RAFAEL ARRABAÇA DE CARVALHO

**Avaliação do método de sequenciamento de
nova geração no diagnóstico genético de
neoplasia endócrina múltipla tipo 1**

Dissertação apresentada à Faculdade de Medicina da
Universidade de São Paulo para obtenção do título de
Mestre em Ciências

Programa de Endocrinologia

Orientador: Prof. Dr. Delmar Muniz Lourenço Jr

Coorientador: Prof. Dr. Alexander Augusto de Lima
Jorge

São Paulo

2016

Este estudo foi realizado na Unidade de Endocrinologia Genética (LIM/25), Disciplina de Endocrinologia, Faculdade de Medicina da Universidade de São Paulo. Sua realização teve colaboração da Unidade de Neurologia (LIM/15), Faculdade de Medicina da Universidade de São Paulo e teve apoio da FAPESP com bolsa de mestrado (2013/15388-4) e Auxílio a Pesquisa Regular (2013/19810-2).

Esta dissertação está de acordo com as seguintes normas, em vigor no momento da sua publicação:

Referências: adaptado de *International Committee of Medical Journals Editors* (Vancouver).

Universidade de São Paulo. Faculdade de Medicina. Divisão de Biblioteca e Documentação. *Guia de apresentação de dissertações, teses e monografias*. Elaborado por Anneliese Carneiro da Cunha, Maria Julia de A.L. Freddi, Maria Fazanelli Crestana, Marinalva de Souza Aragão, Suely Campos Cardoso, Valéria Vilhena. 3. ed. São Paulo: Divisão de Biblioteca e Documentação; 2011.

Abreviatura dos títulos dos periódicos de acordo com *List of Journals Indexed in Index Medicus*.

Agradecimentos

Agradeço a Deus por me acompanhar em todo o meu caminho, me amparando em momentos difíceis, até quando eu duvidava de mim mesmo.

Gostaria de agradecer aos meus pais, Marta e Marivaldo, por me ensinarem o verdadeiro significado de caráter, integridade, respeito e, acima de tudo amor, o maior dos ensinamentos. Obrigado pelo apoio incondicional.

Agradeço a Bruna, minha noiva, por todo o carinho, compreensão e cumplicidade ao longo de todos estes anos ao meu lado.

Agradeço ao meu orientador, Prof. Dr. Delmar, por me receber no serviço e por todos os ensinamentos e conhecimentos a mim transmitidos.

Agradeço a todos do LIM 25, em especial Prof. Dr. Alexander, Elisângela, Lucas, Lilian, Betsaida, Eliete, Graça e Geni pela ajuda na execução do projeto, pelas horas de descontração no laboratório e pela amizade verdadeira.

Agradeço a todos do LIM 15 e do SELA, em especial a Profa. Dra. Suely Marie e Profa. Dra. Sueli Shinjo pela ajuda e apoio na execução do projeto.

Agradeço a todos os residentes e assistentes que colaboraram, de alguma forma, direta ou indiretamente para a execução do projeto.

Agradeço a cada professor que já tive e terei, pois sem eles não chegaria aqui.

Agradeço a FAPESP pelo apoio financeiro destinado a este trabalho.

SUMÁRIO

SUMÁRIO

LISTA DE ABREVIATURAS

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE GRÁFICOS

RESUMO

ABSTRACT

INTRODUÇÃO

1. Neoplasia Endócrina Múltipla tipo 1.....	2
1.1 Tumores principais associados a NEM1.....	2
1.1.1 Hiperparatireoidismo	2
1.1.2 Adenomas hipofisários.....	3
1.1.3 Tumores neuroendócrinos enteropancreáticos.....	3
1.2 Aspectos genéticos.....	4
1.3 Rastreamento gênico e clínico.....	6
2. Sequenciamento gênico Sanger e de nova geração.....	7
3. Justificativa.....	13
4. Escolha da plataforma NGS.....	14
5. Objetivo.....	18
6. Material e métodos.....	19
6.1 População e desenho do estudo.....	19
6.2 Extração de DNA e enriquecimento da região alvo.....	20
6.2.1 Extração de DNA.....	20
6.2.2 Enriquecimento do gene <i>MEN1</i> para SS.....	20
6.2.3 Enriquecimento da região genômica do gene <i>MEN1</i> para NGS...22	
6.3 Sequenciamento pelo método de Sanger (SS).....	23
6.4 Sequenciamento de nova geração (NGS).....	24
6.5 Análise dos dados – bioinformática.....	26
6.6 MLPA.....	27
6.7 Análise de custo-efetividade.....	28
7. Resultados.....	30

7.1	Quantificação e qualificação dos gDNAs extraídos por técnica de <i>Salting-out</i> e resultados de amplificações.....	30
7.2	Fase 1: Resultados e métricas.....	31
7.3	Fase 2: resultados e métricas.....	32
7.4	Fase 1: determinação da concordância, sensibilidade e especificidade do método de NGS.....	34
7.5	Fase 2: determinação da concordância, reprodutibilidade, sensibilidade e especificidade do método de NGS.....	35
7.6	Variantes.....	36
7.7	Análise dos resultados.....	43
7.8	Análise de custo.....	43
7.9	Práticas complementares: MLPA.....	44
8.	Discussão.....	47
8.1	Diagnóstico Genético na prática clínica: A migração do SS para a aplicação de NGS.....	47
8.2	Determinação de métricas e parâmetros mínimos.....	52
8.2.1	Amplificação por PCR.....	52
8.2.2	Estabelecimento das métricas ideais para NGS.....	53
8.3	Normalização das amostras.....	60
8.4	Reprodutibilidade, repetitividade, sensibilidade e especificidade.....	61
8.5	Ferramentas, fluxo de análise e rotinas complementares.....	64
8.6	Determinação de custo.....	66
8.7	EVEREST e STARD.....	67
9.	Conclusão.....	72

REFERÊNCIAS

LISTA DE ABREVIATURAS

NEM1 – Neoplasia Endócrina Múltipla tipo 1

HPT – Hiperparatireoidismo

DNA – Ácido desoxirribonucleico (do inglês *Deoxyribonucleic acid*)

MLPA – reação de amplificação de sonda ligação-dependente (do inglês *Multiplex Ligation-dependent Probe Amplification*)

ddNTPs – Dideoxynucleotídeos

SS – Sequenciamento Sanger

Pb – Pares de bases

NGS – Sequenciamento de nova geração (do inglês *Next-Generation Sequencing*).

GA – Genome Analyzer

Ion PGM – *Personal Genome Machine*

SBS – Sequenciamento Baseado em Síntese

PCR – Reação em Cadeia de Polimerase (do inglês *Polymerase Chain Reaction*)

dsDNA – DNA de fita dupla (do inglês *double stranded DNA*)

ssDNA – DNA de fita simples (do inglês *single stranded DNA*)

MSR – *MiSeq Reporter Software*

USH – Síndrome de Usher

MPS – Sequenciamento paralelo em massa (do inglês *Massive Parallel Sequencing*)

EC – Eletroforese Capilar

SNVs- Variações de um nucleotídeo (do inglês *Single Nucleotide Variation*)

mRNA – RNA mensageiro

CAPPesq – Comissão de Ética para Análise de Projetos de Pesquisa

FMUSP – Faculdade de Medicina da Universidade de São Paulo

PIT – Adenoma Pituitário

PET – Tumores Neuroendócrinos Pancreáticos

gDNA – DNA genômico

qPCR – PCR quantitativo

IGV – *Integrative Genomics Viewer*

LOVD - *Leiden Open Variation Database*

COSMIC - *Catalogue of Somatic Mutations in Cancer*

HGMD – *Human Gene Mutation Database*

HSF – *Human Splicing Finder*

EVS - *Exome Variant Server*

ExAC - *Exome Aggregation Consortium*

Mreads – Milhões de leituras

KCluster –quilo *Clusters*

Mbases – Milhões de bases

IC 95% - Intervalo de confiança de 95%

VPP – Valor preditivo positivo

VPN – Valor preditivo negativo

UTRs – Regiões não traduzidas (do inglês *Untranslated Regions*)

VCF – Arquivo *Variant Call Format*

In/Dels – Inserções e deleções de nucleotídeos

STARD – *Statement for Reporting Studies of Diagnostic Accuracy*

LISTA DE FIGURAS

Figura 1: Princípio de preparação de biblioteca utilizando plataforma de NGS – Illumina

Figura 2: Illumina MiSeq e sua *Flowcell*

Figura 3: Ion PGM e seus *microchips*

Figura 4: Área amplificada e sequenciada por SS do gene *MEN1*, considerando transcrito NM_130799, 5'-3'

Figura 5: Área amplificada por PCR longa e sequenciada por NGS do gene *MEN1*, considerando transcrito NM_130799, 5'-3'

Figura 6: Processo de montagem da biblioteca Illumina – Nextera

Figura 7: Representação da localização das sondas de MLPA dos kits utilizados no estudo dentro da região do gene *MEN1* (Representação na fita senso)

Figura 8: Imagem de Gel de agarose – PCR longa.

Figura 9: Porcentagem dos *reads* de cada amostra com relação ao total da corrida

Figura 10: Imagem da posição da variante NM_130799.2, exon2, c.201delC: p.A68Pfs*51, não chamada durante a corrida #1, mas visível no programa IGV

Figura 11: Variante NM_130799.2, exon2, c.201delC: p.A68Pfs*51, encontrada na corrida #2 e #4, respectivamente.

Figura 12: Representação de grandes deleções detectadas por MLPA em 4% dos indivíduos

Figura 13: Representação gráfica das leituras obtidas ao longo da região genômica do gene *MEN1*, no sentido 3' – 5' do gene

LISTA DE TABELAS

Tabela 1: Comparação entre Ion PGM e Illumina MiSeq

Tabelas 2: Sequências iniciadoras utilizadas em PCRs para SS

Tabela 3: Análise comparativa entre o rendimento esperado para kit MiSeq v2 (500 ciclos) e o rendimento da corrida realizada durante a fase 1

Tabela 4: Análise comparativa entre o rendimento esperado para kit MiSeq Nano v2 (500 ciclos) e o rendimento obtido com cada uma das três corridas realizadas durante a fase 2

Tabela 5: Métricas de NGS (cobertura e profundidade médias $\pm$ DP) calculadas, tendo como região alvo todo o gene *MEN1*

Tabela 6: Métricas de NGS (profundidade média $\pm$ DP) calculadas, tendo como região alvo somente a região codificadora do gene *MEN1*

Tabela 7: Valores de precisão, exatidão, sensibilidade, especificidade, valor preditivo positivo (VPP) e valor preditivo negativo (VPN) considerando a média obtida na análise e reprodução dos 76 pacientes com as corridas #1, #2 e #4

Tabela 8: Variantes encontradas nos 44 indivíduos portadores de diagnóstico genético da síndrome NEM1 que potencialmente afetam funcionamento da proteína Menin e suas frequências intrapopulacionais. **Variantes reportadas neste estudo e descrita no banco de dados com posições coincidentes

Tabela 9: Descrição de variantes benignas, localização, frequência de indivíduos portadores, distribuição alélica e referência em bancos de dados

Tabela 10: Custeio da aplicação do método de NGS para sequenciamento do gene *MEN1* (7189pb) vs método de SS para sequenciamento dos exons e regiões de fronteira do gene (4106pb).

Tabela 11: Dados de quantificação por qPCR

LISTA DE GRÁFICOS

Gráfico 1: Comparação entre as corridas #2 e #4 com relação ao cálculo da profundidade média esperada baseando-se na capacidade da flowcell e no volume de reads obtidos em cada uma das corridas, com a profundidade média realmente obtida nas duas corridas

Gráfico 2: Correlação entre quantidade de dados gerados com quantidade de clusters obtidos na corrida

Gráfico 3a: Correlação entre leituras obtidas e profundidade média de

Gráfico 3b: Correlação entre profundidade média e porcentagens de bases com profundidades mínimas obtidas

Gráfico 3c: Correlação entre leituras obtidas por amostra e porcentagens de profundidades mínimas obtidas

Carvalho RA. *Avaliação do método de sequenciamento de nova geração no diagnóstico genético de neoplasia endócrina múltipla tipo 1* [Dissertação]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2016.

A neoplasia endócrina múltipla tipo 1 (NEM1) é uma doença genética, de herança autossômica dominante, caracterizada pelo desenvolvimento de tumores endócrinos acometendo, principalmente, hipófise, paratireoide e pâncreas/duodeno endócrinos. É causada, principalmente, por mutação germinativa no gene supressor tumoral *MEN1* (11q13). A tumorigênese segue o modelo de Knudson (1971). O diagnóstico genético de famílias com NEM1 reconhece os portadores assintomáticos de mutação *MEN1*, permite o diagnóstico e tratamento precoce de tumores, promove a redução da morbimortalidade relacionada à NEM1 e exclui familiares não portadores de mutação do rastreamento clínico periódico. O diagnóstico genético de NEM1 tem sido realizado por meio da técnica de sequenciamento Sanger. Entretanto, limitações desta técnica a tornam menos custo efetiva, devido a sua reduzida capacidade de geração de dados, que leva a necessidade de obtenção de produtos de PCR de até 700 pb para adequada leitura do sequenciamento. Além disto, condições específicas do gene *MEN1*, como a ausência de "hot spots" mutacionais, levam a necessidade de sequenciamento de toda sua extensão (7Kb) e contribuem para tornar esta técnica laboriosa e dispendiosa. A subdivisão do gene para sequenciamento Sanger pode ocultar informações, principalmente de regiões intrônicas, que podem ser importantes para o desenvolvimento da doença. Tais dificuldades impedem a incorporação do diagnóstico gênico de NEM1 na prática clínica. Desde 2005, estão disponíveis tecnologias denominadas NGS (Next-Generation Sequencing), que consistem em ferramentas para o sequenciamento genético com capacidade aumentada de geração de dados, tornando-as mais atrativas e de melhor custo-benefício. O NGS confere, ainda, maior velocidade ao processo de obtenção de dados e detém a capacidade de realizar a leitura completa do gene, incluindo regiões promotoras e intrônicas. Por isto, torna a leitura mais ampla e informativa, sem desconsiderar aspectos qualitativos. Dentre várias opções de NGS disponíveis, plataformas leves são consideradas mais adequadas para aplicação clínica, destacando-se as plataformas Ion PGM e Illumina MiSeq. Uma forte tendência tem sido mostrada de migração do sequenciamento Sanger para o NGS, incluindo a aplicação da mesma em diagnóstico genético de doenças complexas e de câncer hereditário. Entretanto, não há estudos prévios envolvendo NGS em NEM1. Diante disto, foi avaliado a qualidade desta técnica como método de diagnóstico genético em NEM1 em comparação ao sequenciamento Sanger. Objetivos: validação da técnica de NGS utilizando como parâmetro o sequenciamento Sanger; avaliação da sensibilidade, especificidade e relação custo-benefício do NGS. Para tal, foram analisados 76 casos-índices com diagnóstico clínico de NEM1 na plataforma Illumina MiSeq. As análises foram subdivididas em duas fases. O enriquecimento da região genômica do gene *MEN1* foi realizado por meio de PCR longa. Com

base nos dados obtidos foi possível aferir 96% de reprodutibilidade entre as diferentes fases do estudo e aproximadamente 99% de precisão para detecção de variantes. Exatidão, sensibilidade e especificidade resultaram em 100%. Não houve falsos-positivos ou negativos. A técnica de NGS também se mostrou mais custo-efetiva do que o sequenciamento Sanger. Este estudo permitiu validar e introduzir esta técnica como ferramenta de diagnóstico gênico de NEM1 para rastreamento genético de casos-índices.

Descritores: Análise de sequência de DNA; diagnóstico; Neoplasia endócrina múltipla tipo 1; Técnicas genéticas; Variação estrutural do genoma.

ABSTRACT

Carvalho RA. *Evaluation of next generation sequencing in genetic diagnosis of multiple endocrine neoplasia type 1* [dissertation]. São Paulo: "Faculdade de Medicina, Universidade de São Paulo"; 2016.

The multiple endocrine neoplasia type 1 (MEN1) is a genetic, autosomal and dominant disease and is correlated with the development of endocrine tumors affecting pituitary gland, parathyroid, endocrine pancreas or duodenum. It is mainly caused by a germinative mutation in tumor suppressor gene *MEN1* (11q13). The tumorigenesis follows the Knudson's model (1971). Genetic diagnosis of families with MEN1 is essential to recognize asymptomatic mutation carriers, and allows an earlier detection and treatment of tumors leading to a reduction of mortality and morbidity associated to MEN1. Furthermore, it can exclude family members that do not carry mutations from the periodical screening. The genetic diagnosis for MEN1 is held using Sanger sequencing. However, limitations of this technique make it less cost-effective, mostly, the less capacity of data generation that leads to the need of PCR products up to 700 bp to obtain a suitable read. Moreover, specific conditions of the *MEN1* gene contribute to make this process more laborious and expensive, like the need to read all gene sequence (7kb) to make a correct analysis due to the absence of "hot spots". This way, the need of "fragmentation" to allow the sequencing can hide important information to disease development, mostly in introns. These limitations preclude the clinical application of genetic diagnosis of MEN1. Since 2005, new technologies are available; they are called Next Generation Sequencing (NGS) and consist in a new tool that allows the same sequencing, but with a larger data generation capacity, making them more attractive and cost-effective. The NGS also gives a higher speed to the process of data acquiring and allows the complete read of gene, including promoters and introns. Therefore, it makes the results more informative, not forgetting quality aspects. Among a lot of options of NGS available, lighter platforms are recommended, for example, Ion PGM and Illumina MiSeq. A strong tendency has been shown in order to change the Sanger sequencing to NGS, including clinical application to genetic diagnosis of complex diseases and inherited cancer. However, there is no previous studies evaluating NGS to MEN1 genetic diagnosis. Thus, present study evaluated NGS as a genetic diagnosis method for MEN1, comparing with Sanger sequencing. This study aimed to validate the NGS method using as model the Sanger sequencing and evaluated sensibility, specificity and cost-effectiveness of NGS. For this purpose, 76 index-cases with clinical MEN1 diagnosis were analyzed on Illumina MiSeq. Analyses were divided in two phases. After analyses, 96% of reproducibility and 99% of precision were calculated. Accuracy, sensibility and specificity were resulted in 100%. There were no false negatives or positives. NGS showed more cost-effectiveness with lower costs. This study allowed validation of genetic screening of MEN1 index-cases applying NGS platform.

Descriptors: Sequencing analysis, DNA; diagnosis; Multiple Endocrine Neoplasia type 1; Genetic techniques; Genomic structural variation.

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Carvalho, Rafael Arrabaça de

Avaliação do método de sequenciamento de nova geração no diagnóstico genético de neoplasia endócrina múltipla tipo 1 / Rafael Arrabaça de Carvalho. -- São Paulo, 2016.

Dissertação(mestrado)--Faculdade de Medicina da Universidade de São Paulo.
Programa de Endocrinologia.

Orientador: Delmar Muniz Lourenço Junior.

Coorientador: Alexander Augusto de Lima Jorge.

Descritores: 1.Análise de sequência de DNA 2.Diagnóstico 3.Neoplasia endócrina múltipla tipo 1 4.Técnicas genéticas 5.Variação estrutural do genoma

USP/FM/DBD-247/16

1 INTRODUÇÃO

1. Neoplasia endócrina múltipla tipo 1

Padrão de herança, diagnóstico clínico e gênico, penetrância, variante fenotípica

A neoplasia endócrina múltipla tipo 1 (NEM1) é uma síndrome genética de padrão de herança autossômico dominante cuja prevalência é estimada em 1:30.000. A NEM1 é causada, principalmente, por um gene supressor de tumor, denominado gene *MEN1*. Os portadores de variante patogênica germinativa neste gene apresentam uma elevada penetrância da doença. Assim, quase 100% dos portadores de variante patogênica terão desenvolvido pelo menos um dos tumores relacionados à síndrome ao redor dos 50-60 anos de idade. A NEM1 pode ter uma apresentação clínica familiar ou esporádica (1-6).

O diagnóstico clínico de NEM1 define-se na apresentação de neoplasias em pelo menos dois dos três principais componentes da síndrome: glândula paratireoide, hipófise anterior e células endócrinas do pâncreas ou do duodeno.

O diagnóstico clínico de NEM1 familiar é estabelecido quando se tem o diagnóstico clínico de NEM1 em um paciente (caso-índice) associado ao diagnóstico clínico em um parente de primeiro grau de pelo menos um dos tumores endócrinos principais relacionados à NEM1 – adenoma hipofisário, de paratireoide ou tumores neuroendócrinos enteropancreáticos (1-6).

A penetrância de cada um dos tumores endócrinos principais relacionados à NEM1 é apresentada na tabela 1 em anexo (anexo 1) (6).

Além de múltiplas glândulas endócrinas serem acometidas por tumores, múltiplos tumores frequentemente ocorrem na mesma glândula endócrina caracterizando a multicentricidade tumoral em NEM1. Além disto, outra característica principal em NEM1 é a enorme variabilidade fenotípica intra e inter-familiar (1-6).

O diagnóstico gênico de NEM1 é estabelecido, em sobreposição ao diagnóstico clínico da doença, quando se identifica um indivíduo portador de uma mutação germinativa no gene *MEN1* (1-6).

1.1 Tumores principais associados à NEM1

1.1.1 Hiperparatireoidismo

O HPT é a manifestação clínica mais frequente em NEM1, com uma penetrância estimada de 40% aos 20 anos e próxima de 100% aos 50-60 anos (7). Estudo recente de nosso grupo revelou informações sobre a história natural da doença renal e óssea em casos com HPT associado à NEM1 (8). Assim, a maioria destes pacientes já apresenta urolitíase antes dos 30 anos de idade e uma prevalência maior destes pacientes evolui com redução

da função renal após os 50 anos (8). Em uma metanálise discutimos as características clínicas específicas do HPT associado a NEM1 em comparação ao HPT esporádico (9). O tratamento do HPT é eminentemente cirúrgico e duas técnicas têm sido abordadas: a paratireoidectomia subtotal e a total. Esta última abordagem é associada com implante de tecido paratireóideo em antebraço. Nosso grupo tem utilizado historicamente como rotina a última abordagem com resultados bastante satisfatórios, embora a maioria dos centros especializados em NEM1 advogue preferência para a paratireoidectomia subtotal (10). Estudos avaliando a evolução do tratamento cirúrgico do HPT associado a NEM1 demonstraram ganho, em curto prazo, de massa óssea na coluna e no fêmur (11, 12).

1.1.2 Adenomas hipofisários

A penetrância de adenomas hipofisários é variável (30-70%) em diferentes casuísticas (13). Dentre os subtipos de tumores hipofisários, os prolactinomas são os mais frequentes em NEM1 seguido dos adenomas co-secretores, principalmente de prolactina e hormônio de crescimento, seguido dos não-funcionantes. Acromegalia e doença de Cushing são menos prevalentes (14). Um estudo francês clássico demonstrou que os tumores hipofisários associados à NEM1 são, geralmente, mais agressivos que os adenomas hipofisários esporádicos (14). Assim, tais tumores apresentam-se mais frequentemente como macroadenomas. Também, nos casos de prolactinomas, tendem a ser mais resistentes ao tratamento clínico medicamentoso (14). Estudo patológico extenso revelou que os adenomas hipofisários associados a NEM1 são, mais frequentemente, plurihormonais e duplos (15). Em geral, o tratamento dos adenomas hipofisários associados a NEM1 não difere daquele habitualmente preconizado para os casos esporádicos (1-6, 13, 14). Nosso grupo descreveu uma extensa família com NEM1, onde foi caracterizado efeito fundador, com prevalência relativamente elevada de adenomas hipofisários, especialmente de prolactinomas (13).

1.1.3 Tumores neuroendócrinos enteropancreáticos

Os tumores neuroendócrinos enteropancreáticos em NEM1 podem ser funcionantes ou não-funcionantes. Dentre os funcionantes, o mais prevalente é o gastrinoma seguido do insulinoma. Outros tumores funcionantes, como vipomas e glucagonomas, têm potencial maligno elevado e são raros em NEM1 (1-6; 16).

Os gastrinomas em NEM1 são, na maioria dos casos, localizados no duodeno, se apresentam como tumores múltiplos, pequenos, de crescimento lento e de comportamento biológico indolente sendo diagnosticados em NEM1 principalmente após 40 anos de

idade (1-7; 16-19). O controle da doença é, geralmente, bem-sucedido com o uso de inibidores de bomba de prótons, que são eficientes no controle da hiperacidez gástrica (1-7; 16-19). O tratamento cirúrgico é controverso. Em geral, cirurgia é recomendada quando os tumores são maiores de 2-3 cm pelo maior risco potencial de malignidade ($> 40\%$). Os insulinomas associados à NEM1 são mais prevalentes abaixo dos 40 anos. Em geral, são diagnosticados em associação com tumores pancreáticos não-funcionantes múltiplos o que, geralmente, implica em uma cirurgia pancreática mais extensa do que a geralmente recomendada para a forma esporádica de insulinoma (6; 16-19). Atualmente, os tumores pancreáticos não-funcionantes são considerados os tumores pancreáticos mais prevalentes em NEM1 e, representam a principal causa de óbito relacionada à síndrome (20). Foi descrito uma associação entre o tamanho destes tumores e o risco de malignidade. Assim, tumores maiores de 2 cm apresentam um risco estimado de malignidade próximo de 20% alcançando 40% quando são maiores do que 3 cm. Por isto, a maioria dos centros no mundo preconiza cirurgia para tumores maiores de 2 cm enquanto alguns preferem indicação de cirurgia para tumores maiores de 1 cm quando o risco potencial de malignidade é próximo de 10% (6; 17-22).

Outros tumores endócrinos podem estar associados à NEM1 em uma penetrância menor (tabela 1). Dentre eles, se destaca o tumor neuroendócrino do timo que apresenta elevado grau de malignidade ($> 80\%$). Por este tumor estar associado a uma baixa taxa de cura, é indicada timentomia profilática quando os pacientes se submetem ao tratamento cirúrgico do HPT (23).

1.2 Aspectos genéticos

A NEM1 é causada, na maioria dos casos, por uma mutação germinativa no gene *MEN1*, um gene supressor de tumor. O gene *MEN1* foi localizado no braço longo do cromossomo 11 (11q13) por dois grupos de pesquisa independentes em 1997, tem 9 Kb, contém 10 exons e codifica uma proteína com função supressora tumoral, denominada menin (24; 25). Mais de 1300 mutações germinativas e somáticas foram identificadas até dez anos após sua associação com NEM1, uma evidência clara da ausência de “hot spot” mutacional. A grande maioria das mutações é inativadora (frameshift ou nonsense). Tais mutações codificam uma proteína truncada e com perda de função (26). Até o presente, não foi identificada qualquer relação genótipo-fenótipo evidente em NEM1. Foi demonstrado que o mecanismo de tumorigênese em NEM1 segue a teoria de Kudson (*two hit*), em acordo com a de outros genes supressores de tumor. Assim, a presença de

mutação germinativa *MEN1* determina, como desdobramento posterior, a perda ou inativação do segundo alelo normal (27; 28).

A proteína menin, de 67 Kd, é constituída de 610 aminoácidos, e está localizada no núcleo celular. A proteína menin é conhecida como uma proteína adaptadora capaz de interagir com várias outras proteínas. A menin é diretamente envolvida com o desenvolvimento e função, participa ativamente de vários processos celulares, como regulação da transcrição gênica, replicação e reparo do DNA e na transdução da sinalização. A menin interage com vários genes envolvidos em processos como apoptose, proliferação celular, reparo do DNA e com genes relacionados com funções metabólicas e endócrinas. A função normal da proteína menin é preservar a diferenciação da célula por modificação de proteínas (histonas) e a transcrição de genes responsáveis por inibirem a divisão celular. Quando a proteína menin é defeituosa, há uma perda deste controle, assim como diminuição da expressão de genes inibidores do ciclo celular, como *SMAD3*, *TGFβ*, *CDKN2C*, *CDKN2B* (p27), *hAsh2*, *caspase 8*, *TNFα* e aumento da expressão de genes ativadores do ciclo celular com *junD* e *hTERT* predispondo à tumorigênese. Foi demonstrado que a proteína menin é capaz de interagir com mais de 20 proteínas implicadas, direta ou indiretamente, com o controle da divisão celular. Recentemente, foi identificada a estrutura em cristal da proteína menin que poderá ampliar a compreensão da interação da menin com estas várias proteínas. As interações mais conhecidas da menin são com o fator de transcrição *JunD*, que é um inibidor de crescimento celular, e com *Smad3* que é inibidora da via de sinalização de *TGFβ* (29).

A presença de mutação germinativa em um dos alelos não é suficiente para o desenvolvimento da doença. Assim, o processo de tumorigênese somente se instala quando uma mutação somática ocorre no alelo normal. Por isto, o padrão de herança é dito “dominante” (mutação germinativa em um dos alelos) mas o mecanismo de tumorigênese é dito como “recessivo” (inativação bialélica tecidual) pela necessidade de ocorrer uma 2ª mutação em nível tecidual. Este mecanismo de tumorigênese foi descrito previamente por Knudson para retinoblastomas (*two hit*) (28).

Mais de 1400 mutações germinativas e somáticas foram identificadas no gene *MEN1* desde sua associação com NEM1, uma evidência clara da ausência de *hot spot* mutacionais. A grande maioria das mutações é inativadora (frameshift ou nonsense) gerando uma proteína menin truncada que perde sua função de controle e regulação do ciclo e da proliferação celular (26, 29, 30). Mutações germinativas NEM1 são identificadas em 75 a 95% dos pacientes com fenótipo NEM1 (31). Foi, também,

demonstrado que algumas famílias podem ter deleções grandes no gene *MEN1* (26, 32) que não são reconhecidas por sequenciamento Sanger e são identificadas pela técnica de MLPA (*multiplex ligation-dependent probe amplification*) (33). Deleções grosseiras por MLPA foram identificadas em 1% e 33% dos casos com fenótipo NEM1 (32, 34).

1.3 Rastreamento gênico e clínico

A descoberta do gene *MEN1*, como causa de NEM1, propiciou uma mudança gradual e decisiva no manejo clínico dos pacientes e de seus familiares (35-38). Assim, a análise genética conduziu à identificação dos familiares sob risco negativos para mutação germinativa *MEN1*, bem como de seus descendentes, permitindo a sua exclusão do rastreamento clínico periódico para tumores NEM1-relacionados. Da mesma forma, possibilitou a identificação de familiares portadores de variantes patogênicas no gene, sintomáticos ou assintomáticos, incluindo-os neste seguimento clínico. Além disto, a análise genética propicia a identificação de membros de família com NEM1 que, eventualmente, são diagnosticados com tumores endócrinos esporádicos ocorrendo em glândulas endócrinas frequentemente acometidas em NEM1. Nestes pacientes, chamados de fenocópias, não são encontradas a variante patogênica de seus familiares acometidos por NEM1. Tais casos, uma vez definidos como não portadores de mutação germinativa *MEN1* recebem o tratamento preconizado para tumor esporádico e podem ser excluídos do seguimento clínico periódico de rastreamento ativo de tumores relacionados à NEM1 indicado para os portadores de mutação germinativa (1; 6; 37; 39-44).

Em 2001, um Consenso sobre manejo clínico de NEM1 foi publicado baseando-se em estudos de fenótipo, história natural e penetrância dos tumores associados a NEM1 (1). Tendo como foco o diagnóstico e tratamento precoce dos tumores relacionados à NEM1, este consenso recomendou um programa de rastreamento clínico com dosagens hormonais e realização de imagens periódicas (1). Desde então, vários outros estudos foram publicados sugerindo diferentes programas de rastreamento clínico (37; 41; 44; 45).

Em 2012, um novo consenso foi publicado baseado nos mesmos critérios adotados pelo anterior (6). O objetivo principal do rastreamento clínico, laboratorial e radiológico de portadores de mutação no gene *MEN1* assintomáticos é diminuir a morbidade e mortalidade relacionada à doença e aumentar a expectativa de vida destes pacientes (34; 36; 38; 39; 43-45).

Em nosso Serviço, um programa de rastreamento clínico de famílias com NEM1 foi iniciado em 1997. Em 2002, as primeiras famílias com NEM1 foram caracterizadas geneticamente (2). Mutações germinativas no gene *MEN1* foram identificadas pelo sequenciamento automático (Sanger) e seus membros sob risco foram rastreados geneticamente pelo método de eletroforese em gel por gradiente de denaturação (2). A partir de 2004, o rastreamento gênico através do sequenciamento automático foi retomado e intensificado, tanto para investigação dos casos-índices como para o rastreamento dos familiares sob risco. Em 2007, o grupo publicou estudo descrevendo mutações germinativas em 14 famílias com NEM1 (46). Paralelamente, nós reportamos o impacto da implantação do programa de rastreamento clínico e gênico direcionado ao atendimento de famílias com NEM1 no aumento do diagnóstico e tratamento precoce bem como na diminuição da morbidade e mortalidade em nosso meio (43). Posteriormente, vários estudos, tanto clínicos como genéticos, se sucederam (8-13; 47; 48).

2 Sequenciamento gênico Sanger e de nova geração

A técnica de sequenciamento gênico, descrita em 1975, determinou o início de uma nova fase da genética, conduzindo, desde então, a um crescimento exponencial do conhecimento nesta área (49). Isto foi possível pela possibilidade de identificar as sequências de nucleotídeos do DNA culminando no início do mapeamento genômico dos seres vivos, na descoberta de genes e de suas funções bem como na associação direta dos genes com as doenças. A técnica clássica de sequenciamento que vem sendo utilizada nas últimas décadas é conhecida como técnica de Sanger e se baseia no uso de sequências de nucleotídeos iniciadores, que se anelam com a sequência alvo, iniciando o processo de replicação do DNA por ação de uma DNA polimerase. A presença de nucleotídeos modificados na reação, didesoxiribonucleotídeos (ddNTPs), promove a parada aleatória da extensão da fita de DNA em diversos pontos distintos. A técnica de Sanger permite o sequenciamento de 400-700 pares de bases (pb) de uma determinada região alvo em cada reação.

O sequenciamento gênico Sanger (SS) propiciou a identificação de variantes germinativas ou somáticas em um determinado gene segregando com um dado fenótipo. Assim, uma causa genética foi estabelecida para um imenso número de doenças. A descoberta de genes causadores de doenças monogênicas, tanto no final da década de 80 e, principalmente, a partir da década de 90, associada a gradual redução dos custos com o sequenciamento gênico propiciou a popularização do sequenciamento gênico Sanger.

Assim, nos últimos 10-15 anos, esta metodologia foi incorporada como teste diagnóstico para várias doenças genéticas. Tais testes genéticos permitiram a identificação de indivíduos portadores, tanto sintomáticos como assintomáticos, de uma determinada mutação germinativa que implicaria em uma maior susceptibilidade para o desenvolvimento de uma dada doença. Também, possibilitou excluir os familiares sob risco não portadores de mutação do risco de doença. Tanto a neoplasia endócrina múltipla tipo 1 como tipo 2 são grandes exemplos da incorporação do sequenciamento genético na prática médica (35-38; 42-46).

Apesar da técnica de Sanger ter permitido a introdução do sequenciamento gênico como ferramenta diagnóstica para muitas doenças, algumas limitações dificultam a implantação desta técnica na rotina laboratorial. Dentre elas, se destacam doenças genéticas em que não se identificam pontos frequentemente susceptíveis ao achado de mutações, denominadas de regiões “hot spot”. Esta heterogeneidade gênica implica na necessidade de realização de sequenciamento de toda a região codificadora do gene envolvido. Esta questão é amplificada se o gene responsável pela doença é relativamente grande e se mutações são comuns em regiões intrônicas ou regulatórias do gene. Como exemplos disto, temos o gene *MEN1* e o da neurofibromatose (*NFI*). Nesta situação, há a necessidade de vários primers para cobertura adequada, tanto de toda a região codificadora do gene bem como das fronteiras intron-exon, encarecendo a técnica de sequenciamento e a tornando mais laboriosa.

Assim, mais recentemente, novas gerações de sequenciadores automáticos têm sido desenvolvidas com o objetivo de genotipagem mais ampla, estendendo a área de cobertura de um determinado gene bem como incluindo a cobertura simultânea de outros genes potencialmente relacionados a uma dada doença. Além disto, estes sequenciadores, conhecidos como sequenciadores de nova geração (NGS) apresentam maior velocidade de obtenção de informações com menor custo estimado. O NGS propicia, também, a avaliação de toda a região codificadora de um determinado gene e até mesmo de vários genes simultaneamente em uma única reação (50).

Em caráter comparativo, em 1987 a Applied Biosystems introduziu o primeiro sequenciador automático, o AB370, que detectava 96 bases de uma única vez com capacidade diária de 500 kb, podendo chegar a leitura de 600 bases. O modelo mais recente deste sequenciador por eletroforese em capilar (ABI3730xl) chega a um limite diário de geração de dados de 2.88 Mb, podendo chegar a leitura de 900 bases. Desde 1998 a eletroforese por capilar permitiu avanços consideráveis, incluindo o projeto

genoma (concluído em 2001) (51). Paralelamente ao projeto, já havia um esforço para desenvolvimento de técnicas que melhorassem tanto o rendimento como o custo das mesmas. Em 2005 foram lançados os sequenciadores de segunda geração; o 454 (criado pela 454 e, posteriormente, adquirida pela Roche), o *Genome Analyzer* (criado pela Solexa e, posteriormente, adquirida pela Illumina) e o SOLiD (criado pela Agencourt e, posteriormente, adquirida pela Applied Biosystems) (50).

Após o desenvolvimento destas três plataformas surgiram outras, voltadas para necessidades mais específicas, como a plataforma Illumina GA/HiSeq (global *Beijing Genomics Institute* foram um dos primeiros a adquirirem a plataforma), criada em 2010, com capacidade maior, gerando 600 Gb por corrida (em 8 dias) e, segundo esta mesma empresa global, a plataforma alcançará em um futuro próximo 1 Tb. O Illumina HiSeq 2000, comparado a outros, constitui o método mais barato de sequenciamento de nova geração em larga escala. Outras plataformas mais compactas, denominadas “de bancada” (*bench top*), como o Illumina MiSeq ou o Ion PGM (*Personal Genome Machine* - criada pela Ion Torrent, classificado como sequenciador de quarta geração) foram desenvolvidas subsequentemente (50).

A plataforma da Illumina (HiSeq ou MiSeq) consiste no sequenciamento baseado em síntese (SBS), na qual uma fita de DNA se liga a adaptadores presentes na placa (*Flowcell*), sendo amplificada clonalmente em fase sólida por PCR em ponte (*Bridge*

Figura 1: Princípio de preparação de biblioteca utilizando plataforma de NGS - Illumina

PCR), de maneira a permanecerem próximas, criando *clusters*, que irão conter, cada um deles, milhões de cópias do mesmo fragmento de DNA. Após aplicação de uma enzima de linearização, as fitas *double stranded* DNA (dsDNA) são convertidas em *single stranded* DNA (ssDNA). Deoxinucleotídeos contendo diferentes marcadores fluorescentes cliváveis e uma região bloqueadora removível são, então, adicionados à placa, permitindo a síntese da nova fita e o sequenciamento, realizado base-a-base (Figura 1).

A diferença básica entre estas duas plataformas consiste na menor capacidade de geração de dados e menor tamanho do MiSeq, sendo este rápido e recomendado para utilização em laboratórios menores, consequentemente, com demanda mais diluída (50).

O Illumina MiSeq pode ser considerado um aparelho versátil já que tem capacidade de realizar todas as etapas do processo em um mesmo aparelho (vai da amostra até a análise dos dados), dentro de aproximadamente 8 horas. Uma de suas aplicações relevantes é o sequenciamento de amplicons. Devido a sua capacidade melhorada de leitura, a etapa de análise para alinhamento dos resultados do sequenciamento ocorre de maneira mais satisfatória em comparação ao HiSeq (50). A tabela 1 ilustra as diferenças entre duas plataformas, o Illumina MiSeq e o Ion PGM.

Tabela 1: Comparação entre Ion PGM e Illumina MiSeq. Fonte: Adaptado de LIU *et al.*, 2012

	Ion PGM	Illumina MiSeq
Geração de dados	80 Mb - 2Gb	300 Mb – 15 Gb
Capacidade de leitura	200 - 400 bp	75 - 600 bp
Tempo de máquina	2 - 7 horas (sequenciamento)	4 - 40 horas (Sequenciamento e análise)
Tempo de preparação da amostra (jornada de 8h)	3 dias	2 dias
Método de sequenciamento	Tecnologia de Chips semicondutores com detecção de íons H ⁺	Sequenciamento baseado em síntese (SBS)
Quantidade de material	µg	ng (Nextera)
Análises de dados	Fora do aparelho (Servidor)	No próprio aparelho (MSR)

Até o presente momento não existem estudos validando NGS como método diagnóstico para NEM1. Entretanto, existem trabalhos sugerindo a aplicação deste método, aliado a uma abordagem mais profunda para definição do diagnóstico genético (52) ou utilizando painel gênico em casos suspeitos da doença (53).

Nos últimos anos, vários estudos têm sido publicados aplicando o NGS em oncologia, incluindo síndromes genéticas de predisposição a câncer (54, 55, 56). Como exemplo disto, um estudo avaliou um *pool* de indivíduos normais (não afetados) em busca de variantes no gene *RET*, desde o intron 9 até o 16 onde, em geral, mais de 98% das variantes germinativas patogênicas são identificadas, utilizando o *Illumina genome analyzer*. Este estudo ressalta a importância do sequenciamento de exomas ou até mesmo de todo o genoma para encontrar e compreender novas alterações. Encontrar novas variantes e analisar sua frequência na população normal pode lançar luz sobre a patogenicidade ou não daquela variante em questão, permitindo um melhor entendimento de doenças complexas (55). Outro estudo igualmente recente também revê a possibilidade da aplicação clínica de métodos de NGS, citando a boa relação custo-benefício que esta metodologia poderia trazer ao estudo de câncer hereditário (54)

Para o diagnóstico de doenças mais complexas, o estudo de Licastro *et al.* (2012) avaliou 12 pacientes para os genes referentes ao desenvolvimento da síndrome de Usher (USH); neste grupo, a maioria deles não detinha mutações conhecidas em genes estabelecidos como causadores da doença, fato detectado após genotipagem por microarray. No estudo foram utilizados diferentes técnicas e aparelhos: o SOLiD foi utilizado para sequenciamento exômico enquanto que o Illumina Genome Analyzer II e o Roche 454 para sequenciamento de amplicons (PCR longa) por NGS. O sequenciamento de PCR longa demonstrou uma cobertura de mais de 90% dos genes relevantes para USH, enquanto que o sequenciamento exômico obteve uma cobertura relativamente baixa (50%), considerando os genes alvo. Desta forma, pode-se dizer que o sequenciamento de amplicons gerados por PCR longa seria mais eficaz para diagnóstico molecular baseado em NGS, especificamente neste caso. Vale ressaltar também a acurácia do NGS, uma vez que foram encontradas 11 novas variantes nos genes relevantes para o curso da doença (58).

Outro recente estudo introduz uma revisão de literatura referente ao sequenciamento de nova geração relacionando-o com a clínica médica como sendo uma tecnologia revolucionária no campo da genômica, permitindo sequenciamento em larga escala, com grande capacidade de geração de dados, com boa relação custo-benefício e

sem sacrificar aspectos qualitativos ou referentes à acurácia. Tal relação se deve a grande capacidade de geração de dados desta nova tecnologia, situada na casa de gigabases por corrida. Nos últimos anos, as plataformas de NGS, também conhecidas como ferramentas de sequenciamento paralelo em massa (MPS – *Massive Parallel Sequencing*) têm sido aplicadas para diversos fins: sequenciamentos de regiões alvo, genomas completos, transcriptomas, marcadores epigenéticos e, inclusive, para fins diagnósticos. (59).

É visível o aumento na utilização destas novas técnicas no campo clínico, principalmente lançando mão de painéis de genes alvo buscando a compreensão de possíveis relações entre eles, ou confirmando suas respectivas independências. A aplicabilidade se torna ainda mais palpável no estudo de câncer, particularmente aqueles relacionados a mutações germinativas, em decorrência da rápida obtenção de dados. Contudo, as falhas também são bem estabelecidas; um exemplo destas é a ineficiência na definição de determinadas regiões (regiões ricas em G-C ou homopolímeros), na qual houve, em uma situação extrema a não captura de toda sequência (54).

Sabe-se que a busca por alterações germinativas conhecidas utilizando como ferramenta NGS pode ser de grande valia ao diagnóstico, especialmente em alguns casos bem definidos, como câncer hereditário de mama e ovário. Walsh *et al.* (2010) confeccionaram um painel customizado para busca de variantes patogênicas germinativas em 21 genes, incluindo BRCA1 e 2, conhecidas por causar tumores nestes dois sítios. O diagnóstico realizado, até então, era direcionado para a análise genética somente dos genes BRCA1 e 2 nos casos onde havia histórico familiar. Neste caso o sequenciador utilizado foi o *Genome Analyzer* com uma profundidade média >1200 X. Não foi obtido nenhum resultado de falso-positivo para quaisquer alterações, todas as substituições, pequenas e grandes deleções, duplicações e inserções foram detectadas de maneira satisfatória, com um bom custo-benefício, inclusive, considerando a região razoavelmente mais abrangente na análise feita por NGS, com relação a abordagem costumeira (60).

Contudo, outro estudo faz ressalvas quanto ao uso desta metodologia, argumentando que a chance de erro no sequenciamento paralelo seria maior do que no de Sanger, mas mesmo assim isso dependeria da profundidade (*read depth*) com a qual se garanta uma boa profundidade média de 40X a 80X (estes valores significam que uma mesma base será lida, paralelamente, de 40 a 80 vezes). Quanto mais elevada a profundidade de leitura, maior a quantidade de dados necessária por amostra, tornando, portanto, o processo mais caro (61). Desta forma, entender o “volume de informação”

necessária para um sequenciamento robusto e resultados confiáveis é fundamental para otimização.

O manejo dos dados obtidos a partir de NGS requer um conhecimento múltiplo e, em determinados casos, muito restrito a área de bioinformática e, em outros, em biologia. Sendo assim, o pesquisador deve deter conhecimento suficiente nestas áreas, principalmente se a aplicação da técnica é direcionada para a clínica.

Além disto, aplicação em paralelo de NGS com as técnicas convencionais permitem melhor compreensão da sua capacidade e confiabilidade que, quando validada em cada campo específico pode complementar o método já existente na definição de diagnóstico e, dependendo da doença, do prognóstico (62).

3. Justificativa

A ausência de “hot spot” em NEM1 determina a necessidade de realização de sequenciamento de toda a região codificadora do gene *MEN1* na busca de mutações germinativas responsáveis pela doença. Além disto, um número significativo de famílias apresenta mutações intrônicas tornando obrigatória a análise de regiões de fronteira intron-exon (32). Foi, também, demonstrado que algumas famílias podem ter grandes deleções do gene *MEN1* (32). Deleções grandes são perdidas pela técnica de sequenciamento automático. Assim, esta informação resulta na necessidade de complementação da análise gênica nos casos em que o sequenciamento Sanger não identificou mutações. Nestes casos, a realização de MLPA ou técnicas similares se faz necessária para exclusão definitiva de envolvimento do gene *MEN1* nestas famílias. Isto é necessário apesar da raridade de grandes deleções no gene *MEN1*.

Todos estes fatores agregados resultam no fato que a análise do gene *MEN1* é difícil e dispendiosa. Além disto, mutações germinativas no gene *MEN1* não são identificadas em até 10% dos pacientes com fenótipo NEM1, sugerindo que mutações germinativas em regiões regulatórias e não codificadoras do gene *MEN1* ou mesmo em outros genes de susceptibilidade poderiam ser causa de NEM1 (63). Em 2007, uma mutação no gene *CDKN1B/p27^{Kip1}* (cyclin-dependent kinase inhibitor 1b) foi relacionada ao fenótipo da NEM1 em família cuja análise prévia do gene *MEN1* foi negativa (64). Em seguida, um segundo caso foi diagnosticado com mutação neste gene apresentando-se com fenótipo NEM1-símile (65). Posteriormente outras 3 famílias foram, também, identificadas com mutações no gene *CDKN1B/p27^{Kip1}* (59). Devido à identificação de pacientes e famílias com mutação germinativa no gene *CDKN1B/p27^{Kip1}* e com tumores

endócrinos múltiplos, uma nova síndrome foi reconhecida e denominada NEM4 pelo “Online Mendelian Inheritance in Man” (OMIM) in 2007. Mutações germinativas foram, também, encontradas nos genes *CDKN2B*, *CDKN2C* e *CDKN1A* em 4 famílias não relacionadas, contudo, não respondem por mais de 3,5% dos casos com o fenótipo NEM1-símile (66). Mais recentemente, um paciente com fenótipo atípico de NEM1 foi identificado com mutação germinativa no gene *AIP* (67). Assim, outros genes ainda não conhecidos poderiam ser responsáveis por NEM1.

Neste contexto, o estudo genético através do NGS poderia auxiliar na identificação de mutações germinativas em regiões regulatórias e intrônicas do próprio gene *MEN1*, teoricamente não cobertas pela análise padrão do gene *MEN1* pelo sequenciamento Sanger, como em outros genes relacionados à NEM1. Soma-se a isto, a maior capacidade de geração de dados, amplificando a capacidade e velocidade de análise destes pacientes, aliado ao menor custo, corroborado nos trabalhos supracitados.

Espera-se que haja uma tendência nos próximos anos de migração da técnica padrão – sequenciamento Sanger – para a de NGS, independente da plataforma de escolha. Com sua boa relação custo-benefício, surge o interesse de avaliar a robustez desta técnica como método de diagnóstico genético em NEM1 comparativamente ao método clássico de sequenciamento automático Sanger, considerando as regiões de difícil acesso (regiões ricas em GC e homopolímeros).

Este projeto pretende contribuir para a incorporação da técnica de NGS em nosso meio pela avaliação da mesma como ferramenta potencial no diagnóstico genético de NEM1.

4. Escolha da plataforma de NGS

Em contraste com o mercado para eletroforese capilar (EC), exclusivamente fornecido pela *Thermo Fisher Scientific* (EUA), as plataformas de NGS são disponibilizadas por diversos fornecedores, desde o surgimento e distribuição da plataforma 454, desenvolvido pela empresa 454 que, posteriormente, foi adquirida pela Roche (Suíça). Na atualidade, as plataformas mais comumente utilizadas são Illumina (EUA), cujas possibilidades incluem sequenciadores denominados de segunda geração; MiniSeq, MiSeq, NextSeq 500 ou 550 e HiSeq, ou Ion Torrent da Thermo Fisher (EUA), cujas plataformas disponíveis, classificadas como quarta geração, são Ion PGM (*Personal Genome Machine*) e Ion Proton (52). Os sequenciadores de bancada (*bench top Next Generation Sequencers*) Illumina MiSeq (sequenciamento baseado em síntese) e Ion

PGM (sequenciador semiconductor) são plataformas com custos e capacidade de geração de dados mais adequadas à prática clínica. Estas plataformas utilizam processos e meios diferentes para realização do sequenciamento massivo paralelo. Assim, algumas características específicas bem como as vantagens e desvantagens de cada uma delas devem ser consideradas durante o processo de seleção da plataforma mais adequada ao objetivo proposto.

A plataforma MiSeq (Figura 2) permite, no máximo, a geração de 15Gb de dados, por meio da utilização do *kit* v3, operando com 600 ciclos (600pb por leitura) em célula de fluxo com 25 milhões de leituras. Com *kits* denominados “v2”, ainda disponíveis para uso, sua capacidade máxima cai para 7,5Gb, com 500 ciclos em célula de fluxo de 15 milhões de leituras. O menor *kit* disponível, denominado *nano*, com 150 ciclos, define a menor capacidade da plataforma, que é de 150Mb. O custo do equipamento é estimado em US\$120.000,00 (68).

Figura 2: Illumina MiSeq e sua célula de fluxo (*Flowcell*).

Já a plataforma Ion PGM (Figura 3), possui um custo operacional e de equipamento menor, sendo estimado em US\$80.000,00. Entretanto, sua capacidade máxima de geração de dados é de 2,2Gb, inferior ao oferecido pelo MiSeq. Também, na maioria de suas leituras com escore *phred* q20 (1 erro a cada 100 bases sequenciadas), enquanto que a plataforma MiSeq tem a maior parte das suas leituras acima do escore *phred* q30 (1 erro a cada 1000 bases). Quando comparados os custos por cada par de base sequenciados por ambas as plataformas, o Ion PGM se revela aproximadamente 2 vezes mais oneroso. Quanto à taxa de erros, o Ion PGM é mais eficiente na leitura de regiões repetitivas GC enquanto que o MiSeq é melhor com sequências AT. Contudo, o Ion PGM possui uma ausência de uniformidade na cobertura de regiões ricas em AT independente do método de fragmentação empregado. São comparáveis, também, taxa de leituras livres de erros, onde o Illumina MiSeq possui maior taxa de leituras livres de erros

(aproximadamente 5 vezes mais). Quando as regiões de homopolímeros são consideradas, o MiSeq gera erros em sequências homopoliméricas acima de 20pb, enquanto que estes erros são gerados em sequências repetitivas acima de 8pb com a plataforma Ion PGM, que também não consegue resolver regiões de homopolímeros maiores que 14pb. Vale ressaltar que, o Illumina MiSeq tende a gerar erros de leitura em regiões onde encontra a sequência GGC (68).

Figura 3: Ion PGM e seus *Microchips*, 314v2, 316v2 e 318v2, respectivamente.

O método de SBS (sequenciamento baseado em síntese), empregado pela plataforma MiSeq é relacionado com erros específicos que podem gerar falhas durante a chamada de variantes. Estes erros ocorrem, principalmente, em variantes de nucleotídeos únicos (SNVs) e, em menor extensão, com chamadas de deleções e inserções. Pode haver, também, ausência de uniformidade nos padrões de cobertura e profundidade em regiões ricas em GC e AT. Os problemas de cobertura e profundidade são causados, provavelmente, por reações de PCR ineficientes em determinadas regiões da sequência. Além disso, podem ocorrer falhas nas últimas leituras devido a fragmentos de tamanho reduzido e por problemas na amplificação da fita antisenso (69).

Vale ressaltar que todas são dotadas de vantagens e desvantagens e que, por isso, a escolha da plataforma a ser utilizada depende das peculiaridades e dos objetivos de cada estudo. De modo geral, busca-se alcançar a maior eficácia aliada ao menor custo. Durante o processo de seleção da plataforma deve ser considerado o tamanho da região alvo a ser enriquecida, o método de fragmentação e de enriquecimento a ser utilizado, composição AT/CG do alvo, a necessidade estimada de profundidade média de leituras, a distribuição das variantes bem como os tipos mais comuns encontrados na região de interesse e a presença de sequências repetitivas.

O gene *MEN1* possui, aproximadamente, 55% de conteúdo GC ao longo de toda sua cadeia genômica e aproximadamente 40% das variantes encontradas são variantes *frameshift* (31).

Assim, a seleção da plataforma de NGS para o estudo foi baseada na somatória das características inerentes ao gene *MEN1* que poderiam, potencialmente, ser mais bem avaliadas com a plataforma Illumina MiSeq, uma vez que a mesma confere maior confiabilidade no sequenciamento de regiões repetitivas, responde melhor a chamada de deleções e inserções com deslocamento de matriz de leitura do mRNA e possui maior qualidade de leitura (escore *phred* >q30).

5. Objetivo

Os objetivos principais deste estudo são:

- Padronização, validação e avaliação da sensibilidade da técnica de sequenciamento de nova geração (NGS) no diagnóstico de neoplasia endócrina múltipla tipo 1 (NEM1) comparada à técnica de sequenciamento Sanger por eletroforese capilar (EC).
- Avaliação do custo-benefício da técnica de NGS no diagnóstico genético de NEM1 comparativamente à técnica de sequenciamento Sanger por EC.

6. Material e métodos

6.1 População e desenho do estudo

O presente estudo teve seu início após assinatura do termo de consentimento livre e esclarecido previamente aprovado pela Comissão de Ética para Análise de Projetos de Pesquisa (CAPPesq) da Faculdade de Medicina da Universidade de São Paulo (FMUSP).

A casuística foi constituída por 76 pacientes, aparentemente não relacionados, que tiveram o diagnóstico clínico de NEM1 definido usando os critérios previamente estabelecidos por Consensos, ou seja, presença de, ao menos, duas neoplasias das três principais relacionados a síndrome. Por sua vez, o quadro de apresentação onde há um indivíduo com NEM1 e, pelo menos, um parente em primeiro grau com ao menos um dos tumores principais (paratireoide, pâncreas ou hipófise) corresponde a NEM1 familiar. Por sua vez, pacientes com NEM1 e história familiar negativa são caracterizados como NEM1 esporádica (1, 6). Inicialmente, foram pré-selecionados 77 indivíduos, contudo, um deles foi excluído do estudo por não preencher critérios clínicos para NEM1. Este indivíduo era portador de tumor endócrino primário em intestino delgado (íleo terminal) e um segundo tumor neuroendócrino em pâncreas.

Todos os 76 pacientes realizam acompanhamento clínico no Ambulatório de Neoplasia Endócrina Múltipla tipo 1 (A1MN 8002) da Disciplina de Endocrinologia e Metabologia do Hospital das Clínicas da FMUSP. NEM1 familiar foi caracterizada, seguindo os critérios estabelecidos por Consensos, em 49 casos-índices enquanto que os 27 casos-índices remanescentes tinham história familiar negativa e foram inicialmente caracterizados como NEM1 aparentemente esporádica.

A prevalência dos tumores MEN1-relacionados para classificação destes indivíduos foi: hiperparatireoidismo primário (HPT) (99%; 75/76), adenoma pituitário (PIT) (75%; 57/76) e tumores neuroendócrinos pancreáticos (PET) (82%; 62/76). A maioria dos pacientes (55%; 42/76) apresentava três tumores MEN1-relacionados. Os demais indivíduos (45%; 34/76) apresentavam dois dos principais tumores MEN1-relacionados com as seguintes associações: HPT/PET (25%; 19/76), HPT/PIT (18%; 14/76) and PET/PIT (1%; 1/76).

Inicialmente uma primeira reação de sequenciamento por NGS, com objetivo de otimização e padronização da rotina dentro do serviço onde a técnica seria validada, foi realizada em 23 indivíduos com genótipo conhecido; 22 casos-índices portadores de variante patogênica no gene *MEN1* e 1 indivíduo caso-índice, não portador de variante patogênica. Tais dados foram previamente obtidos por sequenciamento Sanger (SS).

Durante esta fase, doravante denominada fase 1, o pesquisador responsável pela execução não foi informado sobre qualquer identificação das amostras de modo a manter a análise cega. O critério primário para esta seleção foi presença de diferentes tipos de variantes (SNVs, deleções e inserções exônicas e em regiões de *splice*) localizadas em diversas posições ao longo do gene *MEN1*.

Posteriormente, na fase 2 do projeto, outras reações de sequenciamento por NGS envolveram os 23 pacientes inicialmente avaliados, juntamente com os outros 53 casos que compõem o restante da casuística. O sequenciamento destes indivíduos foi realizado ao longo de 3 corridas subsequentes realizadas na plataforma de NGS. Em paralelo, estes indivíduos foram sequenciados e analisados utilizando plataforma de SS para preservação do teste cego. Somente ao término das análises, os dados puderam ser confrontados.

Como complementação ao diagnóstico genético, todos os indivíduos, cuja análise genética por SS e por NGS foi negativa para identificação de variantes patogênicas no gene *MEN1*, foram submetidos à investigação de microdeleções no gene *MEN1* usando a técnica de *Multiplex Ligation-dependent Probe Amplification* (MLPA).

6.2 Extração de DNA e enriquecimento da região alvo

6.2.1 Extração de DNA

O DNA genômico (gDNA) foi extraído de leucócitos de sangue periférico pelo método *salting out*, descrito previamente (70), e, então, usado nas etapas subsequentes de SS e NGS descritas adiante. A purificação do gDNA extraído pelo método supracitado foi realizada por meio do *kit DNA Illustra Blood GenomicPrep Mini Spin Kit* (GE Healthcare, Reino Unido) quando dificuldades foram encontradas na fase de amplificação gênica. Todos os gDNAs extraídos foram quantificados por espectrofotometria (Nanovue, GE Healthcare, Reino Unido) e tiveram suas razões registradas.

6.2.2 Enriquecimento do gene *MEN1* para SS.

As regiões alvo localizadas no gene *MEN1* foram enriquecidas exon a exon, a partir do gDNA extraído, incluindo regiões de fronteira intron-exon. Este enriquecimento foi realizado por meio de reação em cadeia de polimerase utilizando *kit GoTaq* (Thermo Fisher Scientific, CA, EUA) para SS. Para tal reação foram usadas sequências iniciadoras customizadas desenhadas para cobrir toda região codificadora do gene *MEN1* bem como regiões fronteiriças intron-exon (tabela 2).

Tabela 2: Sequências iniciadoras utilizadas em PCRs para SS.

Nome	Sequências (5'-3')	Produto
Exon 2	Fw: GGGCGGGTGGAAACCTTAG Rv: CGAACCTCACAAGGCTTACAGTTC	653 bp
Exon 3	Fw: ATCACTACCTGGCCCCCTTT Rv: GGGTGTGGCCCAAGAAAATG	366 bp
Exon 4	Fw: ACTGTAGTAGCCCAAGCCAC Rv: TCCTTAGCCTATCGGGCAGA	585 bp
Exon 5-6	Fw: TAAAGGGCTGGTCCCAGAAT Rv: GTGAGACTGGATGGGCGATA	470 bp
Exon 7	Fw: AAGAGTGTCTGAGAAAGAGAAGGG Rv: GGACTGGATGGAAAGGGGATG	522 bp
Exon 8	Fw: AACCACCATCATCCAGCAGTGG Rv: CCATCCCTAATCCCGTACATGC	451 bp
Exon 9	Fw: AGGGGTCTTGCCTAGGTG Rv: CTGGGCCAGAAAAGTCTGA	337 bp
Exon 10	Fw: GTGCCGATGGGACTGAGAC Rv: GGGAACCTAGGGTTTGGGTA	722 bp

Considerando o genoma GRCh37/Hg19, bem como o transcrito NM_130799/ENST00000312049, os *primers* iniciadores supracitados envolvem uma área total de 4106pb, com tamanho médio dos fragmentos de 513pb. As áreas amplificadas e sequenciadas do gene *MEN1* são representadas na figura 4.

Figura 4: Área amplificada e sequenciada por SS do gene *MEN1*, considerando transcrito NM_130799, 5'-3'.

Para tal amplificação foi utilizado termociclador programado para seguinte ciclagem:

- 95°C – 10 minutos
 - 94°C – 45 segundos
 - 64°C – 40 segundos
 - 72°C – 60 segundos
 - 72°C – 10 minutos
 - 4°C – *Hold*
-
- 35 ciclos**

Após cada reação de PCR os materiais foram submetidos à eletroforese em gel de agarose para confirmação de amplificação. A concentração de agarose utilizada variou entre 1% e 2% de acordo com a necessidade exigida em decorrência do tamanho de cada produto de PCR. Foi utilizado o padrão de peso molecular (*DNA Ladder*) com padrão de bandas de 100pb (GeneRuler 100 bp Plus DNA Ladder, Thermo Scientific, EUA). A voltagem variava de acordo com o tamanho da cuba utilizada, seguindo recomendações do fabricante.

6.2.3 Enriquecimento da região genômica do gene *MEN1* para NGS

O enriquecimento do gDNA extraído foi realizado a partir de *Long Range PCR* (reação longa em cadeia de polimerase) (Thermo Fisher, CA, EUA). A região englobada possui, segundo GRCh37/hg19, 7187pb, tendo suas coordenadas situadas entre as posições 64.571.274 e 64.578.460 no cromossomo 11. Os oligonucleotídeos iniciadores utilizados para amplificação desta região foram: sequencia senso 5' TAACAGACACTGATACCCAGCTAAAGC 3' e sequencia anti-senso 5' GGGATACGAAGGAGAGGAACTAGG 3'.

O produto de PCR amplificado de 7187pb engloba toda a região genômica do gene *MEN1* que inclui: todos os exons codificadores (parte do exon 2, exons 3 a 9 e parte do exon 10), exons não codificadores (exon 1 e parte dos exons 2 e 10), toda a região intrônica e a região promotora (localizada dentro do exon 1), como representado na figura 5.

Figura 5: Área amplificada por PCR longa e sequenciada por NGS do gene *MEN1*, considerando transcrito NM_130799, 5'-3'.

A ciclagem utilizada em termociclador para amplificação do material foi:

- 94°C – 4 minutos
 - 95°C – 15 segundos
 - 61,5°C – 30 segundos
 - 68°C – 7:50 minutos
 - 68°C – 10 minutos
 - 4°C – *Hold*
- } 39 ciclos

As condições de ciclagens e os parâmetros de reação para cada *kit* seguiram os protocolos estabelecidos pelo fabricante. Após cada reação de PCR os materiais foram submetidos à eletroforese em gel de agarose para confirmação de amplificação. A concentração de agarose utilizada variou entre 1% e 2% de acordo com a necessidade exigida em decorrência do tamanho de cada produto de PCR. Foi utilizado o padrão de peso molecular (*DNA Ladder*) com padrão de bandas de 1Kb (GeneRuler 1 Kb Plus DNA Ladder, Thermo Scientific, EUA). A voltagem variava de acordo com o tamanho da cuba utilizada, seguindo recomendações do fabricante.

6.3 Sequenciamento pelo método de Sanger (SS)

Inicialmente, todos os produtos de PCR gerados (tabela 2), quando amplificados com sucesso, foram submetidos a purificação enzimática com *kit* Illustra ExoProStar 1 step (GE Healthcare, UK), como recomendado. Em seguida, para SS, a reação de sequenciamento foi realizada com *kit* ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction 3.1 (Applied Biosystems, EUA), seguindo as condições de ciclagem estabelecidas pelo fabricante.

Após incorporação aleatória de dideoxinucleotídeos marcados os fragmentos gerados foram submetidos à eletroforese capilar no sequenciador ABI Prism 3130xl Genetic Analyzer (16 capilares) (Applied Biosystems, EUA). Subsequentemente à separação dos fragmentos, o programa CodonCode 5.0 (corporação CodonCode) foi

empregado para alinhamento e leitura dos dados gerados com sequência referência do gene *MEN1* (ENSG00000133895 ou RefSeq NG_008929.1), considerando o NEM1-001 (ENST00000312049 ou NM_130799.2) como transcrito.

6.4 Sequenciamento de nova geração (NGS)

Tendo como referência o objetivo do presente estudo, foram avaliadas plataformas de NGS com menor capacidade de geração de dados e de menor custo, ditas “plataformas de bancada” (*bench top sequencers*). Dentre as opções, foi selecionada a plataforma MiSeq (Illumina, EUA).

Assim, a biblioteca de gDNA foi preparada para NGS utilizando o *kit* Nextera XT (fragmentação enzimática) (Illumina, EUA) a partir do material enriquecido obtido pela PCR longa que contempla toda região genômica do gene *MEN1*. Tais amostras foram progressivamente diluídas em água ultrapura e aferidas por método fluorimétrico Qubit (Invitrogen, EUA) até alcançarem concentração de 0,2ng/ul.

O processo de tagmentação (figura 6) (fragmentação por tranposases seguida da colocação de adaptadores), colocação do par de índices (tecnologia *dual-indexing*) bem como normalização das bibliotecas foram realizados seguindo os procedimentos estabelecidos pelo fabricante a partir de uma concentração total de 1ng de gDNA por amostra previamente enriquecida.

Figura 6: 1) Ilustração do processo de fragmentação enzimática baseada em transposases. A transposase hidrolisa uma ligação fosfodiéster, clivando a dupla fita de gDNA e disponibilizando um sítio 3'-OH livre, catalisando, portanto, uma transesterificação, levando a "inserção" do fragmento referente ao adaptador na sequência alvo. 2 e 3) O processo realizado pela transposase é denominado "tagmentação", referente ao processo de fragmentação e simultânea adição das *tags*, ou seja, dos adaptadores. 4) A ação de transesterificação promovida pela transposase gera extremidades com quebras, que necessitam de reparo, a ser realizado por uma PCR, que acaba, também, aumentando em número aqueles fragmentos devidamente "etiquetados" com adaptadores. 5) colocação dos índices, referente a tecnologia *dual-indexing*, que permite a combinação entre 2 sequências para reconhecimento do fragmento. 6) Fragmento pronto para ser submetido ao processo de sequenciamento SBS (Sequenciamento Baseado em Síntese - Illumina), realizado na modalidade *paired-end*, onde o mesmo fragmento é lido em ambos os sentidos, aumentando a profundidade de leitura. São sequenciados em 4 etapas; Exemplo: kit 500 ciclos; #1 250 ciclos, leitura da sequência com primer *read 1*; #2 8 ciclos para leitura do primeiro índice com primer *index 1* (i7); Intervalo para re-síntese dos fragmentos por PCR em ponte; #3 8 ciclos para leitura do segundo índice com primer *index 2* (i5); #4 250 ciclos, leitura da sequência, utilizando a fita reversa como molde, por intermédio do primer *read 2*. Adaptado de Pekka Ellonen, FMIM, 2013.

A normalização das amostras foi realizada utilizando microesferas (*beads*) magnéticas com capacidade limitada de carregamento, de modo a “nivelar” a concentração de todas as bibliotecas. Após tal normalização, mesmo que não fosse necessário (etapa opcional), as amostras foram separadamente quantificadas em triplicata por PCR em tempo real quantitativa com o *kit* Kapa SYBR Fast Universal qPCR (Kapa Biosystems, EUA) sob diluição de 1:1000 em tampão Tris-HCl pH 8,5 + 0,01% de Tween 20. Os equipamentos de PCR em tempo real utilizados foram ECO (Illumina, CA, EUA), StepOne Plus (Applied Biosystems, MA, EUA) e AriaMX (Agilent, CA, EUA). A montagem das reações seguiu o padrão para volume final de 10ul, seguindo protocolo Kapa, para cada equipamento específico, bem como o cálculo para acerto do coeficiente de diluição para correta quantificação das amostras.

Após quantificação as amostras foram diluídas em Tris-HCl pH 8,0 para concentração final de 0,325nM e distribuídas igualmente em um *pool* final para sequenciamento. O material final foi, então, sequenciado sob concentração de 13pM com fragmentos entre 400 e 500pb seguindo as recomendações do fabricante.

Uma vez que se tratava de uma biblioteca de baixa complexidade dado o método de enriquecimento (PCR longa), foram adicionadas ao volume final da biblioteca a ser carregada no sequenciador, 25% (150ul) do controle Phix v3 (Illumina, EUA) a 12,5pM (GC 55% e AT 45%).

O sequenciamento foi realizado na plataforma Illumina MiSeq (Illumina, EUA) utilizando *kit* de sequenciamento V2, bem como o kit de sequenciamento Nano V2, ambos em 500 ciclos com sequenciamento bi-direcional (2x 252 ciclos).

6.5 Análise dos dados – Bioinformática

Para facilitar a análise dos dados, foi selecionada, ainda sob o próprio sequenciador, a modalidade de resequenciamento, que utiliza como referência e região alvo todo o genoma humano (GRCh37/hg19). Nesta modalidade, o alinhamento e chamada de variantes, sem realização de anotações, são executados dentro do sequenciador (*stand alone*) por meio do programa MSR (MiSeq Reporter Software). Os arquivos FASTq, obtidos após a *demultiplexação* (obtidos a partir do programa BCL2FASTq; consiste na identificação dos índices e agrupamento das sequências dentro da pasta de cada amostra), as sequências alinhadas (.BAM e .BAI) e os arquivos de

chamada de variantes (.VCF) são disponibilizados em nuvem de armazenamento (*Basespace*) (ferramenta gratuita para usuários).

Além da ferramenta do MSR e do *Basespace*, os dados brutos (FASTq) foram processado *in-house*. O alinhamento e chamada de variantes utilizando BWA (71), GATK (72) e Samtools (73) e anotações pelo Annovar (74) e SnpEff (75). Ambos os resultados foram analisados e comparados. O programa IGV (*Integrative Genomics Viewer* – Broad Institute) (76-77) foi utilizado para visualização dos alinhamentos por meio do arquivo .BAM.

Análises fidedignas das métricas de qualidade não foram possíveis uma vez que os arquivos gerados diretamente pelo método de resequenciamento tinham como região alvo todo o genoma humano (hg19). Assim, o programa Qualimap (78) foi utilizado para tais análises baseados nos dados da segunda análise supracitada.

Para determinar o impacto das variantes sobre o fenótipo dos indivíduos analisados foram utilizados bancos de dados e ferramentas predição *in silico* abertas, ou somente abertas para uso em pesquisa científica, sendo estes dbSNP, *Leiden Open Variation Database* (LOVD), Ensembl, *Catalogue of Somatic Mutations in Cancer* (COSMIC), HGMD (banco de dados público), Sift, Polyphen, Mutation Assessor, Mutation Taster, SNPS&GO, ClinVar. Para avaliação dos efeitos das variantes de sítios de excisão foram consultados *Human Splicing Finder* (HSF 3.0). Como banco de dados para populações normais houve consulta ao banco de dados 1000 genomas. Já para bancos de dados clínicos, foram utilizados *Exome Variant Server* (EVS) e *Exome Aggregation Consortium* (ExAC).

6.6 MLPA

O rastreamento de microdeleções e microduplicações, realizado por *Multiplex Ligation-dependent Probe Amplification* (MLPA), seguiu todas as especificações do fabricante. Cada reação contou com a utilização de 5 controles normais. O *kit* utilizado contém sondas especificamente para o gene *MEN1* (P017 – são aplicadas 1 sonda para cada exon, sendo 2 sondas nos exons 2, 7 e 10) ou para os genes *CDKN1B*, *MEN1* e *AIP* (P244 – contém 11 sondas para os 10 exons do gene *MEN1*), além do *kit* de reagentes padrão da técnica. Os fragmentos foram separados utilizando padrão de peso molecular ROX 500 em dispositivo para eletroforese capilar (ABI 3130xl – Applied Biosystems, MA, EUA).

A figura 7 ilustra o padrão de distribuição de sondas de ambos os *kits* de MLPA. Após obtenção dos dados provenientes da análise dos picos, o programa Coffalyser (MRC-Holland, AM, HOL) foi utilizado para determinação do resultado final.

Figura 7: Representação da localização das sondas de MLPA dos kits utilizados no estudo dentro da região do gene *MEN1* (Representação na fita sendo).

6.7 Análise de custo-efetividade

Para avaliação de custo-efetividade da aplicação de NGS foram levados em conta aspectos como custo do equipamento, custo de insumos necessários para aplicação de cada técnica, reagentes necessários, levando-se em conta sua capacidade de geração de resultados, demanda periódica relativo ao método em questão, taxa de repetição de reações, custo com manutenções, depreciação, tempo de trabalho (bancada e análise) e custo de equipamento ocioso. Além do custo básico, calculado por paciente, foram analisados, também, o custo por par de base sequenciado, comparando-se as regiões obtidas por SS e por NGS. Os testes foram avaliados em três cenários distintos: Custo operacional com capacidade máxima, custo operacional sob demanda (casos-índices) e custo para inclusão de rastreamento de familiares afetados na plataforma de NGS.

7 RESULTADOS

7. Resultados

7.1 Quantificação e qualificação dos gDNAs extraídos por técnica de *Salting-out* e resultados de amplificações.

A técnica de *Salting-out* oferece grande volume de gDNA total extraído, havendo, contudo, problemas quanto a pureza do material em determinados casos.

Os valores obtidos para as quantificações das amostras foram variáveis, com valor médio de 40ng/ul. Contudo, das 76 amostras utilizadas, 24 delas obtiveram valor da razão 260/280 menor do que 1, indicando elevada concentração de contaminantes. Mesmo assim, amostras com valor de razão 260/280 menor que 1 foram amplificadas satisfatoriamente pela PCR longa envolvendo a região do gene *MEN1*. Das 76 amostras, 8 delas (11%) não obtiveram amplificação na primeira reação, sendo necessário purificação do material em coluna. Após a purificação, a quantidade de gDNA perdido foi, em média, de 80%, com relação ao valor inicial, sendo que somente 6 delas tiveram melhora na relação 260/280. Mesmo assim, a PCR para amplificação do gene *MEN1* foi repetida em todas elas. Após a repetição da PCR com material purificado houve amplificação.

Após padronização por gradiente de temperatura (temperatura de anelamento em 61°C), a eficiência da PCR foi comprovada por eletroforese em gel de agarose (figura 8). O fragmento amplificado para a região se posiciona no gel em torno de 7000Kb.

Figura 8: Imagem de eletroforese em gel de agarose. PCR longa para amplificação da região genômica do gene *MEN1*. Vinte e quatro primeiras amplificações após a padronização da reação para ser utilizada como método de enriquecimento para sequenciamento por NGS.

7.2 Fase 1: Resultados e métricas

A primeira fase do estudo envolveu análise de 23 pacientes com NEM1 não relacionados e genótipo conhecido, contudo, mantendo o estudo cego. A corrida foi realizada utilizando um *kit* de reagentes v2 (500 ciclos). Neste ensaio, voltado para otimização e padronização de rotina, foram geradas 35 milhões de leituras (*Mreads*) com um total de 17 Gbases sequenciadas. Oitenta e oito por cento destas sequências foram mapeadas na região alvo.

Considerando toda a região genômica do gene *MEN1* (hg19 chr11:64571274-64578460), a cobertura foi de 100% da região alvo, com profundidade média de $253.400X \pm 11.600$. As médias das porcentagens de profundidade mínima foram: $99,8\% \pm 0,4 >50x$, $99,9\% \pm 0,2 >20x$ e $100\% \pm 0,1 >10x$. Com a região alvo considerando somente os exons do gene, tais valores passam a ser de $100\% \pm 0 >50x$, $>20x$ e $>10x$.

Nesta corrida, foram identificados 21 portadores de variante sabidamente patogênica no gene *MEN1*: 52% de variantes do tipo *frameshift*, 19% de variantes *missense*, 10% *nonsense*, 10% *in-frame* e 10% de variantes que alteram sítios de *splice*.

Duas das amostras foram caracterizadas como não portadoras de variantes patogênicas. Entretanto, em uma destas, a análise visual por meio do programa IGV (*Broad Institute*), revelou presença de uma deleção, gerando um *frameshift* e o aparecimento de um códon de parada prematuro (NM_130799.2, exon2, c.201delC: p.A68Pfs*51) em heterozigose, com o alelo deletado presente em 18% das leituras na posição (chr11:64577381).

A “não chamada” desta variante supracitada resultou em uma concordância entre NGS e SS de 96%, uma vez que o resultado negativo da outra amostra para presença de variante patogênica no gene *MEN1* foi confirmada por SS.

Já com relação as métricas da corrida, com a utilização do kit de reagentes MiSeq V2 (500 ciclos), onde são esperados, aproximadamente, 15 milhões de sequencias lidas paralelamente, foram obtidos os resultados demonstrados na tabela 3.

Tabela 3: Análise comparativa entre o rendimento esperado para kit MiSeq v2 (500 ciclos) e o rendimento da corrida realizada durante a fase 1.

	Rendimento esperado	Corrida #1
Kcluster/mm²	865-965	1.388
Milhões de <i>reads</i>	15	34,5
Mbases	7.500	17.200
% de <i>reads</i> que passam filtro	70-80%	77.08%
% de <i>reads</i> >q30	80-100%	67,5%

Considerando todas as 23 amostras sequenciadas durante esta fase, a porcentagem média de *reads* de cada amostra foi de 4,1% $\pm$ 4 do total (34,5 milhões de *reads*). A amostra mais representada obteve 22% dos *reads*, enquanto que a amostra menos representada somou 0,2%.

7.3 Fase 2: resultados e métricas

A fase 2 do projeto envolveu o sequenciamento dos 76 indivíduos subdivididos em 3 novas corridas para verificação de reprodutibilidade, repetitividade, especificidade e sensibilidade. Nestas três novas corridas de NGS realizadas foi utilizado o kit de reagentes Nano V2 (500 ciclos).

Nesta etapa, o teste manteve-se cego. As métricas obtidas nestas três corridas, quando comparados aos dados esperados, são ilustradas na tabela 4.

Tabela 4: Análise comparativa entre o rendimento esperado para kit MiSeq Nano v2 (500 ciclos) e o rendimento obtido com cada uma das três corridas realizadas durante a fase 2.

	Rendimento esperado	Corrida #2	Corrida #3	Corrida #4
Kcluster/mm²	~500	620	280	400
Mreads	1	1,26	0,57	0,65
Mbases	500	629,3	286	323,5
% Reads Passing Filter	80-70	93,6	96,4	96,2
% reads >q30	80-100	85,3	85,5	80,4

Nas corridas #2, #3 e #4, as coberturas e profundidades médias da região alvo, envolvendo todo o gene *MEN1* (hg19 chr11:64571274-64578460) ou somente os exons, são apresentadas nas tabelas 5 e 6 respectivamente.

Tabela 5: Métricas de NGS (cobertura e profundidade médias $\pm$ DP) calculadas, tendo como região alvo todo o gene *MEN1*.

	Corrida #2	Corrida #3	Corrida #4
Cobertura média (%)	99,8 $\pm$ 0,6	85,9 $\pm$ 14,9	95,6% $\pm$ 15,6
Profundidade média (X)	564,1 $\pm$ 255,3	33,97 $\pm$ 18,8	136,8 $\pm$ 69,9
% bases >10X	98,7 $\pm$ 2,2	58,3 $\pm$ 37,9	85,7 $\pm$ 25,4
% bases >20X	97,2 $\pm$ 4,5	46,8 $\pm$ 38,6	79,2 $\pm$ 30,1
% bases >50X	92,3 $\pm$ 13,4	25,6 $\pm$ 31,4	67,5 $\pm$ 35,4

Tabela 6: Métricas de NGS (profundidade média $\pm$ DP) calculadas, tendo como região alvo somente a região codificadora do gene *MEN1*.

	Corrida #2	Corrida #3	Corrida #4
% bases >10X	99,5 $\pm$ 2,6	59,7 $\pm$ 38	86,2 $\pm$ 25,4
% bases >20X	98,2 $\pm$ 5,8	48,8 $\pm$ 40,7	79,7 $\pm$ 30,4
% bases >50X	92,9 $\pm$ 13,9	27,3 $\pm$ 33,9	66,5 $\pm$ 36,2

Da mesma forma, a representação de cada amostra nas corridas foi variável na fase 2 (figura 9). Na corrida #2, a amostra mais representada obteve 6,4% dos *reads*, enquanto que a menos representada, obteve 0,08%. Já na corrida #3, a amostra mais representada alocou 13,5% dos *reads*, enquanto que a menos representada obteve 0,01%, sendo considerada insatisfatória. Por fim, a corrida #4 teve sua amostra mais representada com 4% e a menor delas 0,3%. A média da porcentagem de *reads* em cada amostra nas corridas #2, #3 e #4 foi de $1,9\% \pm 1,3$, $4,2\% \pm 2,9$ e $1,8\% \pm 0,8$.

Figura 9: Porcentagem dos *reads* de cada amostra com relação ao total da corrida

7.4 Fase 1: determinação da concordância, sensibilidade e especificidade do método de NGS

Na fase 1, 100% (N=22) das variantes patogênicas identificadas pelo método de NGS, foram confirmadas por SS, inclusive com relação a distribuição alélica. Da mesma forma, a amostra, sabidamente negativa, também teve seu resultado corroborado.

Contudo, a concordância obtida durante a fase 1 foi 95,7%, já que 1 das variantes, presentes em um dos indivíduos, considerada patogênica (NM_130799.2, exon2, c.201delC: p.A68Pfs*51) não foi chamada e, portanto, não rastreada (Figura 10). Após análise visual por meio do programa IGV, foi possível identificar a variante em questão, presente em 18% dos *reads* totais na base específica, o que culminou na ausência de chamada neste caso, já que o limite era 20% dos *reads*.

Mesmo com 100% das variantes causadoras tendo sido, mesmo que uma delas indiretamente, corretamente triadas pelo método de NGS, quando consideradas todas as

variantes encontradas (variantes causadoras e variantes polimórficas) a sensibilidade do método de NGS passa a ser de 99,6% (escore $Kappa=0,852$ $p<0,001$), já que duas variantes polimórficas obtiveram achados discrepantes com relação ao SS e NGS, uma delas não confirmada por SS (c.1254C>T:p.D418D) e outra (c.216T>A:p.P72P) não confirmada por NGS.

Figura 10: Imagem da posição da variante NM_130799.2, exon2, c.201delC: p.A68Pfs*51, não chamada durante a corrida #1, mas visível no programa IGV.

Nesta etapa, considerando os resultados expostos, o teste diagnóstico pela plataforma NGS em NEM1 teria como valor de sensibilidade 95,5% (intervalo de confiança de 95% ((IC 95%)) = 77,2 - 99,9%), especificidade em 100% (intervalo de confiança de 95% = 2,5 - 100%), valor preditivo positivo 100% (IC 95% = 83,9 - 100%) e valor preditivo negativo de 50% (IC 95% = 1,3 - 98,7%).

7.5 Fase 2: determinação da concordância, reprodutibilidade, sensibilidade e especificidade do método de NGS

Na segunda fase foram sequenciados todos os 76 indivíduos ao longo de 3 corridas (kit de reagentes MiSeq Nano v2). Dentre estes, 29% foram sequenciados ao menos 3 vezes para avaliação de reprodutibilidade. A corrida #3 apresentou qualidade inferior a desejada devido problemas e falhas no processo de formação de *clusters*, impedindo a utilização dos seus resultados.

Nesta fase, 100% dos resultados encontrados foram concordantes entre as técnicas de NGS e SS, resultando em um escore $Kappa=1,0$ $p<0,001$. O resultado discrepante reportado no item 6.3 acerca da variante NM_130799.2, exon2, c.201delC: p.A68Pfs*51

foi corretamente identificado nas corridas #2 e #4, onde o indivíduo portador foi inserido (figura 11).

Figura 11: Variante NM_130799.2, exon2, c.201delC: p.A68Pfs*51, encontrada na corrida #2 e #4, respectivamente.

Considerando todos os 76 indivíduos, a precisão, exatidão, sensibilidade (bem como IC 95%), especificidade (bem como IC 95%), os valores preditivos positivo e negativo, repetitividade (variação de resultado baseado em um mesmo operador) e reprodutibilidade (variação de resultado baseado em diferentes operadores) são expostos na tabela 7.

Tabela 7: Valores de precisão, exatidão, sensibilidade, especificidade, valor preditivo positivo (VPP) e valor preditivo negativo (VPN) considerando a média obtida na análise e reprodução dos 76 pacientes com as corridas #1, #2 e #4. ND: Não disponível

N = 76	Valor	IC 95%
Reprodutibilidade (N=24)	1	ND
Repetitividade (N=24)	0,986	ND
Precisão	0,991	ND
Exatidão	1	ND
Sensibilidade	1	0,973-1
Especificidade	1	0,824-1
VPP	1	0,973-1
VPN	1	0,824-1

7.6 Variantes

Ao todo, 75% dos pacientes com NEM1 (57/76) foram identificados com variantes patogênicas no gene *MEN1*. Nenhum paciente apresentou mais de uma variante patogênica. Neste grupo, 37% eram deleções ou inserções do tipo *frameshift* (N=21), 25%

alteravam sítio de *splice* (N=14), 17% eram do tipo nonsense (N=11), 12% do tipo missense (N=7) e 9% deleções ou inserções do tipo *in-frame* (N=4) (tabela 7).

Foram identificadas 38 variantes germinativas patogênicas diferentes no gene *MEN1*, sendo que 6 delas foram recorrentes: c.654+1G>T (IVS3), encontrada em 10 casos (17,5% dos casos), c.249_252delGTCT (exon 2, p.I85Sfs*33), encontrada em 5 casos índices (8,8% dos casos), c.1546_1547insC (exon 10, p.R516fs*15) encontrada em 4 casos (7% dos casos) e as variantes c.1579C>T (exon 10, p.R527*), c.1238T>G (exon 9, p.L413R) e c.1243C>T (exon 9, p.R415*) encontradas em 2 casos cada.

As variantes patogênicas estavam distribuídas ao longo do gene *MEN1*, sem caracterização de *hot spots* mutacionais. As mutações foram mais frequentes nos exons 2, 9 e 10. Considerando os exons do gene, bem como suas regiões de fronteira com introns, 29% (N=11) situavam-se no exon 2, 13% (N=5) no exon 3, 3% (N=1) no exon 4, 5% (N=2) no exon 5, 8% (N=3) no exon 6, 5% (N=2) no exon 7, 5% (N=2) no exon 8, 16% (N=6) no exon 9 e 16% (N=6) no exon 10. Desta forma, somente algumas regiões configuram regiões mais frequentes, por exemplo, o exon 2. Todas as variantes encontradas são descritas na tabela 8.

Tabela 8: Variantes encontradas nos 44 indivíduos portadores de diagnóstico genético da síndrome NEM1 que potencialmente afetam funcionamento da proteína Menin e suas frequências intrapopulacionais. **Variantes reportadas neste estudo e descrita no banco de dados com posições coincidentes.

Exon	cDNA	Efeito	dbSNP	HGMD	Referência	Tipo	Frequência no estudo %
10	c.1382_1404del23	p.E461Gfs*62	ND	ND	Este trabalho	Frameshift	2,3
10	c.1546_1547insC	p.R516Pfs*15	rs767319284	CI972640	Agarwal (1997) Hum Mol Genet 6, 1169	Frameshift	9,1
10	c.1554_1578del25	p.P519Efs*32	ND	CD991792**	Poncin (1999) Hum Mutat 13, 54	Frameshift	2,3
10	c.1564_1565delGT	p.V522Rfs*8	ND	ND	Este trabalho	Frameshift	2,3
10	c.1579C>T	p.R527*	rs104894261	CM970942	Chandrasekharappa (1997) Science 276, 404	Nonsense	4,5
10	c.1669A>T	p.K557*	ND	CM066128**	Ozturk (2006) J ENAocrinol Invest 29, 523	Nonsense	2,3
2	c.122T>C	p.L41P	ND	ND	Este trabalho	Missense	2,3
2	c.201_201delC	p.A68Pfs*51	ND	CD075463	Toledo (2007) Clin Endocrinol (Oxf) 67, 377	Frameshift	2,3
2	c.249_252delGTCT	p.I85Sfs*33	rs587776841	CD075467/CD972304**	Chandrasekharappa (1997) Science 276, 404/Tham (2007) J Clin Endocrinol Metab 92, 3389	Frameshift	11,4
2	c.266_286del21	p.L89_A95del	ND	ND	Este trabalho	In Frame	2,3
2	c.315_316insCCTC	p.Y106Pfs*12	ND	ND	Este trabalho	Frameshift	2,3
2	c.358_360delAAG	p.Lys120del	rs794728657	CD972307	Chandrasekharappa (1997) Science 276, 404	In Frame	2,3
2	c.377G>A	p.W126*	ND	CM981253	Bassett (1998) Am J Hum Genet 62, 232	Nonsense	2,3
2	c.385_385delC	p.L129Sfs*56	ND	ND	Este trabalho	Frameshift	2,3
2	c.415C>T	p.H139Y	ND	CM970921	Agarwal (1997) Hum Mol Genet 6, 1169	Missense	2,3
2	c.431_432delinsAA	p.F144*	ND	CX055790	Ellard (2005) Clin Endocrinol (Oxf) 62, 169	Nonsense	2,3
2	c.439A>T	p.I147F	ND	CM074341	Toledo (2007) Clin Endocrinol (Oxf) 67, 377	Missense	2,3
3	c.548G>A	p.W183*	rs794728650	CM981259	Lemmens (1997) Hum Mol Genet 6, 1177	Nonsense	2,3
3	c.558delT	p.F186Lfs*38	ND	CD035821	Cebrian (2003) J Med Genet 40, e72	Frameshift	2,3
3	c.628_631delACAG	p.T210Sfs*13	rs794728640	CD972310**	Chandrasekharappa (1997) Science 276, 404	Frameshift	2,3

3	c.652_652delC	p.R218Gfs*6	ND	ND	Este trabalho	Frameshift	2,3
3/IVS4	c.654+1G>T		rs794728622	CS982266	Teh (1998) J Clin Endocrinol Metab 83, 2621	Splice site	22,7
4/IVS5	c.783+1G>A		ND	CS003705	Morelli (2000) Eur J ENAocrinol 142, 131	Splice site	2,3
IVS5/5	c.784-2A>C		ND	CS094282**	Franchi (2009) Cancer Genet Cytogenet 192, 99	Splice site	2,3
IVS5/5	c.784-9G>A		rs794728625	CS991446	Gortz (1999) Int J Cancer 80, 373	Splice site	2,3
5	c.802_803insGGCTGCTCT	p.W265_L267dup	ND	ND	Este trabalho	In Frame	2,3
6/IVS5	c.825-2_830deletionAGGTACCC	p.275_277del	ND	CS973732/CS075179**	Toliat (1997) Lancet 350, 1223/Hai (2000) Clin Endocrinol (Oxf) 52, 509	In Frame	2,3
6	c.903_909delCTACCAC	p.Y302Rfs*64	ND	ND	Este trabalho	Frameshift	2,3
7	c.1015C>T	p.Q339*	ND	CM053978	Klein (2005) Genet Med 7, 131	Nonsense	2,3
7	c.951_952insC	p.I318Hfs*49	ND	ND	Este trabalho	Frameshift	2,3
8	c.1087G>T	p.E363*	ND	ND	Este trabalho	Nonsense	2,3
8	c.1133delA	p.E378Gfs*67	ND	CD075465	Toledo (2007) Clin Endocrinol (Oxf) 67, 377	Frameshift	2,3
9	c.1238T>G	p.L413R	ND	CM074346	Toledo (2007) Clin Endocrinol (Oxf) 67, 377	Missense	4,5
9	c.1241T>C	p.L414P	ND	CM010342	Asteria (2001) Hum Mutat 17, 237	Missense	2,3
9	c.1243C>T	p.R415*	ND	CM970937	Lemmens (1997) Hum Mol Genet 6, 1177	Nonsense	4,5
9	c.1252G>T	p.D418Y	ND	CM981277**	Bassett (1998) Am J Hum Genet 62, 232	Missense	2,3
9	c.1268G>A	p.W423*	ND	CM004419	Bergman (2000) Br J Cancer 83, 1009	Nonsense	2,3
9/IVS9	c.1350+1_1350+11del11		rs794728644	ND	ND	Splice site	2,3

ND: Não disponível

MAF: Frequência do menor alelo (*Minor Allele Frequency*)

Trinta e sete diferentes variantes genéticas benignas foram encontradas ao longo do gene *MEN1*: regiões não traduzidas (UTRs) (19%, 7), IVS1 (11%, 4), IVS2 (35%, 13), IVS6 (5%, 2), IVS7 (5%, 2), IVS9 (3%, 1) e 5,4% (N=2) nos exons 2, 3, 9 e 10. Nove destas variantes não tinham sido descritas previamente. Os detalhes de cada variante são descritos na tabela 9.

Tabela 9: Descrição de variantes benignas, localização, frequência de indivíduos portadores, distribuição alélica e referência em bancos de dados.

SNVs encontrados	Localização	% dos casos com o SNV	% Heterozigose	% Homozigose	dbSNP	COSMIC	1000 Genomes	6500 exomes	ExAC
c.*307T>G	3'UTR	1,32	100,00	0,00	rs1804848	ND	MAF:G=0,9%	ND	ND
c.*185C>T	3'UTR	1,32	100,00	0,00	rs111895237	ND	MAF:T=0,5%	ND	ND
c.*306_*308delCTC	3'UTR	1,32	100	0	rs143341556	ND	MAF:DEL=8%	ND	ND
c.*302C>T	3'UTR	3,9	100,0	0,0	rs1804849	ND	MAF:T=1,2%	ND	ND
c.-35A>C	5'UTR	9,21	100,00	0,00	rs679946	ND	MAF:C=20%	ND	ND
c.-35A>T	5'UTR	44,74	70,59	29,41	rs679946	ND	MAF:T=23%	ND	ND
c.-6G>A	5'UTR	1,3	100,0	0,0	rs768088337	ND	MAF:A=0,2%	ND	ND
c.-23-16C>G	IVS1	13,16	70,00	30,00	rs509606	ND	MAF:G=17%	ND	30%
g.1139T>C	IVS1	3,9	100,0	0,0	ND	ND	ND	ND	ND
g.890G>A	IVS1	2,6	100,0	0,0	ND	ND	ND	ND	ND
g.894G>A	IVS1	3,9	33,3	66,7	ND	ND	ND	ND	ND
c.445+177G>A	IVS2	1,32	100,00	0,00	rs192961797	ND	MAF:A=0,3%	ND	ND
c.445+183G>A	IVS2	18,42	85,71	14,29	rs624975	ND	MAF:A=17%	ND	ND
c.445+306G>A	IVS2	2,63	100,00	0,00	rs34047150	ND	MAF:A=0,6%	ND	ND
c.445+401_445+402insT	IVS2	25,00	100,00	0,00	rs11404218	ND	ND	ND	ND
c.445+401_445+402insTT	IVS2	26,32	100,00	0,00	rs11404218	ND	ND	ND	ND
c.445+539G>A	IVS2	17,11	92,31	7,69	rs67808744	ND	MAF:A=15%	ND	ND
c.445+719G>A	IVS2	13,16	80,00	20,00	rs11231873	ND	MAF:A=13%	ND	ND
c.446-198G>A	IVS2	1,32	100,00	0,00	rs11607366	ND	MAF:A=0,6%	ND	ND
c.446-347G>A	IVS2	1,32	100,00	0,00	rs75989060	ND	MAF:A=0,7%	ND	ND
c.446-662A>G	IVS2	1,32	100,00	0,00	rs374393702	ND	ND	ND	ND

c.445+709_710delTT	IVS2	52,6	100,0	0,0	ND	ND	ND	ND	ND
c.445+710delT	IVS2	39,5	100,0	0,0	ND	ND	ND	ND	ND
g.2505_2505delT	IVS2	1,3	100	0	ND	ND	ND	ND	ND
c.912+283C>T	IVS6	94,74	16,67	83,33	rs504195	ND	MAF:C=16%	ND	ND
c.912+295_912+297delTGA	IVS6	86,8	18,2	81,8	rs2307769	ND	MAF:TAG=20%	ND	ND
c.1049+115A>G	IVS7	15,79	83,33	16,67	rs669976	ND	MAF:G=17%	ND	ND
g.5276_5276delT	IVS7	34,2	100,0	0,0	ND	ND	ND	ND	ND
c.1350+103G>C	IVS9	59,21	75,56	24,44	rs654440	ND	MAF:C=23%	ND	ND
c.1624A>G;p.T541A	EXON 10	96,1	17,8	82,2	rs2959656	COSM255213	MAF:A=15%	MAF:A=9%	0,94
c.1508G>A;p.G503D	EXON 10	1,3	100,0	0,0	ND	COSM1188290	ND	ND	0,0005
c.71G>C (EXON 2)	EXON 2	27,6	100,0	0,0	ND	ND	ND	ND	ND
c.202G>A;p.A68T	EXON 2	1,3	100,0	0,0	ND	ND	ND	ND	ND
c.597C>T;p.H204H	EXON 3	3,95	0,00	100,00	rs150512958	ND	MAF:T=0,1%	0,02%	0,06%
c.512G>A;p.R171Q	EXON 3	3,9	100,0	0,0	rs607969	COSM23054	MAF:A=0,8%	MAF:A=1,5%	0,00083%
c.1254C>T;p.D418D	EXON 9	52,6	82,5	17,5	rs2071313	ND	MAF:T=30%	MAF:T=39%	0,39
c.1299T>C;p.H433H	EXON 9	88,2	9,0	91,0	rs540012	ND	MAF:T=2%	MAF:T=2,5%	0,993

ND: Não disponível

MAF: Frequência do menor alelo (*Minor Allele Frequency*)

7.7 Análise dos resultados

Cada paciente foi analisado com os resultados obtidos a partir do processamento automatizado (*standalone*), por meio do MSR e, paralelamente, através de um arquivo VCF único, obtido a partir de um *pipeline* de processamento de dados desenvolvido *in house*. Cem por cento das variantes foram catalogadas em ambas as frentes.

Entretanto, o *pipeline* desenvolvido no serviço aplicou, na chamada de variantes (GATK), a unificação dos dados em um arquivo VCF único, gerando uma redução de 25% no tempo de análise, enquanto que os VCFs obtidos a partir do MSR eram individualizados.

7.8 Análise de custo

Todos os custos com sequenciamento dos 76 pacientes foram calculados tendo como referência as plataformas ABI3500xl (Applied Biosystems, EUA) para SS, uma vez que a plataforma ABI3130xl (Applied Biosystems, EUA) foi descontinuada, e a plataforma Illumina MiSeq (Illumina, EUA) para NGS. Foram levados em conta os valores de mercado dos sequenciadores, bem como custos inerentes a manutenção, depreciação, ociosidade de equipamento, taxa de repetição, tempo/custo de mão-de-obra e custos de reagentes e insumos. Os cálculos englobaram 3 diferentes cenários: capacidade máxima de operação, operação sob demanda e absorção de todos os rastreamentos familiares. Todos estes pontos, considerando o sequenciamento de toda região genômica do gene *MEN1* em NGS e somente as regiões exônicas e peri intrônicas descritas anteriormente (material e métodos) em SS (tabela 10).

Os procedimentos para NGS foram, aproximadamente, 60% mais rápidos que aqueles empregados para SS, já considerando as taxas de repetição (5,2% para SS vs 2,7% para NGS):

- 1) Considerando o sequenciamento de 96 casos-índices (máxima capacidade), o custo estimado por paciente usando a técnica NGS foi 71% inferior ao do SS.
- 2) Já assumindo o custo referente a demanda média mensal do serviço em absorver casos-índices (N=20/mês), o custo médio da plataforma NGS se posiciona 63% menor que o custo para desenvolvimento de SS.
- 3) Por fim, assumindo a absorção total de todos os indivíduos familiares de casos-índices para rastreamento de variante conhecida, executando os mesmos processos (sequenciamento total de todo o gene *MEN1* por NGS), e,

consequentemente, elevando a demanda média mensal para 40 indivíduos/mês, o método de NGS se posiciona com custo 29% menor que o método de SS.

Tabela 10: Custeio da aplicação do método de NGS para sequenciamento do gene *MEN1* (7189pb) vs método de SS para sequenciamento dos exons e regiões de fronteira do gene (4106pb).

<u>Custos SS</u>	<u>R\$</u>		
	Custo efetivo	Custo demanda	Custo Rast. Familiar
Custo p/ paciente	R\$ 1.267,76	R\$ 2.503,86	R\$ 672,92
Custo p/ par de base	R\$ 0,31	R\$ 0,61	R\$ 0,16
<u>Custos NGS</u>			
	Custo efetivo	Custo demanda	Custo Rast. Familiar
Custo p/ paciente	R\$ 364,25	R\$ 929,36	R\$ 479,40
Custo p/ par de base	R\$ 0,05	R\$ 0,13	R\$ 0,06
Diferença (%)	-71,3	-62,9	-28,8

7.9 Práticas complementares: *Multiplex Ligation-dependent Probe Amplification* (MLPA)

A análise de deleções/duplicações maiores (não detectáveis ao sequenciamento) foi realizada por MLPA nos 19 pacientes cuja análise genética por SS e por NGS foi negativa. O estudo por MLPA resultou na identificação de deleções em heterozigose em 3 pacientes (4%; 3/76).

O primeiro deles apresentou deleção completa do gene *MEN1*, com região mínima afetada de 6615pb (GRCh37/hg19 – chr11:64571868-64578482) ou máxima de 223147pb (GRCh37/hg19 – chr11:64571868-64795014).

O segundo paciente é portador de uma deleção, em heterozigose, das regiões promotora, exon 2 e, possivelmente, parte do exon 3 do gene *MEN1* cujo tamanho mínimo e máximo estimados foram, respectivamente, de 1066pb (hg19 chr11:64577395-64578460) e de 3038pb (chr11:64575423-64578460).

Já o terceiro caso apresentou uma deleção acometendo a região promotora do gene *MEN1*, de tamanho mínimo e máximo de 342pb (hg19 chr11:64578119-64578460) e 216.995 bp (GRCh37/hg19 – chr11:64578019-64795013), respectivamente.

Nos três casos (figura 12), somente o ponto de quebra 3' do gene *MEN1* foi determinado, pois não é possível afirmar a ocorrência ou não de deleções além da extremidade 5' do gene *MEN1* devido a ausência de sondas na região

Figura 12: Representação de deleções grandes e grosseiras detectadas em 4% dos indivíduos (MLPA); 1) Tamanho estimado mínimo de 6.615pb (GRCh37/hg19 – chr11:64571868-64578482) e máximo 223.147pb (GRCh37/hg19 – chr11:64571868-64795014). 2) Tamanho estimado mínimo de 1.066 bp (GRCh37/hg19 – chr11:64577395-64578460) e máximo de 219.523pb (GRCh37/hg19 – chr11:64575491-64795013). 3) Tamanho estimado mínimo de 342pb (GRCh37/hg19 – chr11:64578119-64578460) e máximo de 216.995 pb (GRCh37/hg19 – chr11:64578019-64795013).

8 DISCUSSÃO

8. Discussão

8.1 Diagnóstico Genético na prática clínica: A migração do SS para a aplicação de NGS.

Desde que o sequenciamento pelo método de Sanger (SS), utilizando ddNTPs (dideoxinucleotídeos) foi descrito em 1975, os métodos para realização de sequenciamento e análise genética vem sendo aprimorados constantemente (79). A determinação da sequência genética por SS foi amplamente utilizada no projeto Genoma Humano concluído em 2001. Esta técnica é, ainda hoje, considerada o método padrão ouro para sequenciamento genético. Um aumento significativo, tanto da capacidade geração, como de análise das sequências genéticas pela técnica de sequenciamento Sanger (SS), cujo princípio básico se mantém o mesmo, se deu após criação e aprimoramento dos sequenciadores por eletroforese capilar (EC), cuja capacidade máxima de leitura é de 700pb a 800pb (80).

O aperfeiçoamento dos dispositivos de EC propiciou um avanço exponencial na compreensão dos estados normais e patológicos relacionados à genética humana. Como exemplos disto, a técnica de SS permitiu definir padrões de herança, possibilitou o aperfeiçoamento do aconselhamento genético, definiu penetrância da doença nos portadores de mutações e estabeleceu correlações genótipo-fenótipo. Como consequência, promoveu o diagnóstico e tratamento precoce de várias condições genéticas, elevando a qualidade de vida e de sobrevivência dos portadores.

O acesso à análise genética permite o reconhecimento de parentes sob risco que não herdaram a mutação, havendo um benefício psicológico, uma vez que recebem aconselhamento genético que os excluem, junto com seus descendentes, da expectativa de desenvolver a doença de seus familiares. Há, também, benefício aos Serviços de Saúde públicos e privados uma vez que estes indivíduos são excluídos da necessidade de vigilância clínica periódica, minimizando os custos de saúde. Além disto, a identificação de portadores assintomáticos de mutação permite a inclusão destes indivíduos em um programa de vigilância clínica periódica que possibilita tanto o diagnóstico como o tratamento precoce.

As neoplasias endócrinas múltiplas tipos 1 e 2 são exemplos do benefício resultante da implementação da análise genética na prática clínica (81-89). Dados indicam que cenários onde são aplicados o diagnóstico genético de NEM1, não há ocorrência de metástases ou mortalidade, enquanto que nos indivíduos não diagnosticados tais taxas resultam em 23% e 15%, respectivamente. Da mesma forma, quando o diagnóstico é puramente clínico, estas taxas resultam em 18% e 7%, respectivamente (9).

Alguns poucos serviços médicos de atenção terciária oferecem o rastreamento genético na prática clínica. A maioria dos recursos voltados para este fim é originado de fundos acadêmicos e para pesquisa, que acabam, indiretamente, beneficiando pacientes nestes serviços. Tal fator se deve, principalmente, ao alto custo para execução dos procedimentos envolvidos e de análise e interpretação dos dados, gerando uma barreira para difusão da análise genética de pacientes com síndromes hereditárias raras e complexas.

Síndromes genéticas que não dispõem de uma clara associação genótipo-fenótipo, que cursam com penetrância e expressividade variáveis e que não apresentam mutações específicas de alta prevalência, ditas *hots spots* mutacionais, amplificam as dificuldades de incorporação da análise genética como rotina na prática clínica. A síndrome NEM1 e Neurofibromatose tipo 1 representam exemplos clássicos de condições genéticas que englobam estas desvantagens. A ausência de *hots spots* mutacionais, por exemplo, implica na obrigatoriedade de sequenciamento de toda região codificadora destes genes. No caso de NEM1, ao menos 8 pares de sequências iniciadoras, 8 reações de PCR e 16 reações de sequenciamento devem ser, no mínimo, consideradas. Em alguns casos, faz-se necessário sequenciamento de regiões não traduzidas (UTRs), promotoras e intrônicas.

O desenho das sequências iniciadoras para determinadas regiões e a utilização das mesmas para reação de sequenciamento se constituem em uma tarefa complexa pela necessidade de alta especificidade destes oligonucleotídeos. A reação de PCR longa, embora também demande desenho de sequências iniciadoras específicas, necessita de somente um par de oligonucleotídeos aliado a uma única reação de PCR. Uma vez amplificado, o material encontra-se pronto para ser processado. Soma-se a isto a eficácia

da enzima utilizada na PCR longa em corrigir falhas de incorporação (sítio de exonuclease). Os sítios de ligação dos oligonucleotídeos não devem conter SNVs conhecidos e devem ser alocados, preferencialmente, em regiões conservadas do genoma, exigindo cautela, uma vez que as regiões UTRs, intergênicas ou intrônicas, em muitos casos, não são bem conhecidas.

O método de NGS permite o sequenciamento e análise de toda região genômica do gene selecionado, além da inclusão, em uma mesma rotina, de diversos indivíduos a partir da técnica de indexamento duplo (*dual indexing* - Illumina) ou aplicação de “códigos de barras” em uma das extremidades da sequência (*Barcoding* – Ion Torrent). Isto é possível mesmo que alvos (painéis ou genes) ou métodos de captura e enriquecimento (PCR longa, PCR *multiplex* ou captura por sondas de DNA ou RNA biotinizadas) sejam distintos. Em decorrência destas características, o método de NGS incorre em um custo menor por cada par de base sequenciadas, sendo mais custo efetivo que SS (90).

O bom custo-benefício da aplicação do método de NGS é corroborado por vários estudos usando esta técnica como método de diagnóstico de várias síndromes genéticas de predisposição a câncer (91-95), incluindo câncer de mama/ovário familiar (84), neoplasia endócrina múltipla tipo 2 (96), casos familiares de paraganglioma e feocromocitoma (97) e outras doenças complexas (98).

Embora não haja, até o momento, nenhum estudo realizando triagem genética para indivíduos portadores da síndrome NEM1 por NGS, Manson-Bahr *et. al.* (2015) estudaram recentemente o gene *MEN1*, dentre vários outros, em amostras de tumores prostáticos fixados, utilizando a plataforma HiSeq 2000 (Illumina, EUA). Neste estudo foi demonstrado que, para maioria dos casos, é possível triar geneticamente materiais provenientes de tumores para avaliação de riscos potenciais (99). A boa aplicação do método de NGS para diagnóstico genético de NEM1 também teria elevado potencial para investigação de outros fatores genéticos atuando, juntamente com a mutação germinativa *MEN1*, no desenvolvimento ou não da doença e/ou modulando o fenótipo (1).

Todos estes estudos envolvendo NGS (90-99) asseguram a boa aplicabilidade e eficácia desta técnica na triagem genética de indivíduos portadores de doenças

monogênicas e complexas, mostrando uma tendência de migração da técnica de sequenciamento Sanger para NGS na prática clínica.

Especificamente no caso de NEM1, 5% a 10% dos indivíduos não possuem variantes patogênicas na região codificadora do gene (100). Em decorrência deste fator, a aplicação do método de NGS com a leitura de toda região genômica do gene, bem como possibilidade de fácil enriquecimento e análise de outros genes possivelmente relacionados ao desenvolvimento da síndrome teria o potencial de revelar variantes genéticas patogênicas nestes casos cuja identificação por SS seria até possível, porém, muito mais dispendiosa e trabalhosa.

Nós não encontramos mutações germinativas em regiões promotoras ou UTR ou definitivamente intrônicas usando NGS em 19 pacientes negativos para mutação *MEN1* em exons ou fronteiras intron-exon. Vários trabalhos sistematicamente sugerem que parte dos pacientes com NEM1 sem mutação poderiam ter mutações patogênicas nestas regiões. Entretanto, até o presente, não há estudos documentando esta hipótese. Embora, nossos dados tenham sido negativos, o número de pacientes analisados em nosso estudo é muito pequeno e, por isto, não se pode excluir que este fenômeno esteja presente em alguns pacientes com NEM1. A técnica aqui utilizada, por sua vez, possibilita que tais regiões sejam facilmente rastreadas com um número maior de pacientes tendo, assim, elevado potencial de responder esta questão em trabalhos subsequentes.

No presente estudo, nós identificamos 3 pacientes com grandes deleções no gene *MEN1* por MLPA, correspondendo a 4% dos casos. O número de casos com deleções grandes é extremamente variável na literatura (1-33%) (33). Provavelmente, isto deve estar relacionado com o fenótipo selecionado para investigação e com o número de pacientes analisados nas diferentes casuísticas.

Dentre os 19 pacientes analisados sem mutação germinativa identificada por SS ou MLPA, havia 8 com NEM1 familiar em nossa casuística. Três deles, tinham deleções grandes por MLPA (38%/3/8). O fenótipo destes pacientes era inequívoco com detecção de NEM1 típica em vários membros destas famílias, incluindo múltiplos tumores e multicentricidade aumentando a expectativa de achado de mutação germinativa. Mesmo se consideramos todos os 19 casos, teríamos 16% com deleções grosseiras. Assim, nós

concordamos que a análise por MLPA deve ser, por ora, recomendada para todos os pacientes com NEM1 negativos para mutação, como recomendado por Consensos. É possível que estudos futuros possam definir mais claramente se todos estes pacientes realmente necessitem de MLPA ou se deveria ser suportada para casos específicos. Estudos avaliando o custo-benefício desta técnica em pacientes com MEN1 considerando subgrupos de maior ou menor risco potencial como aqui sugerido são necessários.

Somente cerca de 2-3% dos pacientes com apresentação esporádica de NEM1 ou NEM1 familiar sem mutação previamente documentada no gene *MEN1* são caracterizados como síndrome NEM1-like. Estes pacientes albergam mutações germinativas em genes inibidores de quinases dependentes de ciclina. Os genes associados com o fenótipo NEM1-like são: *CDKN2B*, *CDKN2C*, *CDKN1A* e *CDKN1B* (*p27*) (100-101). O gene *p27* foi o primeiro deste grupo de genes supressores tumorais atuando na via de quinases dependentes de ciclinas a ser associado com o fenótipo MEN-like e pacientes com mutações neste gene foram classificados como portadores de síndrome NEM4 (101). O fenótipo dos pacientes com MEN-like não é ainda bem caracterizado. Fenótipos NEM1 associados com mutações em um destes genes parecem ser excepcionalmente raros uma vez que 7-10 anos após associação com este fenótipo somente 10 probandos foram reportados com MEN4 e 8 com mutações germinativas em um dos outros genes CDKIs Não *CDKN1B* dentre cerca de 525 pacientes investigados com fenótipo MEN1.

Da mesma forma que sugerido para análise por MLPA, são necessários estudos para avaliar melhor se realmente todos os casos com fenótipo NEM1 ou NEM-like sem mutação germinativa por SS ou por MLPA deveriam ser submetidos à análise de todos os outros genes MEN1-relacionados. Como esta associação parece ser extremamente rara a avaliação de custo-benefício da análise destes múltiplos genes deve ser investigada em estudos futuros.

Baseado nestes dados, nós preferimos estabelecer e validar a técnica NGS somente para o gene *MEN1* uma vez que não há dúvida do valor de sua indicação para todos os pacientes com fenótipo MEN1. Por outro lado, a técnica aqui validada permite

que se explore futuramente o custo-benefício da análise combinada do gene *MEN1* e dos outros genes *MEN1*-relacionados.

Fenocópias correspondem aos pacientes portadores de mutação somática ocorrendo de forma casual e independente em diferentes tecidos relacionados com predisposição ao desenvolvimento de tumores em NEM1. Como exemplo de fenocópia, podemos citar um paciente com diagnóstico clínico de NEM1, caracterizado pela associação de hiperparatireoidismo e acromegalia, ausência de mutação germinativa no gene *MEN1*, ausência de deleções por MLPA, ausência de mutação germinativa nos genes: *CDKN2B*, *CDKN2C*, *CDKN1A* e *p27* e estudo genético dos tumores revelando mutações somáticas. No caso exemplificado, com a presença de mutação inativadora em heterozigose no gene *HRPT2* e perda de heterozigose tecidual (inativação bialélica tecidual) no adenoma de paratireoide e mutações somáticas ativadoras da proteína $G\alpha$ no tumor hipofisário. Ao contrário de NEM1 familiar clássica, estes pacientes não albergam qualquer risco de transmissão genética para seus descendentes.

Finalmente, não pode ser excluído que um grupo restrito de pacientes com NEM1 possam ter variantes patogênicas germinativas ou de significado desconhecido em genes ainda não relacionados à síndrome. Outra possibilidade seria a de que alguns pacientes com *MEN1* esporádica poderiam apresentar mutações somáticas pós-zigóticas no gene *MEN1* ou em outros genes configurando em possíveis mosaicos com é descrito, por exemplo, na síndrome de Mccune Albright.

8.2 Determinação de métricas e parâmetros mínimos

8.2.1 Amplificação por PCR

Inicialmente, um produto de PCR de aproximadamente 7Kb, representando toda a região genômica do gene *MEN1*, foi obtido a partir de um par de primers. Entretanto, para que tal reação fosse bem-sucedida, esperava-se uma boa qualidade de gDNA. Sabe-se que a razão 260/280 é utilizada para indicar pureza do material, sendo que, quando abaixo de 1,8, infere à amostra uma possível contaminação de substâncias carregadas da fase de extração que absorvem na faixa de comprimento de onda de 280 (102). Mesmo com 31,6% das amostras abaixo da razão adequada esperada para amplificação, 11% destas não funcionaram de acordo, tendo de ser purificadas. Tal fator demonstra que,

mesmo na presença de contaminantes há boas chances de amplificação dado ao bom desempenho do kit de PCR longa utilizado no processo. Além disto, a não necessidade de gDNAs de elevada qualidade neste processo auxiliam na rapidez do processo de sequenciamento NGS, contribuindo para sua custo-efetividade.

Além disto, a quantificação do produto de PCR longa por fluorimetria constitui o primeiro ponto de checagem da técnica antes do prosseguimento da reação de construção de biblioteca, auxiliando na diminuição do risco de perdas.

8.2.2 Estabelecimento das métricas ideais para NGS

Para estabelecimento das métricas ideais a corrida #3 foi excluída devido a falhas em seu processo de formação de *clusters*, o que levou a uma subutilização do kit e baixo rendimento, de modo que algumas amostras tiveram de ser repetidas na corrida #4.

Durante o processo conhecido como carregamento do “cartucho” (colocação dos reagentes de corrida, se uso único, e preparo das amostras para iniciar o sequenciamento) a quantidade de biblioteca é crucial. Estabelecer uma quantidade de DNA (em nanomolar – nM) ideal para que não haja sobreposição de *clusters* nem baixo rendimento é fundamental. Outros pontos podem, ainda, fazer com que este processo falhe, levando a perda da corrida, por exemplo, falha na reação de quantificação por qPCR, discutido adiante.

Diante da sequência selecionada e enriquecida, foi escolhido o Kit “nano” para condução do estudo molecular pela sua menor capacidade de geração de dados e pelo melhor custo-benefício. Outro fator preponderante para seleção desta célula de fluxo baseado na demanda de 20 pacientes/mês, foi a falha na detecção de uma variante patogênica durante o uso do kit convencional, com capacidade de gerar 7,5Gb (a ser discutido posteriormente). Esta célula de fluxo “nano” permite o sequenciamento de 1 milhão de leituras em 500 ciclos, que foi realizado de maneira bi-direcional, com o objetivo de aumentar a profundidade de leitura. Com este Kit, a plataforma seria capaz de gerar, no máximo, 500Mb e de analisar, com profundidade média de leitura de 500x, 140 indivíduos, desconsiderando a utilização de controles. Entretanto, analisando as corridas #2 e #4, que o utilizaram, tal métrica pode ser atingida com uma média de 50

indivíduos, revelando uma perda de aproximadamente 75%, em média, das sequências obtidas. Esta perda pode ser resultante de diversos aspectos, dentre eles, falhas na colocação dos índices, duplicatas de PCR descartadas no alinhamento, locação de DNA controle na *flowcell*, dentre outros. O gráfico 1 representa a profundidade média obtida na região alvo esperada e a obtida de fato em cada corrida.

Gráfico 1: Comparação entre as corridas #2 e #4 com relação ao cálculo da profundidade média esperada baseando-se na capacidade da *flowcell* e no volume de reads obtidos em cada uma das corridas, com a profundidade média realmente obtida nas duas corridas.

O número de ideal de *clusters*, para que seja possível extrair, em sua totalidade, a capacidade de dados da *flowcell*, de modo a garantir, com um maior número de amostras por corrida, um menor custo (gráfico 2). Com estas corridas, sabe-se que o número de *clusters* deve ser maior que 600K*clusters*/mm² ($r^2=0,935$). Neste gráfico, a corrida #3 foi incluída devido a falha na formação de *clusters*, de modo que possa contribuir para estabelecimento deste valor ideal.

Fica evidente a relação direta entre quantidade de *clusters* com geração de dados. Porém, um aumento excessivo no número de *clusters* passaria a inverter o gráfico (declínio) devido ao aumento de sobreposição de *clusters*.

Gráfico 2: Correlação entre quantidade de dados gerados com quantidade de *clusters* obtidos na corrida.

Tomando-se como exemplo a corrida #2, que envolveu o maior número de indivíduos, é possível determinar parâmetros mínimos como, número de *reads* ou profundidade mínima de leitura para obtenção de bons resultados utilizando um mesmo protocolo de processamento de amostras. Os gráficos do grupo 3 ilustram correlações entre número de leituras e profundidade média de leitura (3a), profundidade média de leitura e porcentagem de bases com profundidades mínimas (3b) e número de leituras com porcentagens de bases com profundidades mínimas (3c).

Gráfico 3a: Correlação entre leituras obtidas e profundidade média de leitura

Gráfico 3b: Correlação entre profundidade média e porcentagens de bases com profundidades mínimas obtidas

Gráfico 3c: Correlação entre leituras obtidas por amostra e porcentagens de profundidades mínimas obtidas

É possível estipular um número mínimo de leituras da região alvo necessárias para obtenção de 100% de cobertura, tendo 90-95% das bases com >10x de profundidade de leituras (103).

Para garantir 100% de cobertura da região alvo, pode-se afirmar que seriam necessários um número mínimo de 28400 leituras por amostra (gráfico 3a). Entretanto, existem determinadas amostras (ex. 6400, 7400 e 8800 leituras) que conseguem atingir 100% de cobertura da região alvo, enquanto que determinadas amostras que contabilizaram entre 8000 e 11000 leituras não foram capazes de atingir esta métrica mínima. Isto condiz com uma variação expressiva na alocação das leituras na região alvo, bem como o tamanho dos mesmos, de modo que amostras com mais leituras, não alcancem cobertura total da região alvo. A variação de distribuição de leituras na região alvo é ocasionada pelo método de cobertura aliado ao método de fragmentação (transposase).

Quando comparadas as profundidades médias de leituras das amostras com sua taxa de cobertura na região alvo, tal fato se repete; há amostras com profundidades médias entre 140x e 150x com 100% de cobertura, enquanto que amostras com 280x de profundidade média atingem 99,1% de cobertura na região alvo. Com isso, pode-se dizer que 100% da região alvo coberta pelo sequenciamento não garante a confiabilidade dos

achados, já que muitas destas regiões podem ter sido cobertas com profundidades menores do que 10 leituras, o que impossibilita análise (103).

Quando avaliados os parâmetros de profundidade médias de leitura em conjunto com o parâmetro de porcentagem de bases com profundidade mínima de 50x, 20x e 10x (gráfico 3b), é observada uma tendência mais linear. Contudo, algumas das amostras também acabam por demonstrar uma variação considerável, com profundidade média maior, contudo, apresentando porcentagens de bases >50x, >20x e >10x menores. Novamente, a utilização de PCRs ao longo do processo de preparo aliado a fragmentação enzimática contribuem para esta variação.

O mesmo evento ocorre quando comparados número de leituras com porcentagem de bases com profundidades mínimas de 50x, 20x e 10x (gráfico 3c).

Estes pontos corroboram a grande variação na distribuição de leituras ao longo da região alvo, dificultando a garantia de bons resultados com a utilização do fluxograma de preparo de biblioteca aqui proposto, tornando necessário um número maior de dados para um sequenciamento robusto.

É fato que a retirada de reações de PCR do processo de enriquecimento e do preparo de biblioteca melhorariam a distribuição dos dados. Porém, seria necessário um volume maior de gDNA para que as reações fossem iniciadas. Além disto, kits comerciais que operam sem necessidade de amplificação de DNA oneram consideravelmente o processo.

Desta forma, pode-se inferir que os parâmetros mais seguros e eficientes para cálculos de número máximo de amostras por célula de fluxo são número de leituras por amostra e cobertura média ($r^2=0,9363$). Entretanto, deve-se levar em conta o padrão de qualidade objetivado para o experimento. Levando-se em conta 95% das bases >10x, necessários para confiabilidade do método (103), obtêm-se um valor de no mínimo 6000 leituras por amostra.

Analisando o gráfico 3a, pode-se afirmar que este número de leituras gera amostras com 150x de profundidade média, ou seja, utilizando uma célula de fluxo com 1 milhão de leituras tem-se um número máximo de 160 indivíduos. Entretanto, a taxa de perda, supracitada deve ser levada em conta (~75%), de modo que este número máximo resulta em 40. Tal número corrobora a eficiência da corrida #2 (98,5% >10x) com 52 indivíduos.

Embora a profundidade média de leitura seja útil para cálculo de número de amostras, o método aplicado para fragmentação tem implicação direta na maneira como estas leituras estão distribuídas ao longo da região alvo, gerando, na métrica de profundidade média, um alto desvio padrão. Métodos por ultrasonicação oferecem uma profundidade de leitura mais homogênea ao longo da região alvo e, desta forma, permitindo a utilização do parâmetro de profundidade média com maior precisão, já que os desvios padrões serão menores e todas as bases terão profundidades semelhantes (104).

Fragmentações por meio de transposases, como a utilizada neste estudo, geram padrões ditos “picos e vales”, que se repetem em todas as amostras, com profundidades elevadas em regiões de fácil ruptura e baixas em regiões difíceis de serem rompidas, em especial, regiões ricas em GC. A figura 13 ilustra este padrão, ressaltando o início do exon 2 (70% GC) e IVS7 (58% GC), onde há uma queda abrupta no número de leituras. Especula-se acerca da dificuldade do método de transposase em romper nucleotídeos nestas regiões, deixando-as presentes somente em regiões centrais de sequencias maiores e, portanto, difíceis de serem lidas uma vez que serão analisadas nos últimos ciclos (69). As regiões ricas em AT e de homopolímeros foram sequenciadas de maneira satisfatória, sendo que homopolímeros T tiveram problemas a partir da décima segunda base.

Existe, ainda, um fator comprobatório sobre a implicação da fragmentação sobre quedas repentinas de leituras em determinadas regiões. Uma vez que foi utilizada PCR longa e toda região de interesse foi amplificada, sabe-se que o processo de enriquecimento gerou amplicons completos, não favorecendo nenhuma região específica do gene *MEN1*.

Figura 13: Representação gráfica das leituras obtidas ao longo da região genômica do gene *MEN1*, no sentido 3' – 5' do gene.

Apesar destes pontos acerca da fragmentação por transposases, esta técnica se configura em um método de fácil e rápida aplicação, além do baixo custo associado para oferecer gDNA prontamente fragmentado e com adaptadores já inseridos para prosseguimento da preparação da biblioteca. Embora a quantidade necessária de gDNA inicial seja baixa, é fundamental que o material tenha de boa qualidade e seja livre de contaminantes. Material em excesso também afeta o processo, concluindo-se que uma pequena parte do gDNA tem acesso a transposase durante a reação, tornando esta uma das etapas mais críticas para qualidade do sequenciamento (105).

Os gDNAs obtidos inicialmente pela técnica de *Salting-out* não tinham qualidade esperada para técnica de NGS, contudo, a PCR longa ocorreu sem maiores problemas. O aumento deste fragmento específico na solução, a despeito dos contaminantes ali presentes pode ter contribuído para sucesso da reação. Mesmo com determinadas regiões em baixa profundidade de leitura, quando proporcionalmente comparadas a outras, foi possível análise satisfatória. Entretanto, com o aumento da região alvo e possível aplicações de painéis gênicos dentro da mesma plataforma, estas regiões podem necessitar de mais leituras, tornando necessário rever o processo de fragmentação e/ou enriquecimento a longo prazo.

Sequenciamentos com leituras bi-direcionais e fragmentos maiores permitem melhorar os parâmetros de profundidade média e cobertura, já que permitem mais leituras de uma mesma região e melhor alinhamento das sequências. Além disto, atingir 90-95% das bases >10x é fundamental para garantir confiabilidade aos achados e eliminar possíveis falsos positivos e negativos (103). Embora este parâmetro tenha sido estabelecido para sequenciamento exômico, pode ser extrapolado para sequenciamento de alvos específicos a partir de uma plataforma menor.

Pode-se admitir que ~50 amostras por corrida podem gerar métricas suficientes para 100% de cobertura da região alvo com níveis satisfatórios de bases >10x de profundidade, permitindo aplicação do método de NGS com custos reduzidos, quando comparados ao método de Sanger.

8.3 Normalização de amostras

É necessário que todas as amostras sejam disponibilizadas ao equipamento em equimolaridade, para melhor aproveitamento dos dados obtidos por meio do

sequenciamento NGS. Para que isto ocorra, o processo de preparo Nextera (Illumina) inclui uma etapa de normalização baseada em *beads* e sugere realização de PCR em tempo real para quantificação. Os dados da quantificação são dispostos na tabela 11.

Tabela 11: Dados de quantificação por qPCR

	Corrida #1	Corrida #2	Corrida #3	Corrida#4
Quantificação média (nM)	0,85	0,73	0,9	1,0
Desvio padrão	0,24	0,16	0,3	0,39

Tais dados demonstram uma variação nas concentrações após a normalização. Contudo, outra possibilidade a ser considerada é a variação na reação de PCR em tempo real. Após análise dos dados da figura 8, nota-se a propagação desta variação na representação de cada amostra. Uma vez que, baseando-se na quantificação por PCR em tempo real, as amostras foram diluídas e equalizadas novamente, tal propagação não deveria ocorrer. Deste modo, conclui-se que o responsável pela variação seria, justamente, a técnica de PCR em tempo real, não sendo recomendável a utilização da mesma, caso a técnica goze de outros meios para normalização das amostras (kits normalizadores), podendo levar, ainda, a uma redução de custos.

8.4 Reprodutibilidade, repetitividade, sensibilidade e especificidade

Durante a primeira fase do projeto não houve devida identificação da variante NM_130799.2, exon2, c.201delC:p.A68Pfs*51, resultando em uma concordância de 95,7%. O elevado número de leituras obtidas foi o agente causador desta falha, já que a célula de fluxo era superdimensionada nesta fase (15 milhões de *reads*) para os necessários 80,5Mbases de dados para sequenciamento dos 23 indivíduos com profundidade média de 500X. Se tratando de uma deleção, o *gap* gerado era considerado penalidade durante o alinhamento, levando a uma perda de leituras deletadas, deixando somente as sequencias maiores, e a uma desproporção ainda maior entre sequencias deletadas *versus* normais, impedindo a chamada da variante.

De maneira geral, níveis elevados de profundidade tendem a dificultar a chamada de variantes, principalmente de In/Dels. Embora esta interferência seja conhecida, não é possível estabelecer uma linha de corte de profundidade para evitar resultados falso-negativos. Outro fator que provavelmente dificultou a chamada da variante está

relacionado com a sua localização em uma região rica em G-C, difícil de ser reconhecida pela plataforma Illumina. É sabido que a sequência GGC é de difícil acesso para sequenciadores Illumina, consistindo em um erro persistente desta plataforma (figura 11) (69).

É importante enfatizar que tal variante patogênica não identificada em virtude de problemas relacionados com algoritmos de bioinformática nas fases de alinhamento e de chamada de variantes foi identificada pela análise visual das leituras. Desta forma, é possível que seja necessária análise visual para a análise do gene *MEN1*, a despeito da diminuição de tempo que traria análise único do arquivo VCF.

Assim, é possível que a melhora dos algoritmos dos programas utilizados, aliada à utilização de plataformas com taxas de erros menores e, consequente, maior confiabilidade, possa diminuir drasticamente a necessidade de confirmação pelo método de Sanger e enaltecer sua boa aplicabilidade no rastreamento de familiares para variantes conhecidas.

Para definição da capacidade do teste para detecção da presença ou ausência da doença (teste qualitativo) por verificação de variantes na sequência de DNA germinativa, os parâmetros como sensibilidade, especificidade, exatidão e precisão são considerados parâmetros imprescindíveis (106). Ressalta-se que avaliar a capacidade do método de NGS em dizer se um caso positivo é detectado positivo (sensibilidade), ou negativo detectado como negativo (especificidade) depende da classificação da variante encontrada. Admitindo que todas as variantes aqui ditas como patogênicas assim o foram descritas decorrente de bancos de dados e informações presentes até o momento tomando-se isto como verdade absoluta, é possível determinar a sensibilidade e especificidade do teste. Desta forma, foi visto que o método de NGS é sensível e específico, de modo a identificar corretamente aqueles portadores ou não de variantes patogênicas no gene *MEN1*.

A reprodutibilidade foi analisada uma vez que operadores diferentes analisaram mesmos resultados, em plataformas diferentes em todos os indivíduos (N=76). Considerando a concordância completa dos resultados finais obtidos a reprodutibilidade resultou em 1 demonstrou a eficiência da plataforma NGS, utilizando o kit “nano”. A

repetitividade foi avaliada em 24 pacientes, analisados por 2 vezes, por um mesmo operador, resultando em 0,986. A variação obtida na repetitividade ocorreu devido ao incidente com a deleção não encontrada na corrida #1. Fica claro que os ajustes e, principalmente a utilização do kit “nano” extinguirá as possibilidades de repetição deste erro.

Outro fator calculado durante estabelecimento de sensibilidade e especificidade é a prevalência de doença na população, que, para este estudo, pode ser classificada como enviesada, já que a população utilizada é clinicamente diagnosticada como portadora da síndrome NEM1, conseqüentemente, a frequência de prevalência da doença na população é 75%, com intervalo de confiança de 95% entre 63,7% e 84,2%.

Os valores de sensibilidade e especificidade somente resultaram nos valores acima assumindo que todas as variantes capazes de classificar um indivíduo como portador ou não portador da síndrome genética NEM1 foram corretamente identificadas quanto sua presença ou ausência. Para definição da patogenicidade das variantes foi aplicado o escore preconizado pelo ACMG (*American College of Medical Genetics*) (107). Com esta análise, 76,3% (29/38) recebem o escore PVS1 (*very strong*) e PS1 (*strong*), sendo classificadas como patogênicas. As outras 9 variantes preenchem os critérios PS1 (*strong*), PS4 (*strong*) e PM1 (*moderate*) e PP3 (*supporting*), sendo, também, classificadas como patogênicas (107).

Embora 99,6% de todas as variantes foram concordantes na técnica de sequenciamento Sanger, duas variantes polimórficas não foram confirmadas pelo NGS, uma delas no exon 9 c.1245C>T:p.D418D em heterozigose e outra no exon 2 c.216T>A:p.P72P em heterozigose. A primeira delas teve frequência de 57% na população do estudo, com 20% destes em homozigose. Na população geral (1000 *genomes*) a frequência do alelo T é de 30%. Além de ser encontrada em homozigose, o alelo alterado tem frequência elevada na população normal, o que classifica esta variante como tolerada. A segunda variante citada não é descrita nos bancos de dados dbSNP, HGMD (banco de dados público), ExAC ou 6500 exomas (EVS), contudo, na ferramenta MutationTaster ela resulta em uma troca sinônima, não sendo relevante para transcrição e tradução proteica, levando a classificação da variante como tolerada.

Em concordância com a literatura científica (6), não houve regiões consideradas pontos quentes para variantes, patogênicas ou toleradas, ao longo da região genômica do gene *MEN1*.

A boa sensibilidade e especificidade pode ser corroborada por diversos estudos com outras condições clínicas e aplicação de painéis gênicos, mas testes visando aplicação para rastreamento de condições que envolvem um ou poucos genes não são explorados (95-97, 108-110). A análise por plataformas de NGS também resultou em especificidade e sensibilidade de 100% para verificação de possíveis aneuploidias em teste pré-natal, onde profundidades de leituras elevadas são necessárias para diagnóstico (108).

Entretanto, o presente estudo teve como objetivo validação por meio da determinação da sensibilidade do método para detecção de indivíduos positivos para variante patogênica no gene *MEN1*, uma vez que problemas maiores seriam causados na ocorrência de falsos negativos em comparação com possíveis detecções de falsos negativos. Apesar disso, pacientes esporádicos foram incluídos, de modo a fortalecer o método como específico, exaurindo as chances de resultados falsos-positivos, que poderiam trazer grande ônus aos serviços de rastreamento clínico.

8.5 Ferramentas, fluxo de análise e rotinas complementares

O sequenciamento massivo paralelo gera, a partir de plataformas da Illumina, arquivos denominados BCL (*Basecall file*). Ainda dentro da própria plataforma, tal arquivo é transposto para arquivos FASTq, cuja configuração detém a sequência obtida a partir das imagens iniciais e escores de qualidade (111).

A partir deste ponto, há possibilidade de processamento em nuvem disponibilizada pelo próprio fabricante (*Basespace*), utilizando BWA (112) para alinhamento das sequências e GATK (72) para chamada de variantes. Ao realizar resequenciamento, tal método não realiza anotações, embora haja um anotador disponível. A utilização deste meio de análise permite que, imediatamente após o término da reação de sequenciamento os dados brutos sejam processados sem intervenção do usuário. Além disso, não é necessário que plataformas de informática com alto poder de processamento estejam disponíveis dentro da infraestrutura do serviço, o que reduz o custo de implantação.

Entretanto, para o presente estudo foram realizadas análises com *pipelines* desenvolvidos dentro do serviço para fatores de comparação. Para tal análise foram realizados processamentos utilizando os programas BWA (112) para alinhamento das sequências, GATK (72) para chamada de variantes, PINDEL (113) para chamada de deleções e inserções, Annovar (114) para anotações e Qualimap (78) para obtenção de métricas.

A partir da geração do arquivo VCF anotado, foram aplicados filtros específicos para variantes *frameshift*, *nonsense* e em sítios de excisão (*splice*). Desta forma, 60% dos indivíduos já puderam ser, de pronto, triados como portadores de variantes patogênicas causadoras da síndrome NEM1. Os demais casos-índices necessitaram de uma análise em banco de dados para classificação das variantes encontradas. Todas as variantes foram visualmente avaliadas pelo IGV (*Broad Institute*), tendo como objetivo avaliar número de leituras, proporção entre alelos, no caso dos heterozigotos, entre outros parâmetros de qualidade. Um dos fatores importantes na avaliação da variante é a presença destes em *reads* em ambos os sentidos, de modo a tornar improvável falhas de incorporação durante a formação dos *clusters*.

Aproximadamente 30% das variantes consideradas causadoras foram do tipo *frameshift*, em concordância com a literatura científica (31). Plataformas que não ofereçam entraves para detecção deste tipo de variante são fundamentais para atuarem no rastreamento destas variantes.

Também em concordância com a literatura (6, 26, 31), a análise por MLPA revelou que aproximadamente 4% (3/76) dos indivíduos deste estudo apresentavam deleções grandes que afetavam a produção da proteína menin em um dos alelos do gene *MEN1*. Mesmo com baixas frequências, é de suma importância a inserção do MLPA na rotina diagnóstica de NEM1 para que não sejam gerados falsos negativos, já que o impacto clínico do achado é relevante. Um dos indivíduos possuem deleção completa do gene *MEN1* (figura 10), aliado a uma clínica de início precoce (8-9 anos).

Desta forma, foi validada a utilização de um fluxo de análise de toda região genômica do gene *MEN1* por NGS seguida da implantação deste fluxograma na prática clínica direcionada ao diagnóstico genético de pacientes com NEM1 e de casos com suspeita clínica de NEM1. Dentro do fluxograma proposto, pacientes cuja análise genética foi negativa para pequenas deleções, inserções ou variantes pontuais causadoras

da síndrome devem ser investigados para ocorrência de deleções grandes pela técnica de MLPA.

8.6 Determinação de custo

A custo-efetividade do método deve ser avaliada para permitir sua aplicação no contexto clínico para rastreamento de indivíduos com diagnóstico clínico da síndrome NEM1.

Estudo prévio aplicando painéis para análise genética de tumor colorretal e síndromes poliposas demonstraram uma boa relação de custo-efetividade do método de NGS (115). Outros painéis genéticos também demonstram a boa custo-efetividade do método de NGS para (116)

Neste estudo, foi feita uma análise comparativa de custo-efetividade entre SS e NGS direcionada para análise do gene *MEN1*.

A custo-efetividade do NGS se baseia, essencialmente, em sua maior capacidade de geração de dados. Assim, o número de bases sequenciadas por paciente foi 42,9% superior ao número de bases por SS (7187pb vs. 4106bp), uma vez que toda a região genômica do gene *MEN1* foi contemplada por NGS, resultando em um custo estimado por par de base sequenciado 60-80% menor por NGS, considerando os três cenários avaliados.

A capacidade que as plataformas de NGS tem em expandir sua capacidade com facilidade, sem mudanças drásticas de investimento, aliado a tecnologias robustas, com baixas taxas de erro traz ao mercado grandes chances de sucesso. Contudo, a necessidade de infraestrutura para se funcionamento acaba dificultando a aplicação.

A plataforma MiSeq permitiria, utilizando *kits* de índices comerciais, 96 identificações independentes, contudo, os 500Mb de sequenciamento disponíveis para a célula de fluxo utilizada permitiriam, aproximadamente, 50 amostras simultâneas. Para o cálculo de custeio foram consideradas 96 amostras como teto, uma vez que o padrão de indexamento levado em conta tinha este número máximo de combinações.

Foram, assim, consideradas 8 reações de PCR para enriquecimento do material e sequenciamento pelo método de Sanger. Como a análise é bidirecional, ao todo, foram realizadas 16 reações de sequenciamento por amostra. Já para a reação de NGS foi

considerada 1 reação de PCR por indivíduo e sua subsequente preparação para o sequenciamento massivo paralelo, com capacidade máxima para 96 amostras.

Certamente a taxa maior de repetição do SS (5,2%) influenciou no custo, sendo consequente da ausência dos pontos de checagens. O SS somente tem checagem após a execução da PCR para amplificação da região alvo. A partir deste ponto, somente se saberá a qualidade do processo ao término deste, havendo dificuldades no rastreamento de falhas. Já o NGS (taxa de repetição – 2,7%) possui pontos de checagem após a amplificação da região alvo (eletroforese em gel de agarose), antes do início do preparo de biblioteca (quantificação por fluorimetria) e após a normalização (quantificação por PCR em tempo real), de modo a identificar falhas antes do sequenciamento, minimizando custos e possibilidades de erros.

Considerando os três cenários avaliados, o NGS se mostra promissor na análise de casos-índices mesclados em rotinas com rastreamento de familiares, sem variação no método de captura. O custo menor do NGS consiste, principalmente, no número de indivíduos por corrida, já que o custo principal está centrado no kit de reagentes de corrida. Amortizar o custo da *flowcell* em um número elevado de pacientes dilui consideravelmente o custo da plataforma. Em contrapartida, a falta de concorrência no mercado de SS eleva os custos da técnica. Soma-se a isto o custo elevado da plataforma de SS e a necessidade de trocas periódicas de reagentes e componentes, mais duradouros no NGS (custo de máquina ociosa 63,7% menor no caso do NGS, comparado ao SS). Vale ressaltar que a padronização do enriquecimento por PCR longa contribuiu para reduzir os custos da técnica de NGS. Com isto, a técnica de NGS se mostrou, aproximadamente, 30-70% mais custo-efetiva, nos respectivos cenários.

Embora o custo para NGS dependa do número de indivíduos a serem sequenciados de uma única vez, a realização da operação baseada em uma plataforma de menor rendimento (Illumina MiSeq) faz com que, mesmo com um número pequeno de pacientes, o sequenciamento seja mais custo-efetivo por NGS.

8.7 EVEREST e STARD

Foram checados pontos de listas específicos para confirmação da boa aplicabilidade e confiabilidade do método de NGS bem como para validar a maneira como o presente estudo foi desenhado.

O primeiro trabalho, denominado STARD (117), lista pontos importantes a serem verificados em estudos que tem por objetivo validação de métodos diagnósticos e aferição da sensibilidade e especificidade destes, conferindo boa integridade aos dados obtidos pelo estudo. Estudos mal desenhados por levar a aplicação de testes diagnósticos não prontos, trazendo graves prejuízos para os pacientes do serviço em questão. Para isto, são listados 23 pontos a serem atendidos para garantir integridade e aplicabilidade dos dados obtidos pelo estudo. São estes os pontos:

- **Determinação de fatores de sensibilidade e especificidade, calculando tais taxas, bem como valores preditivos positivos e negativos para o teste denominado índice (teste sob validação, comparado a uma ferramenta padrão ouro). Tais termos devem estar explícitos no trabalho.**
- **Classificar como objetivo determinação de “acurácia” do teste X, comparando-o com o método padrão ouro utilizado até o momento**
- **Determinação e descrição criteriosa da população do estudo bem como seus critérios de inclusão e exclusão**
- **Descrição da maneira como população foi selecionada**
- **Descrição da maneira como os indivíduos foram distribuídos durante o estudo**
- **Descrição do desenho do estudo (retrospectivo ou prospectivo)**
- **Descrição do racional por trás do funcionamento de cada método**
- **Citar as referências utilizadas para cada ensaio e como os dados foram obtidos**
- **Descrever treinamentos para obtenção de “*know how*” para realização dos experimentos**
- **Descrever testes cegos comparativos entre ambos os métodos**
- **Citar quais os cálculos utilizados para obtenção de sensibilidade e especificidade**
- **Descrever métodos para cálculo de reprodutibilidade (se usado)**
- **Citar quando o estudo começou e foi encerrado**
- **Descrever a situação clínica dos casos-índices**
- **Citar quantos casos-índices satisfizeram os critérios de inclusão**

- **Descrever os caracteres da doença**
- **Comparar os resultados entre ambos os métodos, mesmo os que não foram possíveis sua obtenção**
- **Reportar problemas ou situações adversas quando da realização do teste índice**
- Reportar taxas de confiabilidade do método índice
- Reportar amostras *outliers* e possíveis remanejamento das mesmas
- Reportar estimativas de variabilidade entre as amostras (se calculada)
- **Reportar reprodutibilidade (se calculada)**
- **Discutir importância e aplicabilidade clínica do teste índice**

Considerando os 23 pontos, 17 deles (73,9%) foram atendidos (em negrito). Entretanto, é sabido que alguns pontos são repetitivos (117) para aumentar a confiabilidade dos resultados obtidos pelo presente trabalho.

Já a lista denominada EVEREST (118) traz 35 pontos importantes para avaliação econômica, ou seja, custo-efetividade do método denominado índice e avaliado neste estudo, de modo a garantir que o estudo econômico tenha sido feito de modo a devolver dados fidedignos. São estes:

- **A questão do trabalho está estabelecida? (Introdução)**
- **A importância econômica está estabelecida? (Introdução)**
- **Há justificativa para análise econômica? (Métodos/discussão)**
- **Racional por trás da escolha do método. (Introdução)**
- **Comparação de métodos para escolha. (Métodos)**
- **Qual a forma de análise econômica? (Métodos)**
- **A forma de análise de custos responde aos questionamentos do trabalho? (Introdução/discussão)**
- **Definir o recurso para estimar a efetividade do método. (Introdução/discussão)**
- **Detalhar o desenho e os resultados de efetividade. (Introdução/métodos)**
- Detalhar o método de síntese ou da metanálise das estimativas obtidas (se necessário). (Não disponível)

- **O resultado primário está estabelecido. (Métodos/análise de custo efetividade)**
- **Definir os benefícios para aplicação do método.**
- **Detalhamento das amostras das quais os dados foram obtidos.**
- **Mudanças de produtividade**
- **A relevância das mudanças de produtividade**
- A quantidade de recursos usados deve ser reportada separadamente dos seus valores unitários.
- Definir quais os métodos de escolha para as quantidades dos insumos utilizados.
- **Moeda utilizada e os dados de preço estão armazenados.**
- Ajustes de moeda e possíveis inflações são descritas.
- **Detalhes sobre modelos utilizados.**
- **Justificar a escolha de modelos baseado em pontos chave.**
- **Benefícios dos custos são determinados.**
- Taxas de desconto.
- Justificativa das taxas de descontos.
- **Detalhar dados de sensibilidade.**
- Detalhar testes estatísticos
- **Detalhar o método para estabelecer sensibilidade.**
- Justificar a escolha de variáveis para calcular sensibilidade.
- Detalhar as taxas de variação.
- **Comparar alternativas relevantes.**
- Análise de incremento.
- **Principais resultados são demonstrados em conjunto e separadamente.**
- **A questão do trabalho é respondida.**
- **Conclusões derivam dos dados reportados?**
- **Estabelecer advertências na conclusão.**

De todos estes pontos, 70% destes foram atendidos, demonstrando a eficiência do trabalho em estabelecer os custos para definição da custo-efetividade do método de NGS.

Dados como taxa de juros e variações cambiais não foram levados em conta pois, no custeio, foram considerados pagamentos à vista, com retorno de capital investido em 36 meses (moeda=R\$). Contudo, uma vez que o tempo de retorno tem como base 3 anos, é possível que este custo oscile devido a aumento de impostos ou variações cambiais incidindo sobre reagentes e insumos necessários, uma vez que todos os itens utilizados e custeados são de origem importada.

9. CONCLUSÃO

Os resultados obtidos pelo método de NGS foram concordantes com aqueles encontrados pelo método de Sanger, corroborando a eficácia do primeiro bem como sua aplicabilidade como teste diagnóstico em NEM1.

Tomando-se NEM1 como modelo, o NGS se mostrou mais custo-efetivo quando comparado ao método de Sanger. Devido à grande capacidade de geração de dados do método NGS, ele poderá ser melhor aproveitado complementando a técnica de Sanger, especificamente em casos-índices, devido a necessidade de leitura de toda região do gene *MEN1*. Por outro lado, o sequenciamento Sanger segue como método escolhido para confirmação de variantes patogênicas identificadas por NGS e para rastreamento de familiares sob risco potencial de terem herdado a variante patogênica encontrada em sua família.

REFERÊNCIAS

1. Brandi ML, Gagel RF, Angeli A, Bilezikian JP, Beck-Peccoz P, Bordi C, Conte-Devolx B, Falchetti A, Gheri RG, Libroia A, Lips CJ, Lombardi G, Mannelli M, Pacini F, Ponder BA, Raue F, Skogseid B, Tamburrano G, Thakker RV, Thompson NW, Tomassetti P, Tonelli F, Wells Jr SA, Marx SJ. Guidelines for diagnosis and therapy of MEN type 1 and type 2. *J Clin Endocrinol Metab.* 2001;86:5658–71.
2. Lourenço DM Jr. *Neoplasia endócrina múltipla tipo 1: estudo clínico e gênico de uma grande família brasileira* [Tese]. São Paulo: Faculdade de Medicina da Universidade de São Paulo (FMUSP); 2001.
3. Lourenço DM Jr, Toledo SP. Neoplasia Endócrina Múltipla Tipo 1. In: Liberman B, Cukiert A editores. *Neuroendocrinologia Clínica e Cirúrgica*. São Paulo: Lemos; 2002. p.577-601.
4. Ezabella MC, Nunes AB, Lourenço DM Jr, Toledo SP. Neoplasias Endócrinas Múltiplas. In: Ferreira CG, Rocha JC, editores. *Oncologia Molecular*. São Paulo: Atheneu; 2004. p. 332-7.
4. Lourenço DM Jr, Toledo RA, Toledo SP, Nunes AB, Ezabella MC, Hayashida CY. Neoplasia Endócrina Múltipla. In: Antonio Carlos Lopes, Vicente Amato Neto editores. Alfredo Halpern (Org.). *Tratado de Clínica Médica*. São Paulo: Roca; 2006. v. 2, p. 3535-48.
5. Toledo, RA; Lourenço, DM Jr, Toledo, SP. Neoplasias Endócrinas Múltiplas. In: Milton de Arruda Martins et al (Org.). *Clínica Médica*. Barueri: Manole; 2009. v.5, p. 388-451.
6. Thakker RV, Newey PJ, Walls GV, Bilezikian J, Dralle H, Ebeling PR, Melmed S, Sakurai A, Tonelli F, Brandi ML; Endocrine Society. Clinical practice guidelines for multiple endocrine neoplasia type 1 (NEM1). *J Clin Endocrinol Metab.* 2012;97:2990-3011.
7. Trump D, Farren B, Wooding C, Pang JT, Besser GM, Buchanan KD, Edwards CR, Heath DA, Jackson CE, Jansen S, Lips K, Monson JP, O'Halloran D, Sampson J, Shalet SM, Wheeler MH, Zink A, Thakker RV. Clinical studies of multiple endocrine neoplasia type 1(NEM1). *QJM.* 1996;89:653-69.
8. Lourenço DM Jr, Coutinho FL, Toledo RA, Montenegro FL, Correia-Deur JE, Toledo SP. Early-onset, progressive, frequent, extensive, and severe bone mineral and renal

complications in multiple endocrine neoplasia type 1-associated primary hyperparathyroidism. *J Bone Miner Res* 2010;25:2382-91.

9. Lourenço DM Jr, Coutinho FL, Toledo RA, Gonçalves TD, Montenegro FL, Toledo SP. Biochemical, bone and renal patterns in multiple endocrine neoplasia type 1-associated hyperparathyroidism. *Clinics (Sao Paulo)*. 2012;67: 99-108.

10. Montenegro FL, Lourenço DM Jr, Tavares MR, Arap SS, Nascimento CP Jr, Massoni Neto LM, D'Alessandro A, Toledo RA, Coutinho FL, Brandão LG, de Britto e Silva Filho G, Cordeiro AC, Toledo SP. Total parathyroidectomy in large cohort of cases with hyperparathyroidism with multiple endocrine neoplasia type 1: experience from a single academic center. *Clinics (Sao Paulo)*. 2012;67:131-9.

11. Coutinho FL, Lourenço DM Jr, Toledo RA, Montenegro FL, Correia-Deur JE, Toledo SP. Bone mineral density analysis in patients with primary hyperparathyroidism associated with multiple endocrine neoplasia type 1 after total parathyroidectomy *Clin Endocrinol (Oxf)*. 2010;72:462-8.

12. Coutinho FL, Lourenco DM Jr, Toledo RA, Montenegro FL, Toledo SP. Post-surgical follow-up of primary hyperparathyroidism associated with multiple endocrine neoplasia type 1 *Clinics (Sao Paulo)*. 2012;67:169-72.

13. Lourenço DM Jr, Toledo RA, Mackowiak II, Coutinho FL, Cavalcanti MG, Correia-Deur JE, Montenegro F, Siqueira SA, Margarido LC, Machado MC, Toledo SP. Multiple endocrine neoplasia type 1 in Brazil: NEM1 founding mutation, clinical features and bone mineral density profile. *Eur J Endocrinol*. 2008;159:259-74.

14. Verges B, Boureille F, Goudet P, Murat A, Beckers A, Sassolas G, Cougard P, Chambe B, Montvernay C, Calender A. Pituitary disease in MEN type 1 (NEM1): data from the France-Belgium NEM1 multicenter study. *J Clin Endocrinol Metab*. 2002;87:457–65.

15. Trouillas J, Labat-Moleur F, Sturm N, Kujas M, Heymann MF, Figarella-Branger D, Patey M, Mazucca M, Decullier E, Verge`s B, Chabre O, Calender A and the GTE (Groupe d'études des Tumeurs Endocrines). Pituitary Tumors and Hyperplasia in Multiple Endocrine Neoplasia Type 1 Syndrome (NEM1): A Case-Control Study in a

Series of 77 Patients Versus 2509 Non-NEM1 Patients. *Am J Surg Pathol.* 2008;32:534-543.

16. Le´vy-Bohbot N, Merle C, Goudet P, Delemer B, Calender A, Jolly D, Thie´fin G, Cadiot G Prevalence, characteristics and prognosis of MEN 1-associated glucagonomas, VIPomas, and somatostatinomas: study from the GTE (Groupe des Tumeurs Endocrines) registry. *Gastroenterol Clin Biol* 2004;28:1075-1081.

17. Tonelli F, Giudici F, Fratini G, Brandi ML. Pancreatic endocrine tumors in multiple endocrine neoplasia type 1 syndrome: review of literature. *Endocr Pract* 2011;17:33-40.

18. Niina Y, Fujimori N, Nakamura T, Igarashi H, Oono T, Nakamura K, Kato M, Jensen RT, Ito T, Takayanagi R The current strategy for managing pancreatic neuroendocrine tumors in multiple endocrine neoplasia type 1. *Gut Liver* 2012;6:287-294.

19. Kouvaraki MA, Shapiro SE, Cote GJ, Lee JE, Yao JC, Waguespack SG, Gagel RF, Evans DB, Perrier ND. Management of pancreatic endocrine tumors in multiple endocrine neoplasia type 1. *World J Surg* 2006;30 643-653.

20. Akerstrom G., Hessman O, Hellman P, Skogseid B. Pancreatic tumours as part of the MEN-1 syndrome. *Best. Pract. Res. Clin. Gastroenterol.* 2005;19:819-30.

21. Triponez F, Dosseh D, Goudet P, Cougard P, Bauters C, Murat A, Cadiot G, Niccoli-Sire P, Chayvialle JA, Calender A, Proye CA. Epidemiology data on 108 MEN 1 patients from the GTE with isolated nonfunctioning tumors of the pancreas. *Ann Surg* 2006;243:265-272.

22. Thomas-Marques L, Murat A, Delemer B, Penfornis A, Cardot-Bauters C, Baudin E, Niccoli-Sire P, Levoir D, Choplin Hdu B, Chabre O, Jovenin N, Cadiot G Groupe des Tumeurs Endocrines (GTE) 2006 Prospective endoscopic ultrasonographic evaluation of the frequency of nonfunctioning pancreaticoduodenal endocrine tumors in patients with multiple endocrine neoplasia type 1. *Am J Gastroenterol* 101:266-273.

23. Goudet P, Murat A, Cardot-Bauters C, Emy P, Baudin E, du Boullay Choplin H, Chapuis Y, Kraimps JL, Sadoul JL, Tabarin A, Vergès B, Carnaille B, Niccoli-Sire P, Costa A, Calender A; GTE network (Groupe des Tumeurs Endocrines). Thymic

neuroendocrine tumors in multiple endocrine neoplasia type 1: a comparative study on 21 cases among a series of 761 NEM1 from the GTE (Groupe des Tumeurs Endocrines). *World J Surg.* 2009;33:1197-207.

24. Chandrasekharappa SC, Guru SC, Manickam P, Olufemi S-E, Collins FS, Emmert-Buck MR, Debelenko LV, Zhuang Z, Lubensky IA, Liotta LA, Crabtree JS, Wang Y, Roe BA, Weisemann J, Boguski MS, Agarwal SK, Kester MB, Kim YS, Heppner C, Dong Q, Spiegel AM, Burns AL, Marx SJ. Positional cloning of the gene for multiple endocrine neoplasia-type 1. *Science* 1997;276: 404-6.

25. Lemmens I, Van de Ven WJ, Kas K, Zhang CX, Giraud S, Wautout V, Buisson N, De Witte K, Salandre J, Lenoir G, Pugeat M, Calender A, Parente F, Quincey D, Gaudray P, De Wit MJ, Lips CJ, Hoppener JW, Khodaei S, Grant AL, Weber G, Kytola S, Teh BT, Farnebo F, Phelan C, Hayward N, Larsson C, Pannet AA, Forbes SA, Bassett JH, Thakker RV et al. Identification of the multiple endocrine neoplasia type 1 (NEM1) gene. *Hum Mol Genet.* 1997;6:1177-83.

26. Lemos MC, Thakker RV. Multiple endocrine neoplasia type 1 (NEM1): analysis of 1336 mutations reported in the first decade following identification of the gene. *Hum Mutat.* 2008;29:22-32.

27. Agarwal SK, Kester MB, Debelenko LV, Heppner C, Emmert-Buck MR, Skarulis MC, Doppman JL, Kim YS, Lubensky IA, Zhuang Z, Green JS, Guru SC, Manickam P, Olufemi SE, Liotta LA, Chandrasekharappa SC, Collins FS, Spiegel AM, Burns AL, Marx SJ. Germline mutations of the NEM1 gene in familial Multiple Endocrine Neoplasia type 1 and related states. *Hum Mol Genet.* 1997;6:1169-75.

28. Agarwal SK, Kester MB, Debelenko LV, Heppner C, Emmert-Buck MR, Skarulis MC, Doppman JL, Kim YS, Lubensky IA, Zhuang Z, Green JS, Guru SC, Manickam P, Olufemi SE, Liotta LA, Chandrasekharappa SC, Collins FS, Spiegel AM, Burns AL, Marx SJ. Germline mutations of the NEM1 gene in familial Multiple Endocrine Neoplasia type 1 and related states. *Hum Mol Genet.* 1997;6:1169-75.

29. Lips CJ, Dreijerink KM, Hoppener JW. Variable clinical expression in patients with a germline NEM1 disease gene mutation: clues to a genotype–phenotype correlation. *Clinics (Sao Paulo)* 2012;67:49-56.

30. Concolino P, Costella A, Capoluongo E. Multiple endocrine neoplasia type 1 (MEN1): An update of 208 new germline variants reported in the last nine years. *Cancer Genet.* 2016;209:36-41.
31. Thakker RV. Multiple Endocrine Neoplasia type 1. *Indian J Endocrinol Metab.* 2012; 16(2): 272-4.
32. Cavaco BM, Domingues R, Bacelar MC, Cardoso H, Barros L, Gomes L, Ruas MM, Agapito A, Garra~ o A, Pannett AA, Silva JL, Sobrinho LG, Thakker RV, Leite V Mutational analysis of Portuguese families with multiple endocrine neoplasia type 1 reveals large germline deletions. *Clin Endocrinol (Oxf)* 2002; 56:465–473
33. Raef H, Zou M, Baitei EY, Al-Rijjal RA, Kaya N, Al-Hamed M, et al. A novel deletion of the NEM1 gene in a large family of multiple endocrine neoplasia type 1 (NEM1) with aggressive phenotype. *Clinical endocrinology.* 2011 Dec;75(6):791–800.
34. Owens M, Ellard S, Vaidya B Analysis of gross deletions in the *MEN1* gene in patients with multiple endocrine neoplasia type 1 2008, 68, 350–354.
35. Dean PG, van Heerden JA, Farley DR, Thompson GB, Grant CS, Harmsen WS, Ilstrup DM 2000 Are patients with multiple endocrine neoplasia type I prone to premature death? *World J Surg* 24:1437-1441.
36. Geerdink EA, Van der Luijt RB, Lips CJ 2003 Do patients with multiple endocrine neoplasia syndrome type 1 benefit from periodical screening? *Eur J Endocrinol* 149:577-582.
37. Lairmore TC, Piersall LD, DeBenedetti MK, Dilley WG, Mutch MG, Whelan AJ, Zehnbauer B 2004 Clinical genetic testing and early surgical intervention in patients with multiple endocrine neoplasia type 1 (MEN 1). *Ann Surg.* 239:637-645.
38. Goudet P, Murat A, Biquet C, Cardot-Bauters C, Costa A, Ruzsniewski P, Niccoli P, Me´ne´gaux F, Chabrier G, Borson-Chazot F, Tabarin A, Bouchard P, Delemer B, Beckers A, Bonithon-Kopp C 2010 Risk factors and causes of death in NEM1 disease. A GTE (Groupe d’Etude des Tumeurs Endocrines) cohort study among 758 patients. *World J Surg* 34:249-255.

39. Schussheim DH; Skarulis MC; Agarwal SK; Simonds WF; Burns AL; Spiegel AM; Marx SJ. Multiple Endocrine Neoplasia Type 1: new clinical and basic finding. *Trends Endocrinol Metab.* 2001;12:173-8.
40. Bassett JH, Forbes SA, Pannett AA, Lloyd SE, Christie PT, Wooding C, Harding B, Besser GM, Edwards CR, Monson JP, Sampson J, Wass JA, Wheeler MH, Thakker RV. Characterization of mutations in patients with multiple endocrine neoplasia type 1. *Am J Hum Genet.* 1998;62:232-44.
41. Dreijerink KM, Lips CJ 2005 Diagnosis and Management of Multiple Endocrine Neoplasia Type 1 (NEM1). *Hered Cancer Clin Pract* 3:1-6
42. Schaaf L, Pickel J, Zinner K, Hering U, Höfler M, Goretzki PE, Spelsberg F, Raue F, von zur Mühlen A, Gerl H, Hensen J, Bartsch DK, Rothmund M, Schneyer U, Dralle H, Engelbach M, Karges W, Stalla GK, Höppner W 2007 Developing effective screening strategies in multiple endocrine neoplasia type 1 (MEN 1) on the basis of clinical and sequencing data of German patients with NEM1. *Exp Clin Endocrinol Diabetes* 115:509-517
43. Lourenço DM Jr, Toledo RA, Coutinho FL, Margarido LC, Siqueira SA, dos Santos MA, Montenegro FL, Machado MC, Toledo SP 2007 The impact of clinical and genetic screenings on the management of the multiple endocrine neoplasia type 1. *Clinics* (Sao Paulo) 62:465-476
44. Pieterman CR, Schreinemakers JM, Koppeschaar HP, Vriens MR, Rinkes IH, Zonnenberg BA, van der Luijt RB, Valk GD 2009 Multiple endocrine neoplasia type 1 (NEM1): its manifestations and effect of genetic screening on clinical outcome. *Clin Endocrinol* (Oxf) 70:575-581
45. Waldmann J, Fendrich V, Habbe N, Bartsch DK, Slater EP, Kann PH, Rothmund M, Langer P. Screening of patients with multiple endocrine neoplasia type 1 (NEM1): a critical analysis of its value. *World J Surg.* 2009;33:1208-18.
46. Toledo RA, Lourenço DM Jr, Coutinho FL, Quedas E, Mackowiack I, Machado MC, Montenegro F, Cunha-Neto MB, Liberman B, Pereira MA, Correa PH, Toledo SP. Novel NEM1 germline mutations in Brazilian families with multiple endocrine neoplasia type 1. *Clin Endocrinol* (Oxf). 2007;67:377-84.

47. Longuini VC. Identificação de moduladores genéticos em uma extensa família com neoplasia endócrina múltipla tipo 1 2011. 81 f. Dissertação (Mestrado) - Faculdade de Medicina da Universidade de São Paulo, São Paulo, 2011.
48. Moraes MB. Análise do status somático dos genes NEM1, AIP e p27kip1. 2012. 100 f. Dissertação (Mestrado) - Faculdade de Medicina da Universidade de São Paulo, São Paulo, 2012.
49. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 1975;94:441-8.
50. Liu L, Li Y, Li S, He Y, Pong R, Lin D, Lu L, Law M. Comparison of Next-Generation Sequencing Systems. *Journal of Biomedicine and Biotechnology* 2012: 11p
51. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science* 2003;300:286-90.
52. Marini F, Giusti F, Brandi ML. Genetic Test in Multiple Endocrine Neoplasia type 1 Syndrome: An Evolving Story. *WJEM* 2015;5(2):124-9.
53. Pigarova EA, Mamedova EO, Przhiyalkovskaya EG, Dzeranova LK, Rozhinskaya LY, Mokrysheva NG, Tiulpakov AN. Molecular Insights into MEN-1 Phenocopies. *Endocrine Society*. 2015.
54. Ku CS, Cooper DN, Iacopetta B, Roukos DH. Integrating next-generation sequencing into the diagnostic testing of inherited cancer predisposition. *Clin Genet*. 2013;83;2-6.
55. Korf BR, Rehm HL. New Approaches to Molecular Diagnosis. *JAMA* 2013;309:1511-1521.
56. Shyr D, Liu Q. Next generation sequencing in cancer research and clinical application. *Biological Procedures Online*. 2013;1:1-11.
57. Margraf RL, Durtschi JD, Stephens JE, Perez M, Voelkerding KV. Determination of RET sequence variation in an MEN2 unaffected cohort using multiple-sample pooling and next-generation sequencing. *Journal of Thyroid Research* 2012: 9p.
58. Licastro D, Mutarelli M, Peluso I, Neveling K, Wieskamp N, Rispoli R, Vozzi D, Athanasakis E, D'eustacchio A, Pizzo M, D'amico F, Ziviello C, Simonelli F, Fabretto

- A, Scheffer H, Gasparini P, Banfi S, Nigro V. Molecular Diagnosis of Usher Syndrome: Application of Two Different Next Generation Sequencing-Based Procedures. *Plos One* 2012; 7(8).
59. Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: Promises and challenges. *Cancer Letters* 2013; 340(2): 284-95
60. Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, Nord AS, Mendell JB, Swisher EM, King MC. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *PNAS* 2010; 107(28); 12629-33.
61. Vanakker OM, De Paepe A. Pharmacogenomics in Children: Advantages and Challenges of Next Generation Sequencing Applications. *International Journal of Pediatrics* 2013.
62. Barzon L, Lavezzo E, Constanzi G, Franchin E, Toppo S, Palu G. Next-generation sequencing technologies in diagnostic virology. *Journal of Clinical Virology* 2013; 58(2); 346-50.
63. Pellegata NS. MENX and NEM4. *Clinics (Sao Paulo)* 2012;67:13-18.
64. Pellegata NS, Quintanilla-Martinez L, Siggelkow H, Samson E, Bink K, Hofler H, Fend F, Graw J, Atkinson MJ. Germ-line mutations in p27Kip1 cause a multiple endocrine neoplasia syndrome in rats and humans. *PNAS* 2006;103:15558–15563.
65. Georgitsi M, Raitila A, Karhu A, van der Luijt RB, Aalfs CM, Sane T, Vierimaa O, Makinen MJ, Tuppurainen K, Paschke R, Gimm O, Koch CA, Gundogdu S, Lucassen A, Tischkowitz M, Izatt L, Aylwin S, Bano G, Hodgson S, De Menis E, Launonen V, Vahteristo P, Aaltonen LA Germline CDKN1B/p27Kip1 mutation in multiple endocrine neoplasia. *J Clin Endocrinol Metab.* 2007;92:3321–3325.
66. Agarwal SK, Mateo CM, Marx SJ. Rare germline mutations in cyclindependent kinase inhibitor genes in multiple endocrine neoplasia type 1 and related states. *J Clin Endocrinol Metab.* 2009;94:1826–34.

67. Belar O, De La Hoz C, Perez-Nanclares G, Castano L, Gaztambide S and Spanish NEM1 Group. Novel mutations in NEM1, CDKN1B and AIP genes in patients with multiple endocrine neoplasia type 1 syndrome in Spain. *Clinical Endocrinology* 2012;76:719–724.
68. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*; 2012, 13:341.
69. Nakamura K, Oshima T, Marimoto T, Ikeda S, Yoshikawa H, Shiwa Y, Linak MC, Hirai A, Takahashi H, Altaf-UI-Amin, Ogasawara N, Kanaya S. Sequence-specific error profile of Illumina sequencers. *Nucleic Acid Research*; 2011, 39(13).
70. Miller S, Dykes D, Poleski H. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16:1215.
71. Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 25:1754-60
72. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297-303.
73. Li H.*, Handsaker B.*, Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G., Durbin R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25, 2078-
74. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data *Nucleic Acids Research*, 38:e164, 2010
75. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3.", Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. *Fly (Austin)*. 2012 Apr-Jun;6(2):80-92.

76. Helga Thorvaldsdóttir, James T. Robinson, Jill P. Mesirov. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 2012.
77. James T. Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov. Integrative Genomics Viewer. Nature Biotechnology 29, 24–26 (2011)
78. García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Götz S, Tarazona S, Dopazo J, Meyer TF, Conesa A. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics* 28, no. 20, 2012: 2678-2679.
79. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 1975;94:441-8.
80. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science* 2003;300:286-90.
81. Dean PG, van Heerden JA, Farley DR, Thompson GB, Grant CS, Harmsen WS, Ilstrup DM 2000 Are patients with multiple endocrine neoplasia type I prone to premature death? *World J Surg* 24:1437-1441.
82. Geerdink EA, Van der Luijt RB, Lips CJ 2003 Do patients with multiple endocrine neoplasia syndrome type 1 benefit from periodical screening? *Eur J Endocrinol* 149:577-582.
83. Lairmore TC, Piersall LD, DeBenedetti MK, Dilley WG, Mutch MG, Whelan AJ, Zehnbauer B 2004 Clinical genetic testing and early surgical intervention in patients with multiple endocrine neoplasia type 1 (MEN 1). *Ann Surg*. 239:637-645.
84. Goudet P, Murat A, Biquet C, Cardot-Bauters C, Costa A, Ruzsniewski P, Niccoli P, Me´ne´gaux F, Chabrier G, Borson-Chazot F, Tabarin A, Bouchard P, Delemer B, Beckers A, Bonithon-Kopp C 2010 Risk factors and causes of death in NEM1 disease. A GTE (Groupe d’Etude des Tumeurs Endocrines) cohort study among 758 patients. *World J Surg* 34:249-255.

85. Schaaf L, Pickel J, Zinner K, Hering U, Höfler M, Goretzki PE, Spelsberg F, Raue F, von zur Mühlen A, Gerl H, Hensen J, Bartsch DK, Rothmund M, Schneyer U, Dralle H, Engelbach M, Karges W, Stalla GK, Höppner W 2007 Developing effective screening strategies in multiple endocrine neoplasia type 1 (MEN 1) on the basis of clinical and sequencing data of German patients with NEM1. *Exp Clin Endocrinol Diabetes* 115:509-517
86. Lourenço DM Jr, Toledo RA, Coutinho FL, Margarido LC, Siqueira SA, dos Santos MA, Montenegro FL, Machado MC, Toledo SP 2007 The impact of clinical and genetic screenings on the management of the multiple endocrine neoplasia type 1. *Clinics* (Sao Paulo) 62:465-476
87. Pieterman CR, Schreinemakers JM, Koppeschaar HP, Vriens MR, Rinkes IH, Zonnenberg BA, van der Luijt RB, Valk GD 2009 Multiple endocrine neoplasia type 1 (NEM1): its manifestations and effect of genetic screening on clinical outcome. *Clin Endocrinol* (Oxf) 70:575-581
88. Waldmann J, Fendrich V, Habbe N, Bartsch DK, Slater EP, Kann PH, Rothmund M, Langer P. Screening of patients with multiple endocrine neoplasia type 1 (NEM1): a critical analysis of its value. *World J Surg.* 2009;33:1208-18.
89. Toledo RA, Lourenço DM Jr, Coutinho FL, Quedas E, Mackowiack I, Machado MC, Montenegro F, Cunha-Neto MB, Liberman B, Pereira MA, Correa PH, Toledo SP. Novel NEM1 germline mutations in Brazilian families with multiple endocrine neoplasia type 1. *Clin Endocrinol* (Oxf). 2007;67:377-84.
90. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology.* 2012;2012:251364.
91. Ku CS, Cooper DN, Iacopetta B, Roukos DH. Integrating next-generation sequencing into the diagnostic testing of inherited cancer predisposition. *Clin Genet.* 2013;83;2-6.
92. Korf BR, Rehm HL. New Approaches to Molecular Diagnosis. *JAMA* 2013;309:1511-1521.
93. Shyr D, Liu Q. Next generation sequencing in cancer research and clinical application. *Biological procedures online.* 2013;15(1):4.

94. Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: Promises and challenges. *Cancer letters*. 2012.
95. Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, et al. Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proceedings of the National Academy of Sciences of the United States of America*. 2010;107(28):12629-33.
96. Margraf RL, Durtschi JD, Stephens JE, Perez M, Voelkerding KV. Determination of *RET* sequence variation in an MEN2 unaffected cohort using multiple-sample pooling and next-generation sequencing. *J Thyroid Res*. 2012:318232.
97. Rattenberry E, Vialard L, Yeung A, et al. A Comprehensive Next Generation Sequencing–Based Genetic Testing Strategy To Improve Diagnosis of Inherited Pheochromocytoma and Paraganglioma. *Journal Clin Endocrinol Metab*; 2013, 98(7), 1248-56.
98. Licastro D, Mutarelli M, Peluso I, Neveling K, Wieskamp N, Rispoli R, Vozzi D, Athanasakis E, D’eustacchio A, Pizzo M, D’amico F, Ziviello C, Simonelli F, Fabretto A, Scheffer H, Gasparini P, Banfi S, Nigro V. Molecular Diagnosis of Usher Syndrome: Application of Two Different Next Generation Sequencing-Based Procedures. *PLoS One*.7:e43799.
99. Manson-Bahr D, Ball R, Gundem G, Sethia K, Mills R, Rochester M, Goody V, Anderson E, O’Meara S, Flather M, Keeling M, Yazbek-Hanna M, Hurst R, Curley H, Clark J, Brewer DS, McDermott U, Cooper C. Mutation detection in formalin-fixed prostate cancer biopsies taken at the time of diagnosis using next-generation DNA sequencing. *J Clin Pathol*. 2015;68(3):212-7.
100. Pellegata NS. MENX and NEM4. *Clinics* (Sao Paulo) 2012;67:13-18.
101. Pellegata NS, Quintanilla-Martinez L, Siggelkow H, Samson E, Bink K, Hofler H, Fend F, Graw J, Atkinson MJ. Germ-line mutations in p27Kip1 cause a multiple endocrine neoplasia syndrome in rats and humans. *PNAS* 2006;103:15558–15563.
102. Nelson DL, Cox MM. Princípios de bioquímica de Lehninger. 6 ed. São Paulo: Artmed; 2015.

103. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hedge MR, Lyon E, ACMG. ACMG clinical laboratory standards for Next-Generation Sequencing. *Genet Med*; 2013, 15(9), 733-47.
104. Volkerding KV, Dames S, Durtschi JD. Next Generation Sequencing for clinical diagnostics-principles and application to targeted resequencing for hypertrophic cardiomyopathy. *Journal of molecular diagnostics*; 2010, 12(5), 539-51.
105. Lamble S, Batty E, Attar M, Buck D, Bowden R, Lunter G, Crook D, El-Fahmawi B, Piazza P. Improved workflows for high throughput library preparation using the transposome-based nextera system. *Biomed Central Biotechnology*; 2013, 13(104).
106. Mattocks CJ, Morris MA, Matthijs G, Swinnen E, Corveleyn A, Dequeker E, Müller CR, Pratt V, Wallace A. A standardized framework for the validation and verification of clinical molecular genetic tests. *European Journal of Human Genetics*; 2010, 18, 1276-88.
107. Richards S, Aziz N, Bale S, Bick D, Das S. ACMG Standards and Guidelines Standards and guidelines for the interpretation of sequence variants : a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;(January):1–20.
108. Crona J, Verdugo AD, Granberg D, Wellin S, Stalberg P, Hellman P, Björklund P. Next-generation sequencing in the clinical screening of patients with pheochromocytoma and paraganglioma. *Endocrine connection*. 2013; 2: 104-11.
109. Welander J, Andreasson A, Juhlin CC, Wiseman RW, Bäckdahl M, Hoög A, et al. Rare germline mutations identified by targeted next-generation sequencing of susceptibility genes in pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab*. 2014;99(7): 10p.
110. Zhou XH, NA Obuchowski, DK McClish (2002) Statistical methods in diagnostic medicine. New York: Wiley.
111. Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*. 2009;38(6):1767–71.
112. Li H. and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 2009; 25:1754-60

113. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009 Nov 1;25(21):2865-71.
114. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Research*, 38:e164, 2010
115. Gallego CJ, Shirts BH, Bennette CS, Guzauskas G, Amendola LM, Horike-Pyne M, et al. Next-Generation Sequencing Panels for the Diagnosis of Colorectal Cancer and Polyposis Syndromes: A Cost-Effectiveness Analysis. *J Clin Oncol* 2015
116. Chen S, Li S, Xie W, Li X, Zhang C, Jiang H, Zheng J, Pan X, Zheng H, Liu JS, Deng Y, Chen F, Jiang H. Performance comparison between rapid sequencing platforms for ultra-low coverage sequencing strategy. *PLOS one*, 2014;9(3)
117. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis C a., Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Clin Chem*. 2003;49(1):7–18.
118. Pashayan N, Lyratzopoulos G, Mathur R. Cost-effectiveness of primary offer of IVF vs. primary offer of IUI followed by IVF (for IUI failures) in couples with unexplained or mild male factor subfertility. *BMC Health Serv Res*. 2006;6:80.