH* present in each mixed DNA solution and the respective expected normalized PMc ratio; last line presents mean and standard deviation (SD) of the observed normalized PMc ratio duplicates. The

results demonstrated that this approach could detect gene dosage alterations even when only 25% of the DNA carried the heterozygous gene deletion (proportion 4 = 1 P : 3 C). **B:** Linear regression of the observed normalized PMc ratio in relation to the expected PMc ratio, demonstrating an  $r^2$  of 0.957.

### **Competing Interests**

The authors GTT, FCCS, AVK, EMMS, BMR and BMC state that they have no competing financial interests.

### **Authors' Contributions**

Conceived and designed experiments: GTT, BMR, DMC; Performed and analyzed experiments: GTT, FCCS, ACVK, DMC; Ascertained and selected patients: EMMS, BMR; Contributed with reagents/materials/analyses tools: BMR, DMC; Wrote/edited the paper: GTT, FCCS, ACVK, DMC. All authors read and approved the final manuscript.

### **Acknowledgments and Funding**

The authors acknowledge the patients for their collaboration, the Nucleic Acid Bank of A. C. Camargo Hospital for sample processing and Bianca C. G. Lisboa for guidance in establishing the sequencing protocols. This work was supported by the National Institute of Science and Technology in Oncogenomics (INCITO) – FAPESP: 2008/57887-9.

### **Description of Additional files**

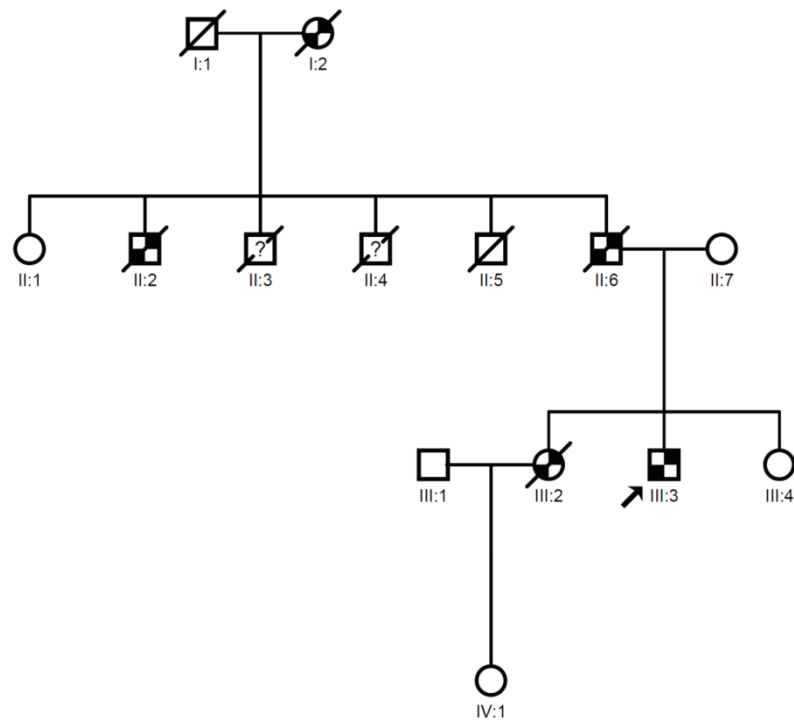
Additional file 1, PDF, Methods, description of the Differential qPCR.

Additional file 1, PDF, Supplementary Figure 1.

## Additional file 1

### Methods:

**Differential qPCR:** 18 ng of each DNA sample was amplified in a 20  $\mu$ L reaction containing 0.3  $\mu$ mol of each primer and 1X SYBR<sup>®</sup> Green PCR MasterMix (Applied Biosystems). qPCR was performed on an ABI Prism 7500 detection system (Applied Biosystems) as following: preincubation of 95°C for 10 minutes and, 40 cycles of denaturation at 95°C for 15 seconds and annealing and extension at 60°C for 1 minute. Samples were amplified in duplicate for both target and normalizer genes. The average  $C_q$  of duplicates was used in the gene dosage ratio calculations. For each exon, four control individuals were used to calculate the average control  $C_q$ . Gene dosage ratios were calculated using the following equation:  $2^{-[\Delta C_q (\text{target}) - \Delta C_q (\text{ref})]}$ , where  $\Delta C_q (\text{target})$  equals the difference between the  $C_q$  values for the patient and the control average for the target exon, and  $\Delta C_q (\text{ref})$  equals the difference between the  $C_q$  values for the patient and the control average for the reference gene. The  $2^{-\Delta\Delta C_q}$  method was performed for two *APC* exons (2 and 15) and for two reference genes: *GAPDH* intron 7 (12p13) and *HPRT1* exon 3 (Xq26.1). Ratios in the range 0.82 - 1.22 were considered normal, ratios between 0.41 and 0.61 indicated one copy deletion, and ratios  $\geq 1.4$  indicated amplifications (10% error in measured concentration).



**Supplementary Figure 1:** Family pedigree of patient FAP02. The arrow indicates the index patient. All affected individuals presented both polyposis and colorectal cancer. The patient presented four affected deceased relatives: grandmother, uncle, father and one sister. One unaffected sister (III:4) and one unaffected niece (IV:1) were tested and neither the deletion nor the missense variant were detected. Genetic testing of the unaffected individual II:1 could not be performed. The symbol (?) indicates unknown cause of death in individuals II:3 and II:4.

## References

- 1 Lipton L, Tomlinson I. **The genetics of FAP and FAP-like syndromes.** *Familial Cancer* (2006) **5**:221–226.
- 2 Half E, Bercovich D, Rozen P. **Familial adenomatous polyposis.** *Orphanet Journal of Rare Diseases* 2009, **4**:1-23.
- 3 Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, Levy BD, Smith KJ, Preisinger AC, Hedge P, McKechnie D. **Identification of FAP locus genes from chromosome 5q21.** *Science* 1991, **253**:661-5.
- 4 Nishisho I, Nakamura Y, Miyoshi Y, Miki Y, Ando H, Horni A, Koyama K, Utsunomiya J, Baba S, Hedge P. **Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients.** *Science* 1991, **253**:665-9.
- 5 Aretz S, Stienen D, Uhlhaas S, Pagenstecher C, Mangold E, Caspari R, Propping P, Friedl W. **Large submicroscopic genomic APC deletions are a common cause of typical familial adenomatous polyposis.** *Journal of Medical Genetics* 2005, **42**:185-192.
- 6 Barrois M, Bièche I, Mazoyer S, Champemeb M-H, Bressac-de Pailleretsa B, Lidereau R. **Real-time PCR-based gene dosage assay for detecting BRCA1 rearrangements in breast-ovarian cancer families.** *Clinical genetics* 2004, **65**:131-6.
- 7 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, **5**:402-408.
- 8 Lindgren V, Bryke CR, Ozcelik T, Yang-Feng TL, Francke U. **Phenotypic, cytogenetic, and molecular studies of three patients with constitutional deletions of chromosome 5 in the region of the gene for familial adenomatous polyposis.** *The American Journal of Human Genetics* 1992, **50**:988–97.
- 9 Raedle J, Friedl W, Engels H, Koenig R, Trojan J, Zeuzem S. **A de novo deletion of chromosome 5q causing familial adenomatous polyposis, dysmorphic features, and mild mental retardation.** *The American Journal of Gastroenterology* 2001, **96**:3016-20.

- 10 Finch R, Jalal S, Moore HG, Markowitz A, Offit K, Lindor N, Jhanwar S, Guillem J. **Familial Adenomatous Polyposis and Mental Retardation Caused by a de novo Chromosomal Deletion at 5q15-q22: Report of a Case.** *Diseases of the Colon & Rectum* 2005, **48**: 2148-2152.
- 11 Mocellin S, Rossi CR, Pilati P, Nitti D, Marincola FM. **Quantitative real-time PCR: a powerful ally in cancer research.** *Trends in Molecular Medicine* 2003, **9**:189-195.
- 12 Ruiz-Ponte C, Carracedo A, Barros F. **Duplication and deletion analysis by fluorescent real-time PCR-based genotyping.** *Clinica Chimica Acta* 2006, **363**:138-46.
- 13 Ruiz-Ponte C, Loidi L, Vega A, Carracedo A, Barros F. **Rapid realtime fluorescent PCR gene dosage test for the diagnosis of DNA duplications and deletions.** *Clinical Chemistry* 2000, **46**:1574–82.



**6 ARTIGO 3: TORREZAN ET AL. 2013 - ORPHANET  
JOURNAL OF RARE DISEASES, 8:54.**

**Mutational spectrum of the *APC* and *MUTYH* genes and genotype–phenotype correlations in Brazilian FAP, AFAP, and MAP patients**

Giovana Tardin Torrezan<sup>1</sup>, Felipe Cavalcanti Carneiro da Silva<sup>1</sup>, Érika Maria Monteiro dos Santos<sup>2,3</sup>, Ana Cristina Victorino Krepschi<sup>2,4</sup>, Maria Isabel Waddington Achatz<sup>2,5</sup>, Samuel Aguiar Junior<sup>3</sup>, Benedito Mauro Rossi<sup>6\*</sup>, Dirce Maria Carraro<sup>1,2\*</sup>

1- Laboratory of Genomics and Molecular Biology, A. C. Camargo Hospital, São Paulo, SP, Brazil.

2- National Institute of Science and Technology in Oncogenomics (INCITO), São Paulo, SP, Brazil

3- Colorectal Tumors Department, A. C. Camargo Hospital, São Paulo, SP, Brazil

4- Laboratory of Structural Genomics, A. C. Camargo Hospital, São Paulo, SP, Brazil

5- Oncogenetics Department, A. C. Camargo Hospital, São Paulo, SP, Brazil

6- Barretos Cancer Hospital, Pio XII Foundation, Barretos, SP, Brazil

**\*CORRESPONDING AUTHORS:**

Benedito Mauro Rossi, MD, PhD - rossibm@gmail.com

Researcher - Barretos Cancer Hospital

Rua Dona Adma Jafet 74 cj. 145 Sao Paulo - SP, Brazil.

Phone and Fax number: +55-11-3259-3433

Dirce Maria Carraro, PhD - dirce.carraro@cipe.accamargo.org.br

Head, Laboratory of Genomics and Molecular Biology

International Center of Research - A. C. Camargo Hospital

Rua Taguá, 440. CEP: 01508-010 São Paulo/SP - BR

Phone and Fax number: 55 11 2189.5023

## ABSTRACT

**Background:** Patients with multiple colorectal adenomas are currently screened for germline mutations in two genes, *APC* and *MUTYH*. *APC*-mutated patients present classic or attenuated familial adenomatous polyposis (FAP/AFAP), while patients carrying biallelic *MUTYH* mutations exhibit *MUTYH*-associated polyposis (MAP). The spectrum of mutations as well as the genotype-phenotype correlations in polyposis syndromes present clinical impact and can be population specific, making important to obtain genetic and clinical data from different populations. **Methods:** DNA sequencing of the complete coding region of the *APC* and *MUTYH* genes was performed in 23 unrelated Brazilian polyposis patients. In addition, mutation-negative patients were screened for large genomic rearrangements by multiplex ligation-dependent probe amplification, array-comparative genomic hybridization, and duplex quantitative PCR. Biallelic *MUTYH* mutations were confirmed by allele-specific PCR. Clinical data of the index cases and their affected relatives were used to assess genotype-phenotype correlations. **Results:** Pathogenic mutations were identified in 20 of the 23 probands (87%): 14 in the *APC* gene and six in the *MUTYH* gene; six of them (30%) were described for the first time in this series. Genotype-phenotype correlations revealed divergent results compared with those described in other studies, particularly regarding the extent of polyposis and the occurrence of desmoid tumors in families with mutations before codon 1444 (6/8 families with desmoid). **Conclusions:** This first comprehensive investigation of the *APC* and *MUTYH* mutation spectrum in Brazilian polyposis patients showed a high detection rate and identified novel pathogenic mutations. Notably, a significant number of *APC*-positive families were not consistent with the predicted genotype-phenotype correlations from other populations.

**Keywords:** *APC*, *MUTYH*, genotype-phenotype, mutation screening, polyposis

## BACKGROUND

Patients with multiple colorectal adenomas are screened for germline mutations in two distinct genes, *APC* and *MUTYH*. According to the polyp number and age of onset, the phenotype of *APC*-mutated patients can be classified as classical familial adenomatous polyposis (FAP: more than 100 polyps, early onset) or attenuated FAP (AFAP: fewer than 100 polyps with later onset) [1-3]. *MUTYH* biallelic mutation carriers usually present 10 to 100 polyps and are categorized as having *MUTYH*-associated polyposis (MAP) [4].

FAP/AFAP (OMIM #175100) is a dominantly inherited colorectal cancer (CRC) predisposing syndrome [1, 2] caused by mutations in the tumor suppressor gene adenomatous polyposis coli (*APC*). The encoded APC protein controls  $\beta$ -catenin turnover in the Wnt pathway [5, 6]. Besides colonic polyposis and colorectal cancer, individuals with FAP can present a number of benign extracolonic features, including multiple osteomas, epidermoid cysts, desmoid tumors, and congenital hypertrophy of the retinal pigment epithelium [2]. Over 1100 different pathogenic *APC* mutations have been reported to date in the Leiden Open Variation Database (<http://www.lovd.nl/2.0/>), the majority of them being nonsense mutations or small insertions or deletions that lead to a truncated protein. Mutations causing AFAP have been reported to occur mainly in three regions of *APC*: at the 5' end (the first five exons), in the alternatively spliced region of exon 9, or at the 3' end (after codon 1580) [7-9].

MAP (OMIM #608456) is a recessively inherited syndrome caused by biallelic mutations in the mutY homolog (*MUTYH*) gene that maps to chromosome 1p34.1 [4]. *MUTYH* encodes a DNA glycosylase that plays a key role in the base

excision repair pathway by removing mispaired bases caused by the oxidation product 8-oxoG [4]. Nearly 300 different sequence variants have been identified in this gene (LOVD Mutation Database), including about 80 pathogenic mutations distributed throughout the gene at positions corresponding to different functional domains of the encoded protein [10]. In contrast to *APC* pathogenic variants, which mostly result in a truncated or absent protein, most *MUTYH* pathogenic variants are missense substitutions and only a minority are splice site or truncating mutations [11].

With regard to clinical features, most *MUTYH*-mutated patients present 10 to 100 colorectal adenomas, usually with later onset compared with FAP patients [4, 12, 13]. *MUTYH* mutation carriers represent approximately 7.5% of patients with more than 100 adenomas without an *APC* mutation, 40% of all patients with 10–100 polyps, and 0.3–1.7% of patients with fewer than 10 polyps and early-onset CRC with no family history [12, 14, 15, 16]. Furthermore, it has also been reported MAP patients having no polyps at the time of CRC diagnosis [17, 18].

Because of the observed overlap between the clinical phenotype of FAP/AFAP and MAP syndromes, the identification of the causative mutation has important implications for family management, allowing effective clinical surveillance and accurate genetic counseling. Moreover, the spectrum of mutations and the genotype–phenotype correlations may have clinical impact and can be population-specific. Therefore, it is important to obtain genetic and clinical data from FAP/AFAP and MAP families in different populations.

The aim of this study was to conduct a comprehensive molecular analysis to determine the spectrum of point mutations and large genomic rearrangements (LGR)

in the *APC* and *MUTYH* genes in a series of 23 Brazilian polyposis patients. This paper summarizes the mutation screening data and outlines the most cost-effective approach to detect *APC* and *MUTYH* mutations in Brazilian polyposis patients. In addition, we discuss the genotype–phenotype associations found in these families in the context of previously described data from the literature.

## **METHODS**

### **Patients**

The study examined the Hereditary Colorectal Cancer Registry of A. C. Camargo Hospital (São Paulo, Brazil) [19], for families clinically suspected for FAP (> 100 colorectal adenomas) or AFAP/MAP (10–100 colorectal adenomas), enrolled between January 1998 and July 2011. Between 2010 and 2011, the genetic test was offered to forty registered unrelated polyposis families, from which 23 were available and willing to undergo genetic testing. Index patients were interviewed after providing informed consent and the family history was obtained through verbal report and, whenever possible, confirmed with clinical or pathological reports.

This study was performed in compliance with the Declaration of Helsinki and was approved by the ethics committee of A. C. Camargo Hospital (approval number: 1169/08-B). Once a mutation was identified in the index case, genetic counseling and molecular testing were offered to relatives.

### PCR and sequence analysis

Mutation screening was performed by capillary sequencing of all coding exons of the *APC* [GenBank:NM\_000038.5] and *MUTYH* [GenBank:NM\_001128425.1] genes, including the intron–exon boundaries. Patients clinically suspected for FAP (> 100 polyps) were first screened for *APC* mutations, while patients with attenuated polyposis (< 100 polyps) were first screened for *MUTYH* mutations. Patients negative for the first screened gene were then screened for the remaining one.

Genomic DNA was obtained from leukocytes using a Puregene Genomic DNA Isolation Kit (Gentra Systems, Minneapolis, MN, USA) according to the manufacturer's instructions. PCR reactions used 25 ng of template and 500 nM of each primer in a final volume of 20 µl with GoTaq Green Master Mix (Promega, Madison, WI, USA). Approximately 200 ng of PCR-amplified fragments were purified with ExoSAP-IT (USB Corporation, Cleveland, OH, USA) and sequenced in both directions. Products were analyzed using an ABI 3130xl DNA sequencer (Applied Biosystems, Foster City, CA, USA) and the resulting sequences were aligned using CLCBio Genomics Workbench Software (Muehltal, Germany). The sequences of primers used for these analyses are available upon request. All mutations were confirmed in a second DNA sample. Mutations were recorded and referenced with respect to the cDNA sequence, using the nomenclature guidelines proposed by the Human Genome Sequence Variation Society ([www.hgvs.org/mutnomen](http://www.hgvs.org/mutnomen)).

### **Allele-specific PCR**

Allele-specific PCR [20] was performed for *MUTYH* mutations to confirm the presence of two heterozygous alleles (compound heterozygosity). Primers were designed to be specific for the wild-type or mutated nucleotide of one of the *MUTYH* mutations. Sequencing of allele-specific PCR amplicons was performed to reveal the haplotype phase of the second mutation. PCR conditions and primers are available upon request.

### **LGR screening**

Seven patients were selected for LGR screening using multiplex ligation-dependent probe amplification (MLPA), array-comparative genomic hybridization (aCGH), and duplex quantitative PCR (qPCR): five negative for *APC* or *MUTYH* point mutations, one with a novel *APC* missense variant, and one with a monoallelic *MUTYH* mutation. All experiments were performed in duplicate.

MLPA was performed using the SALSA P043-C1 *APC* Probemix kit (MRC Holland, Amsterdam, The Netherlands) following the manufacturer's protocol. PCR products were analyzed using an ABI 3130xl DNA sequencer (Applied Biosystems - Foster City, CA, USA), and gene dosage was calculated using Coffalyser V9.4 software (MRC Holland, Amsterdam, The Netherlands).

The aCGH platform used in this study was the SurePrint G3 Human CGH Microarray Kit 4×180k (G4449A; Agilent Technologies, Santa Clara, CA, USA), which has an average resolution of 18 kb, with 13 and three probes located within *APC* and *MUTYH*, respectively. Briefly, samples were labeled with Cy3- or Cy5-dCTPs by random priming. Purification, hybridization, and washing were performed



as recommended by the manufacturer. Data extraction was conducted using Feature Extraction software (Agilent Technologies - Santa Clara, CA, USA). Genomic Workbench software (Agilent Technologies - Santa Clara, CA, USA) was applied to identify constitutive genomic imbalances using the statistical algorithm ADM-2, with a sensitivity threshold of 6.7, and a threshold  $\log_2$  ratio of 0.4 for duplication and – 0.4 for deletion.

Genomic alterations identified by MLPA and aCGH were validated using the duplex qPCR method previously established by our group [21].

### **Variant analysis**

Mutations in the *APC* or *MUTYH* genes were considered deleterious if they: a) were classified as pathogenic in LOVD database; b) introduced a premature stop codon in the protein sequence (nonsense or frameshift mutation); c) occurred at donor or acceptor splice sites; or d) were whole-exon deletions or duplications. To establish the pathogenicity of one novel missense variant, web-based programs that predict the effect of an amino acid substitution were applied (SIFT, Polyphen, and MutationTaster). In addition, the frequency of this variant was assessed in 95 healthy Brazilian individuals.

### **Clinical features and genotype-phenotype correlations**

The following clinical and pathological data were obtained from all families from the Hereditary Colorectal Cancer Registry of A. C. Camargo Hospital [19]: number of affected individuals, age at diagnosis, number of patients with extracolonic features, and primary sites of extracolonic tumors. The extent of polyposis burden (number of adenomas) was assessed for the index cases through

colonoscopy records and/or pathological report from surgical specimens. For most family members this information was unavailable. Patients and their families were grouped according to the affected gene and the index case polyposis burden into five categories: group 1, *APC*-mutated families with fewer than 100 colorectal adenomas (attenuated polyposis); group 2, *APC*-mutated families with 100–1000 adenomas (intermediate polyposis); group 3, *APC*-mutated families with more than 1000 adenomas (severe polyposis); group 4, *MUTYH*-mutated families; and group 5, mutation-negative families.

Genotype–phenotype correlations in the three *APC*-mutated groups were compared with those previously described, as reviewed by Nieuwenhuis and Vasen (2007) [22]. This review evaluated a large number of studies in FAP patients and proposed a categorization of the phenotypes according to the severity of the polyposis and the associated site of the *APC* mutation.

Statistical evaluation was performed using the Student's *t*-test using Prism 5 software (GraphPad, San Diego, CA, USA). Statistical significance was set at a *p*-value < 0.05.

## RESULTS

Twenty-three Brazilian families with a clinical diagnosis of classical or attenuated polyposis were included in this study. The majority of the index cases (15) presented an intermediate or severe FAP phenotype (> 100 polyps) and 13 of them harbored an *APC* pathogenic mutation, while one patient was mutation-negative and one had a monoallelic *MUTYH* mutation. The remaining eight patients presented an

attenuated polyposis burden (< 100 polyps), among whom five carried biallelic mutations in the *MUTYH* gene, one carried a novel *APC* duplication of exons 1–3, one presented a novel *APC* missense variant, and one was mutation-negative. Seven novel germline mutations (six pathogenic and one variant of unknown significance) were detected in this cohort, and two of them have been recently published by our group [21, 23]. The *APC* and *MUTYH* mutation spectrum, including information about previous reports of the detected mutations, is summarized in Table 1.

### ***APC* mutations**

Fourteen pathogenic *APC* mutations were identified in this series: three small duplications, five small deletions, four nonsense mutations, one multiple exon duplication, and one whole-gene deletion. Six of them were novel mutations (Table 1). All patients presented distinct mutations, except for two unrelated probands that presented the hotspot mutation at codon 1309 (c.3927–3931delAAAGA; p.Glu1309Aspfs\*4).

One patient (ID13) presented a novel missense *APC* variant of unknown significance: c.5365G>C (p.Val1789Leu). This patient was diagnosed with attenuated polyposis at the age of 56 years, presenting around 20 polyps at the time of clinical diagnosis. *In silico* studies using three different functional prediction programs (Polyphen, SIFT and MutationTaster), which all classified the p.Val1789Leu variant as having minimal or no effect on protein function, with the following scores: 0 (Polyphen); 0.30 (SIFT); 0.87 (*P*: 0.99, MutationTaster). Because the proband was the only affected member of the family, it was not possible to perform co-segregation analysis of the variant with the disease within the family;

nevertheless, this variant was not detected in a control population of 95 healthy individuals.

In this series, two patients presented *APC* LGRs, identified by MLPA and/or aCGH and confirmed by gene dosage qPCR. Patient ID02 presented a 5.2-Mb deletion at 5q21.3–q22.3 that encompassed the entire *APC* gene and 19 additional genes, which have been previously published by our group [21]. The second patient (ID17) presented a duplication of *APC* exons 1–3 that was identified by MLPA (Figure 1A) and validated by duplex qPCR [21] (Figure 1B).

### ***MUTYH* mutations**

Biallelic germline mutations in the *MUTYH* gene were identified in five patients, among whom two were homozygotes for the causative mutation and the remaining three were compound heterozygotes for two distinct pathogenic variants. A monoallelic mutation was identified in one patient.

One patient (ID19) and her brother were homozygous for the p.Arg241Trp missense mutation, because of a consanguineous marriage between the parents. The second homozygous patient presented a deletion of exons 4–16 (c.348+33\_\*64+146del4285insTA) on both alleles, and stated having no known inbreeding in her family. This 4,285 bp deletion was the first LGR to be described in *MUTYH*, recently published by us [23] and by an independent group that found this deletion in a French patient [24].

For the three patients harboring two distinct pathogenic variants (ID16, ID18, and ID447), we used allele-specific PCR to confirm the biallelic nature of the mutations. All cases presented the hotspot missense mutation p.Tyr179Cys in one

allele accompanied by a second truncating mutation in the remaining allele (one deletion, one duplication, and one splice site mutation).

One patient (ID24) was a monoallelic carrier of the hotspot missense mutation p.Gly396Asp, and no other mutation could be identified.

### **Clinical Features**

Clinical records and verbal reports obtained from the 23 index patients and their relatives revealed 113 affected individuals among all families; their summarized clinical data are described in Table 2.

#### **Polyposis/CRC age at diagnosis**

Across the entire series, the average age at diagnosis and first symptoms of CRC and/or polyposis was 32.6 years (range 7–67 years). The mean age of onset in *APC*-positive families was 46.3 years (range 35–56) for group 1 (attenuated FAP); 35.7 (range 18–67) for group 2 (intermediate FAP); and 29.2 (range 7–58) for group 3 (severe FAP). MAP families presented a mean age of 37.9 years (range 27–53 years); while the average age of onset in families with no identified mutation was 27.5 years (range 26–29 years). Comparison among the five groups revealed that the *APC*-positive group 1, group 2, and MAP patients demonstrated a later age of onset compared with the severe FAP patients (group 3) (Figure 2).

#### **Extracolonic manifestations**

Extracolonic manifestations were reported in all *APC*-mutated families (Table 2). Gastric and duodenal polyps (upper gastrointestinal polyps) were the most common extracolonic manifestations observed in these patients, and occurred in

11/14 families (79%) and across all three *APC* groups. Osteomas were observed most often in the severe FAP patients (6/9), and epidermoid cysts/lipomas occurred in the intermediate and severe FAP patients (5/13).

Desmoid tumors were observed in 8/14 *APC*-positive families (57%) and were associated with different mutation sites, with only two of them occurring after codon 1444. Five families had more than one individual affected by desmoid tumors.

MAP families had fewer extracolonic manifestations: only one family presented upper gastrointestinal polyps and none presented desmoid tumors.

Regarding other tumor sites, papillary thyroid carcinoma appeared in one family of *APC* group 3; liver and breast cancers were reported in two families each: one from group 2 and one from group 3; uterine cancer was reported in one group 1 and one group 4 (MAP) family. Finally, lung, hematologic, brain, or skin cancer and melanoma were reported in one family each.

### **Comparison with described *APC* genotype–phenotype correlations**

Genotype–phenotype correlations in polyposis syndromes have been evaluated in several studies, and a general association between the location of the mutation and the clinical manifestation has been observed, albeit with some inconsistencies [7, 21, 25-27]. Recently, Nieuwenhuis and Vasen (2007) [22] performed a meta-analysis and proposed categorization of the phenotypes into three degrees of polyposis severity and the associated site of *APC* mutation. Attenuated FAP was associated with mutations before codon 157, after codon 1595, and in the alternatively spliced region of exon 9; severe polyposis was related to mutations

between codons 1250 and 1464; and an intermediate phenotype was associated with *APC* mutations located in the remaining sequence of the gene (Figure 3).

We compared the clinical and genetic data from our cohort with the *APC* codon limits defined by Nieuwenhuis and Vasen (2007) [22]. Figure 3A and B show the distribution of the polyposis phenotype of the index cases according to the location of their *APC* mutation, and compare it with the genotype–phenotype correlations previously proposed. Nine (64%) of 14 FAP families with an *APC* pathogenic mutation presented the expected polyposis severity according to the location of the *APC* mutation, while the remaining five families exhibited discordant results from the anticipated phenotype. Nine families presented a profuse polyposis burden (> 1000 polyps) — two of them carried a mutation in codon 1309, two in codon 1367, and one in codon 1294, while the four remaining patients carried mutations in codons usually associated with an intermediate number of polyps (codons 283, 302, 1017, and 1465) (Figure 3A and B).

### **Particular phenotypes**

Two of the herein identified mutations presented a remarkably more aggressive phenotype than expected given their location and the phenotypes reported in the literature. The mutation c.447dupC (p.Lys150GlnfsX18), identified in family ID05, is located in exon 4; the 5' region of the *APC* gene is generally associated with attenuated polyposis and later age of onset. While the mean age of onset in this family (40 years old) was within the predicted range, the polyposis phenotype caused by this novel mutation was more aggressive than expected, since several members of the family presented more than 100 adenomatous polyps. Furthermore, three

relatives had developed a desmoid tumor, usually not observed in patients with a mutation at the 5' end of the *APC* gene.

An aggressive phenotypic expression was also observed for the mutation c.3050–3053delATGA (p.Asn1017MetfsX4), identified in family ID10 (Figure 4). Although the mutation is located in the region associated with an intermediate FAP phenotype, the proband presented a high number of polyps (> 1000) at the age of 15, a desmoid tumor at the age of 20, and thyroid carcinoma and jaw keratocysts at 21 years. Her brother and seven cousins also developed polyps at early ages (7, 14, 15, 19, 17, 22, and two at 29 years old). Desmoid tumors were described in another three relatives.

## DISCUSSION

This is the first report of a comprehensive mutational analysis and genotype–phenotype correlation in Brazilian polyposis families. Through direct sequencing of the *APC* and *MUTYH* genes, MLPA, aCGH, and duplex qPCR, we were able to identify pathogenic mutations in 20 of 23 index cases — a detection rate of 87%. Of the remaining three patients, two were mutation-negative and one harbored a novel *APC* missense variant (p.Val1789Leu). Because of the lack of affected relatives for co-segregation analysis and the inconclusive results given by *in silico* analysis and control population screening, the clinical significance of this alteration is yet to be determined. Interestingly, a parallel study from our institution, performed in high risk cancer patients, revealed that this patient also presents a rare germline microdeletion of the *PIP* gene possibly associated with an increased cancer risk (Silva 2013,



unpublished observations), suggesting that these two alterations may be acting in synergy.

The detection rate in polyposis patients from other populations varies markedly, ranging from 39 to 90%; the variation reflects different selection criteria for testing and diverse sensitivity of screening strategies [4, 28-32]. Our data reinforce the need to apply a combination of mutation-screening methods to detect the disease-causing mutation in polyposis patients efficiently. Most *APC* mutations previously described were identified with conventional methods, for instance denaturing high performance liquid chromatography or the protein truncation test, which can have a relatively low detection rate. Nowadays, the gold standard detection method for polyposis patients is direct DNA sequencing of all *APC* and *MUTYH* coding exons (including intron–exon boundaries), accompanied by screening for LGR, as was performed here.

The majority of mutations identified in our cohort were distinct, except for two families who shared the codon 1309 hotspot *APC* mutation and three families who presented the p.Tyr179Cys hotspot mutation in one of the *MUTYH* alleles. The absence of the commonly reported *APC* mutation at codon 1061, and the relatively low frequency of the hotspot mutations *APC* 1309, *MUTYH* p.Tyr179Cys, and p.Gly396Asp are consistent with the fact that Brazilian patients represent an admixed population, probably lacking a founder *APC* or *MUTYH* mutation.

The p.Tyr179Cys and p.Gly396Asp *MUTYH* mutations were identified in three families and one family, respectively, and corresponded to 44% of all mutated alleles identified in this gene (4/9). These are also the most prevalent mutations in populations of European origin, probably because of a founder effect, and account

for approximately 80% of all reported mutant alleles [12]. A recent report described a screen for these two variants in 30 Brazilian patients with clinical phenotypes of MAP and FAP; 5/30 patients were identified as carrying one of these two hotspot mutations, and four of them were in a biallelic state [33]. However, because the entire coding sequence of *MUTYH* was not evaluated in all patients in this study, we cannot perform comparisons with the frequency found in our patients.

In our series, we could not identify a causative mutation in two index cases—one with attenuated polyposis and one with an intermediate polyposis phenotype. A possible explanation for the polyposis phenotype in these patients is the presence of unusual mutations in the *APC* or *MUTYH* genes, such as intron or promoter point mutations, epimutations or genetic mosaicism. In this sense, a recent study demonstrated that up to 8% of *APC/MUTYH*-negative polyposis patients presented a deep intronic *APC* variant that led to an aberrant transcript [34]. A second possibility is the existence of other susceptibility genes, and with the current possibility of screening all coding genes by next generation exome sequencing, it can be anticipated that novel polyposis-predisposing genes will be identified.

Considering that this study is the first comprehensive analysis of *APC* and *MUTYH* mutations in Brazilian polyposis patients, we attempted to determine the most cost-effective approach to detect the causative mutation in this population. In patients presenting fewer than 100 polyps (N = 8), 62% carried a biallelic mutation in the *MUTYH* gene. Among patients with more than 100 polyps (15 cases), three cases presented a mutation in *APC* exon 8 and eight cases (53%) exhibited a mutation between codons 1017 and 1650 of *APC* exon 15. Interestingly, this initial region of exon 15 comprises only 16% of the coding sequence of *APC*, yet presented a high

mutation rate in our cohort. Therefore, based on our results, an optimized scheme for the molecular diagnosis of *APC* and *MUTYH* mutations in the Brazilian population might be obtained as follows: for patients presenting > 100 polyps, codons 1017–1650 of *APC* exon 15 should be sequenced first, followed by exon 8, and then the remaining *APC* exons; for patients presenting fewer than 100 polyps, *MUTYH* should be firstly screened. Because none of the studied *MUTYH* patients presented only the hotspot mutations p.Tyr179Cys or p.Gly396Asp, the whole gene should be sequenced, instead of undertaking an initial search for these variants, as recommended for other populations [18].

Genotype–phenotype correlations in polyposis syndromes are of great clinical interest, because they can contribute to better genetic counseling and simplify mutation screening. In several studies, an association between the location of the mutation and the clinical manifestations has been observed [7, 20, 25–28]. However, since several contradictions have also been reported [22], it remains unclear whether the genetic information should guide clinical decision-making, such as the extent of the prophylactic colectomy or the protocol for clinical surveillance [7, 35 36, 37, 38, 39].

Regarding the age at which clinical surveillance should begin, the established guidelines suggest that classical FAP patients should start endoscopic surveillance from the early teens, while AFAP and MAP families could start surveillance at age 18–20 [38]. Similar to the literature, our results demonstrated that severe FAP patients had an earlier age of onset (on average 10 years younger than AFAP or MAP patients). However, the most premature case in our series was a 7-year-old patient from a family with an *APC* mutation at codon 1017 — a region usually associated

with an intermediate phenotype. A particularly aggressive phenotypic expression of this mutation was observed in this family; several relatives presented a high number of polyps ( $> 1000$ ), an early age of onset, and desmoid tumors. This case demonstrates the importance of considering the family clinical history when planning the surveillance of other family members.

One of the most clinically important discrepancies observed in our study concerns desmoid tumors, which, even if histologically benign, can lead to life-threatening complications through their size and impingement on vital structures. Indeed, desmoid tumors represent the second leading cause of death in FAP patients [38,40] and were identified at high frequency in our cohort, occurring in 57% (8/14) of *APC*-mutated families. Although described as usually associated with mutations after codon 1444 [22], only two out of eight families affected by desmoid tumors in our study harbored mutations after this codon. In concordance with our findings, previous studies with large cohorts have also failed to confirm this association [41], or described different boundaries for the increased risk of these tumors, such as codon 1310 [42] or 1395 [26]. Furthermore, besides the location of *APC* mutations, several other risk factors are suggested to be related with desmoid tumor development, such as surgical trauma [43], pregnancy [44] and especially positive family history for desmoid tumors [41]. In this regard, the last appears to be the most important risk factor for our population, since five of eight families with desmoid tumors presented more than one relative affected by this tumor.

The differences observed in certain phenotypic features between our series and those of others may be because of our relatively small number of FAP and MAP families, may reflect some selection or data collection bias, or may be related to

phenotypic peculiarities in this specific population. In this sense, the majority of FAP genotype-phenotype studies were performed in Europeans cohorts [25-27, 29-32, 35, 39, 41] and the self-declared ethnic origin of most families from our study was also European (Portuguese, Italian and Spanish, mainly) - except from two Japanese and one Arabian families. However, it is important to highlight that Brazilians represent an extremely admixed population, with most individuals presenting some degree of African and Amerindian ancestry [45].

Furthermore, intra and interfamilial variations in the FAP phenotype are also well documented in other populations, and it is likely that modifying genes and environmental factors, as well as functional polymorphisms of the normal *APC* allele, play a crucial role in determining the clinical course of disease [46, 47]. In this regard, the different genetic background and/or environmental factors of our population could be responsible for the phenotypic differences observed in our study. For instance, in our series, desmoid tumors were much more prevalent in *APC*-mutated patients (57%) than in others previous studies (10-15%) [22, 41, 46], indicating that perhaps our set of patients represents a distinct group regarding this extracolonic feature.

Finally, the lack of a clear phenotypic expression of the mutations identified in our study complicates clinical predictions based on knowledge of the mutation site, and as a result, we can make no specific surveillance and management recommendations for the Brazilian population. In order to accomplish that, larger studies need to be carried out. Instead, we recommend that clinical decisions regarding an individual patient should not be based strictly on the genotype, but mainly on the colonic phenotype and family clinical history.

## CONCLUSIONS

In this comprehensive investigation of the *APC* and *MUTYH* mutation spectrum in Brazilian polyposis patients, we identified a high frequency of germline mutations, allowing the identification of several novel pathogenic variants and the proposal of a cost-effective screening approach for this population. Notably, a significant number of *APC* mutation-positive families were not consistent with predicted genotype–phenotype correlations, and this should be taken into consideration for genetic counseling and patient management of our population.

## Figures and Legends

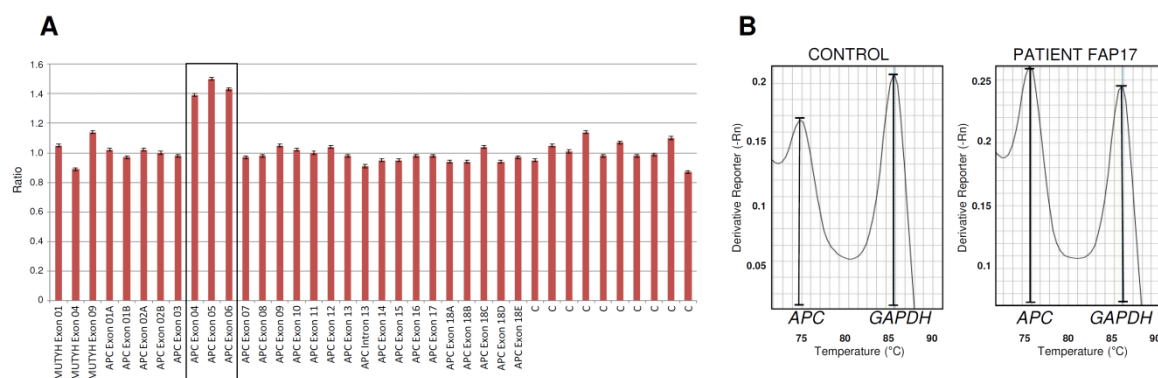


Figure 1: Duplication encompassing exons 1 to 3 of the *APC* gene (patient ID17). A: MLPA graphic showing normalized ratios of probes ordered by genomic position; the box marks the probes that indicate duplication of *APC* exons 1, 2, and 3 (exons 4, 5, and 6 according to MLPA exon numbering [GenBank:NG\_008481]). B: Melt curve of duplex qPCR of *APC* exon 2 and *GAPDH* intron 7 (reference gene). The ratio of *APC*/*GAPDH* peaks of the melting curve was 0.72 in the control sample and 1.06 in patient ID17, leading to a normalized ratio of 1.47 for the FAP patient, which confirms the duplication.

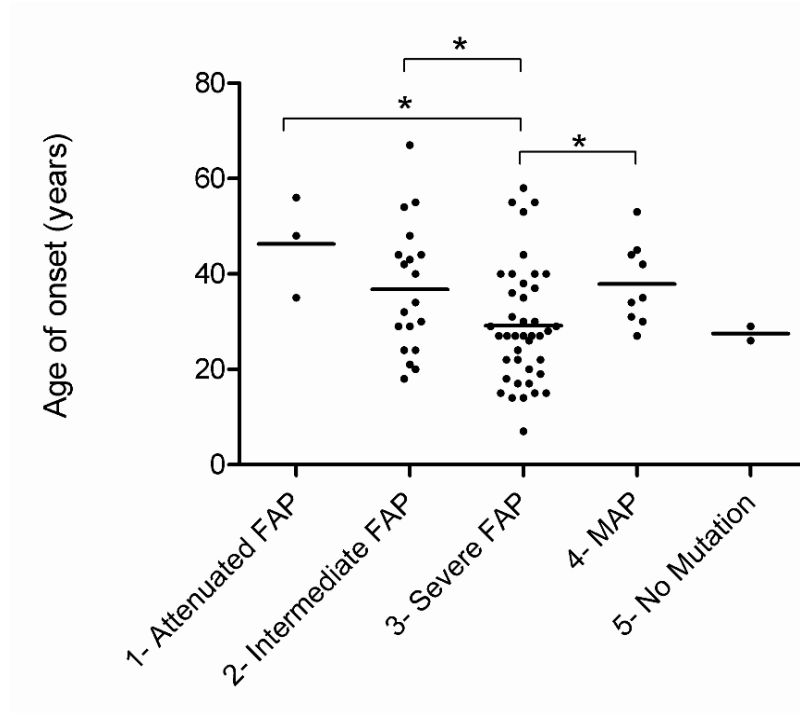


Figure 2: Age of onset per group. The graph shows the distribution and the mean (horizontal line) age of onset for each of the five defined groups. Groups 1, 2, and 4 patients had a significantly later age of onset than group 3 patients ( $t = 2.35$   $p = 0.024$ ;  $t = 2.15$   $p = 0.04$ ;  $t = 2.01$   $p = 0.05$ , respectively).

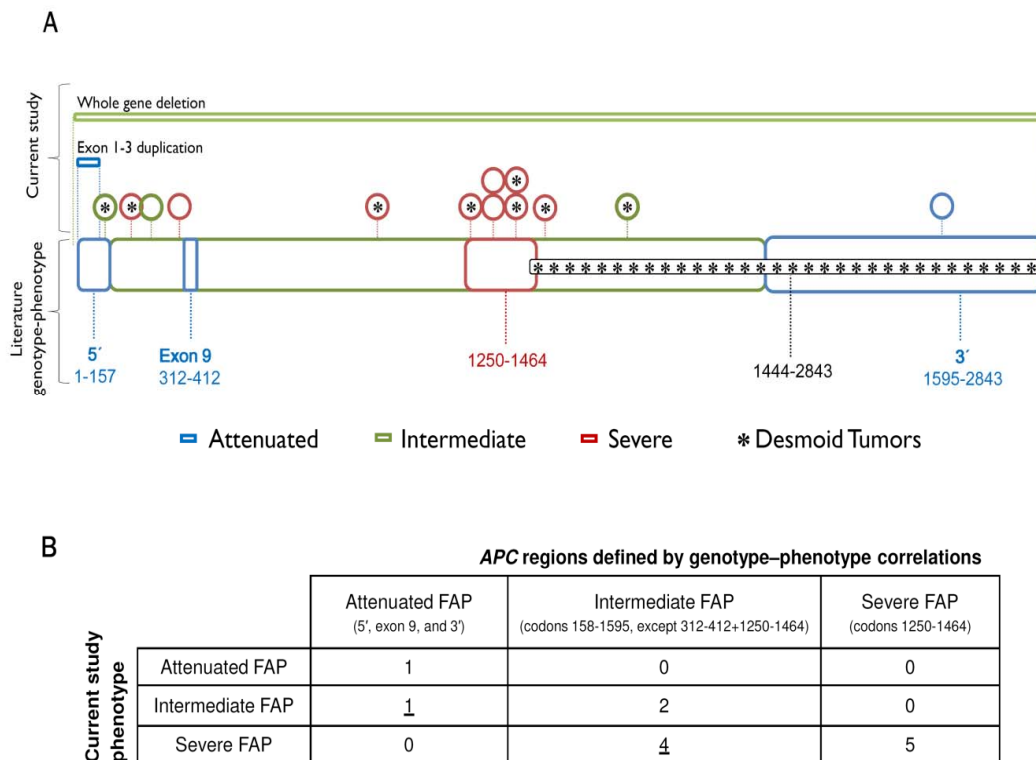


Figure 3: Genotype–phenotype correlation. **A:** Distribution of the polyposis phenotype of the index case and the presence of desmoid tumors along the *APC* gene. Schematic representation of the two large genomic rearrangements (top bars) and the 13 point mutations or small insertions/deletions (circles) identified in this series (including the variant of unknown significance – last circle). The asterisk inside the circles denotes patients with desmoid tumors. The lower, thick bar represents the *APC* regions defined by genotype–phenotype correlations proposed by Nieuwenhuis and Vasen (2007) [22]. Genotype–phenotype correspondence between our results and those previously published is indicated by concordant colors (blue/green/red/black). Numbers represent *APC* codons. **B:** Number of families presenting *APC* point mutations (N = 13) according to the index case polyposis phenotype and the *APC* codon limits, showing that five individuals (underlined numbers) presented a polyposis burden different from that predicted.



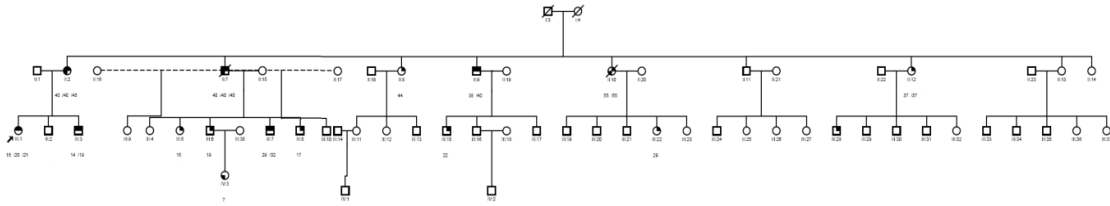


Figure 4: Family tree of ID10 family. This family harbored a truncating mutation at codon 1017 located in a region usually associated with an intermediate FAP phenotype. This mutation displayed an aggressive phenotypic expression: the proband (individual III:1, indicated by an arrow) presented her first polyposis symptoms at the age of 15, a desmoid tumor at age 20, and a thyroid carcinoma at age 21. When available, the ages of onset are presented under each individual. Her brother (III:2) and six first cousins (III:5, 6, 7, 8, 15, and 22) also developed polyps at early ages (14 to 29 years old). The most prematurely affected was a second cousin (individual IV:3), who was diagnosed with polyps at the age of 7. Desmoid tumors were described in another three relatives: two uncles at the age of 40 (II:7 and II:9) and one cousin at 32 years old (III:7).

Table 1: *APC* and *MUTYH* mutation spectrum in Brazilian polyposis patients

ID	Gene	Mutation	Exon	Type	Ref. <sup>a</sup>	reported N times <sup>b</sup>
Mutation-positive patients with > 100 polyps						
02	<i>APC</i>	del 5q21.3-q22.3 (chr5:107916475-113079330 Hg19)	1-15	gene deletion	Torrezan et al [21]	1
04	<i>APC</i>	c.856_859dupCATG (p.Glu287Alafs*2)	8	duplication	Current study	0
05	<i>APC</i>	c.447dupC (p.Lys150Glnfs*18)	4	duplication	Current study	0
23	<i>APC</i>	c.4097dupC (p.Gln1367Serfs*8)	15	duplication	Current study	0
01	<i>APC</i>	c.904C>T (p.Arg302*)	8	nonsense	Mandl et al [48]	22
03	<i>APC</i>	c.4348C>T (p.Arg1450*)	15	nonsense	Miyaki et al [49]	40
06	<i>APC</i>	c.3880-3881delCA (p.Gln1294Glyfs*6)	15	deletion	Miyaki et al [49]	1
07	<i>APC</i>	c.847C>T (p.Arg283*)	8	nonsense	Mandl et al [48]	49
08	<i>APC</i>	c.4099C>T (p.Gln1367*)	15	nonsense	Friedl and Aretz [50]	8
10	<i>APC</i>	c.3050-3053delATGA (p.Asn1017Metfs*4)	15	deletion	Vandrovcová et al [28]	2
11	<i>APC</i>	c.3927-3931delAAAGA (p.Glu1309Aspfs*4)	15	deletion	Miyoshi et al [51]	304
14	<i>APC</i>	c.4393-4394delAG (p.Ser1465Trpfs*3)	15	deletion	Miyoshi et al [51]	40
21	<i>APC</i>	c.3927-3931delAAAGA (p.Glu1309Aspfs*4)	15	deletion	Miyoshi et al [51]	304
24	<i>MUTYH</i>	c.[1187G>C];[=] (p.[Gln396Asp];[=])	13	missense	Al-Tassan et al [4]	532
Mutation-positive patients with < 100 polyps						
15	<i>MUTYH</i>	c.[348+33_*64+146del4285insTA]; [348+33_*64+146del4285insTA]	4-16	large deletion	Torrezan et al [23]	2
13	<i>APC</i>	c.5365G>C (p.Val1789Leu) <sup>c</sup>	15	missense	Current study	0
17	<i>APC</i>	Exon 1-3 duplication	1-3	large duplication	Current study	0
16	<i>MUTYH</i>	c.[536A>G];[1147delC] (p.[Tyr179Cys];[Ala385Profs*23])	7;12	missense; deletion	Al-Tassan et al; [4] Eliason et al [52]	532; 71
18	<i>MUTYH</i>	c.[389-1G>C];[536A>G] (p.[Val130GlnfsX98;p.(spl?)];[Tyr179Cys])	i4;7	splice site; missense	Olschwang et al; [53] Al-Tassan et al [4]	2; 532
19	<i>MUTYH</i>	c.[721C>T];[721C>T] (p.[Arg241Trp];[Arg241Trp])	9	missense	Fleischmann et al [54]	11
447	<i>MUTYH</i>	c.[536A>G];[1227-1228dup] (p.[Tyr179Cys];[Glu410Glyfs*43])	7;13	missense; duplication	Al-Tassan et al [4]; Baglioni et al [55]	506; 31

<sup>a</sup> The following databases were consulted: LOVD: <http://www.lovd.nl/2.0/>; dbSNP: <http://www.ncbi.nlm.nih.gov/projects/SNP/>; and 1000 Genomes: <http://www.1000genomes.org/>;

<sup>b</sup> Number of times reported at LOVD database; <sup>c</sup> Variant of unknown clinical significance.

Table 2: Clinical features of the 23 Brazilian polyposis families.

Family ID	Affected codon	Polyposis burden <sup>a</sup>	N <sup>b</sup>	Age of onset		Extracolonic Manifestations						
				Range	Mean	Gastric polyps	Duodenal polyps	Desmoid tumor(n)	Osteoma	Epidermoid cysts/lipoma	Other tumor sites	
<b>1 - Attenuated FAP</b>	<b>APC</b>			35-56	46.3							
ID13	1789 <sup>c</sup>	attenuated	1	56	56	yes	yes	no	no	NA	Uterus	
ID17	Exon 1-3 dup	attenuated	2	35-48	41.5	yes	No	no	no	NA	Lung	
<b>2 - Intermediate FAP</b>	<b>APC</b>			18-67	35.7							
ID02	Exon 1-15 del	intermediate	5	40-44	42	yes	yes	no	no	NA		
ID03	1450	intermediate	7	21-54	34.2	yes	yes	yes (2)	no	NA	Liver; Hematologic	
ID04	287	intermediate	14	18-44	29.6	no	no	no	no	NA	Bilateral breast; Stomach; BCC	
ID05	150	intermediate	18	29-67	40	yes	yes	yes (3)	no	yes	Breast; Melanoma	
<b>3 - Severe FAP</b>	<b>APC</b>			7-58	29.2							
ID01	302	profuse	5	14-27	23.6	yes	no	no	no	NA	Brain	
ID06	1294	profuse	11	20-35	27.7	yes	yes	yes (1)	yes	yes		
ID07	283	profuse	8	27-58	37	yes	no	yes (2)	yes	yes	Liver	
ID08	1367	profuse	3	15-55	35	yes	yes	yes (1)	yes	NA		
ID10	1017	profuse	15	7-55	28.1	no	no	yes (4)	yes	yes	Thyroid	
ID11	1309	profuse	3	30-40	35	yes	yes	no	yes	NA		
ID14	1465	profuse	8	17-36	26.5	no	no	yes (2)	no	yes		
ID21	1309	profuse	1	18	18	yes	yes	no	yes	no		
ID23	1367	profuse	1	22	22	NA	NA	yes(1)	NA	NA		
<b>4 - MAP</b>	<b>MUTYH</b>			27-53	37.9							
ID15	Exon 4-16 del	attenuated	2	42-44	43	no	no	no	no	NA		
ID16	179;385	attenuated	2	30-53	41.5	yes	no	no	no	NA	Uterus	
ID18	130;179	attenuated	1	45	45	no	no	no	no	NA		
ID19	241;241	attenuated	2	27-31	29	no	no	no	no	NA		
ID24	396;[=]	intermediate	1	55	55	NA	NA	no	NA	NA		
ID447	179;410	attenuated	2	34-35	34.5	no	no	no	no	no		
<b>5 - No Mutation</b>				26-29	27.5							
ID20		attenuated	1	26	26	no	no	no	yes	no		
ID22		intermediate	1	29	29	no	no	no	no	no		

<sup>a</sup> Polyposis burden of the index case. The remaining columns are based on all affective relatives with available clinical information; <sup>b</sup> N = number of clinically affected family members; BCC = basal cell carcinoma; NA = data not available; <sup>c</sup> Variant of unknown clinical significance.

## **CONFLICT OF INTEREST STATEMENT**

The authors declare that they have no competing interests.

## **ACKNOWLEDGMENTS**

The authors acknowledge the patients and families for their collaboration, the staff of the Nucleic Acid Bank of A. C. Camargo Hospital (São Paulo, Brazil) for sample processing, Israel Gomy (CIPE - A. C. Camargo Hospital, São Paulo, Brazil) for patient referral and Bianca C. G. Lisboa (Molecular Diagnostic Lab - A. C. Camargo Hospital, São Paulo, Brazil) for guidance in establishing the sequencing protocols.

## **FUNDING**

This work was supported by the National Institute of Science and Technology in Oncogenomics (INCITO) (FAPESP: 2008/57887-9) and the European Union FP7 Genetic study of Common Hereditary Bowel Cancers in Hispania and the Americas (CHIBCHA) project.

## **AUTHORS' CONTRIBUTIONS**

GTT, DMC, and BMR conceived and designed experiments. GTT, FCCS, ACVK, and DMC performed and analyzed experiments. EMMS, MIA, SAJ, and BMR assessed clinical data and selected patients. BMR and DMC contributed reagents/materials/analysis tools. GTT, FCCS, ACVK, BMR, and DMC wrote/edited the paper. All authors have read and approved the final manuscript.

## REFERENCES

- 1 Gardner EJ, Burt RW, Freston JW. **Gastrointestinal Polyposis: Syndromes and Genetic Mechanisms.** *The Western Journal of Medicine* 1980, **132**:488-99.
- 2 Leppert M, Dobbs M, Scambler P, O'Connell P, Nakamura Y, Stauffer D, Woodward S, Burt R, Hughes J, Gardner E. **The gene for familial polyposis coli maps to the long arm of chromosome 5.** *Science* 1987, **238**:1411-3.
- 3 Knudsen AL, Bisgaard ML, Bülow S. **Attenuated familial adenomatous polyposis (AFAP). A review of the literature.** *Familial Cancer* 2003, **2**:43-55.
- 4 Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR, Cheadle JP. **Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors.** *Nature Genetics* 2002, **30**:227-32.
- 5 Kinzler KW, Nilbert MC, Su LK, Vogelstein B, Bryan TM, Levy DB, Smith KJ, Preisinger AC, Hedge P, McKechnie D, Finniea R, Matrkam A, Groffen J, Boguski MS, Altschul SF, Horii A, Ando H, Miyoshi Y, Miki Y, Nishisho I, Nakamura Y. **Identification of FAP locus genes from chromosome 5q21.** *Science* 1991, **253**:661-5.
- 6 Nishisho I, Nakamura Y, Miyoshi Y, Miki Y, Ando H, Horii A, Koyama K, Utsunomiya J, Baba S, Hedge P. **Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients.** *Science* 1991, **253**:665-9.
- 7 Soravia C, Berk T, Madlensky L, Mitri A, Cheng H, Gallinger S, Cohen Z, Bapat B. **Genotype-phenotype correlations in attenuated adenomatous polyposis coli.** *American Journal of Human Genetics* 1998, **62**:1290-301.
- 8 Young J, Simms L a, Tarish J, Buttenshaw R, Knight N, Anderson GJ, Bell A, Leggett B. **A family with attenuated familial adenomatous polyposis due to a mutation in the alternatively spliced region of APC exon 9.** *Human Mutation* 1998, **11**:450-5.
- 9 Davidson S, Leshanski L, Rennert G, Eidelman S, Amikam D. **Maternal mosaicism for a second mutational event--a novel deletion--in a familial adenomatous polyposis family harboring a new germ-line mutation in**

- the alternatively spliced-exon 9 region of APC. *Human Mutation* 2002, **19**:83-4.
- 10 Venesio T, Balsamo A, Agostino VGD, Ranzani GN. **MUTYH-associated polyposis (MAP), the syndrome implicating base excision repair in inherited predisposition to colorectal tumors.** *Frontiers in Oncology* 2012, **2(83)**:1-9.
- 11 Out AA, Tops CMJ, Nielsen ÅM, Weiss MM, van Minderhout IJ, Fokkema IF, Buisine MP, Claes K, Colas C, Fodde R, Fostira F, Franken PF, Gaustadnes M, Heinimann K, Hodgson SV, Hogervorst FB, Holinski-Feder E, Lagerstedt-Robinson K, Olschwang S, van den Ouweland AM, Redeker EJ, Scott RJ, Vankeirsbilck B, Grønlund RV, Wijnen JT, Wikman FP, Aretz S, Sampson JR, Devilee P, den Dunnen JT, Hes FJ. **Leiden Open Variation Database of the MUTYH Gene Human Mutation.** *Human Mutation* 2010, **31**:1205-1215.
- 12 Aretz S, Genuardi M, Hes FJ. **Clinical utility gene card for: MUTYH-associated polyposis (MAP), Autosomal recessive colorectal adenomatous polyposis , Multiple colorectal adenomas , Multiple adenomatous polyps (MAP) - update 2012.** *European Journal of Human Genetics* 2012, **8**:1-4.
- 13 Sampson JR, Jones S, Dolwani S, Cheadle JP. **MutYH (MYH) and colorectal cancer.** *Biochemical Society Transactions* 2005, **33**:679-683.
- 14 Sieber OM, Lipton L, Crabtree M, Heinimann K, Fidalgo P, Phillips RK, Bisgaard ML, Orntoft TF, Aaltonen LA, Hodgson SV, Thomas HJ, Tomlinson IP. **Multiple Colorectal Adenomas, Classic Adenomatous Polyposis and Germ-Line Mutations in MYH.** *The New England Journal of Medicine* 2003, **9**:791-799.
- 15 Filipe B, Baltazar C, Albuquerque C, Fragoso S, Lage P, Vitoriano I, Mão de Ferro S, Claro I, Rodrigues P, Fidalgo P, Chaves P, Cravo M, Nobre Leitão C. **APC or MUTYH mutations account for the majority of clinically well-characterized families with FAP and AFAP phenotype and patients with more than 30 adenomas.** *Clinical Genetics* 2009, **76**:242-255.

- 16 Casper, M., Plotz, G., Juengling, B., Zeuzem, S., Lammert, F., Raedle, J. **MUTYH hotspot mutations in unselected colonoscopy patients.** *Colorectal Disease*, 2012 **14**:238–244.
- 17 Nielsen M, Morreau H, Vasen HFA, Hes FJ. **MUTYH-associated polyposis (MAP).** *Critical Reviews in Oncology / Hematology* 2010:1-16.
- 18 Sampson JR, Dolwani S, Jones S, Eccles D, Ellis A, Evans DG, Frayling I, Jordan S, Maher ER, Mak T, Maynard J, Pigatto F, Shaw J, Cheadle JP. **Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH.** *Lancet* 2003, **362**:39-41.
- 19 Santos EMM, Ferreira FO, Aguiar S, Nakagawa WT, Lopes A, Medeiros A, Rossi BM. **Familial Adenomatous Polyposis: Data from the Hereditary Colorectal Cancer Registry (HCCR).** *Applied Cancer Research* 2008; **28(1)**:17-23.
- 20 Newton CR, Graham A, Heptinstal LE, Powell SJ, Summers C, Kalsheker N, Smith JC, Markham AF. **Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS).** *Nucleic Acids Research* 1989, **17**:2503-2516.
- 21 Torrezan GT, Silva FC da, Krepischi ACV, dos Santos EM, Rossi BM, Carraro DM. **A novel SYBR-based duplex qPCR for the detection of gene dosage: detection of an APC large deletion in a familial adenomatous polyposis patient with an unusual phenotype.** *BMC Medical Genetics* 2012, **13(55)**.
- 22 Nieuwenhuis MH, Vasen HFA. **Correlations between mutation site in APC and phenotype of familial adenomatous polyposis ( FAP ): A review of the literature.** *Critical Reviews in Oncology/Hematology* 2007, **61**:153-161.
- 23 Torrezan GT, Silva FC da, Krepischi ACV, Santos ÉM, Ferreira Fde O, Rossi BM, Carraro DM. **Breakpoint characterization of a novel large intragenic deletion of MUTYH detected in a MAP patient : Case report.** *BMC Medical Genetics* 2011, **12(128)**.
- 24 Rouleau E, Zattara H, Lefol C, Noguchi T, Briaux A, Buecher B, Bourdon V, Sobol H, Lidereau R, Olschwang S. **First large rearrangement in the**

- MUTYH gene and attenuated familial adenomatous polyposis syndrome.** *Clinical Genetics* 2011:301-303.
- 25 Bunyan DJ, Shea-Simonds J, Reck C, Finnis D, Eccles DM. **Genotype-phenotype correlations of new causative APC gene mutations in patients with familial adenomatous polyposis.** *Journal of Medical Genetics* 1995, **32**:728-31.
- 26 Wallis YL, Morton DG, McKeown CM, Macdonald F. **Molecular analysis of the APC gene in 205 families: extended genotype-phenotype correlations in FAP and evidence for the role of APC amino acid changes in colorectal cancer predisposition.** *Journal of Medical Genetics* 1999, **36**:14-20.
- 27 Groves C, Lamlum H, Crabtree M, Williamson J, Taylor C, Bass S, Cuthbert-Heavens D, Hodgson S, Phillips R, Tomlinson I. **Mutation cluster region, association between germline and somatic mutations and genotype-phenotype correlation in upper gastrointestinal familial adenomatous polyposis.** *The American Journal of Pathology* 2002, **160**:2055-2061.
- 28 Vandrovcová J, Štekrová J, Kebrdlová V, Kohoutová M. **Molecular Analysis of the APC and MYH Genes in Czech Families Affected by FAP or Multiple Adenomas : 13 Novel Mutations.** *Human Mutation* 2004, **23**:397-404.
- 29 Gómez-fernández N, Castellví-bel S, Fernández-rozadilla C, Balaguer F, Muñoz J, Madrigal I, Milà M, Graña B, Vega A, Castells A, Carracedo A, Ruiz-Ponte C. **Molecular analysis of the APC and MUTYH genes in Galician and Catalanian FAP families: a different spectrum of mutations ?** *BMC Medical Genetics* 2009, **10**:1-12.
- 30 Fostira F, Thodi G, Sandaltzopoulos R, Fountzilas G, Yannoukakos D. **Mutational spectrum of APC and genotype-phenotype correlations in Greek FAP patients.** *Molecular Biology* 2010, **10**.
- 31 Rivera B, Gonzalez S, Sánchez-Tomé E, Blanco I, Mercadillo F, Letón R, Benítez J, Robledo M, Capellá G, Urioste M. **Clinical and genetic characterization of classical forms of familial adenomatous polyposis : a Spanish population study.** *Annals of Oncology* 2011:903-909.



- 32 Kanter-Smoler G, Fritzell K, Rohlin A, Engwall Y, Hallberg B, Bergman A, Meuller J, Grönberg H, Karlsson P, Björk J, Nordling M. **Clinical characterization and the mutation spectrum in Swedish adenomatous polyposis families.** *BMC Medicine* 2008, **6**:1-14.
- 33 Pitroski CE, Cossio SL, Koehler-Santos P, Graudenz M, Prolla JC, Ashton-Prolla P. **Frequency of the common germline MUTYH mutations p.G396D and p.Y179C in patients diagnosed with colorectal cancer in Southern Brazil.** *International Journal Of Colorectal Disease* 2011.
- 34 Spier I, Horpaopan S, Vogt S, Uhlhaas S, Morak M, Stienen D, Draaken M, Ludwig M, Holinski-Feder E, Nöthen MM, Hoffmann P, Aretz S. **Deep Intronic APC Mutations Explain a Substantial Proportion of Patients with Familial or Early-Onset.** *Human Mutation* 2012, **33**:1-6.
- 35 Newton K, Mallinson E, Bowen J, Lalloo F, Clancy T, Hill J, Evans DG. **Genotype – phenotype correlation in colorectal polyposis.** *Clinical Genetics* 2011:1-11.
- 36 Vasen HF, Luijt RB van der, Slors JF, Buskens E, de Ruyter P, Baeten CG, Schouten WR, Oostvogel HJ, Kuijpers JH, Tops CM, Meera Khan P. **Molecular genetic tests as a guide to surgical management of familial adenomatous polyposis.** *Lancet* 1996, **348**:433-435.
- 37 Wu JS, Paul P, McGannon E, Church JM. **APC genotype, polyp number, and surgical options in familial adenomatous polyposis.** *Annals of Surgery* 1998, **227**:57-62.
- 38 Vasen HFA, Möslein G, Alonso A, Aretz S, Bernstein I, Bertario L, Blanco I, Bülow S, Burn J, Capella G, Colas C, Engel C, Frayling I, Friedl W, Hes FJ, Hodgson S, Järvinen H, Mecklin JP, Møller P, Myrhøi T, Nagengast FM, Parc Y, Phillips R, Clark SK, de Leon MP, Renkonen-Sinisalo L, Sampson JR, Stormorken A, Tejpar S, Thomas HJ, Wijnen J. **Guidelines for the clinical management of familial adenomatous polyposis ( FAP ).** *Gut* 2008, **57**:704-713.
- 39 Friedl W, Caspari R, Sengteller M, Uhlhaas S, Lamberti C, Jungck M, Kadmon M, Wolf M, Fahnenstich J, Gebert J, Möslein G, Mangold E, Propping P. **Can APC mutation analysis contribute to therapeutic**

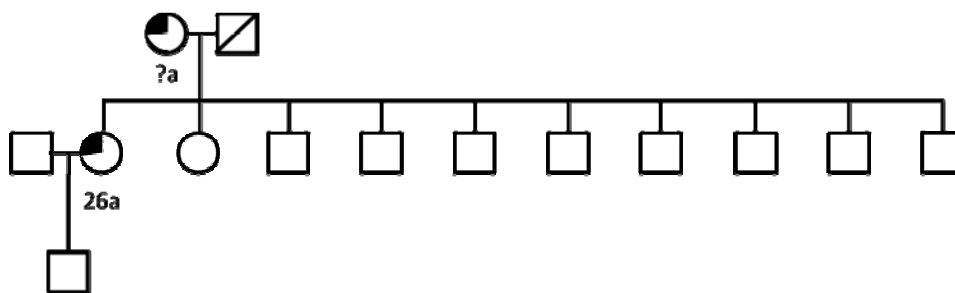
- decisions in familial adenomatous polyposis? Experience from 680 FAP families.** *Gut* 2001, **48**:515-521.
- 40 Half E, Bercovich D, Rozen P. **Familial adenomatous polyposis.** *Orphanet Journal of Rare Diseases* 2009, **4(22)**1-23.
- 41 Nieuwenhuis MH, De Vos Tot Nederveen Cappel W, Botma A, Nagengast FM, Kleibeuker JH, Mathus-Vliegen EM, Dekker E, Dees J, Wijnen J, Vasen HF. **Desmoid tumors in a dutch cohort of patients with familial adenomatous polyposis.** *Clin Gastroenterol Hepatol* 2008, **6(2)**:215-219.
- 42 Bertario L, Russo A, Sala P, Eboli M, Giarola M, D'amico F, Gismondi V, Varesco L, Pierotti MA, Radice P. **Genotype and phenotype factors as determinants of desmoid tumors in patients with familial adenomatous polyposis.** *Int J Cancer* 2001, **95(2)**:102-107.
- 43 Gurbuz AK, Giardiello FM, Petersen GM, Krush A J, Offerhaus G J, Booker S V, Kerr M C, Hamilton S R. **Desmoid tumours in familial adenomatous polyposis.** *Gut* 1994, **35**:377–381.
- 44 Clark SK, Neale KF, Landgrebe JC, Phillips RK. **Desmoid tumours complicating familial adenomatous polyposis.** *Br J Surg* 1999, **86**:1185–1189.
- 45 Pena SDJ, Bastos-Rodrigues L, Pimenta JR, Bydlowski SP. **DNA tests probe the genomic ancestry of Brazilians.** *Braz J Med Biol Res* 2009, **42(10)**: 870-876.
- 46 Sturt NJ, Gallagher MC, Bassett P, Philp CR, Neale KF, Tomlinson IP, Silver AR, Phillips RK. **Evidence for genetic predisposition to desmoid tumours in familial adenomatous polyposis independent of the germline APC mutation.** *Gut* 2004, **53(12)**:1832-1836.
- 47 Crabtree MD, Tomlinson IP, Hodgson SV, Neale K, Phillips RK, Houlston RS. **Explaining variation in familial adenomatous polyposis: relationship between genotype and phenotype and evidence for modifier genes.** *Gut* 2002, **51(3)**:420-3.
- 48 Mandl M, Paffenholz R, Friedl W, Caspari R, Sengteller M, Propping P. **Frequency of common and novel inactivating APC mutations in 202**

- families with familial adenomatous polyposis.** *Human Molecular Genetics* 1994, **3(1)**:181-184.
- 49 Miyaki M, Konishi M, Kikuchi-Yanoshita R, Enomoto M, Igari T, Tanaka K, Muraoka M, Takahashi H, Amada Y, Fukayama M, Maeda Y, Iwama T, Mishima Y, Mori T, Koike M. **Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors.** *Cancer Research* 1994, **54(11)**:3011-20.
- 50 Friedl W, Aretz S. **Familial adenomatous polyposis: experience from a study of 1164 unrelated german polyposis patients.** *Hered Cancer Clin Pract* 2005, **3(3)**:95-114.
- 51 Miyoshi Y, Ando H, Nagase H, Nishisho I, Horii A, Miki Y, Mori T, Utsunomiya J, Baba S, Petersen G, Hamilton SR, Kinzler KW, Vogelstein B and Nakamura Y. **Germ-line mutations of the APC gene in 53 familial adenomatous polyposis patients.** *Proc Natl Acad Sci* 1992, **89**:4452-4456.
- 52 Eliason K, Hendrickson BC, Judkins T, Norton M, Leclair B, Lyon E, Ward B, Noll W, Scholl T. **The potential for increased clinical sensitivity in genetic testing for polyposis colorectal cancer through the analysis of MYH mutations in North American patients.** *Journal Medical Genetics* 2005, **42(1)**:95-6.
- 53 Olschwang S, Blanché H, de Moncuit C, Thomas G. **Similar colorectal cancer risk in patients with monoallelic and biallelic mutations in the MYH gene identified in a population with adenomatous polyposis.** *Genet Test* 2007, **11(3)**:315-20.
- 54 Fleischmann C, Peto J, Cheadle J, Shah B, Sampson J, Houlston RS **Comprehensive analysis of the contribution of germline MYH variation to early-onset colorectal cancer.** *International Journal of Cancer* 2004, **109(4)**:554-8.
- 55 Baglioni S, Melean G, Gensini F, Santucci M, Scatizzi M, Papi L, Genuardi M. **A kindred with MYH-associated polyposis and pilomatricomas.** *American Journal of Medical Genetics part A* 2005, **134A(2)**:212-4.

## **7 ARTIGO 4: IDENTIFICAÇÃO DE NOVOS GENES DE SUSCEPTIBILIDADE À POLIPOSE INTESTINAL POR SEQUENCIAMENTO DE EXOMA (EM PREPARAÇÃO).**

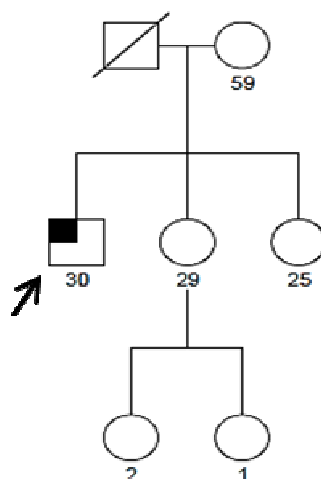
Após a completa investigação de mutações nos genes *APC* e *MUTYH* nos pacientes do estudo, dois probandos (ID20 e ID22) não apresentaram mutações em nenhum dos genes avaliados e foram investigados para mutações em outros genes através do sequenciamento completo do exoma.

Em relação às características clínicas deste probandos, um dos pacientes (ID20) era portador de polipose atenuada (menos que 100 pólipos), com manifestação aos 26 anos de idade. Nenhum dos nove irmãos dessa paciente realizou rastreamento por colonoscopia, não sendo possível saber se estes indivíduos são portadores de polipose. Segundo relato pessoal da paciente, sua mãe retirou parte do intestino recentemente (colectomia parcial), porém, a paciente não soube confirmar o fenótipo de polipose da mãe (Figura 6).



**Figura 4** - Heredograma família ID20. A probanda ID20, indicado pela seta, era portadora de polipose atenuada, manifestada aos 26 anos. Os nove irmãos e o filho da probanda nunca realizaram exame colonoscópico, e seus fenótipos são desconhecidos. Sua mãe realizou uma colectomia parcial recentemente (idade desconhecida), porém, a paciente não soube confirmar o fenótipo de polipose.

Já o segundo paciente (ID22) apresentou polipose clássica intermediária (entre 100 e 1000 pólipos) aos 30 anos de idade. Este indivíduo é o único afetado da família, sendo que seus parentes próximos realizaram o exame colonoscópico e não apresentam pólipos (Figura 7).



**Figura 5** - Heredograma família ID22. O probando ID22, indicado pela seta, era portador de polipose clássica intermediária, manifestada aos 30 anos. A mãe e as duas irmãs do probando realizaram exame colonoscópico e não apresentam polipose. O pai faleceu de diabetes.

O detalhamento dos métodos utilizados para o sequenciamento do exoma está descrito no Anexo 1. Em resumo, para o enriquecimento das porções exônicas foi utilizado o kit de captura direcionada para todo o exoma humano *TargetSeq™ Exome Enrichment Kit* (LifeTechnologies). Esta abordagem realiza a captura de 45,1 Mb do genoma, sendo 37,3 Mb correspondentes a regiões codificadoras, e representando mais de 98% das sequencias codificantes descritas nos banco de dados VEGA (<http://www.sanger.ac.uk/resources/databases/vega/>) e RefSeq ([www.ncbi.nlm.nih.gov/RefSeq/](http://www.ncbi.nlm.nih.gov/RefSeq/)).

Após o preparo da biblioteca, as amostras foram então submetidas ao sequenciamento de alto desempenho na plataforma SOLiD™ 5500xl System (LifeTechnologies). Foi utilizada a estratégia de sequenciamento *paired-end*, com uma leitura de 75 pares de bases a partir de uma extremidade e 50 pares de bases da outra extremidade dos fragmentos capturados. Mais de 55 milhões de leituras (*reads*) foram geradas para cada paciente, resultando em aproximadamente 70% das bases alvo cobertas mais de 20 vezes (Tabela 1).

**Tabela 1** - Cobertura do sequenciamento de exoma

<b>Amostra</b>	<b>Nº Reads</b>	<b>Reads on target (%)</b>	<b>Target bp coverage ≥20X</b>
ID20	55.014.151	45,29	68,61
ID22	58.820.164	56,47	74,08

\*bp: *base pair*

As variantes identificadas foram inicialmente filtradas e selecionadas para avaliação de acordo com os seguintes critérios: variantes novas (ausentes no banco de dados dbSNP), com possível impacto na função da proteína (*missense*, *nonsense*, *splice site* e *indels*), mais de 5 *reads* reportando a base variante e base variante

presente em pelo menos 20% do total de *reads*. Com estes critérios, mais de 300 variações de nucleotídeos únicos (SNVs) presentes em regiões codificantes foram identificadas: 10 *nonsense*, 4 *splice site*, e 302 *missense*. Foram ainda identificadas 10 novas indels (alteração de inserção ou deleção) em regiões codificantes (Tabela 2).

**Tabela 2** - Número e tipo das variantes identificadas por sequenciamento de exoma

Amostra	SNVs			INDELS
	<i>Nonsense</i>	<i>Splice site</i>	<i>Missense</i>	-
ID20	9	4	217	6
ID22	9	4	227	8
Comuns	8	4	144	4
Total	10	4	302	10

SNVs: variações de nucleotídeos únicos; Indels: alteração de inserção ou deleção.

Inicialmente, as variantes selecionadas foram avaliadas quanto ao gene em que foram identificadas, buscando-se por alterações que afetem genes relacionados ao câncer colorretal hereditário. Para identificar candidatos mais robustos, elaboramos um banco de dados contendo genes relacionados ao câncer colorretal hereditário que foram identificados através de busca na literatura. Foram identificados 36 genes comprovadamente associados ao câncer colorretal hereditário, além de mais de 50 genes descritos recentemente como possíveis genes de susceptibilidade em trabalhos de sequenciamento de exoma (OMIN - <http://omim.org/>; ARETZ et al. 2013; DERYCKE et al. 2013; GYLFE et al. 2013; PALLES et al. 2013). O Quadro 1 apresenta a lista dos genes selecionados como possivelmente relacionados ao câncer colorretal hereditário.

**Quadro 1** - Genes relacionados ao câncer colorretal hereditário segundo dados de literatura

Genes associados ao CCR hereditário*	Genes possivelmente associados ao CCR hereditário				
	Exoma de CCR sem polipose <sup>#</sup>				Exoma de polipose <sup>§</sup>
<i>AKT1</i>	<i>MLH1</i>	<i>ACHE</i>	<i>HLA-DRB6</i>	<i>NCAPG2</i>	<i>DSC2</i>
<i>APC</i>	<i>MLH2</i>	<i>ADAM6</i>	<i>HSP90AB2P</i>	<i>NUDT7</i>	<i>HEATR5A</i>
<i>AXIN1</i>	<i>MLH3</i>	<i>AKR1C4</i>	<i>KCNJ12</i>	<i>PCDHB17</i>	<i>PDE4DIP</i>
<i>AXIN2</i>	<i>MSH2</i>	<i>ANKRD30BL</i>	<i>KIF23</i>	<i>POLA1</i>	<u><i>POLD1</i></u>
<i>BCR</i>	<i>MSH3</i>	<i>AQP7P1</i>	<i>KIR2DL</i>	<i>PRADC1</i>	<u><i>POLE</i></u>
<i>BLM</i>	<i>MSH6</i>	<i>ARSD</i>	<i>KIR2DS4</i>	<i>PRSS37</i>	<i>ZSWIM7</i>
<i>BMPR1A</i>	<i>MUTYH</i>	<i>BAGE</i>	<i>KIR3DL3</i>	<i>PSPH</i>	
<i>BRCA1</i>	<i>MYH11</i>	<i>BCL8</i>	<i>KRT16P2</i>	<i>RACGAP1P</i>	
<i>BRCA2</i>	<i>PMS1</i>	<i>C6orf10</i>	<i>KRTAP5-4</i>	<i>ROCK1P1</i>	
<i>CDH1</i>	<i>PMS2</i>	<i>CCDC18</i>	<i>LOC642846</i>	<i>RPGR</i>	
<i>CDH18</i>	<i>PTCH1</i>	<i>CDC27</i>	<i>MED12</i>	<i>SFXN4</i>	
<i>CHEK2</i>	<i>PTEN</i>	<i>CENPE</i>	<i>MRPL3</i>	<i>SIGLEC16</i>	
<i>EPCAM</i>	<i>SMAD4</i>	<i>CFTR</i>	<i>MST1P2</i>	<i>TBC1D3P2</i>	
<i>FLCN</i>	<i>SMAD7</i>	<i>CROCCP2</i>	<i>MST1P9</i>	<i>TTY23</i>	
<i>GALNT12</i>	<i>STK11</i>	<i>CTBP2</i>	<i>MUC12</i>	<i>TWSG1</i>	
<i>GREM1</i>	<i>TGFB1</i>	<i>FAM104B</i>	<i>MUC3A</i>	<i>UACA</i>	
<i>LKB1</i>	<i>TGFBR2</i>	<i>FANK1</i>	<i>NBPF1</i>	<i>WASH2P</i>	
<i>MCC</i>	<i>TP53</i>	<i>FRG1B</i>	<i>NBPF10</i>	<i>ZNF490</i>	
		<i>HLA-DRB1</i>	<i>NBPF12</i>	<i>ZNF717</i>	
		<i>HLA-DRB5</i>	<i>NBPF9</i>		

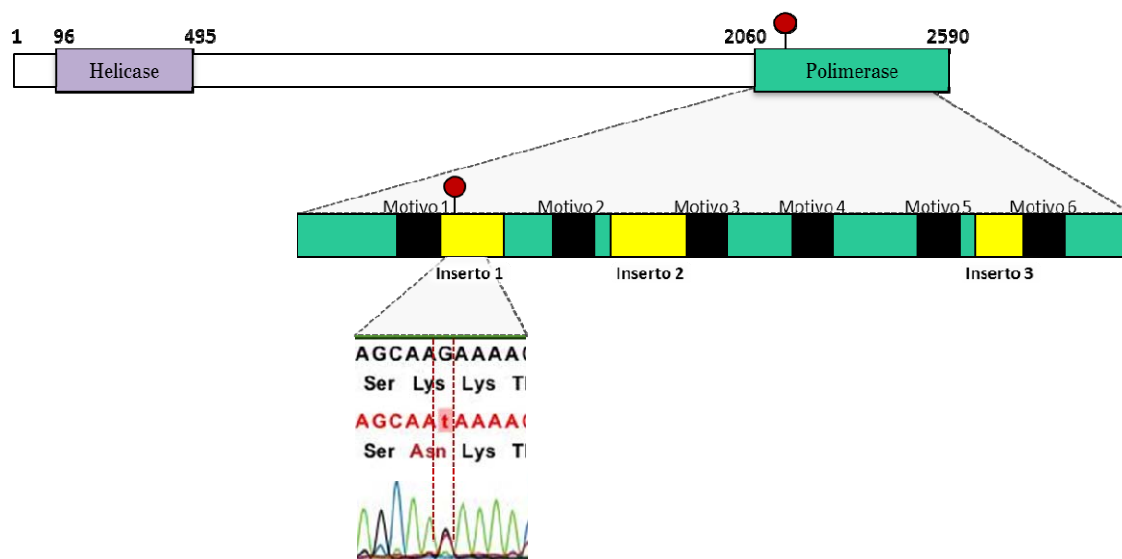
\* Genes selecionados segundo OMIN - <http://omim.org/>; DERYCKE et al. 2013; GYLFE et al. 2013;  
 # Genes identificados como candidatos nos trabalhos de DERYCKE et al. 2013 e GYLFE et al. 2013;  
 § Genes identificados como candidatos nos trabalhos de PALLES et al. 2013 e ARETZ et al. 2013.  
 Os genes sublinhados (*POLD1* e *POLE*) representam genes identificados através de sequenciamento de exoma de pacientes com polipose e para os quais extensa validação já foi realizada, sendo considerados novos genes de susceptibilidade para CCR (PALLES et al. 2013).

Não foi identificada nenhuma alteração nos genes já relacionados ao CCR hereditário descritos no Quadro 1 em nenhum dos probandos. Uma segunda busca foi feita avaliando as vias ou famílias gênicas mais afetadas no CCR hereditário. Nesse sentido, uma das famílias que avaliamos foi a das polimerases, visto que recentemente foram identificadas duas polimerases como sendo novos genes de alta penetrância para predisposição de adenomas e câncer colorretal, as polimerases *POLD1* e *POLE* (PALLES et al. 2013).

Interessantemente, nós identificamos e validamos por sequenciamento capilar no probando ID22 uma alteração *missense* no gene da DNA polimerase *Theta*



(*POLQ* - c.6465G>T; p.Lys2155Asn). Esta alteração não está descrita nos bancos de dados de variações humanas consultados (dbSNP: <http://www.ncbi.nlm.nih.gov/projects/SNP> e 1000Genomes: <http://www.1000genomes.org/>). Este gene codifica uma polimerase da família A que tem como principal função atuar na manutenção da integridade do genoma, através do reparo de quebras de dupla fita e de excisão de bases do DNA (HOGG et al. 2011; LI et al. 2011). Esta alteração foi identificada no domínio de polimerase, mais especificamente em uma região denominada inserto 1, que está relacionada à eficiência de processamento dessa enzima (HOGG et al. 2011) (Figura 6). Análises *in silico* com o software de predição de patogenicidade Polyphen2 classificaram esta alteração como provavelmente patogênica. A presença desta variante foi avaliada por sequenciamento capilar em uma população controle de 280 indivíduos Brasileiros sem histórico de câncer, sendo que nenhum indivíduo controle apresentou a alteração.



**Figura 6** - Localização da variante de POLQ p.Lys2155Asn identificada no paciente ID22. O primeiro desenho esquemático é uma dos domínios da proteína POLQ. A alteração p.Lys2155Asn (identificada pelo marcador vermelho) está localizada entre no domínio de polimerase, mais especificamente no inserto 1. Números acima do esquema da proteína representam os códons. A primeira região mostrada em maior aumento representa o domínio da polimerase, com seus seis motivos proteicos e três insertos característicos da polimerase POLQ. A segunda região mostrada em maior aumento representa o resultado da validação por sequenciamento capilar desta alteração, mostrando uma troca G>T no DNA do paciente ID22, resultando em uma troca de aminoácido p.Lys2155Asn.

Para a validação de variantes identificadas em genes ainda não relacionados ao CCR, as alterações serão priorizadas de acordo com o seu impacto na função da proteína. Inicialmente focaremos na validação daquelas variantes que causem truncamento da proteína (indels e *nonsense*) e alterações em *splice site* e que estão descritas nas Tabelas 3 e 4. É importante ressaltar que a maioria das alterações *nonsense* e *splice site* está presente nos dois probandos. Visto que estes probandos não são aparentados e, ainda, apresentam um fenótipo clínico distinto, tanto do ponto de vista de extensão da polipose quanto do ponto de vista de herança do fenótipo,

não é esperado que exatamente a mesma alteração seja a causadora da síndrome nestes indivíduos. Além disso, muitas destas alterações comuns aos probandos também foram identificadas em indivíduos de outros projetos do grupo que foram sequenciados na mesma corrida (9 indivíduos), representando mais provavelmente erros de sequenciamento ou de mapeamento (estas alterações estão marcadas com \* nas Tabelas 3 e 4). Assim, priorizaremos para a validação aquelas alterações sejam únicas de um dos probandos.

Posteriormente, serão também selecionadas alterações *missense* que ocorram em genes relacionados ao câncer, selecionados a partir de levantamentos bibliográficos e de painéis gênicos comerciais.

**Tabela 3** - Variantes indels identificadas nos pacientes ID20 e ID22 por sequenciamento de exoma

Crom	Gene	Posição gen	Posição do aa	Tipo	Tamanho (pb)	ID20	ID22	ID20 Cref/Cvar	ID22 Cref/Cvar
chr1	PRG4	186277230	793	del	135	SIM	SIM	12/1	17/15
chr12	AMDHD1	96360224	377	ins	2	SIM	NÃO	25/13	57/0
chr15	THSD4*	72063462	943	del	6	NÃO	SIM	109/0	50/31
chr16	NFAT5*	69726416	802	del	3	NÃO	SIM	21/0	10/11
chr19	ZNF83*	53116933	295	del	84	NÃO	SIM	20/0	7/10
chr3	MUC4	195509346	3035	del	48	SIM	SIM	10/10	2/15
chr3	KIAA0226	197427676	357	del	3	NÃO	SIM	23/0	25/10
chr4	MAML3	140811081	503	ins	3	SIM	SIM	12/7	3/19
chr4	ARAP2	36130306	1163	del	3	SIM	NÃO	18/16	70/0
chr8	ODF1*	103573010	217	del	27	SIM	SIM	0/14	8/19

Crom: cromossomo; Posição gen: Posição genômica; Posição aa: Posição aminoácido; Cref: cobertura base referência; Cvar: cobertura base variante; del: deleção; ins: inserção. \*Alterações também identificadas em indivíduos de outros projetos sequenciados na mesma corrida, representando mais provavelmente erros de sequenciamento ou de mapeamento.

**Tabela 4** - Variantes *nonsense* e *splice site* identificadas nos pacientes ID20 e ID22 por sequenciamento de exoma

Crom	Gene	Posição gen	Base ref	Base var	Tipo	Posição aa	ID20	ID22	ID20 Cref/Cvar	ID22 Cref/Cvar
chr1	F5	169511240	G	A	NS	1030	SIM	NÃO	16/11	40/0
chr1	PDE4B*	66798201	T	G	NS	215	SIM	SIM	55/28	88/24
chr1	LRRIQ3	74648516	C	T	NS	93	SIM	SIM	17/5	27/12
chr10	IPMK	59997567	G	A	NS	67	SIM	SIM	27/17	42/6
chr14	KIAA0586*	58955570	T	A	NS	1177	SIM	SIM	27/11	25/21
chr17	C17orf80*	71238475	G	T	NS	536	SIM	SIM	26/6	20/16
chr19	NLRP13	56419270	G	A	NS	779	NÃO	SIM	21/0	16/12
chr3	CP*	148920006	G	A	NS	411	SIM	SIM	34/20	34/31
chr3	SMC4*	160141380	T	G	NS	729	SIM	SIM	20/11	21/8
chrX	DOCK11*	117700112	C	T	NS	280	SIM	SIM	31/7	94/41
chr3	CLASP2*	33600878	C	G	SS	-	SIM	SIM	16/11	17/9
chr3	CLASP2*	33600879	T	A	SS	-	SIM	SIM	16/11	17/9
chr5	GDNF*	37816239	T	C	SS	-	SIM	SIM	16/10	24/11
chr8	PRKDC*	48869993	T	C	SS	-	SIM	SIM	22/12	39/8

Crom: cromossomo; Posição gen: Posição genômica; Posição aa: Posição aminoácido; Cref: cobertura base referência; Cvar: cobertura base variante; NS: *nonsense*; SS: *splice site*. \*Alterações também identificadas em indivíduos de outros projetos sequenciados na mesma corrida, representando mais provavelmente erros de sequenciamento ou de mapeamento.

## 8 COMENTÁRIOS FINAIS

A Polipose Adenomatosa Familiar (FAP) é uma doença autossômica de alta penetrância marcada pelo surgimento de centenas a milhares de pólipos adenomatosos colorretais durante a segunda ou terceira década de vida. Pacientes com FAP tem uma forte predisposição para início precoce de CCR, assim como outras malignidades. Nos últimos anos tem se dado muita atenção para as formas atenuadas da doença que podem ser ocasionadas por mutações nas extremidades 3', 5' e no éxon 9 do gene *APC* (AFAP) ou por mutações bialélicas no gene *MUTYH* (MAP), apresentando um padrão de herança recessivo.

A identificação das alterações genéticas responsáveis por estas síndromes permitiu o desenvolvimento de testes genéticos, beneficiando uma crescente proporção de casos familiares através de melhores estratégias de rastreamento e prevenção, e poderá também melhorar as opções de tratamento. Além disso, o diagnóstico molecular da FAP clássica/atenuada e MAP tem importantes implicações clínicas no acompanhamento de indivíduos em risco, visto que a polipose associada ao *APC* é uma condição dominante (parentes de primeiro grau de um paciente tem a priori 50% de risco de herdarem a mutação), enquanto que a polipose associada ao *MUTYH* é uma condição recessiva, na qual parentes de primeiro grau de um paciente apresentam risco apenas se os dois pais apresentarem pelo menos uma mutação no gene. Em algumas famílias onde não se encontram mutações em *APC* e *MUTYH*, o defeito genético responsável pelo fenótipo pode estar em regiões regulatórias desses

genes, em outros genes de reparo de excisão de bases, genes envolvidos na via Wnt ou ainda em genes não identificados.

No Brasil, uma das primeiras iniciativas para o estudo de famílias com tumores colorretais hereditários ocorreu no A. C. Camargo Cancer Center, com a implementação do Registro de Câncer Colorretal Hereditário (RCCH). O RCCH foi criado em 1992, com o objetivo de arquivar dados de famílias com predisposição ao CCR hereditário e possibilitar a utilização destes dados na prática clínica (para rastreamento, tratamento e seguimento de indivíduos em risco), na educação (disseminando informações entre pacientes, familiares e profissionais de saúde) e na pesquisa (gerando subsídios por meio de dados confiáveis e atualizados) (Rossi, 2008).

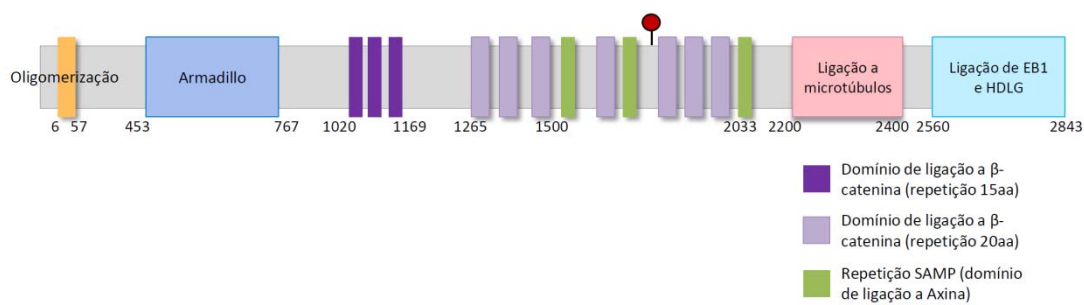
Historicamente, os primeiros pacientes Brasileiros rastreados para mutações em genes de susceptibilidade para CCR foram selecionados a partir deste registro, sendo que a primeira mutação identificada foi no gene *APC* em uma família com FAP (ROSSI et al. 1997) e a segunda em *MSH2* em uma família com síndrome de Lynch (CORVELLO et al. 1999). Atualmente o RCCH tem mais de trezentas famílias cadastradas, e foi a base de dados utilizada para a seleção dos pacientes com polipose para este estudo. Entretanto, a maior parte das famílias cadastradas no RCCH preenche critérios para a Síndrome de Lynch, a mais prevalente síndrome de CCR hereditário. Recentemente, trabalhos com mais de 100 famílias com suspeita para síndrome de Lynch foram desenvolvidos em uma parceria do RCCH com o nosso grupo. Esta parceria possibilitou rastreamento de genes de reparo de DNA (*mismatch repair genes*) associados com a síndrome de forma abrangente, começando pelos genes mais prevalentemente mutados (*MLH1* e *MSH2*) e

investigando também genes com menor prevalência de mutações (*MSH6*, *PMS2*, *TP53* e *CHEK2*) (DOMINGUEZ-VALENTIN et al, 2011; SILVA 2011). A completa caracterização molecular destes pacientes levou a um incremento nos registros do RCCH além de uma adequada assistência para as famílias afetadas, e gerou material de enorme valia para estudos detalhados da síndrome de Lynch.

O presente estudo representa a análise molecular mais abrangente já realizada em famílias Brasileiras com polipose. Além do objetivo inicial de caracterizar mutações germinativas nos genes *APC* e *MUTYH*, este trabalho ainda buscou identificar outros genes associados com a síndrome por sequenciamento do exoma. Como resultados finais, identificamos mutações patogênicas em um dos dois genes em 87% dos pacientes (20/23): seis pacientes apresentaram mutações patogênicas em *MUTYH* e 14 apresentaram mutações patogênicas em *APC*. Um paciente foi portador de uma nova variante *missense* de significado clínico desconhecido em *APC* (c.5365G>C p.Val1789Leu), e dois pacientes foram negativos para ambos os genes.

Esta nova variante foi identificada no paciente ID13, único portador de polipose atenuada nesta família (impossibilitando uma avaliação de segregação). A presença desta alteração já foi avaliada em 96 amostras de indivíduos controles saudáveis, não sendo encontrada em nenhum outro indivíduo. A alteração p.Val1789Leu está localizada entre o segundo domínio SAMP (domínio de ligação à Axina) e o quinto domínio rico em cisteína (domínio de ligação à  $\beta$ -catenina) da proteína APC (Figura 7). As análises realizadas *in silico* utilizando softwares de predição de patogenicidade (SIFT e Polyphen2) classificaram esta variante como benigna.





Fonte: Modificado de BARROS (2012).

**Figura 7** - Localização da variante p.Val1789Leu identificada no paciente ID13. Representação esquemática dos diversos domínios da proteína APC. A alteração p.Val1789Leu (identificada pelo marcador vermelho) está localizada entre o segundo domínio SAMP (domínio de ligação à Axina) e o quinto domínio rico em cisteína de 20 aminoácidos (domínio de ligação à β-catenina)

Devido à sua ausência nos bancos de dados e na população controle e à impossibilidade de realizarmos estudos de segregação, decidimos realizar ensaios funcionais *in vitro* para avaliar o efeito da variante p.Val1789Leu do gene *APC*. Estas análises serão realizadas em conjunto com a avaliação funcional de duas outras alterações *missense* do gene *APC* identificadas por nosso grupo em tumores de Wilms (BARROS 2012). Os plasmídeos que serão utilizados para estas análises, assim como os iniciadores necessários para a mutação sítio-dirigida, já foram adquiridos e os ensaios funcionais estão sendo desenvolvidos pela Dra. Adriana M. Nakahata e Dra. Bruna D. Barros.

Por fim, este estudo permitiu o diagnóstico molecular em 20 dos 23 pacientes avaliados. Para os demais três pacientes, a continuidade da investigação é necessária para comprovar a associação das variantes encontradas com o fenótipo. Os resultados de exoma irão necessitar de extensas análises para que alguma alteração possa ser confirmada como associada à síndrome. Dentre as estratégias previstas para realizar

esta validação estão: quando possível avaliar a segregação nos membros afetados e não afetados de cada família; investigar a presença das alterações em uma população controle; investigar a presença das alterações em indivíduos com CCR em idade precoce ou múltiplos pólipos adenomatosos na ausência de história familiar; realizar estudos funcionais.

O contínuo avanço da genômica na última década possibilitou que grandes descobertas fossem realizadas no campo da medicina molecular, levando a identificação de diversos novos genes relacionados às doenças humanas. Se por um lado as metodologias de análises em larga escala, em especial o sequenciamento de nova geração, possibilitam identificação de genes causadores de doenças em tempo recorde, por outro lado geram uma grande quantidade de dados de significado ainda desconhecido. O maior desafio agora reside em desenvolver ferramentas e métodos capazes de descobrir a relevância destas alterações, para que a promessa da medicina genômica personalizada seja concretizada.

## 9 CONCLUSÕES

- Foi observada uma alta taxa de detecção (87%) e diversas novas mutações nos genes *APC* e *MUTYH* em pacientes brasileiros com polipose. Entre os afetados, dois casos interessantes foram caracterizados, como a identificação da primeira grande deleção descrita para o gene *MUTYH*, revelando que grandes rearranjos representam um possível mecanismo de perda de função deste gene; e também a identificação de um paciente FAP portador de uma deleção de 5,2 Mb em 5q21.3-q22.3, levando a perda do gene *APC* e de mais 19 genes, e com o fenótipo não esperado de ausência de características dismórficas e de déficit cognitivo;
- A classificação das alterações encontradas permitiu o diagnóstico molecular em 20 dos 23 pacientes avaliados. A única variante *missense* de significado clínico incerto encontrada foi classificada como benigna por programas de predição *in silico* e estudos adicionais serão necessários para classificar corretamente esta alteração;
- Famílias positivas para mutações em *APC* não apresentaram uma forte correlação genótipo-fenótipo, e este achado deve ser considerado ao realizar o aconselhamento genético e acompanhamento dessas famílias;
- O sequenciamento do exoma identificou uma variante *missense* no gene *POLQ* como uma forte candidata a estar associada à polipose.

## 10 REFERÊNCIAS BIBLIOGRÁFICAS

Al-Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. **Nat Genet** 2002; 30:227-32.

Aretz S, Uhlhaas S, Caspari R, et al. Frequency and parental origin of de novo APC mutations in familial adenomatous polyposis. **Eur J Hum Genet** 2004; 12:52-8.

Aretz S, Drichel D, Kerick M, et al. Exome sequencing identifies potential new candidate genes for unexplained colorectal adenomatous polyposis. In: **63<sup>rd</sup> Annual Meeting of the American Society of Human Genetics**; 2013 Oct 22-26; Boston(MA).

Azzopardi D, Dallosso AR, Eliason K, et al. Multiple rare nonsynonymous variants in the adenomatous polyposis coli gene predispose to colorectal adenomas. **Cancer Res** 2008; 68:358-63.

Barros BDF. **Definição do perfil de alterações das sequências dos genes APC, CTNNB1, WT1, WTX e PLCG2 em tumores de São Paulo**. São Paulo; 2012. [Tese de Doutorado-Fundação Antônio Prudente].

Beggs AD, Hodgson SV. The genomics of colorectal cancer: state of the art. **Curr Genomics** 2008; 9:1-10.

Bülow S. Results of national registration of familial adenomatous polyposis. **Gut** 2003; 52:742-6.

Burt RW, Leppert MF, Slattery ML, et al. Genetic testing and phenotype in a large kindred with attenuated familial adenomatous polyposis. **Gastroenterology** 2004; 127:444-51.

Campos AH, Carraro DM, Dias-Neto E. Preservando para o futuro: a importância de um banco de tumores para a pesquisa sobre o câncer. **Ciência Hoje** 2010; 277:44-8.

Cheadle JP, Sampson JR. Exposing the MYTH about base excision repair and human inherited disease. **Hum Mol Genet** 2003; 12:159-65.

Cheadle JP, Sampson JR. MUTYH-associated polyposis-from defect in base excision repair to clinical genetic testing. **DNA Repair (Amst)** 2007; 6:274-9.

Corvello CM, Bevilacqua RAU, Rossi BM, et al. A novel germline mutation at exon 7 of the MSH2 gene (1249delG) in a large HNPCC Brazilian kindred. **Hum Mutat** 1999; 13:506.

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

Dominguez-Valentin M, Silva FCC, Santos EMM, et al. Characterization of germline mutations of MLH1 and MSH2 in unrelated South American suspected Lynch syndrome individuals. **Fam Cancer** 2011; 10:641-647.

Fodde R, Smits R, Clevers H. APC, signal transduction and genetic instability in colorectal cancer. **Nat Rev Cancer** 2001; 1:55-67.

Gebert JF, Dupon C, Kadmon M, et al. Combined molecular and clinical approaches for the identification of families with familial adenomatous polyposis coli. **Ann Surg** 1999; 229:350-61.

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

Goss KH, Groden J. Biology of the adenomatous polyposis coli tumor suppressor. **J Clin Oncol** 2000; 18:1967-79.

Grover S, Kastrinos F, Steyerberg EW, et al. Prevalence and phenotypes of APC and MUTYH mutations in patients with multiple colorectal adenomas. **JAMA** 2012; 308:485-92.

Heinimann K, Thompson A, Locher A, et al. Nontruncating APC germ-line mutations and mismatch repair deficiency play a minor role in APC mutation-negative polyposis. **Cancer Res** 2001; 61:7616-22.

Hogg M., Seki M., Wood RD, et al. Lesion bypass activity of DNA polymerase  $\theta$  (POLQ) is an intrinsic property of the pol domain and depends on unique sequence inserts. **J Mol Biol** 2011; 405:642-52.

Isidro G, Laranjeira F, Pires A, et al. Germline MUTYH (MYH) mutations in Portuguese individuals with multiple colorectal adenomas. **Hum Mutat** 2004; 24:353-4.

Kinzler KW, Nilbert MC, Su LK, et al. Identification of FAP locus genes from chromosome 5q21. **Science** 1991; 253:661-5.

Klaus A, Birchmeier W. Wnt signalling and its impact on development and cancer. **Nat Rev Cancer** 2008; 8:387-98.

Knudsen AL, Bisgaard ML, Bülow S. Attenuated familial adenomatous polyposis (AFAP): a review of the literature. **Fam Cancer** 2003; 2:43-55.

Lamlum H, Ilyas M, Rowan A, et al. The type of somatic mutation at APC in familial adenomatous polyposis is determined by the site of the germline mutation: a new facet to Knudson's 'two-hit' hypothesis. **Nat Med** 1999; 5:1071-5.

Leppert M, Dobbs M, Scambler P, et al. The gene for familial polyposis coli maps to the long arm of chromosome 5. **Science** 1987; 238:1411-3.

Li Y, Gao X, Wang JY. Comparison of two POLQ mutants reveals that a polymerase-inactive POLQ retains significant function in tolerance to etoposide and  $\gamma$ -irradiation in mouse B cells. **Genes Cells** 2011; 16:973-83.

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.

Markowitz SD, Bertagnolli MM. Molecular origins of cancer: molecular basis of colorectal cancer. **N Engl J Med** 2009; 361:2449-60.

Ministério da Saúde. Instituto Nacional do Câncer. **Estimativa/2010 incidência de câncer no Brasil**. Rio de Janeiro: INCA; 2009.

Morin PJ, Sparks AB, Korinek V, et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. **Science** 1997; 275:1787-90.

Newton CR, Graham A, Heptinstal LE. et al. Analysis of any point mutation in DNA: the amplification refractory mutation system (ARMS). **Nucleic Acids Res** 1989; 17:2503-16.

Nielsen M, Hes FJ, Nagengast FM, et al. Germline mutations in APC and MUTYH are responsible for the majority of families with attenuated familial adenomatous polyposis. **Clin Genet** 2007; 71:427-33.

Nieuwenhuis MH, Vasen HF. Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. **Crit Rev Oncol Hematol** 2007; 61:153-61.

Nishisho I, Nakamura Y, Miyoshi Y, et al. Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. **Science** 1991; 253:665-9.

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.

Pineda M, González S, Lázaro C, Blanco I, Capellá G. Detection of genetic alterations in hereditary colorectal cancer screening. **Mutat Res** 2010; 693:19-31.

Powell SM, Petersen GM, Krush AJ, et al. Molecular diagnosis of familial adenomatous polyposis. **N Engl J Med** 1993; 329:1982-7.

Pritchard CC, Smith C, Salipante SJ, et al. ColoSeq provides comprehensive Lynch and polyposis syndrome mutational analysis using massively parallel sequencing. **J Mol Diagn** 2012; 14:357-66.

Rossi BM, Corvello CM, Anelli A, et al. Hereditary colorectal tumors: routine care and the multidisciplinary therapeutic approach. **South Am J Cancer** 1997; 1:191-7.

Rossi BM. **Contribuição para o estudo do câncer colorretal com predisposição hereditária no Brasil**. São Paulo; 2008. [Tese de Livre-docência-Faculdade de Medicina, Universidade de São Paulo].

Rozen P, Macrae F. Familial adenomatous polyposis: The practical applications of clinical and molecular screening. **Fam Cancer** 2006; 5:227-35.

Sampson JR, Jones S, Dolwani S, Cheadle JP. MutYH (MYH) and colorectal cancer. **Biochem Soc Trans** 2005; 33:679-83.



[SEER] The Surveillance, Epidemiology, and End Results. **SEER stat fact sheets: colon and rectum cancer.** Available from: <URL:<http://seer.cancer.gov/statfacts/html/colorect.html>> [2013 dez 12]

Sieber OM, Lipton L, Crabtree M, et al. Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. **N Engl J Med** 2003; 348:791-9.

Sieber OM, Tomlinson IP, Lamlum H. The adenomatous polyposis coli (APC) tumour suppressor--genetics, function and disease. **Mol Med Today** 2000; 6:462-9.

Silva FC. **Frequência de mutações nos genes *MSH6*, *PMS1*, *PMS2*, *TP53* e *CHEK2* em pacientes suspeitos de Síndrome de Lynch.** São Paulo; 2011. [Tese de Doutorado-Fundação Antônio Prudente].

Sjöblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. **Science** 2006; 314:268-74.

Strate LL, Syngal S. Hereditary colorectal cancer syndromes. **Cancer Causes Control** 2005; 16:201-13.

Torrezan GT, Silva FCC, Krepischi ACV, et al. (2011). Breakpoint characterization of a novel large intragenic deletion of MUTYH detected in a MAP patient: case report. **BMC Med Genet** 2011; 12:128.

Torrezan GT, Silva FCC, Krepischi ACV, Santos EMM, Rossi BM, Carraro DM. A novel SYBR-based duplex qPCR for the detection of gene dosage: detection of an APC large deletion in a familial adenomatous polyposis patient with an unusual phenotype. **BMC Med Genet** 2012; 13:55.

Torrezan, GT, Silva FCC, Santos EMM, et al. Mutational spectrum of the APC and MUTYH genes and genotype–phenotype correlations in Brazilian FAP, AFAP, and MAP patients. **Orphanet J Rare Dis** 2013; 8:54.

Tsao J, Shibata D. Further evidence that one of the earliest alterations in colorectal carcinogenesis involves APC. **Am J Pathol** 1994; 145:531-4.

van der Luijt RB, Khan PM, Vasen HF, et al. Molecular analysis of the APC gene in 105 Dutch kindreds with familial adenomatous polyposis: 67 germline mutations identified by DGGE, PTT, and southern analysis. **Hum Mutat** 1997; 9:7-16.

Varesco L. Familial adenomatous polyposis: genetics and epidemiology. **Tech Coloproctol** 2004; 8 Suppl 2:s305-8

Vasen HFA, Möslein G, Alonso A, et al. Guidelines for the clinical management of familial adenomatous polyposis (FAP). **Gut** 2008; 57:704-13.

Young J, Simms LA, Tarish J, et al. A family with attenuated familial adenomatous polyposis due to a mutation in the alternatively spliced region of APC exon 9. **Hum Mutat** 1998; 11:450-5.

## **Anexo 1 - Detalhamento dos Métodos Utilizados**

### **1. CASUÍSTICA E CRITÉRIOS DE INCLUSÃO**

Foram revistos os registros do banco de dados do Registro das Famílias de Câncer Colorretal Hereditário do A.C. Camargo Cancer Centerem busca de pacientes com critérios clínicos para FAP (mais de 100 pólipos adenomatosos) e AFAP/MAP (10-100 pólipos adenomatosos). Foram incluídos no estudo de rastreamento de mutações 44 indivíduos de 23 famílias portadoras de FAP e AFAP - 23 probandos clinicamente afetados, 8 familiares afetados e 13 não afetados. Foram excluídos pacientes com outras síndromes de câncer colorretal hereditárias não FAP/AFAP. Os prontuários dos pacientes selecionados foram revistos para a coleta de informações clínicas. Além dos dados clínicos dos 44 indivíduos recrutados, informações clínicas também foram obtidas para os demais familiares afetados de cada família durante a entrevista do probando com o aconselhador genético para montagem do heredograma, totalizando 113 pacientes afetados por polipose neste estudo.

A coleta de sangue periférico foi feita após o recrutamento de pacientes e familiares interessados em realizar o teste genético. As informações sobre os objetivos da pesquisa foram apresentadas aos pacientes para a assinatura do Consentimento Livre e Esclarecido (Anexo 1).

Para avaliação em uma população controle, foram selecionadas 96 amostras a partir de uma grande casuística composta de 1000 indivíduos coletados para outro projeto realizado na nossa instituição, cujo DNA processado pelo banco de macromoléculas (Biobanco Hospital A. C. Camargo) está disponível para utilização de outros projetos aprovados pelo Comitê de Ética em Pesquisa (CEP).

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa do Centro de Tratamento e Pesquisa do Hospital A.C. Camargo, sob o nº 1169/08-B.

### **2. EXTRAÇÃO DO DNA SANGUÍNEO**

As amostras de sangue periférico de cada paciente foram coletadas em dois tubos K2 VACUETTE com EDTA anticoagulante e encaminhadas ao Banco de Macromoléculas do Hospital A. C. Camargo para extração de DNA (CAMPOS et al. 2010). O processamento do sangue periférico baseia-se na lise das células vermelhas (hemácias) com Tris-EDTA 1X pH 8.0, seguida de centrifugações para separação dos

leucócitos até a obtenção de um *buffy coat* ou *pellet*. Um dos *pellets* de leucócitos foi submetido diretamente à extração de DNA genômico usando Puregene Genomic DNA Isolation kit (Gentra Systems, USA) de acordo com as recomendações do fabricante e o outro *pellet* foi estocado a -20°C para posterior confirmação dos resultados obtidos com o sequenciamento.

### **3. DETECÇÃO DE MUTAÇÕES POR SEQUENCIAMENTO SANGER DOS GENES *APC* E *MUTYH***

Inicialmente, um probando da família que tenha sido diagnosticado clinicamente com polipose foi investigado por sequenciamento direto para avaliação da presença de mutação nos genes *APC* e *MUTYH*. Após a identificação da mutação presente no probando, demais familiares (no máximo seis) que tiveram interesse em realizar o teste genético foram investigados para a presença da mutação encontrada.

#### **Desenho dos Iniciadores**

As sequências genômicas dos genes *APC* (NM\_000038) e *MUTYH* (NM\_001128425.1) foram obtidas a partir do banco NCBI/Nucleotide e utilizadas como base para o desenho dos iniciadores, os quais foram desenhados nas regiões intrônicas a aproximadamente 70 bases do limite éxon-intron. Os critérios utilizados para o desenho dos iniciadores foram: proporção de bases C e G em torno de 40 a 60%, comprimento de 19 a 23 bases, temperatura de anelamento (por %CG) entre 65 e 70°C. O programa OLIGOTECH 1.0 (Oligos Etc. Inc. & Oligo Therapeutics Inc, 1995) foi utilizado para verificar a presença de estruturas secundárias (*stem loop* e *homodimer*) e a temperatura de anelamento. Os oligoiniciadores *forward* e *reverse* foram desenhados a distâncias de no máximo 2000pb para as amplificações de PCR e no máximo 600pb para as reações de sequenciamento.

Para o gene *APC* foram desenhados 34 pares de iniciadores. Todos os éxons foram amplificados com um único par de iniciadores, exceto o éxon 15 que devido a sua extensão (8867pb) foi dividido em cinco regiões para a amplificação por PCR (fragmentos A-E) e em 20 fragmentos para o sequenciamento. O Quadro 1 apresenta a localização dos iniciadores do gene *APC*, o tamanho do produto amplificado na PCR, os iniciadores utilizados no sequenciamento de cada fragmento e a temperatura de pareamento padronizada.

**Quadro 1** - Características dos iniciadores utilizados para amplificação e sequenciamento do gene *APC*.

Iniciadores PCR	Iniciadores sequenciamento	Tamanho do <i>amplicon</i> (pb)	Tm (C)
1F-1R	1F,1R	570	62
2F-2R	2F,2R	303	62
3F-3R	3F,3R	581	62
4F-4R	4F,4R	365	62
5F-5R	5F,5R	455	62
6F-6R	6F,6R	416	62
7F-7R	7F,7R	362	62
8F-8R	8F,8R	279	TD 65-60
9F-9R	9F,9R	670	62
10F-10R	10F,10R	315	62
11F-11R	11F,11R	597	62
12F-12R	12F,12R	421	62
13F-13R	13F,13R	400	62
14F-14R	14F,14R	570	62
15A1F-15A4R	15A1F,15A1R, 15A2F,15A2R, 15A3F,15A43, 15A4F,15A4R	1708	60
15B1F-15B4R	15B1F,15B1R, 15B2F,15B2R, 15B3F,15B43, 15B4F,15B4R	1694	62
15C1F-15C4R	15C1F,15C1R, 15C2F,15C2R, 15C3F,15C43, 15C4F,15C4R	1703	62
15D1F-15D4R	15D1F,15D1R, 15D2F,15D2R, 15D3F,15D43, 15D4F,15D4R	1705	62
15E1F-15E4R	15E1F,15E1R, 15E2F,15E2R,	1643	62

**Legenda:** os números descritos para os iniciadores das reações de PCR e sequenciamento representam o éxon em que estão localizados; F = iniciador *forward*, R = iniciador *reverse*; TD = PCR *touchdown*.

Os 16 éxons do gene *MUTYH* foram amplificados em 5 fragmentos na PCR e sequenciados com 11 pares de iniciadores adicionais internos aos fragmentos de PCR. AO Quadro 2 apresenta a localização dos iniciadores do gene *MUTYH*, o tamanho do produto amplificado na PCR, os iniciadores internos utilizados no sequenciamento de cada fragmento e a temperatura de anelamento padronizada.

**Quadro 2** - Características dos iniciadores utilizados para amplificação e sequenciamento do gene *MUTYH*.

Iniciadores PCR	Iniciadores sequenciamento	Tamanho do amplicon (pb)	Tm (C)
1F-1R	1F,1R	459	62
2F-3R	2F,2R,3F,4R	1276	62
4F-11R	4-6F,4-6R,7-8F,7-8R, 9-10F, 9-10R, 11F, 11R	1501	62
12F-15R	12-13F,12-13R,14F,14R, 15F, 15R	1537	62
16F-16R	16F,16R	440	62

**Legenda:** os números descritos para os iniciadores das reações de PCR e sequenciamento representam o exon em que estão localizados; F = iniciador *forward*, R = iniciador *reverse*.

### PCR e sequenciamento Sanger

As reações de amplificação dos fragmentos dos genes *APC* e *MUTYH* foram realizadas contendo 1X GoTaq Green Master Mix (Promega, USA), 0,4µmol de cada iniciador e 25ng de DNA genômico, em um volume final de 20µl. As reações foram realizadas em um termociclador segundo a seguinte ciclagem: 1 ciclo inicial de desnaturação a 94°C por 2 minutos; 40 ciclos de amplificação (94°C por 20 segundos, 58-62°C por 1 minuto, 72°C por 1-2 minutos) e 1 ciclo de extensão final a 72°C por 5 minutos.

Os produtos amplificados foram verificados em gel de agarose 1%. Para reação de sequenciamento direto, alíquotas de 2,0 a 5,0µL dos produtos de PCR foram purificadas com exonuclease I e fosfatase alcalina na proporção de 5µl de PCR para 1µl de ExoSAP-IT (USB Corporation). A reação foi incubada por 30 minutos a 37°C, seguida da inativação das enzimas a 80°C por 15 minutos, conforme protocolo recomendado pelo fabricante.

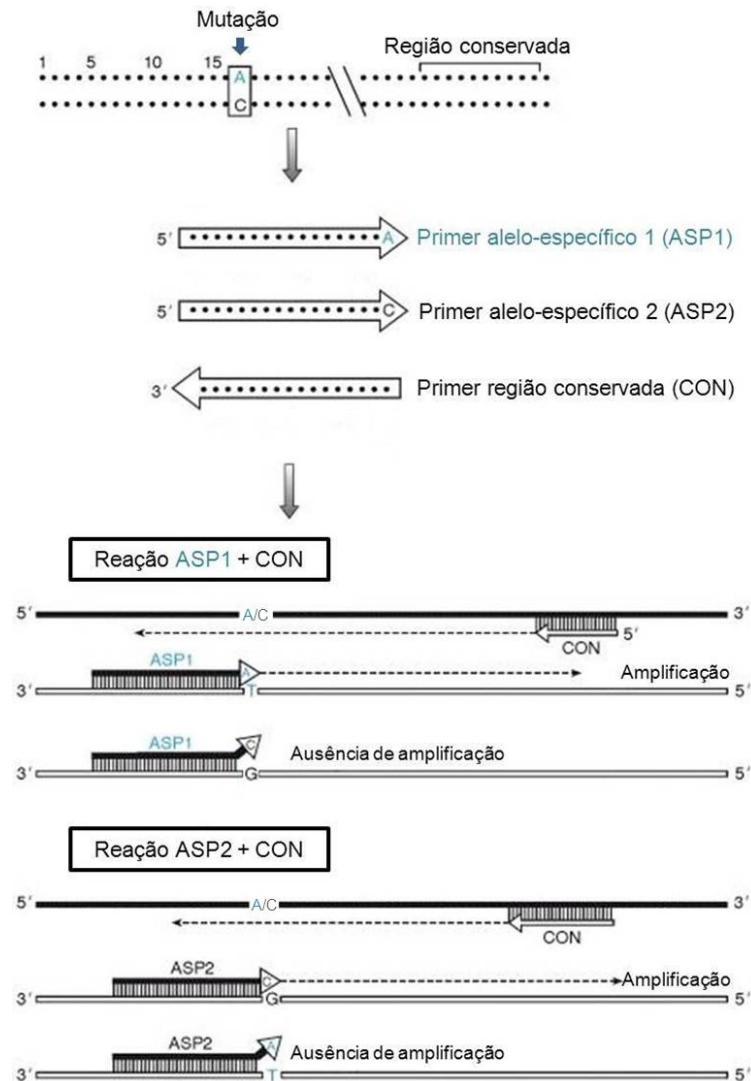
As reações de sequenciamento foram realizadas com 2,5µl do produto de PCR purificado, 0,5µl de BigDye® Terminator v3.1 (Applied Biosystems), 1,5µl de tampão e

0,5µmol de cada iniciador específico em um volume final de 10µl. As reações de sequenciamento foram realizadas segundo a seguinte ciclagem: 1 ciclo inicial de desnaturação a 95°C por 2 minutos, seguido por 40 ciclos de amplificação (95°C por 18 segundos, 54°C por 18 segundos, 60°C por 4 minutos). Todos os fragmentos foram sequenciados nos sentidos *forward* e *reverse*. A precipitação da reação de sequenciamento foi realizada com metodologia baseada em etanol. As amostras foram ressuspendidas em 13µl de formamida, desnaturadas a 95°C por 3 minutos e imediatamente acondicionadas em gelo por 5 minutos. O sequenciamento foi realizado no equipamento ABI3130XL (Applied Biosystems).

#### **4. PCR Alelo-específica**

A metodologia de PCR Alelo-específica (AS-PCR – *Allele-specific PCR*) ou ARMS (*Amplification refractory mutation system*) permite a discriminação de qualquer mutação em ponto ou polimorfismo de base única (NEWTON et al. 1989). Esta técnica se baseia na dependência da PCR do correto pareamento da extremidade 3' do oligo iniciador para que ocorra a síntese da fita complementar. Para isso, utilizam-se dois iniciadores diferentes, um deles apresentando a última base da extremidade 3' complementar a um dos nucleotídeos da alteração e o outro iniciador sendo complementar ao outro nucleotídeo possível (Figura 1).

Em uma amostra de um indivíduo heretozigoto, cada um dos alelos será amplificado apenas na reação contendo o iniciador que apresente o nucleotídeo complementar a este alelo. A amplificação de apenas um dos alelos em cada reação permite a caracterização de haplótipos, permitindo assim a comprovação de mutações bialélicas, também chamadas heterozigotas compostas, em síndromes recessivas.



Fonte: modificado de Human Molecular Genetics (1999).

**Figura 1** - Representação esquemática da PCR Alelo-Específica. Esta técnica se baseia no desenho de dois iniciadores diferentes, um deles apresentando a última base da extremidade 3' complementar a um dos nucleotídeos da alteração (ASP1) e o outro iniciador sendo complementar a outro nucleotídeo possível (ASP2). Utilizando um iniciador reverso comum desenha região conservada entre os dois alelos (CON), duas reações de PCR podem ser realizadas separadamente com cada um dos iniciadores específicos, levando a amplificação de cada alelo de modo individual.

Para confirmar que as duas alterações em heterozigose encontradas no gene *MUTYH* de três probandos estavam presentes cada uma em um alelo (mutação bialélica ou em heterozigose composta), novos iniciadores alelo-específicos foram desenhados para a mutação *missense* c.536A>G p.Tyr179Cys, comum aos três pacientes (Quadro 3).



**Quadro 3:** Iniciadores alelo-específicos para a mutação c.536A>G p.Tyr179Cys no gene *MUTYH*.

ID	Mutação 1	Mutação 2	ASP1	ASP2	CON	Tam
CCR447	c.536A>G (éxon 7)	c.1227-1228dup (éxon 13)	5'CTGGGCTGG CCTGGGCTA3' (forward)	5'CTGGGCTGG CCTGGGCTG3' (forward)	5'GTGGACAACC TGGAGGAAG3' (reverse)	1495
FAP16-1	c.536A>G (éxon 7)	c.1147delC (éxon 12)				
FAP18-1	c.389-1G>C (ínton 4)	c.536A>G (éxon 7)	5'CCGGCCACG AGAATAGT3' (reverse)	5'CCGGCCACG AGAATAGC3' (reverse)	5'CCATGGAGAA GACGGGTAG3' (forward)	641

**Legenda:** ID = código do paciente; ASP1 = iniciador alelo-específico 1 – alelo selvagem; ASP2 = iniciador alelo-específico 2 – alelo mutado; CON = iniciador conservado; Tam = tamanho do *amplicon*, em pares de base.

As condições da AS-PCR e do sequenciamento de seus *amplicons* foram as mesmas que as descritas acima para as PCRs comuns.

## 5. ANÁLISE DE GRANDES REARRANJOS GENÔMICOS

Os cinco pacientes que não apresentaram nenhuma alteração pontual em *APC* e *MUTYH* foram investigados para a presença de grandes deleções e ampliações. Para isto, foram empregadas as técnicas de CGH-array (*Microarray-based Comparative Genomic Hybridization*) e MLPA® (*Multiplex Ligation-Dependent Probe Amplification*). As alterações identificadas com estas metodologias foram validadas através da técnica de duplex PCR quantitativa (qPCR duplex) que foi desenvolvida neste estudo.

### 5.1. CGH-array

A técnica de Hibridização genômica comparativa baseada em microarranjos (CGH-array) foi realizada na plataforma de 180K para genoma completo da OGT Technologies. As amostras foram marcadas com Cy3- e Cy5-dCTPs através de *random-priming*. As etapas de purificação, hibridização e lavagem foram realizadas como recomendado pelo fabricante. Os dados foram extraídos utilizando o software Feature Extraction (Agilent Technologies) e as análises foram realizadas no programa Genomic Workbench (Agilent Technologies) utilizando o algoritmo ADM-2. O *threshold* de sensibilidade usado foi 6,7. Os valores de *threshold* para a definição de alterações no número de cópias genômicas foram: 0,4 para duplicações, 1,1 para ganhos de múltiplas cópias, -0,4 para deleções de uma cópia e -1,1 para perdas em homozigose.

### 5.2. Multiplex Ligation-Dependent Probe Amplification (MLPA)

A técnica de MLPA foi empregada para avaliar a presença de deleções e ampliações de éxons inteiros, não detectáveis pela técnica de sequenciamento. O kit utilizado foi o SALSA P043-C1 APC probemix kit (MRC Holland, Amsterdam, The Netherlands), que apresenta sondas para os promotores e para cada éxon do gene *APC* (sendo que o éxon 15, devido ao tamanho, apresenta 5 sondas). Além disso, este kit também contém sondas para três éxons do gene *MUTYH* (éxons 1, 2 e 7). As reações de MLPA foram realizadas segundo recomendado pelo fabricante. Após a amplificação das sondas, os produtos de PCR foram analisados no sequenciador ABI3130xl (Applied Biosystems - Foster City, CA, USA) e a dosagem de cada éxon foi calculada utilizando o programa Coffalyser V9.4 program (MRC Holland, Amsterdam, The Netherlands).

### 5.3. Dosagem gênica por PCR quantitativa em tempo real

Para confirmar a presença de grandes duplicações e deleções identificadas por CGH-array ou MLPA, foram realizados dois métodos de dosagem gênica por PCR quantitativa em tempo real (qPCR): o  $2^{-\Delta\Delta Ct}$  (LIVAK e SCHMITTGEN 2001) e a técnica de qPCR Duplex utilizando SYBR Green<sup>®</sup> que foi desenvolvida neste trabalho. Além de representar um método alternativo simples e robusto de dosagem gênica, essa técnica apresenta a grande vantagem de minimizar erros de dosagem ocasionados por variações

nas quantidades iniciais de DNA-molde (TORREZAN et al. 2012). Para ambas as metodologias foram utilizados iniciadores para o íntron 7 do gene *GAPDH* (12p13) e éxon 3 do gene *HPRT1* (Xq26.1) como genes referência.

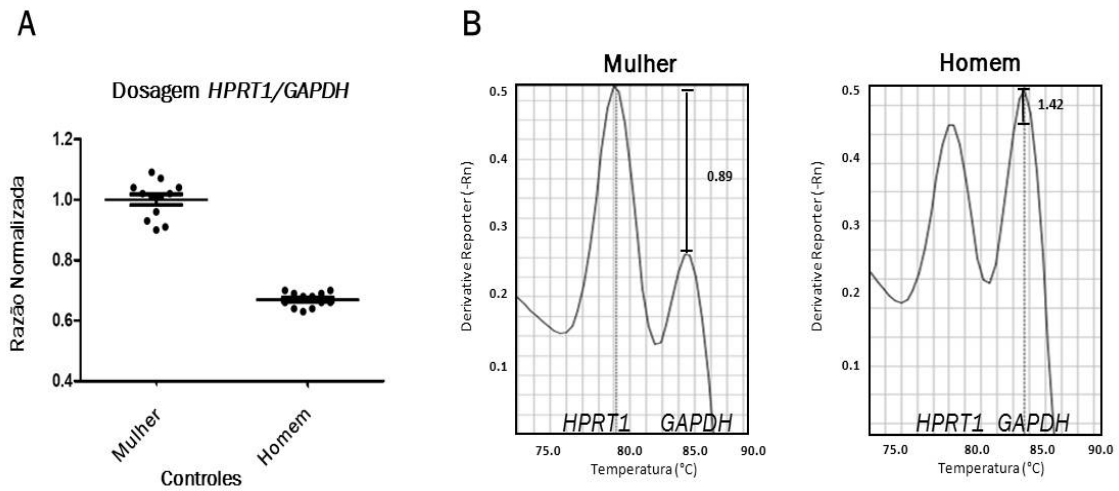
Para a realização do método  $2^{-\Delta\Delta C_t}$ , 18ng de DNA foram amplificados em uma reação de 20 $\mu$ L contendo 0,3  $\mu$ mol de cada iniciador e 1X SYBR® Green PCR MasterMix (Applied Biosystems). A reação de qPCR foi realizada no equipamento ABI Prism 7500 (Applied Biosystems) com a seguinte ciclagem: 95°C por 10 minutos, seguido de 40 ciclos de desnaturação a 95°C por 15 segundos e anelamento e extensão a 60°C por um minuto.

As amostras dos pacientes e controles foram amplificadas em duplicata para os fragmentos alvo e referência. Para os cálculos de dosagem gênica foi utilizada a média dos  $C_t$ s das duplicatas. Para cada fragmento, quatro indivíduos normais foram usados para calcular o  $C_t$  médio dos controles. A dosagem gênica foi calculada utilizando a seguinte equação:  $2^{-[\Delta C_t(\text{alvo}) - \Delta C_t(\text{ref})]}$  onde  $\Delta C_t(\text{alvo})$  refere-se à diferença entre os valores de  $C_t$  entre o paciente e a média dos controles para o éxon alvo, e o  $\Delta C_t(\text{ref})$  é igual à diferença entre os valores  $C_t$  para o paciente e a média dos controles para o éxon referência. Valores entre 0,82 e 1,22 foram considerados normais, entre 0,41 e 0,61 deleções de um alelo e acima de 1,4 ampliações (permitindo erro de 10% na concentração medida).

As condições da qPCR Duplex baseada em SYBR® Green foram basicamente as mesmas descritas acima para o método  $2^{-\Delta\Delta C_t}$ , com a seguinte diferença: éxons alvo e referência foram amplificados em duplex com 0,15 $\mu$ mol de cada iniciador por 25 ciclos. Após a coamplificação, os produtos de PCR foram submetidos a uma temperatura de transição de 60°C a 95°C, e foram geradas curvas de *melting* da queda da fluorescência do SYBR Green. Os picos de *melting* (PM) foram então obtidos a partir da primeira derivada negativa da variação da fluorescência (-dF/dT) versus a temperatura, permitindo a discriminação de dois picos – um para o éxon alvo e outro para o éxon referência. A razão entre a altura do PM do éxon alvo e do éxon referência determina a concentração relativa do éxon alvo na amostra, e normalizando a razão do paciente pela razão obtida em amostras controle é possível determinar o número de cópias deste fragmento no paciente. Todos os ensaios foram realizados em duplicata e três indivíduos controle foram avaliados para cada fragmento.

## Validação do método de dosagem gênica por qPCR duplex

Antes desta metodologia ter sido utilizada nos pacientes do estudo, a mesma foi validada através da quantificação do número de cópias do cromossomo X. Para isso, utilizou-se um grupo controle composto por 12 homens e 12 mulheres, no qual foi avaliada a razão entre o éxon 3 do gene *HPRT1* (Xq26.1) e o íntron 7 do gene referência *GAPDH* (12p13). Utilizando uma amostra feminina como normalizadora, a razão obtida entre os homens variou de 0,63 a 0,70, enquanto a razão entre as mulheres variou de 0,90 a 1,09 (Figura 2A). A comparação visual da altura dos picos da curva de *melting* (PM) entre os dois grupos evidencia uma redução do PM do gene *HPRT1* nas amostras masculinas em relação às femininas (Figura 2B).



**Figura 2** - Validação da qPCR duplex com SYBR Green. A: Distribuição das razões entre *HPRT1* e *GAPDH* em 24 amostras controles (12 mulheres e 12 homens); nos indivíduos do sexo masculino a razão normalizada variou de 0,63 a 0,70, enquanto entre indivíduos do sexo feminino a razão normalizada variou de 0,90 a 1,09. B: Comparação visual entre os dois sexos, evidenciando uma redução do PM de *HPRT1* nas amostras masculinas em relação às femininas.

A validação deste método também foi realizada em genes adicionais (*ATM*, *PTEN*, *BRCA1* e *PIP*) em outros pacientes com síndromes de câncer hereditário e com alterações previamente identificadas por CGH-array.

## 6. AVALIAÇÃO FUNCIONAL DA VARIANTE p.Val1789Leu

A alteração *missense* p.Val1789Leu identificada neste estudo será avaliada em relação ao seu impacto funcional através de estudos *in vitro*. Resumidamente, células de linhagens tumorais de tumores de cólon serão transfectadas com plasmídeos expressando o gene *APC* em sua forma selvagem, com uma mutação sabidamente patogênica e com a alteração p.Val1789Leu. Serão realizados ensaios de imunoprecipitação e localização celular, além de outros ensaios que estão sendo estudados e discutidos pelo grupo.

Serão utilizados plasmídeos pCMVs contendo o cDNA do gene *APC* selvagem (pCMV-Neo-Bam *APC*) e com uma mutação patogênica de códon de parada prematuro (pCMV-Neo-Bam *APC* 1-1309). Os plasmídeos descritos acima foram construídos por B. Vogelstein e K. Kinzler (Johns Hopkins University, Baltimore, MD) e adquiridos por nosso grupo através da organização sem fins lucrativos Addgene (<http://www.addgene.org/>). O plasmídeo contendo a alteração p.Val1789Leu será construído por meio de mutagênese sítio-dirigida, com o kit QuikChange II-E Site-Directed Mutagenesis Kit (Cat #200555).

Em resumo, o plasmídeo contendo o gene *APC* selvagem será utilizado como molde para a reação de mutagênese. Os iniciadores *forward* e *reverse* (Quadro 4), ambos desenhados com a mutação de interesse, anelam-se a sequência alvo no vetor e com o aumento da temperatura ocorre à extensão dos iniciadores pela polimerase PfuUltra high fidelity (Stratagene) gerando um plasmídeo com a alteração de interesse. A reação de mutagênese é então tratada com a endonuclease DpnI (sequência de reconhecimento: 5' Gm6ATC-3'), uma enzima de restrição que é utilizada porque reconhece e digere DNA metilado ou hemimetilado, clivando desta forma apenas o vetor parental. O vetor contendo a mutação desejada recém sintetizado não é metilado, e por isso não é digerido pela enzima. A reação é então purificada utilizando a resina Strataclean e o produto purificado é transformado por eletroporação em XL1-blue. A confirmação da presença das alterações será feita por sequenciamento direto, com iniciadores do inserto que flanqueiam a região da mutação.

**Quadro 4:** Iniciadores para mutação sítio dirigida no gene *APC*

Alteração	Iniciador	Sequência	Tamanho
c.5365 G>C p.Val1789Leu	APC- g5365c_F	5'-caaaatactgaatataggacacgt <b>cta</b> agaa aaaatgcagactcaaaaa-3'	49
	APC - g5365c_R	5'- ttttgagtctgcatttttctta <b>gac</b> gtgctctatatt cagtattttg-3'	49

A terceira coluna apresenta a sequencia dos iniciadores *forward* e *reverse* e as bases correspondentes a mutação de interesse são representadas em negrito e vermelho.

## 7. DETECÇÃO DE MUTAÇÕES POR SEQUENCIAMENTO EM LARGA ESCALA

Os pacientes negativos para mutações em *APC* e *MUTYH* pelas técnicas de sequenciamento, CGH-array e MLPA foram rastreados para mutações germinativas em outros genes por sequenciamento de nova geração do exoma. Para o enriquecimento das porções exônicas foi utilizado o kit de captura direcionada para todo o exoma humano *TargetSeq™ Exome Enrichment Kit* (LifeTechnologies, Foster City, USA), seguindo as especificações do fabricante. Esta abordagem realiza a captura de 45,1 Mb do genoma, sendo 37,3 Mb correspondentes a regiões codificadoras, e representando mais de 98% das sequencias codificantes descritas nos banco de dados VEGA (<http://www.sanger.ac.uk/resources/databases/vega/>) e RefSeq ([www.ncbi.nlm.nih.gov/RefSeq/](http://www.ncbi.nlm.nih.gov/RefSeq/)).

A captura é feita através de mais de dois milhões de sondas de DNA biotinizadas que são hibridadas contra o DNA genômico após quebra, seleção de tamanho, ligação de *barcodes* específicos e adaptadores para o sequenciamento. As moléculas híbridas DNA-sonda são selecionadas por incubação com *beads* magnéticos de streptavidina, seguindo as recomendações do fabricante e o DNA capturado é amplificado em poucos ciclos de PCR. Uma das vantagens desta plataforma, em relação a outras disponíveis, é que a ligação dos *barcodes* específicos para cada amostra é realizado antes da etapa de enriquecimento do exoma, e até 4 amostras podem ser adicionadas em um *pool* para a realização da etapa de captura, diminuindo o tempo de preparo e a utilização de alguns reagentes.

Após o preparo da biblioteca, as amostras enriquecidas foram então submetidas ao sequenciamento de alto desempenho na plataforma SOLiD™ 5500xl System

(LifeTechnologies, Foster City, USA). A tecnologia de sequenciamento do SOLiD utiliza a amplificação em emulsão do DNA ligado a esferas. Após a amplificação e enriquecimento, as esferas com fragmentos de DNA modificados na extremidade 3' são depositadas em uma *FlowChip* de vidro que contém 6 *lanes* individuais. O sequenciamento ocorre por múltiplas etapas de hibridização e ligação. Iniciadores são hibridados aos adaptadores no DNA molde e octâmeros fluorescentes hibridizam com a região adjacente do iniciador. Através da ação de uma ligase ocorre a ligação destes fragmentos e então é feita a detecção da fluorescência presente no octâmero. A marcação fluorescente do octâmero depende do par de bases inicial, permitindo a interrogação das duas primeiras bases nas quais ele se hibrida. Após a hibridização e detecção, as três bases finais do octâmero são clivadas e ocorrem múltiplos ciclos de ligação, detecção e clivagem. Posteriormente, o produto de extensão é removido e um iniciador complementar à região n-1 do DNA molde inicia um novo ciclo de ligações. Cinco ciclos de iniciadores são completados para cada fragmento de DNA molde. A principal vantagem desta metodologia é que cada base é determinada em duas ligações independentes, resultando em uma leitura precisa (99,9%) que requer pouca cobertura para atingir um valor confiável de chamada de bases.

Foi utilizada a estratégia de sequenciamento *paired-end*, com uma leitura de 75 pares de bases a partir de uma extremidade e 50 pares de bases da outra extremidade dos fragmentos capturados (Figura 3). As duas amostras foram sequenciadas juntamente com 10 amostras de outro projeto do grupo. Este *pool* contendo 12 exomas foi sequenciado em sete *lanes*, e com essa estratégia estimamos obter em torno de 4,5 Gb para cada exoma.



**Figura 3** - Representação esquemática de sequenciamento *paired-end*. Neste trabalho utilizamos a estratégia de sequenciamento *paired-end*, com uma leitura de 75 pares de bases a partir de uma extremidade e 50 pares de bases da outra extremidade dos fragmentos capturados (*enriched DNA*). Além disso, também é realizada a leitura de 5 pares de bases para identificar o *barcode* (BC) de cada amostra. P1 = adaptador P1; P2 = adaptador P2; IA = *internal adapter*.

## 8-ANÁLISE DOS RESULTADOS

### Descrição e Interpretação das Mutações Encontradas

A nomenclatura utilizada para descrição das mutações encontradas nos genes *APC* e *MUTYH* foi baseada no trabalho de DEN DUNNEN e ANTONARAKIS (2000), conforme recomendado pela Human Genome Variation Society ([www.hgvs.org/mutnomen](http://www.hgvs.org/mutnomen)).

Os bancos de dados LOVD (Leiden Open Variation Database), InSIGHT, The Human Gene Database, dbSNP e 1000genomes foram consultados para interpretação e determinação da significância patogênica das mutações encontradas. Foram classificadas como mutações patogênicas as mutações dos tipos *nonsense* (que geram códon de parada prematuro); as pequenas duplicações, deleções e inserções (que ocasionem mudança no quadro de leitura); e as mutações que afetaram sítios doadores ou receptores de *splicing*, (afetam o processo de *splicing* do RNAm, alterando a estrutura da proteína) e deleções/duplicações de éxons inteiros. Mutações *missense* não descritas nos bancos de dados foram investigadas quanto a sua patogenicidade em algoritmos de predição online (Polyphen, SIFT e GVGD).

### Análise de Bioinformática

Softwares específicos foram utilizados para converter as imagens capturadas pelos equipamentos em escores de intensidade, chamada de bases, escores de qualidade de alinhamentos e outras análises para converter os dados brutos em informação biológica relevante. A avaliação e edição dos resultados do sequenciamento direto foram realizadas utilizando o CLC Main Workbench Software (CLC Bio, Denmark). Os dados CGH-array foram extraídos utilizando o software Feature Extraction (Agilent Technologies, Santa Clara, USA) e as análises foram realizadas no programa Genomic Workbench (Agilent Technologies, Santa Clara, USA) utilizando o algoritmo ADM-2. Os resultados de MLPA foram avaliados utilizando o programa Coffalyser V9.4 program (MRC Holland, Amsterdam, The Netherlands).

As sequências geradas no SOLiD 5500xl foram alinhadas através do programa BioScope 2.0 (Life Technologies) a um genoma humano de referência (UCSC hg19) e as coordenadas genômicas das regiões alvos do kit *TargetSeq* (*TargetSeq™ Exome Enrichment Kit*). O programa Bioscope já apresenta parâmetros definidos para



mapeamento de sequências curtas geradas pelo equipamento SOLiD e realiza o mapeamento na sequência genômica em color-base. Além disso, o programa apresenta ferramentas para a busca de tipos específicos de alterações como SNPs, CNVs, indels, translocações e outras.

Todas as mutações encontradas pelo sequenciamento SOLiD e selecionadas como candidatas serão confirmadas por sequenciamento pelo método de Sanger.

**Anexo 2 - Termo de Consentimento Livre e Esclarecido para Indivíduos com Câncer Colorretal e Polipose**

(Obrigatório para Pesquisa Clínica em Seres Humanos – Resolução N. 196/96 e resolução CNS 252/97 do Ministério da Saúde)

**PROJETO: Estudo Genético de Câncer Colorretal na Hispânia e nas Américas**

**DADOS DE IDENTIFICAÇÃO DO PACIENTE OU RESPONSÁVEL LEGAL**

NOME DO PARTICIPANTE: \_\_\_\_\_

Sexo:  masculino  feminino Data de nascimento: \_\_\_/\_\_\_/\_\_\_\_\_

Documento de identidade n.: \_\_\_\_\_

Endereço: \_\_\_\_\_

Número: \_\_\_\_\_ Complemento: \_\_\_\_\_

Cidade: \_\_\_\_\_ Estado: \_\_\_\_\_

CEP: \_\_\_\_\_ TEL: \_\_\_\_\_

RESPONSÁVEL: \_\_\_\_\_

Sexo:  masculino  feminino Data de nascimento: \_\_\_/\_\_\_/\_\_\_\_\_

Documento de identidade n.: \_\_\_\_\_

Endereço: \_\_\_\_\_

Número: \_\_\_\_\_ Complemento: \_\_\_\_\_

Cidade: \_\_\_\_\_ Estado: \_\_\_\_\_

CEP: \_\_\_\_\_ TEL: \_\_\_\_\_

**II. OBJETIVOS DO ESTUDO**

Você foi convidado para participar deste estudo porque você tem câncer colorretal e familiar. O objetivo deste estudo é pesquisar por alterações genéticas que possam estar relacionadas a um aumento no risco de câncer do intestino grosso.

**III. PROCEDIMENTOS**

Após a assinatura deste termo, será preenchido um formulário com características como sexo, idade e história familiar. Depois você será encaminhado ao laboratório para a retirada da amostra de sangue. No laboratório serão retirados 8ml de sangue (aproximadamente duas colheres de sopa) através de uma agulha colocada no seu braço. O sangue será colocado em quatro tubos, identificado e enviado ao laboratório para

análise. Um dos tubos de sangue ficará no Brasil e o outro tubo será enviado para a Inglaterra para a realização da pesquisa. Nesta pesquisa será realizado um teste que irá buscar alterações genéticas que possam aumentar o risco de câncer de intestino grosso. As características clínicas e os resultados dos testes na amostra de sangue dos participantes com câncer será comparado com as características de participantes sem câncer.

Nós realizaremos no Brasil um estudo para verificar se você tem alterações nos genes que estão relacionadas com o câncer familiar. Cerca de 10% dos tumores de intestino grosso são hereditários, ou seja, são causadas por alterações genéticas que podem ser transmitidas dos pais aos filhos. Iremos realizar um teste nos genes (código genético) chamados APC e MYH. Os possíveis resultados deste teste são:

1 – você pode ter a alteração nos genes APC ou MYH que está associada o aparecimento de pólipos no intestino e risco de câncer de intestino grosso. É o que chamamos de resultado positivo; se você tem a alteração no gene, tem 50% de chance de passar a alteração para os seus filhos. ^

2 – o exame não identificou alterações nos genes APC ou MYH. Isto é chamado de resultado inconclusivo. Isto pode ocorrer em duas situações: você tem alterações em outros genes que não foram estudados, ou o teste de laboratório utilizado não conseguiu encontrar a alteração;

3 – o exame encontrou uma alteração que não é possível determinar se ela está associada a um risco maior de câncer.

Os resultados deste teste serão divulgados para você em uma consulta.

Os resultados deste estudo serão utilizados para apresentação de artigos e trabalhos científicos. Na divulgação dos resultados não serão divulgados os nomes ou qualquer forma que possa identificar os participantes da pesquisa.

Também solicitamos sua autorização para guardar parte da amostra de sangue (dois tubos) que não será utilizada nesta pesquisa. Esta amostra será guardada no banco de tumores do Hospital A.C. Camargo, e poderá ser utilizada em pesquisas futuras. Todas as pesquisas são aprovadas pelo Comitê de Ética em Pesquisa antes do seu início, e se necessário solicitaremos sua autorização para uso. Você pode decidir participar da pesquisa, mas não permitir que a amostra seja guardada após o final desta pesquisa.

#### IV - BENEFÍCIOS

A participação estudo pode não trazer benefícios para você ou para os membros de sua família; mas pode ajudar os médicos a compreender como o câncer se desenvolve.

#### V - RISCOS

Os riscos associados com sua participação nesse estudo são dor ou queimação no local da retirada do sangue. Pode ocorrer que após o resultado do teste genético você tenha sentimentos como ansiedade, raiva e tristeza. Caso seja necessário você poderá receber o apoio de uma psicóloga.

#### VI – ASPECTOS ÉTICOS

Esta pesquisa foi aprovada pelo Comitê de Ética do Hospital A.C. Camargo. A confidencialidade das suas informações será mantida. Apenas as pessoas envolvidas diretamente nesse estudo ou o Comitê de Ética em Pesquisa poderão ter acesso as informações se necessário.

A sua participação neste estudo é voluntária, tendo o direito de retirar-se do estudo a qualquer momento. Sua recusa ou desistência não irá prejudicar seu o tratamento.

Qualquer dúvida sobre o estudo, você poderá entrar em contato com a enfermeira Erika no telefone 2189-5000 ramal 1080. Se o pesquisador principal não fornecer as informações/esclarecimentos suficientes, entre em contato com o Coordenador do Comitê de Ética do Hospital do Câncer – SP, pelo telefone 21895020.

Declaro que fui esclarecido: sobre os procedimentos, riscos e benefícios sobre este estudo; que tenho liberdade em retirar o meu consentimento a qualquer momento, sem que isto traga prejuízo a continuidade do tratamento do meu familiar; que não haverá remuneração financeira para este estudo; sobre a segurança de que minha identidade será preservada, mantendo-se todas as informações em caráter confidencial.

( ) concordo participar do estudo

( ) \_\_\_\_\_ guarda da amostra de sangue

São Paulo, \_\_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_

---

Assinatura do Participante ou Responsável

---

Nome do Pesquisador ou Representante