ANÁLISE DA VARIAÇÃO NO NÚMERO DE CÓPIAS (CNV) DE DNA EM UMA FAMÍLIA COM SÍNDROME NEM 2A E MUTAÇÃO P.G533C NO GENE *RET*: IDENTIFICAÇÃO DE REGIÕES ASSOCIADAS À PREDISPOSIÇÃO PARA O DESENVOLVIMENTO DE METÁSTASE AOS LINFONODOS

Dissertação de Mestrado apresentada ao Curso de Pós-Graduação em Biologia Estrutural e Funcional da Universidade Federal de São Paulo para obtenção do título de Mestre em Ciências.

São Paulo

2013

ANÁLISE DA VARIAÇÃO NO NÚMERO DE CÓPIAS (CNV) DE DNA EM UMA FAMÍLIA COM SÍNDROME NEM 2A E MUTAÇÃO P.G533C NO GENE *RET*: IDENTIFICAÇÃO DE REGIÕES ASSOCIADAS À PREDISPOSIÇÃO PARA O DESENVOLVIMENTO DE METÁSTASE AOS LINFONODOS

Dissertação de Mestrado apresentada ao Curso de Pós-Graduação em Biologia Estrutural e Funcional da Universidade Federal de São Paulo para obtenção do título de Mestre em Ciências.

Orientadora: Prof^a Dr^a Janete Maria Cerutti

São Paulo

2013

Araujo, Aline Neves

Análise da variação no número de cópias (CNV) de DNA em uma família com síndrome NEM 2A e mutação p.G533C no gene *RET*: identificação de regiões associadas à predisposição para o desenvolvimento de metástase aos linfonodos / Aline Neves Araujo. -- São Paulo, 2013

xvi, 50f.

Dissertação (Mestrado) – Universidade Federal de São Paulo. Escola Paulista de Medicina. Programa de Pós-Graduação em Biologia Estrutural e Funcional.

Titulo em inglês: Copy number variation (CNV) analysis of DNA in a family with MEN2A syndrome and a *RET* p.G533C mutation: identification of regions associated with a predisposition to lymph node metastasis development.

1. Neoplasia Endócrina Múltipla Tipo 2A. 2. RET. 3. CNV. 4. Carcinoma Medular da Tiroide.

Banca Examinadora da Dissertação de Mestrado

Orientador: Prof^a. Dr^a. Janete Maria Cerutti

Membros:

1. Dr^a. Ana Amélia Fialho de Oliveira Hoff

2. Prof^a. Dr^a. Ileana Gabriela Sánchez de Rubió

3. Prof^a. Dr^a. Maria Isabel de Souza Aranha Melaragno

Suplente:

1. Prof^a. Dr^a. Rosa Paula Mello Biscolla

Programa de Pós-Graduação em Biologia Estrutural e Funcional

Universidade Federal de São Paulo.

Data: São Paulo, Fevereiro de 2013.

Aos meus pais, **Edna** e **Edmar**, pelo amor maior, dedicação e ensinamentos. Por serem meu porto seguro, pela cumplicidade, pelo exemplo. Por acreditarem em mim e, principalmente, por serem o que tenho de mais precioso, meu maior tesouro.

Ao meu noivo, **Christian**, pelo companheirismo e cumplicidade, por entender minhas escolhas e apoiá-las. Por ser, acima de tudo, meu melhor amigo. Por dividirmos a mesma jornada e, com ela, nossas conquistas e desafios.

À Prof^a. Dr^a. **Janete Maria Cerutti**, orientadora e amiga, pelo exemplo profissional. Pela oportunidade e confiança, e pela indispensável participação em meu aprendizado. Agradeço a Deus pelo maravilhoso dom da vida, por cuidar dos meus passos, me mostrar a direção a seguir e sempre me proteger, por ser minha alegria verdadeira e conforto em momentos difíceis, por ser tão presente na minha família e em minha vida, e por me mostrar quais são os reais valores da vida.

A todos meus familiares, avós, tios e primos, e principalmente aos meus pais, pelo amor e dedicação incondicionais; aos meus irmãos, Edmar Junior e Julia, simplesmente pelo fato de existirem em minha vida e serem tão importantes nela, por me permitirem ajudar e cuidar, por serem meus amigos de sangue; à minha madrinha Nadir e minhas tias Ivanete e Patrícia, por se preocuparem comigo e com meu futuro profissional. Em especial à tia Janete, por ser, além de tudo, minha grande amiga.

Aos amigos do Laboratório de Bases Genéticas dos Tumores da Tiroide, André, Bruno, Maria José e Marta, pelo auxílio e aprendizado, pelas discussões de protocolos e resultados, e pelos momentos, vários, de descontração; vocês são parte do meu crescimento profissional e pessoal. Agradecimento especial às queridas Maria Isabel e Larissa, pela amizade e convivência, dentro e fora do laboratório, vocês tornaram meus dias mais descontraídos. E, por último, não menos especial, muito pelo contrário, agradeço a Laís, por toda dedicação e empenho ao me auxiliar em experimentos, discussão de resultados e intermináveis análises estatísticas, com certeza, sem sua amizade e auxílio, esse trabalho não teria a mesma qualidade, muito obrigada!

Às minhas amigas Tamires Gustavo, Eriane Eller, Ana Beatriz Mansano, Bruna Miollo, Zaira Magalhães, Flávia Sato, Amanda Campos, Suane Sousa, Paloma Besson, Daniela Oliveira, Nágila Furlan e Rithiely Campos, por ouvirem meus desabafos, e dividirem suas novidades, conquistas e planos. Por se mostrarem perto, mesmo quando distante. Por serem tão especiais pra mim, cada uma do seu jeito, pela espontaneidade... Por toda diversão!

8

À Dr^a. Janete Cerutti, Dr^a. Maria Inez, Dr^a. Ileana Rubió, Dr^a. Gianna Carvalheira, Dr^a. Isabel Melaragno, Dr. Rui Maciel e Dr. Magnus, pelo exemplo, apoio e ensinamento, com certeza, contribuíram muito para meu aprendizado e crescimento.

A todos os amigos, médicos, alunos e funcionários do Laboratório de Endocrinologia Molecular e Translacional. Em especial à secretária Ângela por todas as conversas e conselhos, por me ouvir e se mostrar uma grande amiga.

Às amigas do Laboratório de Investigação das Neoplasias da Infância e Adolescência, Bruna Mascaro e Gabriela Rampazzo, pela descontração e apoio mútuo nas etapas do mestrado, disciplinas, congresso, qualificação e, por fim, dissertação. Por ouvirem minhas preocupações, partilharem as suas, e por nos ajudarmos para que tudo desse certo, e deu.

À Mariana Moyses e Renata Pellegrino pela disposição em auxiliar as análises pelos softwares e dúvidas.

À Fapesp, pela contribuição financeira através da concessão da bolsa de mestrado.

ARTN	artemina					
ATA	Sociedade Americana de Tiroide (do inglês <u>American Thyroid Association</u>)					
bp	pares de base					
CAT	carcinoma anaplásico da tiroide					
CFT	carcinoma folicular da tiroide					
CGH	hibridização genômica comparativa					
CMT	carcinoma medular da tiroide					
CMTF	carcinoma medular da tiroide familial					
CNV	variação no número de cópias de DNA					
CPT	carcinoma papilífero da tiroide					
СТ	calcitonina					
DECIDIED	<u>D</u> atabase of <u>C</u> hromosomal <u>I</u> mbalance and <u>P</u> henotype in <u>H</u> umans Using <u>E</u> nsembl					
DECIPHER	<u>R</u> esources					
DGV	Database of Genomic Variation					
DNA	ácido desoxirribonucleico					
FEO	feocromocitoma					
FISH	hibridização fluorescente in situ					
FMN2	<u>Formin 2</u>					
GFLs	<u>G</u> NDF <u>F</u> amily <u>L</u> igands					
GFRα	<u>G</u> DNF- <u>F</u> amily a <u>R</u> eceptors					
GNDF	<u>G</u> lial <u>N</u> eurotrophic <u>D</u> erived <u>F</u> actor					
HPTP	hiperparatiroidismo primário					

HSCR	Hirschsprung					
kb	quilobase					
kDa	quilodalton					
LCA	líquen cutâneo amiloidótico					
LCE3C	Late <u>C</u> ornified <u>E</u> nvelope <u>3C</u>					
LOH	perda de heterozigosidade					
NEM	síndrome de neoplasia endócrina múltipla					
NEM 2	síndrome de neoplasia endócrina múltipla do tipo 2					
NEM 2A	síndrome da neoplasia endócrina múltipla do tipo 2A					
NEM 2B	síndrome da neoplasia endócrina múltipla do tipo 2B					
NTN	neurturina					
OMIM	<u>O</u> nline <u>M</u> endelian <u>I</u> nheritance in <u>M</u> an database					
p21	proteína inibidora de CDKs (<u>Cyclin-Dependent K</u> inases)					
PCR	reação em cadeia da polimerase					
qPCR	PCR quantitativa					
PSP	persefina					
RET	<u>RE</u> arranged during <u>T</u> ransfection					
RMND5A	<u>R</u> equired for <u>M</u> eiotic <u>N</u> uclear <u>D</u> ivision <u>5</u> homolog <u>A</u>					
RTK	receptor tirosina quinase					
SNP	polimorfismo de único nucleotídeo					
ТК	tirosina quinase					
VNTR	número variável de repetições em tandem					

ÍNDICE

APRESENTAÇÃO	13
RESUMO	16
INTRODUÇÃO	17
REFERÊNCIAS BIBLIOGRÁFICAS	39
OBJETIVO	45
MANUSCRITO	46
DISCUSSÃO E CONCLUSÕES	47

Esta dissertação de mestrado, realizada com apoio financeiro da FAPESP (bolsa de mestrado 2011/04143-5, projetos 2006/60402-1, 2011/10787-2, 2012/02902-9), intitulase "Análise da Variação no Número de Cópias (CNV) de DNA em uma Família com Síndrome NEM 2A e Mutação p.G533C no Gene *RET*: Identificação de Regiões Associadas à Predisposição para o Desenvolvimento de Metástase aos Linfonodos".

De acordo com as recomendações da comissão especial do programa de Pós-Graduação em Biologia Estrutural e Funcional da Universidade Federal de São Paulo (UNIFESP), apresentamos um trabalho que será submetido à publicação na revista The Journal of Clinical Endocrinology & Metabolism (último fator de impacto: 5.967), que compõe essa dissertação de mestrado, conforme descrito:

<u>Aline N. Araujo</u>, Lais S. Moraes, Maria Inez C. França, Jin Lee, Renata Pellegrino, Rui M. B. Maciel and Janete M. Cerutti. Genome-Wide Copy Number Analysis in a Family with p.G533C *RET* Mutation and Medullary Thyroid Carcinoma Identified Regions Associated with Higher Predisposition to Lymph Node Metastasis.

Este trabalho foi desenvolvido no laboratório de Bases Genéticas dos Tumores da Tiroide, Disciplina de Genética, Departamento de Morfologia e Genética da UNIFESP e Laboratório de Endocrinologia Molecular e Translacional da Disciplina de Endocrinologia Molecular do Departamento de Medicina da UNIFESP.

Em estudos anteriores de nosso grupo foi descrita uma nova mutação no exon 8 do gene *RET* em uma família com carcinoma medular da tiroide familial, uma substituição do nucleotídeo guanina (G) por uma timina (T) na posição 1597, que corresponde à substituição do aminoácido glicina (Gly) pela cisteína (Cys) no códon 533 (1), mutação essa que foi relacionada à patogênese do carcinoma medular de tiroide (CMT) hereditário.

Apesar da forte correlação genótipo-fenótipo, que possibilitou o desenvolvimento de um consenso (2) com a estratificação de risco para CMT e condutas no diagnóstico e tratamento para a maioria das diferentes mutações identificadas no *RET*, tem sido observado uma variabilidade clinica dentro de famílias e entre as famílias com a mesma mutação no gene.

De fato, o seguimento dos pacientes desta família revelou grande heterogeneidade clínica em relação à idade de início e evolução da doença, apresentando indivíduos em idade tardia com CMT menos agressivo, bem como indivíduos em idade precoce que apresentaram o CMT e metástase aos linfonodos (1, 3), sugerindo que possíveis genes moduladores poderiam estar associados à heterogeneidade clínica observada. Além disso, identificamos um caso de feocromocitoma em um dos pacientes portadores da mutação p.G533C, que já havia sido diagnosticado com CMT, modificando, desta forma, a apresentação clinica para a síndrome da neoplasia endócrina múltipla do tipo 2A (NEM 2A)(4).

Recentemente, inúmeros trabalhos verificaram que a presença de variações no número de cópias (CNVs, do inglês <u>Copy Number Variations</u>) de segmento do DNA, principal fonte de variabilidade genética entre os indivíduos, podem influenciar no risco de desenvolvimento de câncer hereditário ou esporádico, bem como, na evolução da doença. Além disso, vários trabalhos investigam o papel dessas variações na etiologia

de doenças como diabetes, doença cardíaca, esquizofrenia, surdez e outras doenças complexas (5, 6).

Desta forma, acreditamos que a investigação da variação do número de cópias global nos pacientes portadores da mutação p.G533C no gene *RET* e variabilidade clinica, principalmente correlacionada à presença de metástase aos linfonodos, será importante para melhor compreendermos os mecanismos moleculares associados à patogênese e à progressão do CMT nesta família. Assim, inicialmente avaliamos a presença de CNVs nas amostras obtidas de sangue periférico (DNA constitucional) de portadores da mutação p.G533C com CMT e presença ou ausência de metástase aos linfonodos.

O carcinoma medular da tiroide (CMT) é um tumor maligno originado das células C da tiroide, que pode ocorrer na forma hereditária como parte da síndrome de neoplasia endócrina múltipla do tipo 2 (NEM 2), ou na forma esporádica. A forma hereditária é dividida nas variantes clínicas NEM 2A, NEM 2B e CMTF (carcinoma medular de tiroide familial). Nosso grupo identificou uma nova mutação no gene *RET* (p.G533C) em uma família com NEM 2A.

Uma vez que a análise de seguimento dos pacientes indicou uma heterogeneidade clínica entre os portadores da mutação, principalmente associada à presença de metástase aos linfonodos, tivemos como objetivo principal investigar se a variação no número de cópias (CNV) de DNA, presente na linhagem germinativa, poderia estar associada à predisposição ao desenvolvimento de metástase aos linfonodos nesta família, bem como correlacionar as CNVs candidatas as demais características clínicas. Para isso, quinze pacientes com CMT e mutação p.G533C foram divididos em dois grupos (com presença de metástase aos linfonodos, n=8, e sem metástase, n=7) e utilizados em um *screening* inicial, pela hibridização individual de seus DNAs em *chips* da plataforma *Genome-Wide Human SNP Array 6.0*.

A análise comparativa entre os grupos resultou na identificação de dez regiões contendo CNVs que poderiam estar associadas à agressividade do CMT. As regiões candidatas, algumas delas contendo genes, foram avaliadas pela reação em cadeia da polimerase quantitativa (qPCR) em um grupo maior de pacientes da família (n=26), composto pelos indivíduos iniciais além de mais membros da família.

A validação dessas CNVs por qPCR identificou uma CNV associada ao fenótipo mais agressivo, observado pela presença de metástase aos linfonodos na família (Teste Exato de Fisher; $\leq 0,05$), que foi a deleção de *FMN2* (*P*=0,0362). As regiões alteradas também foram analisadas em conjunto, e o Teste Exato de Fisher revelou que as alterações (deleções/amplificações) foram diferentes entre o grupo metastático e não metastático (*P*=0,0310), sendo que as deleções no grupo metastático (*P*=0,0375) apresentaram-se como mais determinantes para diferenciar os grupos do que as amplificações no grupo não metastático (*P*=0,0873).

Os dados sugerem que, em geral, as deleções foram mais associadas ao pior curso clínico da doença, como observado pela CNV de deleção que contém o gene *FMN2* que foi relacionada à presença de metástase aos linfonodos, assim como o agrupamento das regiões deletadas foi também relacionado à metástase. Em conclusão, nós definimos um perfil de CNVs relacionadas à presença de metástase aos linfonodos em pacientes com CMT e mutação p.G533C no gene *RET*.

Tumores da Tiroide

A glândula tiroide é composta por dois tipos celulares, células foliculares, que compõem a maior parte do epitélio e são responsáveis pela produção dos hormônios da tiroide, e células C ou parafoliculares, responsáveis pela síntese do hormônio calcitonina (7). Os carcinomas da tiroide podem ser originados de algum desses tipos celulares, sendo que mais de 95% são derivados das células foliculares e a minoria é originada das células C. Os carcinomas derivados das células foliculares são divididos com base em parâmetros clínicos e histológicos em: bem diferenciado, que inclui o carcinoma papilífero da tiroide (CPT) e o carcinoma folicular da tiroide (CFT), indiferenciado que inclui o carcinoma anaplásico da tiroide (CAT), e pobremente diferenciado que inclui os carcinomas com morfologia e comportamento intermediários ao bem diferenciado e o indiferenciado (7). Já o carcinoma originado das células C, corresponde ao carcinoma medular da tiroide (CMT).

Uma vez que as células C se originam das células da crista neural, o CMT pertence ao grupo dos tumores neuroendócrinos, diferentemente dos outros tumores bem diferenciados da tiroide (CPT e CFT) que derivam de células foliculares originadas na endoderme (8).

As neoplasias da tiroide possuem padrões de metástases distintos. O carcinoma papilífero da tiroide (CPT) preferencialmente se dissemina para linfonodos, o carcinoma folicular da tiroide (CFT) para órgãos distantes, e o carcinoma medular da tiroide (CMT) metastatiza para linfonodos e órgãos distantes (9). O CMT possui um pior prognóstico devido ao seu comportamento biológico mais agressivo, quando comparado aos outros tumores bem diferenciados da tiroide (8).

Carcinoma Medular da Tiroide

O carcinoma medular da tiroide (CMT) ocorre com a frequência aproximadamente igual entre mulheres e homens (7, 10). Corresponde a 5-10% de todas as neoplasias da glândula, e pode ocorrer sob a forma esporádica (75-80%) ou hereditária (20-25%) (2). Na forma hereditária, o CMT é transmitido como um caráter autossômico dominante com penetrância quase completa e expressividade variada, podendo apresentar-se isoladamente na forma de carcinoma medular da tiroide familial (CMTF) ou como parte das síndromes de neoplasia endócrina múltipla (NEM) do tipo 2A ou 2B (11).

A NEM 2A, o subtipo clínico mais comum da síndrome NEM 2, é definida pela presença da tríade de características: CMT, feocromocitoma (FEO), e hiperparatiroidismo primário (HPTP) (2, 12). Outras associações raras com NEM 2A incluem a presença de líquen cutâneo amiloidótico e a doença de Hirschsprung.

A definição de CMTF pode seguir um critério mais rígido, que consiste na presença de CMT em pelo menos duas gerações e ausência de feocromocitoma ou hiperparatiroidismo, e uma definição menos rígida, que consiste na presença de CMT em quatro indivíduos de uma família sem outras manifestações de NEM 2A (2, 12).

A NEM 2B é a forma mais rara e agressiva (2), sendo definida pela presença de CMT, feocromocitoma, ganglioneuromatose e hábitos marfanóides. A idade típica de acometimento da síndrome é na terceira e quarta décadas de vida nas formas de NEM 2A e CMTF, e, mais precocemente, na NEM 2B, geralmente na primeira década de vida (2) (**Tabela 1**).

Subtipo	СМТ	FEO	HPTP	Doenças Associadas
CMTF	90-100%	-	-	-
NEM 2A	90-100%	50%	20-30%	Líquen Cutâneo Amiloidótico Doença de Hirschsprung
NEM 2B	90-100%	50%	-	Ganglioneuromatose Hábito Marfanóide

Tabela 1. Manifestação Clínica e Prevalência nas variantes da Síndrome NEM 2 (2, 10, 13)

CMT, carcinoma medular da tiroide; FEO, feocromocitoma; HPTP, hiperparatiroidismo primário; CMTF, carcinoma medular da tiroide familial; NEM, neoplasia endócrina múltipla do tipo 2.

Embora o padrão de herança seja autossômico dominante, a apresentação da NEM 2 é bastante heterogênea, sendo geralmente o CMT a primeira manifestação clínica. Estes dados sugerem que as células C são mais susceptíveis à transformação induzida por *RET* do que as células da adrenal (feocromocitoma) e da paratiroide (hiperparatiroidismo) (11).

Diferenciar NEM 2A de CMTF tem sido um desafio na literatura. O CMTF é uma variante clínica de NEM 2 que ocorre pela presença de CMT isolado sem evidencia de outras doenças endócrinas (2, 12). Uma vez que o feocromocitoma pode ser diagnosticado tardiamente na forma NEM 2A, pode–se ter dificuldade em caracterizar uma família como portadora de CMTF ou NEM 2A com baixa penetrância de feocromocitoma (12).

Gene RET e Proteína

A forma hereditária do CMT está associada a mutações de ganho de função no gene *RET*. O gene *RET* está localizado no braço longo do cromossomo 10, *locus* 10q11.2 (43,572,516bp - 43,625,798bp), tendo aproximadamente 51 Kb e sendo constituído por 21 exons que codificam uma proteína transmembrana (RET), membro

da superfamília de receptores tirosina quinase (RTK, do inglês <u>Receptor Tyrosine</u> <u>Kinase</u>). A expressão de RET ocorre principalmente durante o desenvolvimento dos sistemas nervoso e urogenital e, em adultos, tem sido observada no cérebro, timo, sistema entérico periférico, neurônios sensoriais e testículo (11).

A proteína RET é constituída por três domínios: 1. domínio extracelular com quatro repetições semelhantes à caderina (*cadherin-like*) e uma região altamente conservada rica em cisteínas, com 27 cisteínas; 2. domínio transmembrana e; 3. domínio intracelular que contêm dois subdomínios tirosina quinase (TK, do inglês *Tyrosine Kinase*) divididos pela inserção de 27 aminoácidos. Os subdomínios são denominados TK1 e TK2 e contribuem para a sinalização mediada por RET (**Figura 1**) (11).

Devido ao processo de *splicing* alternativo na região 3' podem ser identificadas três isoformas da proteína, que diferem em sua região C-terminal pela presença de 9, 51 e 43 resíduos. As isoformas RET 9 e RET 51, de 1072 aminoácidos (RET 9; NM_020630.4) e 1114 aminoácidos (RET 51 NM_020975.4) respectivamente, são as principais variantes identificadas (11), sendo geralmente coexpressas e predominantes. Além dessas, uma terceira isoforma que contêm 1.106 aminoácidos (RET 43) também pode ser gerada (**Figura 1**).



Figura 1. Representação esquemática de RET tirosina quinase. A proteína é constituída por uma região extracelular com quatro repetições semelhantes à caderina e um domínio rico em cisteína, uma região transmembrana única, e dois domínios tirosina quinase (TK1 e TK2) localizados na região intracelular. As três isoformas de RET (9, 43 e 51) estão indicadas. aa, aminoácidos. Figura adaptada da revisão de Groot *et a*l. (11).

Alguns estudos têm demonstrado que podem existir diferenças funcionais entre as isoformas e que seus níveis de expressão podem variar durante o desenvolvimento embrionário e pós-natal. A proteína RET, após modificação por glicosilação, apresenta um peso molecular de 170 kDa, embora ainda não esteja claro a qual das isoformas este peso molecular está associado.

Mecanismos de Ativação de RET e Vias de Sinalização

Sob condições fisiológicas normais, a ativação da proteína RET depende de um complexo de ligantes e correceptores pertencentes a dois grupos de proteínas: os ligantes GFLs (do inglês <u>GNDF Family Ligands</u>) que incluem o GNDF (do inglês glial neurotrophic derived factor), a neurturina (NTN), a artemina (ARTN) e a persefina

(PSP); e correceptores glicosil-fosfatidilinositol, designados GFR α (do inglês *GDNF-family* α *receptors*), GFR α 1, GFR α 2, GFR α 3, e GFR α 4. Assim, um dos quatro GFLs se liga a um dos GFR α para formar um complexo GFR α /GFL (11). GDNF se associa preferencialmente ao correceptor GFR α -1, neurturina utiliza GFR α -2, artemina utiliza GFR α -3 e persefina utiliza GFR α -4 (14).

A associação de RET ao complexo GFR α /GFL pode ocorrer sob a forma *cis* ou *trans* (**Figura 2**). O modelo *cis* sugere que GFL se liga inicialmente ao GFR α , que se encontra ancorado a membrana no micro-domínio lipídico denominado "jangadas lipídicas". Subsequentemente, o complexo atrai duas moléculas da proteína RET resultando na fosforilação das tirosinas intracelulares e ativação da cascata de sinalização. O modelo *trans* sugere que o GFL pode se associar a GFR α na forma solúvel, não associada à membrana, para formar o complexo GFL/GFR α . O complexo se associa a proteína RET, que se encontra ligada a membrana na forma inativa (monômeros fora do micro-domínio lipídico), promovendo a sua dimerização e ativação da cascata de sinalização (11). O mecanismo de ativação depende do estado de ativação de RET, tipo de complexo formado, processo envolvido e do tecido.



Figura 2. Ativação da proteína RET por diferentes mecanismos, sugerido pelo modelo *cis* (A) e *trans* (B). Figura adaptada da revisão de Groot *et a*l. (11).

Uma vez ativado, resíduos específicos de tirosinas no domínio intracelular tornam-se autofosforilados, levando à subsequente fosforilação de outros substratos protéicos e consequente ativação de diversas cascatas de sinalização associadas à sobrevivência celular, diferenciação, proliferação, migração e quimiotaxia (11).

Mutações em RET e Síndrome MEN2 – Correlação Genótipo-Fenótipo

Em 1987, o lócus associado à NEM 2A foi localizado no cromossomo 10 (15). Em 1993, foi demonstrado que o gene *RET* está associado à predisposição à síndrome NEM 2A e CMTF, mais especificamente a mutações no gene *RET* presente nas células da linhagem germinativa (16). Subsequentemente, mutações ativadoras no gene *RET* foram associadas à predisposição herdada a NEM 2B (12, 17). Aproximadamente 95% dos pacientes com NEM 2A e NEM 2B, e 88% daqueles com CMTF possuem mutação no *RET* (2).

Mutações que afetam os exons 5-16 do *RET* foram identificadas em famílias com síndrome NEM 2 (**Figura 3**). Nos casos de NEM 2A e CMTF, as mutações mais frequentemente identificadas ocorrem no domínio rico em cisteínas e menos frequentemente no domínio intracelular (11). Nas famílias com NEM 2A, o códon 634 é o mais frequentemente afetado (85%), sendo a substituição C634R a mais comumente identificada (14); tal substituição nunca foi descrita em uma família com CMTF. Em aproximadamente 10% dos casos de NEM 2A as mutações causam a substituição das cisteínas C609, C611, C618, C620, C630 e C634 por outros aminoácidos. Os restantes 5% ocorrem em outros códons no domínio extracelular ou menos frequentemente no domínio intracelular.

As mutações em famílias com CMTF podem ocorrer no domínio extracelular ou no domínio intracelular, sendo principalmente distribuídas nos códons 609, 611, 618 e 620 (14). Outras mutações aparentemente específicas para CMTF estão nos códons 768, 790, 791, 804, 844, 891 e 912 do domínio intracelular (18). Duplas mutações no gene *RET* têm sido identificadas em pacientes com CMTF e NEM 2A, além de pequenas inserções nos códons 531, 532, 633, 634, 635 e 637 e deleções no códon 616 que também foram associadas ao fenótipo NEM 2A ou CMTF.

A NEM 2B é usualmente causado por mutações no subdomínio tirosina quinase, envolvendo o códon 918, com uma mutação específica M918T, em 95% dos casos e o códon 883 em 5% (11).



Figura 3. Sinopse das mutações identificadas no gene *RET* (exons 1-21), códons associados ao carcinoma medular da tiroide hereditário. Adaptado de Nikiforov *et al.* (10).

Todas as mutações encontradas em NEM 2A, NEM 2B e CMTF possuem um efeito de ganho de função (19), e desencadeiam o processo neoplásico por meio da ativação permanente do receptor, independente de ligante (14).

Mutações nos códons ricos em cisteínas, no domínio extracelular de RET, que correspondem à grande maioria de casos de NEM 2A e CMTF, ativam a proteína RET pela indução da dimerização constitutiva das moléculas (**Figura 4**). A proteína RET com cisteína mutada possui resíduos de cisteína não pareados que podem formar

ligações dissulfeto com outra molécula de RET mutada, ocasionando a dimerização sem a necessidade de ligante (10, 19).

A mutação no códon 918, que é associada à NEM 2B, identificada no domínio quinase intracelular, é responsável pela ativação constitutiva de RET na forma monomérica (**Figura 4**). Essa mutação ocasiona a alteração conformacional no sítio de ligação do domínio tirosina quinase, que resulta na alteração da especificidade do substrato de RET. Alguns casos de CMTF também possuem mutações no domínio tirosina quinase, mas os mecanismos de ativação associados a essas mutações ainda não estão bem definidos (10, 19).



Ativação Neoplásica de RET

Figura 4. Mecanismo neoplásico de ativação de RET, mostrando a ativação por dimerização constitutiva em NEM 2A e CMTF e ativação monomérica em NEM 2B. Adaptado de Nikiforov *et al.* (10).

Mutações em RET e Implicações Clínicas - Estratificação de Risco

Desde o 7th International Workshop on Multiple Endocrine Neoplasia, 1999, existe um consenso mundial (13) de que a decisão de submeter um indivíduo à tiroidectomia profilática deve ser baseada, predominantemente, na presença da mutação no gene *RET* e estratificação dessa informação genética. Esta resolução é fundamentada no fato de que as mutações em *RET* apresentam forte correlação genótipo-fenótipo, e que a análise de DNA resulta em menor número de falsos negativos e falsos positivos que os testes bioquímicos previamente utilizados para diagnóstico e tratamento do CMT. Assim, as mutações no *RET* foram inicialmente estratificadas em três níveis de risco para CMT (níveis 1-3). As três categorias foram inicialmente definidas de acordo com o códon afetado e predizem a variante clínica, idade de manifestação e agressividade do CMT (13).

Mais recentemente, com base nos dados da literatura, no conhecimento e na experiência clínica de um grupo de especialistas, em junho de 2009 a Sociedade Americana de Tiroide (ATA, <u>American Thyroid Association</u>) publicou um consenso para condutas no diagnóstico e tratamento dos pacientes com CMT (2). De acordo com este consenso, o risco de agressividade para o CMT na síndrome NEM 2 pode ser dividido em quatro níveis (A-D) (**Tabelas 2 e 3**). O nível D inclui as mutações com maior risco de desenvolvimento de formas agressivas e precoces do CMT (inclui preferencialmente as mutações no códon 918 e 883). O nível C inclui as mutações no códon 634 que foram classificadas com alto risco de desenvolvimento da forma agressiva do CMT. O nível B engloba as mutações com risco de desenvolvimento de formas agressivas de CMT inferior ao nível C (inclui preferencialmente as mutações no

códon 609, 611, 618, 620, 630) e o nível A corresponde as mutações com o menor risco de desenvolvimento de CMT agressivo (**Tabelas 2 e 3**).

Algumas mutações descritas na **Tabela 2**, embora classificadas neste consenso da ATA, ainda não possuem consistência na literatura quanto à idade de indicação da tiroidectomia profilática. Como, por exemplo, as mutações G321, C515S, G533C, R600Q, K603E, Y606E, D631Y, S649L, K666E E768D e N777S.

Além das mutações mencionadas na **Tabela 2**, algumas mutações raras ou novas mutações têm sido descritas (20). Algumas dessas novas mutações foram associadas à CMTF ou NEM 2A, enquanto outras representam alelos de baixa penetrância ou alelos modificadores que conferem baixo risco de desenvolvimento de CMT. Ainda, outros podem representar polimorfismos sem significado clínico (21).

Um grande número de estratégias, incluindo análise de segregação, análise *in silico* (22), análise de associação, bem como, estudos funcionais podem ser utilizados para determinar a significância destas variantes. Além dessas análises, um seguimento a longo prazo de famílias com estas mutações menos prevalentes é fundamental.

, , , , , , , , , , , , , ,		Nivel		NEM 2A				NEM 2B		
Mutação	Exon	Risco	CMTF	CMT	HPTP	FEO	LCA	HSCR	CMT	FEO
G321R	5	А	+	MA	-	_	—		-	-
531/9 duplicação de par de base	8	A	+	MA	—	—	—	-	-	-
532 duplicação	8	А	+	?	-	-	-	—	-	—
C515S	8	Α	+	MA	-		_	_		-
G533C	8	А	+	MA	-	R	-	—	—	-
R600Q	10	Α	+	MI	_	-	-		-	-
K603E	10	Α	+	MI			-	—	-	-
Y606C	10	Α	+	?	-	_	-	-	—	-
C609F/R/G/S/Y	10	В	+	MA	MI	R	-	+	—	—
C611R/G/F/S/W/Y	10	В	+	MA	MI	R	-	+	-	-
C618R/G/F/S/Y	10	В	+	MA	MI	MI	-	+		—
C620R/G/F/S/W/Y	10	В	+	MA	MI	MI	-	+	—	_
C630R/F/S/Y	11	В	+	MA	R	R		<u></u>		_
D631Y	11	В	+	?	-	<u> </u>	-	-		_
633/9 duplicação de par de base	11	В	+	MA	MI	MI	-	_	-	-
C634R	11	С	-	MA	МІ	MA	+	_	_	-
C634G/F/S/W/Y	11	С	+	MA	MI	MA	+	-	-	_
634/12 duplicação de par de base	11	В	+	MA	MI	-	-	-	-	-
635/ inserção ELCR;T636P	11	A	+	MA	-	-	-		-	-
S649L	11	Α	+	MI	R	<u> </u>	-	_	<u> </u>	_
K666E	11	А	+	MI/MA	-	MI	_	<u> </u>	—	_
E768D	13	Α	+	MA	R	R	_	-	_	
N777S	13	А	+	MI	-	-	-	_	—	-
L790F	13	Α	+	MA	R	R/MI	_	_	-	
Y791F	13	А	+	MA	MI	MI	-	_	-	-
V804L	14	Α	+	MA	MI	R	-	—		_
V804M	14	А	+	MA	R	R	<u> </u>			<u> </u>
V804M+V778I	13/14	В	+	MA	_	_	_	_		_
V804M+E805K	14	D		-	-			·	MA	MA
V804M+Y806C	14	D		0	_		<u></u>	·	MA	MA
V804M+S904C	14/15	D		<u> </u>	MI	-	-	-	MA	-
G819K	14	Α	+	?		_	_	-		_
R833C	14	A	+	?	—	—	-	—	-	-
R844Q	14	Α	+	?	-	_	_		-	_
R866W	15	A	+	MA	-	-	-	-	-	-
A883F	15	D	-	_	-		_	-	MA	MA
S891A	15	А	+	MA	R	R		-	-	-
R912P	16	A	+	MI	_	-	-	-		-
M918T	16	D	_	_	_	_	_	_	MA	MA

Tabela 2. Correlação genótipo-fenótipo e níveis de risco para CMT. Adaptado de Kloos *et al.* 2009, consenso da Sociedade Americana de Tiroide (ATA) (2).

CMTF, carcinoma medular da tiroide familial; HPTP, hiperparatiroidismo primário; FEO, feocromocitoma; LCA, líquen cutâneo amiloidótico; HSCR, Hirschsprung; MA, maioria, MI, minoria, R, raro.

A região destacada consiste na mutação existente na família em estudo (p.G533C).

Nível de Risco	Idade de Teste de <i>RET</i>	Idade do Primeiro Ultrassom Exigido	Idade da Primeira Dosagem de Calcitonina Sérica	Idade da Cirurgia Profilática		
D	O mais rápido possível, dentro do primeiro ano de vida.	O mais rápido possível, dentro do primeiro ano de vida.	Seis (6) meses, se a cirurgia não foi realizada.	O mais rápido possível, dentro do primeiro ano de vida.		
С	< 3-5 anos	> 3-5 anos	> 3-5 anos	Antes dos 5 anos de idade.		
В	< 3-5 anos	> 3-5 anos	> 3-5 anos	Considerar cirurgia antes dos 5 anos. Pode ser adiada para além dos 5 anos, se alguns critérios rigorosos forem cumpridos**.		
A	< 3-5 anos	> 3-5 anos	> 3-5 anos	A cirurgia pode ser adiada para além dos 5 anos, se alguns critérios rigorosos forem cumpridos**.		

Tabela 3. Estratificação de Risco, Indicações dos Testes e Tiroidectomia Profilática

**Níveis normais de Calcitonina sérica basal/estimulada, ultrassom cervical anual normal, história familiar de CMT indolente e desejo da família.

Calcitonina como marcador de CMT

A calcitonina (CT) é um hormônio tecido-específico, produzido quase exclusivamente pelas células C ou parafoliculares da tiroide. Sua secreção pode ser estimulada de acordo com a concentração extracelular de cálcio ou, ainda, por substâncias exógenas como a pentagastrina, utilizada em testes de estímulo. Esse hormônio é produzido em excesso no CMT, uma vez que o tumor é resultado da hiperplasia das células C produtoras de calcitonina, que não perdem suas características bioquímicas e patológicas (8). Além da realização do exame físico, o diagnóstico de CMT pode ser suspeitado pela combinação a níveis aumentados de calcitonina basal ou após testes de estímulo específicos. Desta forma, a calcitonina pode ser considerada um importante marcador tumoral para diagnóstico e seguimento de pacientes com CMT (12), considerando que níveis elevados poderiam indicar a presença de tumor primário ou metástase.

Identificação da mutação p.G533C no gene *RET* em uma família com NEM 2A: história e dados do grupo

A família analisada neste trabalho, descrita pelo nosso grupo em 2003 (1), é de origem caucasiana, com ascendentes de Barcelona, Espanha, que imigraram para o Brasil no século 19 e residem na região sudeste do Brasil. Em 2003, a família de 6 gerações apresentava um total de 229 indivíduos, sendo que o sequenciamento do *RET* a partir do DNA genômico dos membros da família revelou que 76 deles apresentavam uma mutação no exon 8 do gene *RET* que leva à substituição de uma glicina por uma cisteína no códon 533 (p.G533C) (1), enquanto que 153 não eram portadores da mutação.

A mutação foi identificada como causadora do CMT e, inicialmente, a família foi diagnosticada como CMTF, de acordo com o consenso para NEM (13), pois não foi observada nos membros da família nenhuma evidencia clínica ou bioquímica de feocromocitoma, doença da paratiroide ou outras características clínicas atribuídas a NEM 2.

Essa mesma mutação foi descrita em duas famílias gregas com CMTF (18). Entretanto, outro estudo identificou feocromocitoma bilateral em um paciente com mutação p.G533C na linhagem germinativa do *RET* como a primeira manifestação clínica. O exame histológico realizado após a tiroidectomia mostrou a presença de CMT, indicando que pacientes com a mutação p.G533C também tem predisposição para NEM 2A (23).

Recentemente, nosso grupo identificou a presença de feocromocitoma em um dos membros da família estudada, confirmando que a mutação p.G533C ocasiona o fenótipo NEM 2A (4).

A evolução clínica do CMT tem sido variável entre os membros dessa família. Observou-se uma variabilidade fenotípica em relação à apresentação clínica da doença, visto que alguns pacientes apresentaram CMT numa idade mais avançada e com menor agressividade, enquanto alguns pacientes manifestaram CMT em idade mais precoce e com metástase aos linfonodos (1).

Embora o conhecimento sobre o gene causador e a patogênese do CMT seja significativo, algumas particularidades da doença ainda são pouco compreendidas, como os fatores que poderiam determinar ou influenciar a idade de acometimento e a agressividade tumoral, demonstrando, assim, a necessidade de avaliar possíveis moduladores de tal variabilidade.

Nos últimos anos, diversos autores têm investigado se a presença de sequências variantes ou SNPs (do inglês, <u>Single Nucleotide Polymorphisms</u>) poderiam estar associados à susceptibilidade para o desenvolvimento ou diferente evolução do CMT. Inclusive nosso grupo, com o objetivo de compreender a heterogeneidade clínica associada à idade de início e agressividade tumoral do CMT, tem investigado se a presença de SNPs no gene *RET* pode atuar como modificadores nesta família com mutação p.G533C (3) ou ainda associado à mutação C634Y/Y791F no gene *RET* (5, 24).

A variabilidade fenotípica pode ser modulada por vários fatores, entre eles a própria variabilidade genômica, que inclui os SNPs e variação no número de cópias de DNA (CNVs, do inglês <u>Copy Number Variations</u>) (5). Portanto, além dos SNPs identificados no gene *RET*, as CNVs poderiam representar uma importante fonte de variabilidade que pode influenciar o curso clínico do CMT.

Variação no Número de Cópias de DNA (CNVs)

A variabilidade genômica pode ocorrer de muitas formas, incluindo os SNPs, pequenos polimorfismos de inserção-deleção, sequencias repetitivas de números variáveis (VNTRs, do inglês <u>Variable Number of Tandem Repeats</u>) e alterações genômicas estruturais (5). As alterações estruturais se referem a alterações genômicas de DNA que usualmente abrangem mais de 1000 bases e incluem variações no número de cópias (CNVs), o tipo mais prevalente, e, menos comum, inversões, inserções e translocações cromossômicas (25).

A CNV é, então, definida pela duplicação ou deleção de um segmento de sequencia de DNA maior que um kilobase (1 kb) quando comparada a um genoma de referencia (5, 26-28). Essa definição de CNVs não está relacionada ao impacto clínico de um dado desequilíbrio genômico (29).

As duplicações gênicas foram primeiramente identificadas na patogênese da doença "Charcot-Marie-Tooth" do tipo 1 em 1991, onde a amplificação no número de cópias do gene *PMP22* mostrou ser suficiente para causar a doença (30).

Acreditava-se que essas regiões de variação eram raras. De fato, quando o Projeto Genoma Humano foi publicado, as principais variações observadas no genoma de indivíduos normais foram os SNPs. Até então, os SNPs eram reconhecidos como a maior fonte de variação genética e, consequentemente, maior influência para a variação fenotípica normal (31). Entretanto, mais recentemente, análises individuais de genomas humanos normais revelaram quantidades inesperadas de CNVs (27, 28) e, desde então, essas observações têm sido replicadas e expandidas.

Desde a descoberta de que as CNVs estariam presentes e espalhadas no genoma, muitos estudos têm mostrado que essas alterações podem representar uma grande porção da diversidade genética. Redon e colaboradores (32) estudaram 270 indivíduos sadios de quatro populações com ancestralidade na Europa, África e Ásia. Utilizando duas plataformas, os autores identificaram 1.477 CNVs, sendo que, aproximadamente metade destas foram detectadas em múltiplos indivíduos. Os autores construíram o primeiro mapa de CNVs, mostrando que essas alterações estão distribuídas pelo genoma e cobrem aproximadamente 12% dele (360 Mb). Entre as CNVs identificadas, Redon et al., verificaram que 285 continham genes associados à doenças humanas e que estão listadas no OMIM (<u>Online Mendelian Inheritance in Man database</u>), sugerindo um possível papel destas CNVs em doenças mendelianas e também a doenças complexas.

Muitas CNVs foram relatadas de afetar a susceptibilidade a doenças. Estudos identificaram associações entre CNVs e diversas doenças, entre elas estão distúrbios neurológicos complexos (doença de Parkinson, doença de Alzheimer, retardo mental e distúrbios do desenvolvimento, autismo, esquizofrenia, doença bipolar), doenças cardiovasculares, doenças infecciosas e autoimune, doenças metabólicas (diabetes do tipo 2, excesso de peso e obesidade, gene da amilase), câncer e outras diversas doenças comuns (susceptibilidade a HIV/AIDS, doença de Crohn, psoríase, lúpus eritematoso sistêmico e glomerulonefrite, asma) (33, 34). Os avanços no estudo das CNVs deixam

clara a necessidade de que estudos genéticos de doenças complexas devam ter uma atenção na contribuição que essas CNVs possam ter na doença.

Uma vez que as CNVs podem conter parte ou todo o gene, ou ainda conter vários genes, elas podem influenciar diretamente a expressão gênica através de efeitos de dosagem, a partir do princípio de que mais cópias do gene resultam em maior expressão. Além disso, podem afetar indiretamente a expressão por incluir elementos regulatórios da transcrição, por exemplo, a deleção de um elemento que regula negativamente a expressão de um determinado gene pode levar a um aumento na sua transcrição. Aproximadamente 17,7% das CNVs estão associadas a alguma alteração na variabilidade da expressão gênica (35).

CNVs em geral são estáveis e podem ser herdadas, podendo ser desenvolvidas espontaneamente durante a meiose. Embora o mecanismo do surgimento de CNVs não esteja totalmente elucidado, deleções, duplicações, duplicações segmentares, inserções, inversões e translocações representam alguns dos processos resultantes em CNVs (33).

As CNVs podem ser detectadas usando uma variedade de plataformas incluindo hibridização genômica comparativa (CGH, do inglês <u>Comparative Genomic</u> <u>Hybridization</u>), array-CGH, SNP arrays de genoma completo e sequenciamento de nova geração. Sendo que as tecnologias mais novas são mais sensíveis à detecção de CNVs, além de demonstrar limites mais precisos das alterações. O Affymetrix Genome-Wide Human SNP Array 6.0, por exemplo, inclui 906.600 SNPs e 940.000 sequências para a identificação de CNVs.

Tendo em vista o papel das CNVs na determinação dos fenótipos de doenças que seguem o padrão de herança mendeliano, como é o caso da síndrome NEM 2, é fundamental caracterizar se alterações na arquitetura genômica podem estar associadas à heterogeneidade clínica observada em famílias com a síndrome. Em outras palavras, se as CNVs podem afetar a função biológica e evolução do CMT, ou ainda se a presença dessas alterações poderia estar associada à resistência a determinadas drogas, que ainda não se sabe.

CNVs e Carcinoma Medular da Tiroide

Alguns grupos tem avaliado o genoma por meio da análise de cariótipo, perda de heterozigosidade (LOH, do inglês, *Loss of Heterozygosity*) e CGH em busca de alterações no número de copias de DNA relacionadas à gênese e/ou progressão dos diferentes subtipos de tumores da tiroide. O primeiro estudo (36), utilizando CGH, identificou alterações no número de cópias em 80% dos casos (16/20), e relatou que as alterações foram mais prevalentes em CFT do que em CPT, sendo a deleção do cromossomo 22 particularmente comum em CFT. Estes dados foram confirmados por hibridização fluorescente *in situ* (FISH, do inglês *Eluorescence In Situ Hybridization*). Perdas e ganhos também foram identificados em CPT, CMT e Carcinoma Indiferenciado. Entretanto, pequenas deleções ou amplificações não puderam ser detectadas por CGH. Particularmente no CMT, a variação no número de cópias ocorreu preferencialmente em oito cromossomos, sendo que as perdas foram mais frequentes que ganhos. A perda do cromossomo 22 foi detectada em dois casos de CMT.

Com o intuito de identificar alterações relacionadas à progressão tumoral Marsh e colaboradores (37) avaliaram a presença de variações no número de cópias no CMT familial e esporádico e no sangue dos pacientes. Utilizando CGH, foram observadas alterações em 25% dos casos de CMT hereditário e 76% dos casos de CMT esporádicos. As alterações mais frequentes foram perdas nos cromossomos 3q26-q27,
1p, 4, 13q, 9q13-q22 e cromossomos 22q. Ganho no cromossomo 19 também foi observado em 10% dos casos.

Mais recentemente, Ye e colaboradores (38), utilizando *array*-CGH, identificaram alterações associadas à gênese do CMT. Os autores avaliaram 30 casos de CMT, sendo 10 esporádicos e negativos para mutações no gene *RET*, 10 casos esporádicos positivos para mutações no *RET* (*RET* M918T e *RET* A883F) e 10 casos hereditários positivos para *RET* (*RET* C620R, *RET* C634R e *RET* V804L, *RET* M918T). Os autores verificaram que 76/78 variações somáticas no número de cópias observadas eram perdas, evidenciando a significância da perda alélica e possível papel dos genes supressores de tumor na gênese do CMT. As perdas foram identificadas nos três grupos, sendo mais comuns as perdas no 12p12.31, 7q3613q12.11 e 19p13. Os autores também observaram grandes perdas no cromossomo 22q e 1p. A similar distribuição entre os três grupos e ausência de anormalidades associadas aos grupos de tumores hereditários, sugere um mecanismo comum associada à gênese do CMT esporádico e hereditário.

Ciampi e colaboradores (39) investigaram variações no número de cópias do cromossomo 10 do gene *RET* em uma série de pacientes com CMT esporádicos (n=53) e familiares (n=53), utilizando FISH e reação em cadeia da polimerase (PCR, do inglês <u>Polymerase Chain Reaction</u>) em tempo real. Os autores observaram que 27,7% (18/65) dos casos de CMT apresentavam amplificação de *RET* ou alteração do número de cópias do cromossomo 10 (aneuploidia). Entre estes, mistura de trissomia e tetrassomia de *RET* foram observados em sete casos e tetrassomia em cinco casos. Além disso, seis casos apresentavam amplificação de *RET* sem alterações no cromossomo 10. Nenhum caso apresentou perda do gene *RET* ou cromossomo 10.

Por fim, Flicker e colaboradores, por meio da técnica de *array*-CGH, investigaram tumores primários e metástases de CMT com o intuito de avaliar se o número de alterações no DNA aumenta em estágios tumorais avançados (40). Os autores identificaram maior número de alterações genéticas nas metástases, quando comparados ao tumor primário do mesmo paciente. Embora ganhos tenham sido raros nos tumores primários, foram frequentemente encontrados nas metástases, entre estes, os mais comuns foram ganhos nos cromossomos 6, 16, 19q, 19p, 7 e 15. Entre as perdas na metástase, as mais frequentemente encontradas envolvem os cromossomos 22q e 13q.

Desta forma, diversos estudos demonstraram o envolvimento de alterações cromossômicas/genômicas no CMT que podem facilitar na identificação de genes ativadores ou supressores do processo tumoral, entretanto essas análises são prejudicadas pela baixa resolução de muitas plataformas.

Até o momento nenhum trabalho na literatura analisou o genoma completo, por SNP *array*, e explorou a associação de CNVs com a heterogeneidade clínica observada dentro das famílias com CMT e mutações no gene *RET* e nas diferentes famílias com a mesma mutação. 1. Da Silva AM, Maciel RM, Da Silva MR, Toledo SR, De Carvalho MB, Cerutti JM. A novel germ-line point mutation in RET exon 8 (Gly(533)Cys) in a large kindred with familial medullary thyroid carcinoma. The Journal of clinical endocrinology and metabolism. 2003;88(11):5438-43.

2. Kloos RT, Eng C, Evans DB, Francis GL, Gagel RF, Gharib H, et al. Medullary thyroid cancer: management guidelines of the American Thyroid Association. Thyroid : official journal of the American Thyroid Association. 2009;19(6):565-612.

3. Tamanaha R, Camacho CP, Pereira AC, da Silva AM, Maciel RM, Cerutti JM. Evaluation of RET polymorphisms in a six-generation family with G533C RET mutation: specific RET variants may modulate age at onset and clinical presentation. Clinical endocrinology. 2009;71(1):56-64.

4. Oliveira MN, Hemerly JP, Bastos AU, Tamanaha R, Latini FR, Camacho CP, et al. The RET p.G533C mutation confers predisposition to multiple endocrine neoplasia type 2A in a Brazilian kindred and is able to induce a malignant phenotype in vitro and in vivo. Thyroid : official journal of the American Thyroid Association. 2011;21(9):975-85.

Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, et al.
 Copy number variation: new insights in genome diversity. Genome research.
 2006;16(8):949-61.

6. Zhang F, Gu W, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Annual review of genomics and human genetics. 2009;10:451-81.

7. Kondo T, Ezzat S, Asa SL. Pathogenetic mechanisms in thyroid follicular-cell neoplasia. Nat Rev Cancer. 2006;6(4):292-306.

8. Elisei R. Routine serum calcitonin measurement in the evaluation of thyroid nodules. Best Pract Res Clin Endocrinol Metab. 2008;22(6):941-53.

9. Machens A, Holzhausen HJ, Lautenschlager C, Thanh PN, Dralle H. Enhancement of lymph node metastasis and distant metastasis of thyroid carcinoma. Cancer. 2003;98(4):712-9.

10. Nikiforov YE, Biddinger PW, Thompson LDR. Medullary Carcinoma. In: Nikiforov YE, Biddinger PW, Thompson LDR, Nikiforova MN, Biddinger PW, editors. Diagnostic Pathology and Molecular Genetics of the Thyroid: a comprehensive guide for practicing thyroid pathology. 1st ed. Baltimore, MD: Lippincott Williams & Wilkins; 2009. p. 249-79.

11. de Groot JW, Links TP, Plukker JT, Lips CJ, Hofstra RM. RET as a diagnostic and therapeutic target in sporadic and hereditary endocrine tumors. Endocrine reviews. 2006;27(5):535-60.

12. Eng C, Mulligan LM. Mutations of the RET proto-oncogene in the multiple endocrine neoplasia type 2 syndromes, related sporadic tumours, and hirschsprung disease. Hum Mutat. 1997;9(2):97-109.

13. Brandi ML, Gagel RF, Angeli A, Bilezikian JP, Beck-Peccoz P, Bordi C, et al. Guidelines for diagnosis and therapy of MEN type 1 and type 2. The Journal of clinical endocrinology and metabolism. 2001;86(12):5658-71.

14. Santoro M, Melillo RM, Carlomagno F, Fusco A, Vecchio G. Molecular mechanisms of RET activation in human cancer. Ann N Y Acad Sci. 2002;963:116-21.

15. Mathew CG, Chin KS, Easton DF, Thorpe K, Carter C, Liou GI, et al. A linked genetic marker for multiple endocrine neoplasia type 2A on chromosome 10. Nature. 1987;328(6130):527-8.

16. Donis-Keller H, Dou S, Chi D, Carlson KM, Toshima K, Lairmore TC, et al. Mutations in the RET proto-oncogene are associated with MEN 2A and FMTC. Human molecular genetics. 1993;2(7):851-6.

17. Santoro M, Carlomagno F, Romano A, Bottaro DP, Dathan NA, Grieco M, et al. Activation of RET as a dominant transforming gene by germline mutations of MEN2A and MEN2B. Science. 1995;267(5196):381-3.

18. Kaldrymides P, Mytakidis N, Anagnostopoulos T, Vassiliou M, Tertipi A, Zahariou M, et al. A rare RET gene exon 8 mutation is found in two Greek kindreds with familial medullary thyroid carcinoma: implications for screening. Clinical endocrinology. 2006;64(5):561-6.

19. Asai N, Jijiwa M, Enomoto A, Kawai K, Maeda K, Ichiahara M, et al. RET receptor signaling: dysfunction in thyroid cancer and Hirschsprung's disease. Pathology international. 2006;56(4):164-72.

20. Muzza M, Cordella D, Bombled J, Bressac-de Paillerets B, Guizzardi F, Francis Z, et al. Four novel RET germline variants in exons 8 and 11 display an oncogenic potential in vitro. Eur J Endocrinol. 2010;162(4):771-7.

21. Schulte KM, Machens A, Fugazzola L, McGregor A, Diaz-Cano S, Izatt L, et al. The clinical spectrum of multiple endocrine neoplasia type 2a caused by the rare intracellular RET mutation S891A. The Journal of clinical endocrinology and metabolism. 2010;95(9):E92-7.

41

22. Cosci B, Vivaldi A, Romei C, Gemignani F, Landi S, Ciampi R, et al. In silico and in vitro analysis of rare germline allelic variants of RET oncogene associated with medullary thyroid cancer. Endocrine-related cancer. 2011;18(5):603-12.

23. Bethanis S, Koutsodontis G, Palouka T, Avgoustis C, Yannoukakos D, Bei T, et al. A newly detected mutation of the RET protooncogene in exon 8 as a cause of multiple endocrine neoplasia type 2A. Hormones (Athens). 2007;6(2):152-6.

24. Tamanaha R, Camacho CP, Ikejiri ES, Maciel RM, Cerutti JM. Y791F RET mutation and early onset of medullary thyroid carcinoma in a Brazilian kindred: evaluation of phenotype-modifying effect of germline variants. Clinical endocrinology. 2007;67(5):806-8.

25. Ionita-Laza I, Rogers AJ, Lange C, Raby BA, Lee C. Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. Genomics. 2009;93(1):22-6.

26. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome research. 2007;17(11):1665-74.

27. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nature genetics. 2004;36(9):949-51.

28. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. Science. 2004;305(5683):525-8.

29. Lee C, Iafrate AJ, Brothman AR. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nature genetics. 2007;39(7 Suppl):S48-54.

30. Lupski JR, de Oca-Luna RM, Slaugenhaupt S, Pentao L, Guzzetta V, Trask BJ, et al. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell. 1991;66(2):219-32.

31. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001;409(6822):928-33.

32. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444(7118):444-54.

33. Almal SH, Padh H. Implications of gene copy-number variation in health and diseases. Journal of human genetics. 2012;57(1):6-13.

34. Breheny P, Chalise P, Batzler A, Wang L, Fridley BL. Genetic association studies of copy-number variation: should assignment of copy number states precede testing? PloS one. 2012;7(4):e34262.

35. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science. 2007;315(5813):848-53.

36. Hemmer S, Wasenius VM, Knuutila S, Franssila K, Joensuu H. DNA copy number changes in thyroid carcinoma. The American journal of pathology. 1999;154(5):1539-47.

37. Marsh DJ, Theodosopoulos G, Martin-Schulte K, Richardson AL, Philips J, Roher HD, et al. Genome-wide copy number imbalances identified in familial and

43

sporadic medullary thyroid carcinoma. The Journal of clinical endocrinology and metabolism. 2003;88(4):1866-72.

38. Ye L, Santarpia L, Cote GJ, El-Naggar AK, Gagel RF. High resolution arraycomparative genomic hybridization profiling reveals deoxyribonucleic acid copy number alterations associated with medullary thyroid carcinoma. The Journal of clinical endocrinology and metabolism. 2008;93(11):4367-72.

39. Ciampi R, Romei C, Cosci B, Vivaldi A, Bottici V, Renzini G, et al. Chromosome 10 and RET gene copy number alterations in hereditary and sporadic Medullary Thyroid Carcinoma. Molecular and cellular endocrinology. 2012;348(1):176-82.

40. Flicker K, Ulz P, Hoger H, Zeitlhofer P, Haas OA, Behmel A, et al. Highresolution analysis of alterations in medullary thyroid carcinoma genomes. International journal of cancer Journal international du cancer. 2012;131(2):E66-73. Identificar variações no número de cópias (CNVs) de DNA em indivíduos de uma família com NEM 2A e mutação p.G533C no gene *RET*, utilizando a técnica de SNP *array*, e investigar se as CNVs candidatas podem estar associadas à agressividade do CMT.

Genome-Wide Copy Number Analysis in a Family with p.G533C *RET* Mutation and Medullary Thyroid Carcinoma Identified Regions Associated with Higher Predisposition to Lymph Node Metastasis

Aline N. Araujo¹, Lais S. Moraes¹, Maria Inez C. França², Jin Lee³, Renata Pellegrino³, Rui M. B. Maciel² and Janete M. Cerutti^{1,2}

Authors' Affiliations: ¹Genetic Bases of Thyroid Tumors Laboratory, Division of Genetics, Department of Morphology and Genetics, Universidade Federal de São Paulo, SP, Brazil; ²Laboratory of Molecular and Translational Endocrinology, Division of Endocrinology, Department of Medicine, Universidade Federal de São Paulo, SP, Brazil; ³Center for Applied Genomics, The Children's Hospital of Philadelphia, Research Institute, Philadelphia, PA, USA.

Requests for reprints should be addressed to:

Janete M. Cerutti, Ph. D. Genetic Bases of Thyroid Tumour Laboratory Rua Pedro de Toledo 669, 11° andar. Universidade Federal de São Paulo 04039-032, São Paulo, SP, Brazil Phone: +55-11- 5576-4979 E-mail: j.cerutti@unifesp.br

Acknowledgment: This work was supported by the São Paulo State Research Foundation (FAPESP) grant numbers 06/54922-2, 11/10787, 12/02902-9. ANA and LSM are scholars from FAPESP. JMC and RMBM are investigator of the Brazilian Research Council (CNPq).

Conflict of interest: The authors declare that they have no affiliations that would constitute a financial conflict of interest relating to the subject matter of this study.

Key Words: RET, MEN 2A, CNVs, p.G533C, MTC, Lymph node metastasis

Abstract

BACKGROUND: Medullary thyroid carcinoma (MTC) is a malignant tumor of parafollicular C cells which occurs as a part of inherited Multiple Endocrine Neoplasia Type 2 (MEN 2) syndrome or in a sporadic form. Our group identified a novel RET mutation (p.G533C) in 76 individuals from a family with MEN 2A. Clinical heterogeneity was observed among the p.G533C-carriers, mainly associated with lymph node metastasis. **OBJECTIVE:** To investigate whether copy number variations (CNVs), present in the constitutional DNA, are associated with higher predisposition to lymph node metastasis in this kindred, as well as to correlate the candidates CNVs to other clinical-pathological features. METHODS: Fifteen p.G533C carriers with MTC have been chosen to perform initial screening. They were divided into two groups according to presence (n=8) or absence of lymph node metastasis (n=7). DNA from peripheral blood was independently hybridized using Genome-Wide Human SNP Array 6.0 platform and the results were analyzed by Genotyping Console and PennCNV softwares. The cases (metastatic MTC) were compared to control (non-metastatic MTC) to eliminate non-pathogenic CNVs and find possible candidate regions associated to aggressiveness. The candidate regions were evaluated by qPCR in an extended cohort (n=26), composed of the initial individuals and additional family members. RESULTS: On the initial screening we identified ten CNVs possibly associated with presence of lymph node metastasis, some of them encompass nonannotated and annotated genes. The validation step by qPCR identified a CNV, the loss of FMN2 gene, associated with a more aggressive phenotype, observed as presence of lymph node metastasis in the family (Fisher's Exact Test; $\leq 0.05 P = 0.0362$). The altered regions were also analyzed as a group, and Fisher's Exact Test revealed that the alterations (losses/gains) are different between metastatic and non-metastatic groups (P=0.0310), being CNV deletion on metastatic group more important (P=0.0375) than CNV amplification on non-metastatic group (P=0.0873). CONCLUSION: The CNV deletion, which contains the FMN2 gene, found through genome-wide copy number analysis may be associated with higher predisposition to lymph node metastasis in this family, as well as the deletion of combined regions observed on metastatic patients.

Introduction

Medullary thyroid carcinoma (MTC) arises from the neural crest-derived calcitoninsecreting cells (C-cells) of the thyroid gland. It represents 3-10% of all thyroid cancer [1]. MTC can occur as a part of an autosomal dominant inherited Multiple Endocrine Neoplasia Type 2 (MEN 2) syndrome (20-25%) or in a sporadic form (75-80%) [2]. Three distinct clinical subtypes of MEN 2 have been characterized [3]. MEN 2A is characterized by the presence of MTC, pheochromocytoma (PHEO), and primary hyperparathyroidism (PHPT). MEN 2B is characterized by the presence of MTC, PHEO, mucosal and intestinal ganglioneuromatosis and marfanoid habitus. Familial MTC (FMTC) is characterized by a high predisposition to MTC in families with absence of other endocrinopathies related to MEN 2 [2, 3].

The *RET* (*<u>RE</u>arrangement during <u>Transfection</u>) gene, which maps on chromosome 10q11.2, contains 21 exons [4, 5] and encodes a tyrosine kinase transmembrane receptor which is primarily expressed in neural crest-derived cell lineages and urogenital progenitor cells during fetal development. Under normal conditions <i>RET* activation depends on the interaction of coreceptors (GDNF Family α Receptors) and their respective ligands (GDNF Family Ligands). Gain-of-function mutations can trigger aberrant activation of oncoprotein *RET*, and activating germline *RET* mutations have been associated with the inherited form of the disease [4, 6, 7]. In MEN 2A and FMTC, mutations occur more frequently on cysteine rich extracellular domain of *RET*, while MEN 2B is usually caused by mutations on tyrosine kinase intracellular domain [4].

It is well known that there is a strong genotype-phenotype correlation, as specific *RET* mutation correlates with the MEN 2 variant, the age of onset and aggressiveness of MTC [1]. In fact, using both available evidence from the literature and the expert opinion, the American Thyroid Association developed the guidelines for the diagnosis and management of MTC [2].

However, clinical heterogeneity has been observed among families or even within families with the same *RET* mutation and also in carriers of the same kindred. Specifically, in our previously reported study, we described a large Brazilian family with FMTC and *RET*

p.G533C mutation, 76 of whom were carrier of this mutation in exon 8 of *RET* gene [8]. This rare *RET* mutation was later found in two Greek kindreds with FMTC [9]. A more complex scenario has emerged from the fact that pheochromocytoma was identified as the first clinical manifestation in unrelated Greek families with *RET* p.G533C mutation. Follow-up of the patients confirmed the presence of MTC, providing the first evidence that *RET* p.G533C mutation leads to MEN 2A phenotype [10, 11]. Recently, we described a case of pheochromocytoma in one of the family members who was *RET* p.G533C carrier, endorsing that this mutation is associated with MEN 2A syndrome [12].

In addition to variability on phenotype among families with *RET* p.G533C mutation, considered variability on the clinical presentation has also been observed among genetically related individuals [8, 13]. This phenotypic variability was mainly related to the age at onset and aggressiveness of MTC (presence of lymph node metastasis). It has been speculated that the intra-familial phenotypic heterogeneity and milder phenotypes may be related to genetic modifiers.

Current studies have demonstrated that copy number variations (CNVs) are the most prevalent type of structural variation in the human genome and represent a significant source of genetic diversity [14]. CNVs are considered as DNA segment that are greater than one kilobase in size, which are present in variable copy number compared with a reference genome [15]. Although they were initially described as benign genetic changes that did not cause a clinically recognized phenotype, recently CNVs have been shown to contribute significantly to phenotypic variation and susceptibility to disease [16, 17]. CNVs are referred, therefore, as being definitely pathogenic, likely pathogenic, uncertain, likely not pathogenic or of little clinical significance, or not pathogenic or of no clinical significance [18]. Although CNVs are found to be common in the genome, their associations with prognosis factors such as presence of lymph node metastasis, tumor grade, tumor size and early age of onset, are yet understudied.

To asses if the presence of CNVs is associated with the clinical-pathological features of the MTC patients harboring the p.G533C *RET* mutation, a set of p.G533C-carriers with presence (*n*=8) or absence (*n*=7) of lymph node metastases were scanned for CNV using Affymetrix 6.0 SNP array. The CNVs detected from genome-wide analysis were confirmed in a larger number of patients using a quantitative PCR method. We here report an association of constitutional CNV with higher risk of lymph node metastasis, which support the hypothesis that CNVs may possibly help to differentiate phenotypic subtypes in family with similar *RET* mutations.

Subjects and Methods

Subjects

For the genome-wide copy number estimation, the casuistic was composed of patients from a family with the germline *RET* p.G533C mutation who developed MTC [8, 13]. Main criteria for sample incorporation into CNV analysis included clinical aggressiveness of the MTC, which was evaluated as the presence/absence of lymph node metastasis. Eligibility criteria included patients who had undergone total thyroidectomy and central neck dissection, were treated only on the basis of clinical and/or biochemical diagnosis (elevated calcitonin levels and/or a palpable thyroid nodule on physical examination or visualized by ultrasound) and followed up annually with clinical examination, basal/stimulated calcitonin, urinary catecholamine, serum ionized calcium and PTH measurements. Previously reported clinical and pathological features such as age at onset (defined as the age at which clinic-pathological diagnosis of MTC was made), gender, tumor size, local invasion (defined as invasion of adjacent structures), presence of lymph node metastasis, pTNM, and serum Calcitonin (sCT) were assessed [8, 12].

According to the above mentioned criteria, p.G533C carriers (n=15) were selected and divided into two groups. The first group comprised 8 metastatic MTC with a mean age of onset of 48.63 (SD \pm 13.05) and the second group included 7 non-metastatic MTC with a mean age of onset of 34.14 (SD \pm 6.12) (**Table 1**). The validation set included the initial individuals and additional family members, in a total of 9 p.G533C carriers with metastatic MTC with a mean age of onset of 47.44 (SD \pm 12.71) and 17 with non-metastatic MTC with a mean age of onset of 40.82 (SD \pm 13.70) (**Table 2**), with ages ranging from 21 to 62 years on metastatic MTC group, and 23 to 72 on non-metastatic MTC group (**Table 2**). Although the mean of both groups was different, as the 95% confidence interval (CI) for the age of onset in metastatic group [37.68; 57.21] and non-metastatic group [33.78; 47.87] overlapped, we assumed that there is no difference between the groups age. The approximately 10 years of patients' follow up indicated no MTC recurrence (lymph node metastasis) or distant metastasis in both groups.

Ethical approval for this study was obtained from the Federal University of São Paulo Research Ethics Committee (0305/11).

DNA Isolation

DNA was extracted from peripheral blood leukocytes using a standard phenol/chloroform method [19]. DNA was quantified using a NanoDrop spectrophotometer (NanoDrop, Thermo Fisher Scientific Inc., Waltham, MA), and the quality of each DNA sample was assessed by 1% agarose gel electrophoresis. Only high quality DNA was incorporated into CNV analysis.

Genotyping and Copy Number Variation Discovery

The Affymetrix Genome-Wide Human SNP Array 6.0 (Affy6) is an array platform that aims to perform both high-density SNP genotyping and high resolution CNV discovery simultaneously (Affymetrix, Santa Clara, CA, USA). This array is designed to interrogate 1,8 million genetic markers, including 906,600 single nucleotide polymorphisms (SNPs) and 946,000 probes for the detection of CNV. The copy number probe set includes 202,000 probes targeting 5,677 CNV regions from the Toronto Database of Genomic Variants and 744,000 encompass other probes spread through the genome.

Genomic DNA passed all the plataform's quality controls (QC) and was genotyped with Affy6. About 250 ng of DNA was digested by restriction enzymes, amplified, purified, labeled, fragmented and hybridized according to the manufacturer's protocol (Affymetrix, Santa Clara, USA). To avoid low-quality genotyping results, only samples that passed through the three Affy6 checkpoints were used. This allows assessing experimental error prior to hybridization on GeneChip arrays. Fluorescent intensities were quantified using the Affymetrix GeneChip Scanner 3000 7G, which is controlled by Affymetrix GeneChip Command Console.

Data analysis

To identify CNV potentially associated with higher risk of developing a more aggressive MTC in patients with MEN 2A and *RET* p.G533C carriers, metastatic MTC (cases) were compared with non-metastatic MTC (controls). For samples' data clustering and normalization we used Affymetrix 6.0 SNP data in 270 HapMap samples.

To maximize the detection rate of these CNVs, we used two freely available software packages (which are based on two different algorithms), that are commonly used for analyzing CNV obtained from the Affymetrix 6.0 SNP platform.

The complete dataset was initially analyzed by visual inspection using the Genotyping Console 4.1 and Chromosome Analysis Suite (ChAS) software (Affymetrix, Santa Clara, USA). Genotyping calls (SNPs and CNVs probes) were generated using Birdseed3 v1 or v2 algorithm. All samples passed the software's QC with the average call rate of 98.6%, higher than the default threshold of 86%. Calculation of copy number (CN) state was established on the Genotyping Console using the Canary algorithm. In order to identify gains or losses, each sample's copy number was estimated in comparison to HapMap diploid samples. To avoid the detection of random events, we considered as allelic alterations only those events (deletion/amplification) occurring in at least 3 out of 8 patients from metastatic MTC group (30% of cases) [15].

In addition, the dataset was also assessed using the PennCNV software (University of Pennsylvania, Philadelphia, and The Children's Hospital of Philadelphia, USA) [20]. Affymetrix CEL files were pre-processed according to protocols on Penncnv website (http://www.openbioinformatics.org/penncnv/penncnv_tutorial_affy_gw6.html) to calculate log R ratio (LRR) and B allele frequency (BAF). The Hapmap CEU samples genotyped on Affymetrix GW6 array were used with our samples to generate canonical genotype clusters. CNV calling was conducted via PennCNV. PennCNV is based on Hidden Markov Model (HMM) that correlates the found intensities with copy number (CN) state of each locus. PennCNV called CNVs that contained at least three SNP and/or CNV probes that were

observed to be deleted or duplicated in our dataset. The files affygw6.hmm and affygw6.hg18.pfb implemented in PennCNV were used in CNV calling. Also, we generated statistical p-values derived from Fisher's Exact test using PLINK and ParseCNV [21] scripts for comparison of copy number based allele frequency difference in cases versus controls.

To determine the copy number changes, we used the HMM model that performs segmentation of the log2 ratio intensity data and, for each segment, predicts copy number up to 5 states representing: homozygous deletion (CN state = 0), hemizygous deletion or single copy loss (CN state = 1), neutral copy number or normal diploid (CN state = 2), single copy gain (CN state = 3), and amplified (CN state = 4). (**Supplementary Figure 1**). Although the algorithms of the SNP array analysis allowed the prediction of copy numbers from 0 up to 4 states, to perform the statistical analysis the samples were classified as loss (CN state = 0 and 1), normal (CN = 2) and gain (CN state = 3 and 4).

Confirmation of germ-line CNV by quantitative PCR

The selected CNVs were validated by quantitative PCR (qPCR) in a larger series of patients (*n*=26) with MTC and *RET* p.G533C mutation (**Table 2**). Primers were designed in a CNV's common region for all the patients, so that the altered region is present in all patients that tested positive for that specific CNV (**Figure 1**). If the CNV location comprised a gene, the gene region was preferably chosen.

The PCR reaction was performed in triplicate using 30 ng of genomic DNA in a 12 μ l PCR volume containing SYBR Green PCR Master Mix 1X (PE Applied Biosystems, Foster City, CA, USA) and 3 pmol of each specific primer for target loci/gene. All PCR reactions were performed using the ABI Prism 7500 Sequence Detection System (PE Applied Biosystems) and followed the cycling conditions. For normalization, *ACTB* (beta-actin) was used as reference gene. Dissociation (melting) curves, performed for all the genes, allowed the detection of primer dimers or inespecific products, and were performed by denaturation at 95°C for 15 sec, followed by a gradual increasing ramp rate of the instrument from 60°C to 95°C. Copy number

was calculated using the comparative Ct method ($2^{(-\Delta Ct)}$, where $\Delta Ct = Ct_{test sample} - Ct_{reference control}$) [22]. Standard curves were initially performed to determine the efficiency of each primer reaction, using the established equation: $E = (10^{-1/slope} -1) \times 100$ [23]. A slope of -3.2 to -3.6 indicated efficiencies of 90-110%, which allowed the comparison between different reactions

Statistical Analysis

Two-tailed Fisher's Exact Test was performed to compare the qualitative clinical features between metastatic and non-metastatic group (gender, local invasion, MTC stage and pheochromocytoma occurrence), as well as to evaluate the association between observed CNV and metastasis, gender, local invasion and stage. For quantitative features Shapiro-Wilk test was used to verify the normality of distribution. As the samples did not show a normal distribution, two-tailed Mann Whitney test was performed to compare the features between metastatic and non-metastatic group (MTC's age of onset, median of tumor size, calcitonin level in preoperative and after total thyroidectomy), and to evaluate the CNV association with the median of tumor size, calcitonin level in preoperative and after total thyroidectomy in the groups with observed CNV. On combining analysis, Fisher's Exact Test (two-sided) was performed to analyze the difference between deletions and amplifications frequencies in metastatic and non-metastatic groups.

The software used for data analysis was StatView 4.5 (Abacus Concepts, Berkeley, CA, USA) and GraphPad Prism 5.01 (GraphPad Software, La Jolla, CA, USA). Only P-values<0.05 were considered statistically significant.

Results

Copy number variation identified

We evaluated separately the total number of CNVs detected on metastatic and non-metastatic groups. The total number of autosomal CNVs identified using Genotyping Console was 328, of which 251 were found on metastatic or non-metastatic group (**Table 3**). More deletions than amplifications were found across the chromosomes (**Figure 2**). The mean of CNVs per patient was 121.5 on metastatic and 125.6 on non-metastatic group, with a mean size of 28.7 kb on metastatic and 28.1 kb on non-metastatic.

The PennCNV analysis resulted in a total of 407 autosomal CNVs, of which 278 were found on metastatic group whereas 273 were found on non-metastatic. As observed in the Genotyping Console, more deletions than amplifications were found. The mean of CNVs per patient was 17.4 on metastatic and 19.5 on non-metastatic group, with a mean size of 261.3 kb on metastatic and 548.6 kb on non-metastatic.

Discovery of CNV associated with presence of lymph node metastasis

To analyze CNVs possibly related to lymph node metastasis we compared the metastatic group with the non-metastatic. Using Genotyping Console 4.1 and Chromosome Analysis Suite (ChAS) software (Affymetrix), based on GRCh37/hg19, we found 8 loci with CNV (**Table 4** and **Supplementary Figure 2**). The most common CNVs were losses (5/8), although 3 CNVs showed additional copies of sequence. The vast majority of candidate CNVs (6/8) contained genes, which may suggest a functional CNV.

Using PennCNV, based on NCBI36/hg18, we found 3 CNV deleted regions (**Supplementary Figure 2**), with deletions in at least 3 cases of metastatic MTC (P<0.05), which were not found in any case of non-metastatic MTC (0/7).

Within comparison of the two algorithms, concordance was observed for two CNVs, located at 1q31.3 and 2q22.3, which were identified using both Genotyping Console 4.1 and

PennCNV/ParseCNV.

Validation of CNV by qPCR in a larger set of p.G533C carriers with MTC

We next sought to validate the selected CNVs on 26 p.G533C carriers with MTC (**Table 2**) by qPCR method. We initially observed concordant results when comparing the changes estimated by the array platform with changes estimated by qPCR.

Although *LCE3C* and 2q22.3 were deleted in most patients with lymph node metastasis (7/9; 77.78%), only loss of *FMN2* correlated with presence of lymph node metastasis at significant levels (P=0.0362) (**Table 5 and Figure 3A**). These CNVs, although more frequently found on metastatic group, were also found on non-metastatic group.

We additionally correlated the candidate CNV with clinical-pathological features (gender, tumor size, local invasion, MTC stage, calcitonin level in preoperative and after total thyroidectomy) (**Supplementary Tables 2-18**). We found an association between loss of *LCE3C* and tumor size (P=0.0327) (**Supplementary Table 2 and Figure 3B**), and loss of *RMND5A* and sCT levels after total thyroidectomy (P=0.0439) (**Supplementary Table 13 and Figure 3C**).

Effect of combining multiples CNVs haplotypes

We additionally evaluated whether grouping multiple CNVs could show a specific pattern associated with a more aggressive MTC. To this end, we used the CN states obtained by qPCR of all patients (*n*=26) and clustered the CN information of each patient (**Figure 4**) using the CN states as follows: deletions (CN=0 or 1), normal (CN=2) and amplifications (CN=3 or 4). This allowed us to classify each patient as having more amplified or deleted CNVs (Patients' CNVs Consensus; **Figure 4**), making it easier to find similarities among patients from the same groups (metastatic and non-metastatic). The *RMND5A* gene was excluded from the cluster analysis, as it showed an opposite behavior when compared to all other genes, being amplified on metastatic group and deleted on the non-metastatic. The statistical analysis revealed a difference between

the two groups regarding deletions and amplifications (P=0.0310), and more importantly, it showed that the CNVs deletions observed on the metastatic group (P=0.0375) were statistically more determinant to this difference when compared to amplification on non-metastatic group (P=0.0873).

Discussion

Since clinical heterogeneity is a common feature observed among families with MEN 2A or even within families with the same *RET* mutation and also in carriers of the same kindred, in this study we investigate if CNVs present in constitutional DNA of a family with MTC and p.G533C *RET* mutation could be related to the clinical heterogeneity observed, specifically associated with lymph node metastasis.

Some studies were performed in thyroid tumors in order to evaluate alterations related to tumorigenesis. These studies, either realized by CGH or array-CGH, aimed to differentiate thyroid tumor types and, regarding DNA copy number changes on MTC, all of them identified a higher number of deletions than amplifications [15, 24, 25]. In our study, the same proportion was observed once we found more deletions (70%) than amplifications (30%) by SNP array analysis (**Table 4**). Other studies analyzed specific gene alterations on MTC [26, 27], or compared primary tumor and metastasis from MTC [28], but none compared peripheral blood DNA from patients with MTC and clinical heterogeneity.

The general analysis in metastatic and non-metastatic groups' genome, by Genotyping Console and PennCNV, revealed that there is almost no difference in number and size of CNVs between groups (**Table 3**). These results, associated to the distribution of CNVs from metastatic and non-metastatic groups across the chromosomes (Figure not shown), which also reveals few differences; suggest that in the genome as a whole there is no difference between groups.

In totality we found more CNVs using PennCNV (**Figure 2**), but considering a single sample (mean CNVs per patient on **Table 3**) more CNVs were found using Genotyping Console. Once the softwares are based on different algorithms, they are expected to detect a different number of CNVs, with more CNVs detected by Genotyping Console when compared to PennCNV [29].

Our initial screening was performed by a high resolution SNP array to find possible candidate CNVs' regions related to lymph node metastasis. The nine altered regions identified (**Table 4**) were then submitted to validation by quantitative PCR on the extended group, and some regions were found to be statistically related to the clinical features: loss of *FMN2* with metastasis (P=0.0362) (**Table 5 and Figure 3A**), loss of *LCE3C* with larger tumor size (P=0.0327) (**Supplementary Table 2 and Figure 3B**) and loss of *RMND5A* with lower serum calcitonin after total thyroidectomy (P=0.0439) (**Supplementary Table 13 and Figure 3C**).

The only identified CNV associated with lymph node metastasis was the loss of FMN2 (Formin 2) gene. This gene encodes several proteins thought to have a role in actin cytoskeletal organization and/or establishment of cell polarity on meiotic process, and are expected to localize in mitochondrion and nucleus [30]. The CNV, of 2.9 kb in size, is localized within the gene (461 kb) and was found deleted on metastatic group. These data combined to gene function allows us to speculate that the deletion could lead to a spliced variant protein, possibly nonfunctional, and then cause cellular disorganization that can contribute to the metastatic process. Interestingly, a recent study [31] identified a linkage region, including Fmn2 and Grem2 as major candidates genes, associated with colon tumor susceptibility on mice. This region is highly conserved between mice and humans, what corroborates to the idea that this gene, and its possible deletion, have an important role in tumorigenesis process. Other recent study involving FMN2 functional characterization [32] demonstrated the participation of FMN2 as a stabilizer of cell-cycle inhibitor p21 protein expression. Presence of FMN2 protein was found to inhibit p21 degradation, resulting in its increase and, consequently, cell-cycle arrest. Once the increased expression of p21 causes the cell-cycle arrest, deletion of FMN2 could result in p21 loss of expression and, consequently, cell-cycle progression. Even though the FMN2 gene was associated with tumor susceptibility on colon, no study has related this gene with the metastatic process. However, these data allows us to suppose that loss of FMN2 function could influence on metastasis, once its deletion could cause cellular disorganization and cell-cycle progression, which are features of an aggressive tumor. In our study we evaluated the aggressiveness as presence of lymph node metastasis, and *FMN2* deletion could be associated with predisposition to lymph node metastasis.

Another finding was the loss of *LCE3C* (Loss of Late Cornified Envelope Protein 3C) gene which was related with a larger tumor size. The deletion of this gene has previously been highly associated with psoriasis [33-35], due to its possible role in keratinization process [30]. As it is already a disease associated gene, it becomes more interesting to evaluate its possible role in other diseases, such as MTC.

Finally, loss of *RMND5A* (Required for Meiotic Nuclear Division 5 homolog A) gene was associated with lower serum calcitonin after total thyroidectomy. Even though there is no described association between this gene and cancer until the present study, its expected role in microtubule dynamics, cell migration, nucleokinesis and chromosome segregation [36] seems to be promising. A study revealed that *RMND5A* duplication could represent an insight into encephalocele formation [37].

Analysis of combined altered regions was performed to estimate the combined effect of variations, as aggressiveness observed as lymph node metastasis might not be explained solely by a single CNV, which is likely to contribute to small or moderate phenotypic effects; but it could be a consequence of combined CNVs, which could result in more considerable effects [38]. These combined analysis revealed different frequency of losses and gains between metastatic and non-metastatic groups (P=0.0310) (**Figure 4**), thus we found regions with different copy numbers between groups, which together could explain genetically what was observed in the clinics, suggesting that losses could be associated with predisposition to lymph node metastasis. We also found that the deletion on metastatic group (P=0.0375) was more determinant to this difference, being possible that loss of found CNVs' regions may be associated with events resulting on advanced tumor stages as well as lymph node metastasis. In conclusion, we found nine significant copy number variations on metastatic MTC and non-metastatic MTC patients' group as amplifications or deletions. These alterations were validated by qPCR and deletions were associated with a poorer clinical course, as *FMN2* loss alone could

be related with presence of lymph node metastasis and the losses of combined regions were able to define a profile of CNVs related to the presence of lymph node metastasis in a family with p.G533C *RET* mutation and MTC. Altogether, these findings support the hypothesis that CNVs may possibly help to define aggressiveness' genetic markers observed both as a single CNV or a profile of CNVs.

References

- 1. Kouvaraki, M.A., et al., *RET proto-oncogene: a review and update of genotypephenotype correlations in hereditary medullary thyroid cancer and associated endocrine tumors.* Thyroid, 2005. 15(6): p. 531-44.
- Kloos, R.T., et al., Medullary thyroid cancer: management guidelines of the American Thyroid Association. Thyroid : official journal of the American Thyroid Association, 2009. 19(6): p. 565-612.
- Brandi, M.L., et al., *Guidelines for diagnosis and therapy of MEN type 1 and type 2*.
 The Journal of clinical endocrinology and metabolism, 2001. 86(12): p. 5658-71.
- 4. de Groot, J.W., et al., *RET as a diagnostic and therapeutic target in sporadic and hereditary endocrine tumors.* Endocr Rev, 2006. 27(5): p. 535-60.
- Machens, A., K. Lorenz, and H. Dralle, *Constitutive RET tyrosine kinase activation in hereditary medullary thyroid cancer: clinical opportunities.* J Intern Med, 2009. 266(1): p. 114-25.
- Santoro, M., et al., *Molecular mechanisms of RET activation in human cancer*. Ann N Y Acad Sci, 2002. 963: p. 116-21.
- Arighi, E., M.G. Borrello, and H. Sariola, *RET tyrosine kinase signaling in development and cancer*. Cytokine Growth Factor Rev, 2005. 16(4-5): p. 441-67.
- Da Silva, A.M., et al., A novel germ-line point mutation in RET exon 8 (Gly(533)Cys) in a large kindred with familial medullary thyroid carcinoma. J Clin Endocrinol Metab, 2003. 88(11): p. 5438-43.
- 9. Kaldrymides, P., et al., A rare RET gene exon 8 mutation is found in two Greek kindreds with familial medullary thyroid carcinoma: implications for screening. Clin Endocrinol (Oxf), 2006. 64(5): p. 561-6.
- Bethanis, S., et al., *A newly detected mutation of the RET protooncogene in exon 8 as a cause of multiple endocrine neoplasia type 2A*. Hormones (Athens), 2007. 6(2): p. 152-6.

- Peppa, M., et al., Multiple endocrine neoplasia type 2A in two families with the familial medullary thyroid carcinoma associated G533C mutation of the RET proto-oncogene. European journal of endocrinology / European Federation of Endocrine Societies, 2008. 159(6): p. 767-71.
- 12. Oliveira, M.N., et al., *The RET p.G533C mutation confers predisposition to multiple endocrine neoplasia type 2A in a Brazilian kindred and is able to induce a malignant phenotype in vitro and in vivo.* Thyroid : official journal of the American Thyroid Association, 2011. 21(9): p. 975-85.
- 13. Tamanaha, R., et al., Evaluation of RET polymorphisms in a six-generation family with G533C RET mutation: specific RET variants may modulate age at onset and clinical presentation. Clinical endocrinology, 2009. 71(1): p. 56-64.
- Iafrate, A.J., et al., *Detection of large-scale variation in the human genome*. Nat Genet, 2004. 36(9): p. 949-51.
- 15. Ye, L., et al., *High resolution array-comparative genomic hybridization profiling reveals deoxyribonucleic acid copy number alterations associated with medullary thyroid carcinoma.* J Clin Endocrinol Metab, 2008. 93(11): p. 4367-72.
- Freeman, J.L., et al., *Copy number variation: new insights in genome diversity*. Genome Res, 2006. 16(8): p. 949-61.
- Zhang, F., et al., *Copy number variation in human health, disease, and evolution*. Annu Rev Genomics Hum Genet, 2009. 10: p. 451-81.
- Plon, S.E., et al., Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat, 2008. 29(11): p. 1282-91.
- Sambrook, J., E.F. Fritsch, and T. Maniatis, *Isolation of High-molecular-weight DNA from Mammalian Cells*, in *Molecular Cloning: A Laboratory Manual*, C. Nolan, Editor 1989, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. p. 9.14-9.23.

- Wang, K., et al., PennCNV: an integrated hidden Markov model designed for highresolution copy number variation detection in whole-genome SNP genotyping data. Genome Res, 2007. 17(11): p. 1665-74.
- 21. Glessner, J.T., J. Li, and H. Hakonarson, *ParseCNV integrative copy number variation association software with quality tracking*. Nucleic Acids Res, 2013.
- 22. Kumari, A., S.K. Yadav, and S. Ali, Organizational and functional status of the Ylinked genes and loci in the infertile patients having normal spermiogram. PLoS One, 2012. 7(7): p. e41488.
- 23. Li, Q.Q., J. Skinner, and J.E. Bennett, *Evaluation of reference genes for real-time quantitative PCR studies in Candida glabrata following azole treatment*. BMC Mol Biol, 2012. 13: p. 22.
- 24. Hemmer, S., et al., *DNA copy number changes in thyroid carcinoma*. Am J Pathol, 1999. 154(5): p. 1539-47.
- Marsh, D.J., et al., *Genome-wide copy number imbalances identified in familial and sporadic medullary thyroid carcinoma*. J Clin Endocrinol Metab, 2003. 88(4): p. 1866-72.
- 26. Ciampi, R., et al., *Chromosome 10 and RET gene copy number alterations in hereditary and sporadic Medullary Thyroid Carcinoma*. Mol Cell Endocrinol, 2012. 348(1): p. 176-82.
- 27. Wu, G., et al., *Uncommon mutation, but common amplifications, of the PIK3CA gene in thyroid tumors.* J Clin Endocrinol Metab, 2005. 90(8): p. 4688-93.
- 28. Flicker, K., et al., *High-resolution analysis of alterations in medullary thyroid carcinoma genomes.* Int J Cancer, 2012. 131(2): p. E66-73.
- 29. Winchester, L., C. Yau, and J. Ragoussis, *Comparing CNV detection methods for SNP arrays*. Brief Funct Genomic Proteomic, 2009. 8(5): p. 353-66.
- 30. Thierry-Mieg, D. and J. Thierry-Mieg, *AceView: a comprehensive cDNA-supported gene and transcripts annotation*. Genome Biol, 2006. 7 Suppl 1: p. S12 1-14.

- 31. Liu, P., et al., *Genome-wide association and fine mapping of genetic loci predisposing to colon carcinogenesis in mice*. Mol Cancer Res, 2012. 10(1): p. 66-74.
- 32. Yamada, K., et al., *Identification and Functional Characterization of FMN2, a Regulator of the Cyclin-Dependent Kinase Inhibitor p21.* Mol Cell, 2013. 49(5): p. 922-33.
- 33. de Cid, R., et al., *Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis.* Nat Genet, 2009. 41(2): p. 211-5.
- 34. Li, M., et al., Deletion of the late cornified envelope genes LCE3C and LCE3B is associated with psoriasis in a Chinese population. J Invest Dermatol, 2011. 131(8): p. 1639-43.
- 35. Ammar, M., et al., *Association analysis of LCE3C-LCE3B deletion in Tunisian psoriatic population*. Arch Dermatol Res, 2012. 304(9): p. 733-8.
- 36. Gonzalez-Fernandez, R., et al., *Changes in leukocyte gene expression profiles induced by antineoplastic chemotherapy*. Oncol Lett, 2012. 3(6): p. 1341-1349.
- 37. Vogel, T.W., S. Manjila, and A.R. Cohen, Novel neurodevelopmental disorder in the case of a giant occipitoparietal meningoencephalocele. J Neurosurg Pediatr, 2012. 10(1): p. 25-9.
- Kuningas, M., et al., *Large common deletions associate with mortality at old age*. Hum Mol Genet, 2011. 20(21): p. 4290-6.

	Metastatic MTC Non Mestastatic			Dualma	
	(n=8)	MTC (n=7)	Total (n=15)	P value	
Age of Onset* (mean ± SD)	48.63 ± 13.05	34.14 ± 6.12	41.87 ± 12.53	0.0176	
Gender					
Male	3/8 (37.5%)	5/7 (71.43%)	8/15 (53.33%)	0.3147	
Female	5/8 (62.5%)	2/7 (28.57%)	7/15 (46.67%)	0.0117	
Tumor Size	1.63 ± 0.61	0.76 ± 0.78	1.22 ± 0.80	0.0205	
Local Invasion					
Yes	1/8 (12.5%)	0/7 (0%)	1/15 (6.67%)	1	
No	7/8 (87.5%)	7/7 (100%)	14/15 (93.33%)	1	
Stage					
I and II	0/8 (0%)	5/5 (100%)	5/13 (38.46%)	0 0008	
III	8/8 (100%)	8/8 (100%) 0/5 (0%)		0.0008	
Pheochromocytoma					
Yes	1/8 (12.5%)	0/7 (0%)	1/15 (6.67%)	1	
No	7/8 (87.5%)	7/7 (100%)	14/15 (93.33%)	I	
Preoperative sCT levels	866 63 ± 602 62	320.86 ± 641.03	616 13 ± 650 18	0.0170	
(ng/mL; mean ± SD)	000.0 <i>3</i> ± 002.02	547.00 £ 041.05	010.13 ± 037.10	V . V14V	
sCT levels after TT	54 96 + 79 93	1 39 + 2 40	29 96 + 62 95	0 0006	
(ng/mL; mean ± SD)	51.70 - 17.75	1.57 ± 2.40	27.70 ± 02.73	0.0000	

 Table 1. Clinical features of patients selected for initial screening

MTC, medullary thyroid carcinoma; sCT, serum calcitonin; TT, total thyroidectomy. *Age of onset was defined as the age at which clinicopathological diagnosis of MTC was made.

	Metastatic MTC Non Mestastatic				
	(n=9)	MTC (n=17)	Total (n=26)	P value	
Age of Onset* (mean ± SD)	47.44 ± 12.71	40.82 ± 13.70	43.12 ± 13.49	0.1378	
Gender					
Male	3/9 (33%)	9/17 (53%)	12/26 (46%)	44 201	
Female	6/9 (66%)	8/17 (47%)	14/26 (54%)	44.291	
Tumor Size	1.76 ± 0.69	0.82 ± 0.72	1.142 ± 0.83	0.0021	
Local Invasion					
Yes	1/9 (11%)	1/17 (6%)	2/26 (8%)	1	
No	8/9 (89%)	16/17 (94%)	24/26 (92%)	1	
Stage					
I and II	0/9 (0%)	13/15 (86.67%)	13/24 (54.17%)	-0.0001	
III	9/9 (100%)	2/15 (13%)	11/24 (46%)	<0.0001	
Pheochromocytoma					
Yes	1/9 (11)	0/17 (0%)	1/26 (4%)	0 2462	
No	8/9 (88%)	17/17 (100%)	25/26 (96%)	0.5402	
Preoperative sCT levels	866.6 ± 607.6	202.1 ± 465.7	501 7 ± 579 9	0.0046	
(ng/mL; mean ± SD)	600.0 ± 002.0	293.1 ± 403.7	501.7 ± 578.8	0.0040	
sCT levels after TT	40.00 + 76.92	2 91 + 7 22	20.10 + 40.02	0.0020	
(ng/mL; mean ± SD)	49.08 ± /0.82	3.81 ± 1.23	20.10 ± 49.92	0.0020	

 Table 2. Clinical-pathological features of validation patients set

MTC, medullary thyroid carcinoma; sCT, serum calcitonin; TT, total thyroidectomy. *Age of onset was defined as the age at which clinicopathological diagnosis of MTC was made.

Group
netastatic
Non-r
and
f Metastatic
' Profile o
S
3.
Table

	Ge	notyping Conse	ole		PennCNV	
	Metastatic N	Von-metastatic	Total	Metastatic	Non-metastatic	Total
Number of patients	8	7	15	8	7	15
Total of CNVs	251	251	328	278	273	407
Losses	207	210	268	146	142	210
Gains	44	41	60	132	131	197
CNVs per patient (mean ± SD)	121.5 ± 7.0	125.6±13.8	123.4±10.5	17.4±5.4	19.5 ± 4.5	18.4 ± 5.0
Losses	105.2 ± 7.0	109.1 ± 14.6	107.1 ± 10.9	18.3±7.0	20.3±6.0	19.2 ± 6.4
Gains	16.2 ± 3.3	16.4 ± 3.8	16.3 ± 3.4	16.5 ± 3.5	18.7 ± 2.5	17.5 ± 3.2
Size of CNVs (kb; mean ± SD)	28.7±63.5	28.1±55.4	26.2 ± 59.5	261.3 ± 576.3	548.6±2653.4	465.3 ± 2191.8
Losses	26.4±58.6	22.9±44.7	23.1±54.7	63.7 ± 133.6	61.9 ± 96.8	70.2±130.9
Gains	39.3±82.8	54.9±88.7	39.8 ± 76.4	479.8±768.6	1076.3 ± 3765.9	886.5±3096.3

							Alteration	n Detected	
	Cytoband	CNV Coordinates	Size (kb)	Marker Count	Reference Genes	Distance from Gene (kb)	Metastatic (n=8)	Non Metastatic (n=7)	P value
Genotyping	Losses								
Console	1q21.3	chr1:152555175-152586528	31.4	23	LCE3C (NM_178434.2)	0	3/8 (37.5%)	0/2 (0%)	0.2000
(GRCh37/hg19)	1q31.3	chr1:196738610-196802060	63.4	78	CFHR3 (NG_015993.1)	0	4/8 (50%)	2/7 (28.57%)	0.6084
					CFHR1 (NG_013060.1)				
	1q43	chr1:240391234-240394127	2.9	8	FMN2 (NM_020066.4)	0	4/8 (50%)	0/2 (0%)	0.0769
	2q22.3	chr2:146865961-146866922	1.0	11	PABPC1P2 (NR_026904.1)	477.7	4/8 (50%)	1/7 (14.29%)	0.2821
	3q22.1	chr3:129775845-129795080	19.2	45	ALG1L2 (NM_001136152.1)	5.6	7/8 (87.5%)	1/7 (14.29%)	0.0101^{*}
	Gains								
	2p11.2	chr2:86948834-86970406	21.6	4	RMND5A (NM_022780.3)	0	4/8 (50%)	1/7 (14.29%)	0.2821
	7q34	chr7:142476706-142490975	14.3	35	PRSS3P2 (NR_001296.3)	0	3/8 (37.5%)	0/ <i>J</i> (0%)	0.2000
					PRSS2 (NG_008322.1)				
	11q11	chr11:55353110-55447453	94.3	54	OR4C11 (NM_001004700.2)	0	5/8 (62.5%)	2/7 (28.57%)	0.3147
					OR4P4 (NM_001004124.1)				
					OR4S2 (NM_001004059.2)				
					OR4C6 (NM_001004704.1)				
PennCNV	Losses								
(NCB136/hg18)	1q31.3	chr1:194994474-195068695	74.2	94	CFHR3 (NG_015993.1)	0	5/8 (62.5%)	0/2 (0%)	0.0447*
					CFHR1 (NG_013060.1)				
	2q22.3	chr2:146580874-146583404	2.5	2	PABPC1P2 (NR_026904.1)	477.7	4/8 (50%)	0/2 (0%)	0.0027*
	4q26	chr4:115394773-115403839	9.1	3	ARSJ (NM_024590.3)	274.4	3/8 (37.5%)	0/7 (0%)	0.0185^{*}
kb, kilobase. The u *Statistically signifi	nderlined coord cant using Fish	dinates indicate that the same ru net's Exact Test (two-tailed)	egion w	as identifi	ed with both softwares				

Table 4. Relevant CNVs detected by comparison of Metastatic and Non-metastatic Group

25

			Copy Number Changes		_
Gene/Locus	CNV Type	Metastasis	Patients without CNV	Patients with CNV	P value
LCE3C	Loss	Yes	2/9 (22.22%)	7/9 (77.78%)	0.01(7
		No	9/17 (52.94%)	8/17 (47.06%)	0.2167
CFHR1	Loss	Yes	7/9 (77.78%)	2/9 (22.22%)	1
		No	14/17 (82.35%)	3/17 (17.65%)	1
CFHR1	Gain	Yes	5/9 (55.56%)	4/9 (44.44%)	0 2341
		No	5/17 (29.41%)	12/17 (70.59%)	0.2341
FMN2	Loss	Yes	1/9 (11.11%)	8/9 (88.89%)	0.0262*
		No	10/17 (58.82%)	7/17 (41.18%)	0.0302*
FMN2	Gain	Yes	8/9 (88.89%)	1/9 (11.11%)	0.2574
		No	11/17 (64.71%)	6/17 (35.29%)	0.3374
2q22.3	Loss	Yes	2/9 (22.22%)	7/9 (77.78%)	0 2080
		No	8/17 (47.06%)	9/17 (52.94%)	0.3989
2q22.3	Gain	Yes	8/9 (88.89%)	1/9 (11.11%)	1
		No	14/17 (82.35%)	3/17 (17.65%)	1
3q22.1	Loss	Yes	5/9 (55.56%)	4/9 (44.44%)	0 1881
		No	14/17 (82.35%)	3/17 (17.65%)	0.1001
3q22.1	Gain	Yes	6/9 (66.67%)	3/9 (33.33%)	0.6828
		No	9/17 (52.94%)	8/17 (47.06%)	0.0828
4q26	Loss	Yes	2/9 (22.22%)	7/9 (77.78%)	0 4177
		No	7/17 (41.18%)	10/17 (58.82%)	0.4177
4q26	Gain	Yes	8/9 (88.89%)	1/9 (11.11%)	0 2709
		No	12/17 (70.59%)	5/17 (29.41%)	0.3798

Table 5.	Frequency	of 10 select	ed CNVs in	the 26 MT	C patients of	obtained by	qPCR
RMND5A	Loss	Yes	8/9 (88.89%)	1/9 (11.11%)	0 1002		
---------	------	-----	----------------	----------------	---------		
		No	10/17 (58.82%)	7/17 (41.18%)	0.1902		
RMND5A	Gain	Yes	3/9 (33.33%)	6/9 (66.67%)	1		
		No	7/17 (41.18%)	10/17 (58.82%)	1		
PRSS3P2	Loss	Yes	7/9 (77.78%)	2/9 (22.22%)	1		
		No	13/17 (76.47%)	4/17 (23.53%)	1		
PRSS3P2	Gain	Yes	4/9 (44.44%)	5/9 (55.56%)	1		
		No	8/17 (47.06%)	9/17 (52.94%)	1		
OR4S2	Loss	Yes	4/9 (44.44%)	5/9 (55.56%)	0 2341		
		No	12/17 (70.59%)	5/17 (29.41%)	0.23 11		
OR4S2	Gain	Yes	7/9 (77.78%)	2/9 (22.22%)	0.11		
		No	7/17 (41.18%)	10/17 (58.82%)	0.11		

Fisher's Exact Test. *Statistically significant

Gene/Loci	Primer Sequence	Product Size
LCE3C	F: 5'- CTCCACCCTCTTCTGACTGT -3'	162
	R: 5'- AGGAGCCCCCGCCTTGCT -3'	
CFHR1	F: 5'- CTTTTAGTGCTGAGTTGTAATC -3'	162
	R: 5'- TTCACTTATTTTCCTCTGCTAC -3'	
FMN2	F: 5'- TGTTGTTGGTGTTTTTACTTGC -3'	178
	R: 5'- GCAAGTGGATTTCAACTCAGA -3'	
2q22.3	F: 5'- GGCAAATCCATCTAAGCAAAC -3'	148
	R: 5'- CAATGAAGAGAAAGTGGTAGTA -3'	
3q22.1	F: 5'- CTGTTGAATCCTCCCACTGT -3'	169
	R: 5'- GTCGTGAACTCCTCATCTGT -3'	
4q26	F: 5'- AACCAAAACCCTATGCTTATCT -3'	156
	R: 5'- TATGATGTTGGCTGGGTGTTT -3'	
RMND5A	F: 5'- GTAAGTTGAAGGGGGGATGTG -3'	164
	R: 5'- CAATACCCCTCAAACCATAAAT -3'	
PRSS3P2	F: 5'- TGAGACTGGGAGAGCACAAT -3'	145
	R: 5'- ATGACGGCAGGTGTGGAGA -3'	
OR4S2	F: 5'- GGGAGTTTTGTTATCTTGCTAA -3'	174
	R: 5'- TGAAAAGGTCGTATCAGGGC -3'	
ACTB	F: 5'-CCCTTCCCTCCTCAGATCATT -3'	108
	R: 5'- CTTGCTGATCCACATCTGCTG -3'	

Supplementary Table 1. Primer sequences and PCR products for selected CNV validations by quantitative PCR

Variable	Copy number of <i>LCE3C</i> (Not Deleted	Copy number of <i>LCE3C</i> (Deleted	P value
	n=11)	n= 15)	
Gender			
Male	5/11 (45.45%)	7/15 (46.67%)	1
Female	6/11 (54.55%)	8/15 (53.33%)	1
Tumor Size (cm; mean ± SD)	0.6818 ± 0.3995	1.48 ± 0.9128	0.0327*
Local Invasion			
Yes	0/11 (0%)	2/15 (13.33%)	0.4923
No	11/11 (100%)	13/15 (86.67%)	
Stage			
Ι	5/9 (55.56%)	4/15 (26.67%)	0.2119
II	1/9 (11.11%)	3/15 (20%)	1
III	3/9 (33.33%)	8/15 (53.33%)	0.4225
Preoperative sCT levels (ng/mL; mean ± SD)	281.6 ± 234.0	627.4 ± 681.8	0.4736
sCT levels after TT (ng/mL; mean ± SD)	24.61 ± 61.60	16.56 ± 40.63	0.5996

Supplementary Table 2. Correlation between loss of *LCE3C* and clinical characteristics

sCT, serum calcitonin; TT, total thyroidectomy.

*Statistically significant

Variable	Copy number of CFHR1 (Not Deleted n=21)	Copy number of CFHR1 (Deleted n= 5)	P value
Gender			
Male	8/21 (38.1%)	4/5 (80%)	0 1 4 7 9
Female	13/21 (61.9%)	1/5 (20%)	0.1478
Tumor Size (cm; mean ± SD)	1.176 ± 0.8420	1.000 ± 0.8660	0.4725
Local Invasion			
Yes	2/21 (9.52%)	0/5 (0%)	1
No	19/21 (90.48%)	5/5 (100%)	1
Stage			
Ι	7/20 (35%)	2/4 (50%)	0.6146
II	4/20 (20%)	0/4 (0%)	1
III	9/20 (45%)	2/4 (50%)	1
Preoperative sCT levels (ng/mL; mean ± SD)	548.5 ± 618.6	342.6 ± 432.9	0.6383
sCT levels after TT (ng/mL; mean ± SD)	15.87 ± 45.90	37.04 ± 67.05	0.706

Supplementary Table 3. Correlation between loss of CFHR1 and clinical characteristics

Variable	Copy number of CFHR1 (Not Amplified	Copy number of CFHR1 (Amplified	P value
	n=10)	n= 16)	
Gender			
Male	7/10 (70%)	5/16 (31.25%)	0 1054
Female	3/10 (30%)	11/16 (68.75%)	0.1034
Tumor Size (cm; mean ± SD)	0.9600 ± 0.7121	1.256 ± 0.9018	0.3419
Local Invasion			
Yes	1/10 (10%)	1/16 (6.25%)	1
No	9/10 (90%)	15/16 (93.75%)	1
Stage			
Ι	4/9 (44.44%)	5/15 (33.33%)	0.6785
П	0/9 (0%)	4/15 (26.67%)	0.2589
III	5/9 (55.56%)	6/15 (40%)	0.6752
Preoperative sCT levels (ng/mL; mean ± SD)	308.6 ± 318.6	662.6 ± 703.1	0.4483
sCT levels after TT (ng/mL; mean ± SD)	41.87 ± 75.34	5.593 ± 7.650	0.5947

Supplementary Table 4. Correlation between gain of *CFHR1* and clinical characteristics

sCT, serum calcitonin; TT, total thyroidectomy.

Variable	Copy number of FMN2 (Not Deleted n=11)	Copy number of FMN2 (Deleted n= 15)	P value
Gender			
Male	4/11 (36.36%)	8/15 (53.33%)	0 4527
Female	7/11 (63.64%)	7/15 (46.67%)	0.4327
Tumor Size (cm; mean ± SD)	1.255 ± 0.9353	1.060 ± 0.7707	0.4657
Local Invasion			
Yes	1/11 (9.09%)	1/15 (6.67%)	1
No	10/11 (90.91%)	14/15 (93.33%)	1
Stage			
Ι	4/10 (40%)	5/14 (35.71%)	1
П	3/10 (30%)	1/14 (7.14%)	0.2721
III	3/10 (30%)	8/14 (57.15%)	0.2397
Preoperative sCT levels (ng/mL; mean ± SD)	382.6 ± 575.2	569.7 ± 591.1	0.4736
sCT levels after TT (ng/mL; mean ± SD)	5.320 ± 8.748	29.96 ± 62.95	0.5565

Supplementary Table 5. Correlation between loss of FMN2 and clinical characteristics

Variable	Copy number of FMN2 (Not Amplified n=19)	Copy number of FMN2 (Amplified n= 7)	P value
Gender			
Male	10/19 (52.63%)	2/7 (28.57%)	0.2012
Female	9/19 (47.37%)	5/7 (71.43%)	0.3913
Tumor Size (cm; mean ± SD)	1.168 ± 0.8622	1.071 ± 0.8036	0.794
Local Invasion			
Yes	1/19 (5.26%)	1/7 (14.29%)	0.4738
No	18/19 (94.74%)	6/7 (85.71%)	
Stage			
Ι	6/17 (35.29%)	3/7 (42.86%)	1
Π	3/17 (17.65%)	1/7 (14.28%)	1
III	8/17 (47.06%)	3/7 (42.86%)	1
Preoperative sCT levels (ng/mL; mean ± SD)	559.8 ± 625.7	240.3 ± 123.1	0.7017
sCT levels after TT (ng/mL; mean ± SD)	25.08 ± 58.22	7.314 ± 9.960	0.5412

Supplementary Table 6. Correlation between gain of FMN2 and clinical characteristics

sCT, serum calcitonin; TT, total thyroidectomy.

Variable	Copy number of 2q22.3 (Not Deleted n=10)	Copy number of 2q22.3 (Deleted n= 16)	P value
Gender			
Male	4/10 (40%)	8/16 (50%)	0 7015
Female	6/10 (60%)	8/16 (50%)	0.7015
Tumor Size (cm; mean ± SD)	0.9700 ± 0.6961	1.250 ± 0.9114	0.634
Local Invasion			
Yes	0/10 (0%)	2/16 (12.5%)	0.5077
No	10/10 (100%)	14/16 (87.5%)	
Stage			
Ι	4/9 (44.45%)	5/15 (33.33%)	0.6785
II	2/9(22.22%)	2/15 (13.33%)	0.6146
III	3/9 (33.33%)	8/15 (53.34%)	0.4225
Preoperative sCT levels (ng/mL; mean ± SD)	607.6 ± 790.2	441.1 ± 440.6	0.9185
sCT levels after TT (ng/mL; mean ± SD)	6.470 ± 9.015	29.19 ± 63.23	0.8887

Supplementary Table 7. Correlation between loss of 2q22.3 and clinical characteristics

Variable	Copy number of 2q22.3 (Not Amplified n=22)	Copy number of 2q22.3 (Amplified n= 4)	P value
Gender			
Male	10/22 (45.45%)	2/4 (50%)	1
Female	12/22 (54.55%)	2/4 (50%)	1
Tumor Size (cm; mean ± SD)	1.182 ± 0.8600	0.9250 ± 0.7182	0.5681
Local Invasion			
Yes	2/22 (09.09%)	0/4 (0%)	1
No	20/22 (90.91%)	4/4 (100%)	1
Stage			
Ι	8/21 (38.09%)	1/3 (33.33%)	1
II	4/21 (19.05%)	0/3 (0%)	-
III	9/21 (42.86%)	2/3 (66.67%)	0.5761
Preoperative sCT levels (ng/mL; mean ± SD)	465.1 ± 502.6	733.3 ± 1073	1
sCT levels after TT (ng/mL; mean ± SD)	21.86 ± 54.24	10.88 ± 13.55	0.9702

Supplementary Table 8. Correlation between gain of 2q22.3 and clinical characteristics

sCT, serum calcitonin; TT, total thyroidectomy.

Variable	Copy number of 3q22.1 (Not Deleted n=19)	Copy number of 3q22.1 (Deleted n= 7)	P value
Gender			
Male	7/19 (36.84%)	5/7 (71.43%)	0 1004
Female	12/19 (63.16%)	2/7 (28.57%)	0.1904
Tumor Size (cm; mean ± SD)	1.116 ± 0.8153	1.214 ± 0.9388	0.9306
Local Invasion			
Yes	2/19 (10.53%)	0/7 (0%)	1
No	17/19 (89.47%)	7/7 (100%)	1
Stage			
Ι	6/17 (35.29%)	3/7 (42.86%)	1
II	4/17 (23.53%)	0/7 (0%)	0.2833
III	7/17 (41.18%)	4/7 (57.14%)	0.6591
Preoperative sCT levels (ng/mL; mean ± SD)	506.1 ± 618.2	492.1 ± 529.8	0.8879
sCT levels after TT (ng/mL; mean ± SD)	6.772 ± 9.022	54.39 ± 88.50	0.445

Supplementary Table 9. Correlation between loss of 3q22.1 and clinical characteristics

Variable	Copy number of 3q22.1 (Not Amplified n=15)	Copy number of 3q22.1 (Amplified n= 11)	P value
Gender			
Male	7/15 (46.67%)	5/11 (45.45%)	1
Female	8/15 (53.33%)	6/11 (54.55%)	1
Tumor Size (cm; mean ± SD)	1.147 ± 0.8026	1.136 ± 0.9102	0.8964
Local Invasion			
Yes	1/15 (6.67%)	1/11 (9.09%)	1
No	14/15 (93.33%)	10/11 (90.91%)	1
Stage			
Ι	5/13 (38.46%)	4/11 (36.36%)	1
II	1/13 (7.69%)	3/11 (27.28%)	0.3002
III	7/13 (53.85%)	4/11 (36.36%)	0.4442
Preoperative sCT levels (ng/mL; mean ± SD)	403.5 ± 432.1	712.0 ± 813.2	0.5728
sCT levels after TT (ng/mL; mean ± SD)	29.67 ± 65.59	7.927 ± 9.823	0.7611

Supplementary Table 10. Correlation between gain of 3q22.1 and clinical characteristics

sCT, serum calcitonin; TT, total thyroidectomy.

Variable	Copy number of 4q26 (Not Deleted n=9)	Copy number of 4q26 (Deleted n= 17)	P value
Gender			
Male	4/9 (44.44%)	8/17 (47.06%)	1
Female	5/9 (55.56%)	9/17 (52.94%)	1
Tumor Size (cm; mean ± SD)	1.078 ± 0.9324	1.176 ± 0.8020	0.6652
Local Invasion			
Yes	1/9 (11.11%)	1/17 (5.88%)	1
No	8/9 (88.89%)	16/17 (94.12%)	1
Stage			
Ι	3/8 (37.5%)	6/16 (37.5%)	1
II	3/8 (37.5%)	1/16 (6.25%)	0.0909
III	2/8 (25%)	9/16 (56.25%)	0.2108
Preoperative sCT levels (ng/mL; mean ± SD)	484.0 ± 625.8	509.9 ± 578.4	0.8879
sCT levels after TT (ng/mL; mean ± SD)	3.750 ± 7.625	27.80 ± 59.29	0.0996

Supplementary Table 11. Correlation between deletion of 4q26 and clinical characteristics

Variable	Copy number of 4q26 (Not Amplified n=20)	Copy number of 4q26 (Amplified n= 6)	P value
Gender			
Male	10/20 (50%)	2/6 (33.33%)	0 6522
Female	10/20 (50%)	4/6 (66.67%)	0.0322
Tumor Size (cm; mean ± SD)	1.200 ± 0.8335	0.9500 ± 0.8735	0.4634
Local Invasion			
Yes	1/20 (5%)	1/6 (16.67%)	0 4154
No	19/20 (95%)	5/6 (83.33%)	0.4154
Stage			
Ι	6/18 (33.33%)	3/6 (50%)	0.6349
II	2/18 (11.11%)	2/6 (33.33%)	0.2513
III	10/18 (55.56%)	1/6 (16.67%)	0.166
Preoperative sCT levels (ng/mL; mean ± SD)	444.6 ± 546.4	758.8 ± 738.8	0.4693
sCT levels after TT (ng/mL; mean ± SD)	24.98 ± 56.54	4.667 ± 8.773	0.2892

Supplementary Table 12. Correlation between loss of 4q26 and clinical characteristics

sCT, serum calcitonin; TT, total thyroidectomy.

Variable	Copy number of <i>RMND5A</i> (Not Deleted n=18)	Copy number of <i>RMND5A</i> (Deleted n= 8)	P value
Gender			
Male	7/18 (38.89%)	5/8 (62.5%)	0.4000
Female	11/18 (61.11%)	3/8 (37.5%)	0.4009
Tumor Size (cm; mean ± SD)	1.172 ± 0.7790	1.075 ± 0.9953	0.2655
Local Invasion			
Yes	2/18 (11.11%)	0/8 (0%)	1
No	16/18 (88.89%)	8/8 (100%)	1
Stage			
Ι	5/17 (29.41%)	4/7 (57.14%)	0.3564
II	2/17 (11.77%)	2/7 (28.57%)	0.552
III	10/17 (58.82%)	1/7 (14.29%)	0.0778
Preoperative sCT levels (ng/mL; mean ± SD)	562.5 ± 555.6	395.3 ± 641.4	0.0945
sCT levels after TT (ng/mL; mean ± SD)	26.93 ± 57.76	2.557 ± 3.781	0.0439*

Supplementary Table 13. Correlation between loss of RMND5A and clinical characteristics

sCT, serum calcitonin; TT, total thyroidectomy.

*Statistically significant

Variable	Copy number of <i>RMND5A</i> (Not Amplified n=10)	Copy number of <i>RMND5A</i> (Amplified n= 16)	P value
Gender			
Male	6/10 (60%)	6/16 (37.5%)	0 4216
Female	4/10 (40%)	10/16 (62.5%)	0.4210
Tumor Size (cm; mean ± SD)	1.100 ± 0.8844	1.169 ± 0.8260	0.5262
Local Invasion			
Yes	1/10 (10%)	1/16 (6.25%)	1
No	9/10 (90%)	15/16 (93.75%)	1
Stage			
Ι	4/9 (44.45%)	5/15 (33.33%)	0.6785
II	2/9 (22.22%)	2/15 (13.33%)	0.6146
III	3/9 (33.33%)	8/15 (53.34%)	0.4225
Preoperative sCT levels (ng/mL; mean ± SD)	405.8 ± 566.9	581.6 ± 601.1	0.2485
sCT levels after TT (ng/mL; mean ± SD)	6.878 ± 9.181	27.54 ± 61.47	0.3608

Supplementary Table 14. Correlation between gain of *RMND5A* and clinical characteristics

sCT, serum calcitonin; TT, total thyroidectomy.

Variable	Copy number of PRSS3P2 (Not Deleted n=20)	Copy number of PRSS3P2 (Deleted n= 6)	P value
Gender			
Male	8/20 (40%)	4/6 (66.67%)	0.2652
Female	12/20 (60%)	2/6 (33.33%)	0.3032
Tumor Size (cm; mean ± SD)	1.215 ± 0.8928	0.9000 ± 0.5831	0.4634
Local Invasion			
Yes	2/20 (10%)	0/6 (0%)	1
No	18/20 (90%)	6/6 (100%)	1
Stage			
Ι	6/18 (33.33%)	3/6 (50%)	0.6349
Π	3/18 (16.67%)	1/6 (16.67%)	1
III	9/18 (50%)	2/6 (33.33%)	0.6494
Preoperative sCT levels (ng/mL; mean ± SD)	502.6 ± 539.7	499.2 ± 730.2	0.9706
sCT levels after TT (ng/mL; mean ± SD)	23.68 ± 57.00	8.783 ± 7.868	0.4601

Supplementary Table 15. Correlation between loss of PRSS3P2 and clinical characteristics

Variable	Copy number of PRSS3P2 (Not Amplified n=12)	Copy number of PRSS3P2 (Amplified n= 14)	P value
Gender			
Male	8/12 (66.67%)	4/14 (28.57%)	0 1121
Female	4/12 (33.33%)	10/14 (71.43%)	0.1151
Tumor Size (cm; mean ± SD)	1.100 ± 0.8581	1.179 ± 0.8396	0.5702
Local Invasion			
Yes	1/12 (8.33%)	1/14 (7.14%)	1
No	11/12 (91.67%)	13/14 (92.86%)	1
Stage			
Ι	5/11 (45.46%)	4/13 (30.77%)	0.6752
Π	2/11 (18.18%)	2/13 (15.38%)	1
III	4/11 (36.36%)	7/13 (53.85%)	0.4442
Preoperative sCT levels (ng/mL; mean ± SD)	400.8 ± 578.0	622.7 ± 586.2	0.1765
sCT levels after TT (ng/mL; mean ± SD)	21.02 ± 45.56	19.39 ± 54.79	1

Supplementary Table 16. Correlation between gain of PRSS3P2 and clinical characteristics

sCT, serum calcitonin; TT, total thyroidectomy.

Variable	Copy number of OR4S2 (Not Deleted n=16)	Copy number of OR4S2 (Deleted n= 10)	P value
Gender			
Male	8/16 (50%)	4/10 (40%)	0 7015
Female	8/16 (50%)	6/10 (60%)	0.7015
Tumor Size (cm; mean ± SD)	1.069 ± 0.8491	1.260 ± 0.8343	0.6529
Local Invasion			
Yes	2/16 (12.5%)	0/10 (0%)	0 5077
No	14/16 (87.5%)	10/10 (100%)	0.3077
Stage			
Ι	5/14 (35.71%)	4/10 (40%)	1
Π	3/14 (21.43%)	1/10 (10%)	0.6146
III	6/14 (42.86%)	5/10 (50%)	1
Preoperative sCT levels (ng/mL; mean ± SD)	398.0 ± 428.7	626.1 ± 724.9	0.4483
sCT levels after TT (ng/mL; mean ± SD)	5.993 ± 8.738	41.27 ± 75.47	0.218

Supplementary Table 17. Correlation between loss of OR4S2 and clinical characteristics

Variable	Copy number of OR4S2 (Not Amplified n=14)	Copy number of OR4S2 (Amplified n= 12)	P value
Gender			
Male	6/14 (42.86%)	6/12 (50%)	1
Female	8/14 (57.14%)	6/12 (50%)	1
Tumor Size (cm; mean ± SD)	1.236 ± 0.8015	1.033 ± 0.8886	0.6422
Local Invasion			
Yes	0/14 (0%)	2/12 (16.67%)	0 2021
No	14/14 (100%)	10/12 (83.33%)	0.2051
Stage			
Ι	4/12 (33.33%)	5/12 (41.67%)	1
Π	1/12 (8.33%)	3/12 (25%)	0.5901
III	7/12 (58.33%)	4/12 (33.33%)	0.4136
Preoperative sCT levels (ng/mL; mean ± SD)	624.4 ± 682.7	286.9 ± 238.0	0.3936
sCT levels after TT (ng/mL; mean ± SD)	30.64 ± 65.22	6.700 ± 9.901	0.5432

Supplementary Table 18. Correlation between gain of OR4S2 and clinical characteristics

Legends

Figure 1. Schematic representation of overlapping CNV calls in three patients. Primers where designed in the overlapping region for all patients.

Figure 2. Comparison of CNVs detected by Genotyping Console and PennCNV along the chromosomes.

Figure 3. (A) Distribution of qPCR results for *FMN2* gene among metastatic and nonmetastatic patients' groups. (B) Mean tumor size (cm) of patients with or without loss of *LCE3C* gene. (C) Mean sCT levels after TT (ng/mL) of patients with or without loss of *RMND5A* gene.

Figure 4. Clustered CN information obtained by qPCR of possible candidate CNVs associated with a more aggressive phenotype.

Supplementary Figure 1. Illustration of Chromosome Analysis Suite software (Affymetrix, Santa Clara, CA, USA), used to visualize Genotyping Console (Affymetrix) results. The figure shows a representation of patients' data for two candidate CNVs, including Log2 Ratio, which represents the normalized intensity related to a reference (diploid) and further correction by sample specific variation; and the Copy Number State (HMM-derived) based on Log2.: (A) deletion on 1q31.3 and (B) amplification on 11q11.

Supplementary Figure 2. Karyotype illustration with summarized CNVs found by Genotyping Console and/or PennCNV. The red arrows indicate losses, while green arrows indicate gains.

Figure 1.



Figure 2.



40

Figure 3.



Figure 4.



CNV Status







Supplementary Figure 2.

Dado que as CNVs representam importante fonte de variabilidade genômica, que, por consequência, podem alterar fenótipos; e, pelo fato de já terem sido relacionadas a diversas doenças, tivemos como objetivo investigar se a presença de CNVs poderia estar relacionada à heterogeneidade clínica observada na família, especificamente associada à metástase aos linfondos. Utilizamos a metodologia do SNP-*array* de alta densidade, por meio da plataforma *Genome-Wide Human SNP Array* 6.0, uma vez que tal metodologia fornece resultados mais refinados quando comparados às metodologias anteriores, como a hibridização genômica comparativa (CGH, do inglês *Comparative Genomic Hybridization*) e o *array*-CGH, que mostram alterações maiores.

Na análise pelos softwares, comparando o grupo dos pacientes com CMT metastático e não metastático, nós encontramos regiões com diferentes números de cópias, entre deleções e amplificações. Dez CNVs diferiram entre os grupos, dentre as quais 70% representavam deleções e 30% as amplificações; as deleções foram encontradas em 1q, 2q, 3q, 4q e 10q, sendo 1q a mais comum, e as amplificações em 2p, 7q e 11q.

As CNVs foram pesquisadas individualmente em bancos de dados, embora a informação gerada seja meramente "ilustrativa", pois uma alteração encontrada em um indivíduo normal pode influenciar em determinado fenótipo alterado quando combinada a outras alterações ou ao meio, agindo, desta forma, como uma CNV potencialmente patogênica, e não como uma alteração comum. Além de que diversos genes relacionados ao câncer, englobados ou sobrepostos por alguma CNV, já foram encontrados na população normal. Essa análise em bancos de dados demonstrou que as CNVs encontradas já foram relacionadas a alguma alteração encontrada na população

normal pelo banco de dados das variações genômicas (DGV, do inglês <u>Database of</u> <u>Genomic Variation</u>, http://projects.tcag.ca/variation), enquanto que nenhuma dessas CNVs analisadas foram encontradas no banco de dados que compila alterações submicroscópicas em pacientes com fenótipos alterados (DECIPHER, do inglês <u>Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl</u> <u>Resources</u>, https://decipher.sanger.ac.uk).

Com exceção das regiões em 2q, 3q e 4q, todas as outras CNVs contêm genes e, dentre os genes, o CFHR1 foi o único encontrado no banco de dados *online* de herança mendeliana humana (OMIM, do inglês <u>Online Mendelian Inheritance in Man</u>, http://www.omim.org/), associado ao risco reduzido de degeneração macular relacionada à idade (603075 - Macular_degeneration_age-related_reduced_risk_of) e à susceptibilidade a síndrome hemolítico-urêmica atípica (235400 – Hemolytic_uremic_syndrome_atypical_susceptibility_to).

Pela análise por qPCR em um grupo ampliado, as regiões alteradas foram validadas e algumas foram estatisticamente significantes relacionadas à característica clínica, como a deleção de *FMN2* com metástase (P=0,0362), deleção de *LCE3C* com tamanho tumoral (P=0,0327), e deleção de *RMND5A* com níveis de calcitonina sérica após tiroidectomia total (P=0,0439).

O gene *FMN2* (*Formin 2*), cuja deleção foi significante relacionada à metástase, codifica uma proteína com função na organização do citoesqueleto e polaridade da célula e, uma vez que a CNV de deleção se encontra dentro do gene em pacientes com CMT metastático, podemos sugerir que a deleção resultaria em uma proteína não funcional, causando a desorganização celular que contribuiria para o processo metastático. Um estudo recente de caracterização funcional de *FMN2* propôs que o gene

é necessário para expressão estável da proteína inibidora de ciclo celular p21, sendo suficiente para o aumento da expressão dessa proteína. Uma vez que a expressão aumentada da p21 ocasiona a parada do ciclo celular na fase G1; a deleção do gene *FMN2* poderia levar à perda da expressão da p21 e, consequentemente, resultando na progressão do ciclo celular.

O gene *LCE3C* (*Late cornified envelope 3C*), também conhecido como *LEP15* ou *SPRL3A*, é bem descrito na literatura associado à psoríase, mais especificamente sua deleção, demonstrando ser interessante avaliar o possível papel do mesmo em outras doenças.

O gene *RMND5A* (*Required for meiotic nuclear division 5 homolog A*), também conhecido como *CTLH*, *GID2*, *RMD5*, *GID2A* ou *p44CTLH*, possui um papel esperado na dinâmica de microtúbulos, migração celular e segregação cromossômica, que pode ser um estudo promissor no câncer.

A análise combinada das regiões revelou diferença na frequência de perdas e ganhos entre os grupos (P=0,0310), estando principalmente associada a esta característica, a deleção no grupo metastático (P=0,0375). Estes dados sugerem que o efeito combinado das variações poderia explicar geneticamente a heterogeneidade observada na clínica.

Em conclusão, nós encontramos e validamos diversas CNVs de amplificação e deleção nos grupos, e os dados resultantes das análises revelaram que as deleções foram, em geral, associadas ao pior curso clínico, sugerindo que essas CNVs de deleção poderiam estar associadas a eventos que resultam em estágios tumorais avançados. Como a deleção de *FMN2* que, sozinha, pode ser relacionada à presença de metástase aos linfonodos, e a combinação das regiões deletadas que permitiu definir um perfil de

CNVs relacionadas à presença de metástase aos linfonodos no CMT, em uma família com NEM 2A e mutação p.G533C no gene *RET*.