

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCÊNCIAS
DEPARTAMENTO DE BIOLOGIA MOLECULAR E BIOTECNOLOGIA
TRABALHO DE CONCLUSÃO DE CURSO

PÂMELLA BORGES

**UTILIZAÇÃO DE BANCOS DE DADOS GENÔMICOS PARA INVESTIGAÇÃO
DA FREQUÊNCIA DE DOENÇAS GENÉTICAS RARAS**

Porto Alegre

2018

Pâmella Borges

**UTILIZAÇÃO DE BANCOS DE DADOS GENÔMICOS PARA INVESTIGAÇÃO
DA FREQUÊNCIA DE DOENÇAS GENÉTICAS RARAS**

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Biociências da Universidade Federal do Rio Grande do Sul (UFRGS) como requisito parcial para a obtenção do título de Bacharel em Biotecnologia.

Orientadora: Prof^a. Dr^a. Ursula Matte

Coorientadora: Prof^a. Dr^a. Gabriela Pasqualim

Porto Alegre

2018

AGRADECIMENTOS

A Professora Dra. Ursula Matte pela oportunidade de desenvolver esse projeto, por toda a confiança e orientação ao longo de toda a minha iniciação científica no Centro de Terapia Gênica, bem como por todas as contribuições para o meu crescimento intelectual e profissional.

A Professora Dra. Gabriela Pasqualim, pela coorientação, ensinamentos e confiança durante todo o período de iniciação científica e nesse trabalho.

A todos os colegas do Centro de Terapia Gênica, obrigada pela amizade, companheirismo e ensinamentos.

Aos meus amigos, pelo apoio durante o percurso, pelos conselhos, conversas, ânimo e amizade por todos esses anos.

Aos meus pais, por todos os ensinamentos ao longo da vida, pelo apoio incondicional, pelas conversas, abraços e lágrimas derramadas em conjunto.

A minha irmã, por sempre estar comigo. Por todo carinho, aventuras, arteirices e momentos que compartilhamos ao longo desses doze anos.

SUMÁRIO

LISTA DE ABREVIATURAS.....	6
RESUMO.....	7
ABSTRACT.....	8
1. INTRODUÇÃO	9
1.1 Mucopolissacaridoses.....	9
1.2 Heparan Sulfato - MPS III.....	10
1.3 Dermatan Sulfato - MPS VI	11
1.4 Dermatan e Heparan Sulfato - MPS I e II.....	11
1.5 Condroitin e Queratan sulfato - MPS IV.....	12
1.6 Dermatan, Heparan e Condroitin sulfato - MPS VII.....	13
1.7 Hialuronana - MPS IX	13
2. DIAGNÓSTICO	13
3. MANEJO CLÍNICO	15
4. FREQUÊNCIAS	16
5. BANCOS DE DADOS GENÉTICOS.....	17
5.1 Variantes de Significado Incerto	18
5.2 Ferramentas <i>In Silico</i>	18
6. JUSTIFICATIVA.....	19
7. OBJETIVOS.....	19
7.1 Objetivo Geral.....	19
7.2 Objetivo Específico	19
ARTIGO	20

DISCUSSÃO	32
CONCLUSÃO	35
REFERÊNCIAS BIBLIOGRÁFICAS.....	36

LISTA DE ABREVIATURAS

CS: condroitin sulfato

DS: dermatan sulfato

GAG: Glicosaminoglicanos

HA: hialuronana

HS: heparan sulfato

MPS: Mucopolissacaridose

MPS I: Mucopolissacaridose I

MPS II: Mucopolissacaridose II

MPS III: Mucopolissacaridose III

MPS IV: Mucopolissacaridose IV

MPS VI: Mucopolissacaridose VI

MPS VII: Mucopolissacaridose VII

MPS IX: Mucopolissacaridose IX

NGS: *Next-Generation Sequencing*

QS: queratan sulfato

TCTH: Transplante de células-tronco hematopoiéticas

TRE: Terapia de reposição enzimática

VUS: *Variants of Uncertain Significance*

WES: *Whole-Exome Sequencing*

WGS: *Whole-Genome Sequencing*

RESUMO

As doenças lisossômicas são um grupo de doenças raras caracterizadas pelo acúmulo de substâncias no lisossomo. As doenças causadas pela acumulação de um ou mais tipos de glicosaminoglicanos são chamadas de mucopolissacaridoses (MPS). As MPS são multissistêmicas, crônicas e progressivas, com amplos espectros de manifestações clínicas, comprometendo desde funções somáticas até funções neurológicas. Não existe cura para as MPS, mas existem tratamentos que melhoram a qualidade de vida dos pacientes. Por serem raras, dados relativos à frequência são difíceis de obter, existindo dados de apenas alguns países e regiões. Os avanços nas tecnologias de sequenciamento possibilitaram a criação de bancos de dados populacionais. Esses bancos de dados oferecem informação valiosa sobre os padrões genéticos e podem ser usados para melhorar o entendimento da relação genótipo/fenótipo nas doenças. O grande número de dados gerados nesses sequenciamentos estimulou a criação de novas ferramentas que possibilitem a análise *in silico* dessas informações. No presente trabalho, as frequências das MPS foram calculadas a partir de dois bancos de dados populacionais disponíveis *online*: gnomAD e ExAC. Analisaram-se 16656 variantes, sendo 2005 de MPS I, 961 de MPS II, 2043 de MPS IIIA, 1048 de MPS IIIB, 1533 de MPS IIIC, 1141 de MPS IIID, 2988 de MPS IVA, 1629 de MPS IVB, 1159 de MPS VI, 1067 de MPS VII e 1082 de MPS IX. Essas variantes foram separadas por tipos, excluindo-se aquelas que foram encontradas em posições menos propensas a serem patológicas. Em seguida, todas as variantes presentes em homozigose foram retiradas, já que nenhum dos bancos de dados incluem amostras de pacientes de MPS ou outras doenças pediátricas graves. As variantes que permaneceram foram separadas para análises. Variantes de ponto na região codificante (*stop gain*, *stop loss*, *start loss*) foram classificadas como patogênicas; as de alteração de sítio de *splice* foram analisadas pelo *Human Splicing Finder*; as de *frameshift* e as deleções e inserções *in-frame* foram analisadas pelo *SIFT Indel*; as *missense* foram analisadas com cinco preditores e o consenso de no mínimo três programas foi considerado. Com os resultados dessas análises, a frequência alélica das variantes consideradas patogênicas foi usada para calcular a frequência das doenças utilizando a equação de Hardy-Weinberg. As frequências encontradas foram, para MPS I, MPS IIIA, B, C e D, MPS IVA e B, MPS VI, MPS VII e MPS IX, respectivamente: 1:18.289,86; 1:58.974,05; 1:101.350,40; 1:68.947,25; 1:263.641,49; 1:53.397,01; 1:69.648,95; 1:117.071,68; 1:137.218,05; 1:201.123,12. A frequência estimada para MPS II é incompatível com o observado na clínica. Porém, as frequências obtidas para as demais MPS seguiram o padrão esperado de apresentarem frequências maiores do que as apresentadas na literatura.

ABSTRACT

Lysosome diseases are a group of rare diseases characterized by an accumulation of substances in the lysosome. Diseases caused by the accumulation of one or more types of glycosaminoglycans are called mucopolysaccharidoses (MPS). The MPS are multisystem, chronic and progressive, with broad spectra of clinical manifestations, compromising from somatic to neurological functions. There is no cure for MPS, but there are treatments that improve patients' quality of life. Because they are rare, it is difficult to obtain frequency data, with data from only a few countries and regions. Advances in sequencing technologies enabled the creation of population databases. These databases provide valuable information about genetic patterns and can be used to improve understanding of the genotype/phenotype relationship in diseases. A large number of data generated in these sequencing stimulated the creation of new tools that allow in silico analysis of this information. In the present work, the frequencies of MPS were calculated from two population databases available online: gnomAD and ExAC. 16656 variants were analyzed, 2005 of MPS I, 961 of MPS II, 2043 of MPS IIIA, 1048 of MPS IIIB, 1533 of MPS IIIC, 1141 of MPS IIID, 2988 of MPS IVA, 1629 of MPS IVB, 1159 of MPS VI, 1067 MPS VII and 1082 MPS IX. The variants found in these databases were separated by types, excluding those that were found in positions less likely to be pathological. After that, we removed all variants that were homozygous, on the basis that both databases did not include MPS patients in their databases. The variants that remained were separated for analysis. Variant point in the coding region (stop gain, stop loss, start loss) were classified as pathogenic; the splice site alterations were analyzed by the Human Splicing Finder; those of frameshift and in-frame deletions and insertions were analyzed by SIFT Indel; the missense were analyzed with five predictors and the consensus of at least three programs was considered. With the results of these analyzes, the allelic frequency of variants considered pathogenic was used to calculate the frequency of diseases using the Hardy-Weinberg equation. The frequencies found were, for MPS I, MPS IIIA, B, C and D, MPS IVA and B, MPS VI, MPS VII and MPS IX, respectively: 1: 18,289.86; 1: 58,974.05; 1: 101,350,40; 1: 68,947.25; 1: 263,641.49; 1: 53,397.01; 1: 69,648.95; 1: 117,071.68; 1: 137,218.05; 1: 201,123.12. The estimated frequency for MPS II is incompatible with that observed in the clinic. However, the frequencies obtained for the other MPS followed the expected pattern of presenting higher frequencies than those presented in the literature.

1. Introdução

1.1 Mucopolissacaridoses

As mucopolissacaridoses (MPS) são um conjunto de sete doenças raras (Tabela 1) caracterizadas pelo acúmulo intralissossomal de glicosaminoglicanos (GAGs). Esse acúmulo anormal ocorre pela degradação parcial ou ausente desses GAGs causada pela deficiência na atividade de alguma das enzimas lisossômicas envolvidas.

Tabela 1: Características gerais das MPS. Adaptado de Neufeld e Muenzer, 2001.

Tipo	Epônimo	Sigla	Gene afetado	Localização Gênica	Enzima deficiente	GAG acumulado
I	Hurler	MPS IH	<i>IDUA</i>	4p16.3	α -L-iduronidase	HS,DS
	Hurler/Scheie	MPS IH/S				
	Sheie	MPS IS				
II	Hunter	MPS II	<i>IDS</i>	Xq28	Iduronato 2-sulfatase	HS,DS
III	Sanfilippo A	MPS IIIA	<i>SCSH</i>	17q25.3	N-Sulfoglicosamina Sulfohidrolase (Sulfaminidase)	HS
	Sanfilippo B	MPS IIIB	<i>NAGLU</i>	17q21.2	α -N-Acetilglicosaminidase	
	Sanfilippo C	MPS IIIC	<i>HGSNAT</i>	14p21	AcetilCoA-alfa-glucosamina acetiltransferase	
	Sanfilippo D	MPS IIID	<i>GNS</i>	12q14	N-Acetilglicosamina-6-sulfatase	
IV	Morquio A	MPS IVA	<i>GALNS</i>	16q24.3	N-acetilgalactosamina-6-sulfato sulfatase	QS,CS
	Morquio B	MPS IVB	<i>GLB1</i>	3p21.3	β -galactosidase	QS
VI	Maroteaux-Lamy	MPS VI	<i>ARSB</i>	5q11-q13	N-Acetilgalactosamina 4-sulfatase (Arilsulfatase B)	DS
VII	Sly	MPS VII	<i>GUSB</i>	7q21.11	β -glicuronidase	HS, DS, CS
IX	Natowicz	MPS IX	<i>HYAL1</i>	3p21.3	Hialuronidase 1	HA

*Abreviações: MPS: mucopolissacaridoses; HS: Heparan sulfato; DS: Dermatan sulfato; QS: Queratan sulfato; CS: Condroitin sulfato; HA: Hialuronana.

O desenvolvimento das MPS é crônico e progressivo, apresentando um amplo espectro de variações fenotípicas em todos os tipos. Essas manifestações apresentam formas mais moderadas e mais graves de acordo

com a atividade residual enzimática e o GAG acumulado. As manifestações clínicas mais recorrentes dessas doenças incluem face grosseira, contraturas articulares, rigidez articular, turvação da córnea, infecções recorrentes no nariz e ouvido e manutenção da inteligência e estatura. Os fenótipos mais graves, além das manifestações anteriores, também podem apresentar deformidades ósseas, hepatoesplenomegalia, baixa estatura, deterioração mental progressiva, disostose múltipla, surdez, cardiopatias, apneia do sono e problemas de comportamento (GIUGLIANI, 2012; NEUFELD; MUENZER, 2001).

Os GAGs, cadeias polissacarídicas não-ramificadas compostas de unidades dissacarídicas não repetitivas, são constituintes da matriz extracelular de diversos tecidos, servindo como suporte. A acumulação dos GAGs leva a um comprometimento da função celular e orgânica em diversas células, tecidos e órgãos, levando a essas manifestações multissistêmicas (MEIKLE et al., 1999). Os órgãos e os sistemas são afetados de acordo com o tipo e a localização celular da acumulação do GAGs. Assim, as MPS são classificadas e agrupadas de acordo com o tipo ou os tipos de GAGs acumulados e com a deficiência enzimática envolvida.

1.2 Heparan Sulfato - MPS III

O heparan sulfato (HS) é um GAG com carga negativa ligado covalentemente a proteínas de superfície celular e da matriz extracelular (DONCEL-PÉREZ et al., 2018). Quando acumulado, compromete principalmente o sistema neurológico por inibir as demais enzimas lisossomais (FEDELE, 2015). A Síndrome de Sanfilippo (MPS III) é o único tipo desse conjunto de doenças que acumula apenas HS e apresenta, nos quadros clínicos dos pacientes, grave degeneração do sistema nervoso central (NEUFELD; MUENZER, 2001).

A MPS III tem quatro subtipos bem definidos com diferenças enzimáticas (Tabela 1). O início dos sintomas ocorre normalmente entre os dois e seis anos de idade, com manifestações neurológicas e quase nenhuma alteração esquelética. O rápido declínio das habilidades sociais, grande retardo mental, hiperatividade e graves problemas de comportamento são manifestações

frequentes. A deterioração do sistema nervoso continua rapidamente, levando os pacientes ao estado vegetativo até a morte, que ocorre por volta dos dez anos nos casos mais graves e entre os dezesseis nos casos mais atenuados (COUTINHO; LACERDA; ALVES, 2012; FEDELE, 2015; NEUFELD; MUENZER, 2001).

1.3 Dermatan Sulfato - MPS VI

O dermatan sulfato (DS) é um GAG constituinte dos tecidos conjuntivos e, quando acumulado, causa deformidades esqueléticas, anormalidades cardíacas e articulares (COUTINHO; LACERDA; ALVES, 2012; MUENZER, 2011; NEUFELD; MUENZER, 2001). A Síndrome de Maroteaux-Lamy (MPS VI) é a única patologia, dentre as MPS, a acumular apenas DS. Os fenótipos mais graves da MPS VI são normalmente diagnosticados antes dos três anos devido as alterações esqueléticas, com comprometimento e prejuízos de locomoção por volta dos dez e sobrevive até a segunda década de vida. Já os fenótipos mais brandos tem um retardo no aparecimento dos sintomas, que normalmente são reconhecidos na adolescência e início da idade adulta, apresentando uma expectativa de vida quase normal (GIUGLIANI et al., 2010; NEUFELD; MUENZER, 2001; VAIRO et al., 2015).

1.4 Dermatan e Heparan Sulfato - MPS I e II

Algumas MPS apresentam acumulação de mais de um tipo de GAG. As MPS I e MPS II acumulam tanto DS, quanto HS. Devido a essa característica, essas MPS apresentam grande variação fenotípica, com fenótipos extremos bem definidos e uma gama de fenótipos intermediários. Os pacientes com os fenótipos atenuados costumam apresentar quadros clínicos mais voltados para comprometimentos somáticos, com pouco ou nenhum comprometimento neurológico (BALDO, 2012; PINTO, 2005; YOUNG et al., 1982). Os fenótipos mais graves tem grande comprometimento neurológico que os levam a um rápido declínio mental e comportamental, bem como graves comprometimentos somáticos e esqueléticos. Os fenótipos não são distinguíveis bioquimicamente e, por isso, são agrupados de acordo com a evolução clínica apresentada pelos pacientes (NEUFELD; MUENZER, 2001).

Esse grande espectro de manifestações clínicas fez com que, inicialmente, alguns casos atenuados de MPS I fossem classificados como uma nova MPS classificada como tipo V. Entretanto, à medida que os casos foram sendo estudados, descobriu-se a confusão (COUTINHO; LACERDA; ALVES, 2012). Com isso, a nomenclatura MPS V deixou de ser usada e hoje a MPS I é dividida em três epônimos: Hurler que é a forma com fenótipos mais grave da doença e a mais comumente encontrada na clínica, chegando a representar 80% dos casos (MUENZER et al., 2009); Hurler-Scheie que é considerada a forma moderada da doença, com aparecimento dos sintomas normalmente entre os três e oito anos, podendo chegar a 20 ou 30 anos de vida; Scheie, a forma atenuada, onde o diagnóstico ocorre normalmente entre os dez e vinte anos, com sobrevida quase normal (GIUGLIANI et al., 2010; MUENZER, 2011; NEUFELD; MUENZER, 2001; SCOTT et al., 1995).

A MPS II, também chamada de Síndrome de Hunter, é o único tipo desse grupo de doenças com herança ligada ao cromossomo X. É uma das MPS mais frequentemente diagnosticadas no Brasil (GIUGLIANI et al., 2017; NEUFELD; MUENZER, 2001; PINTO, 2005). O critério fundamental para a divisão dos pacientes entre a forma mais grave e a mais atenuada do espectro é a regressão neurológica (YOUNG et al., 1982).

1.5 Condroitin e Queratan sulfato - MPS IV

O condroitin e queratan sulfato (CS, QS) são importantes para formar associações e apresentar os perfis osmóticos necessários para formar a resistência das cartilagens e córneas (HORKAY et al., 2017; VALAYANNOPOULOS; WIJBURG, 2011). A Síndrome de Morquio (MPS IV) apresenta dois subtipos (A e B, Tabela 1) classificados pelas deficiências enzimáticas (MUENZER, 2011). Os pacientes apresentam inteligência preservada, baixa estatura e uma displasia esquelética distinta das demais MPS (GUARANY, 2011). A MPS IVA foi inicialmente considerada a forma grave desse tipo de MPS. Entretanto, o surgimento de casos de MPS IVB com quadros clínicos tão severos quanto MPS IVA acabou com essa divisão. O início dos sintomas, diagnóstico e expectativa de vida são variados. A característica clínica mais grave apresentada pelos pacientes é a hipoplasia da odontoide e desenvolvimento de mielopatia cervical precoce. Nesses casos, os

pacientes não chegam à segunda ou terceira década de vida (NEUFELD; MUENZER, 2001).

1.6 Dermatan, Heparan e Condroitin sulfato - MPS VII

A Síndrome de Sly (MPS VII) é a única com acúmulo de três GAGs: condroitin sulfato (CS), HS e DS. Essa combinação faz com que a MPS VII seja uma das formas mais raras do grupo (KHAN et al., 2017; NEUFELD; MUENZER, 2001). Na forma mais grave da doença, os pacientes desenvolvem hidropsia fetal e morrem durante a gestação ou ao nascimento, enquanto as formas mais atenuadas têm sobrevida até o início da fase adulta (BALDO, 2012; NEUFELD; MUENZER, 2001).

1.7 Hialuronana - MPS IX

Hialuronana (HA) é um polímero de cadeia linear, encontrado em maiores concentrações em tecidos conjuntivos moles, estimando-se que 50% da HA esteja na pele, 25% no esqueleto e nas articulações, e a maior parte restante esteja no cérebro, rim, pulmão e músculos (TRIGGS-RAINE; NATOWICZ, 2015). Devido à localização do GAG, seu acúmulo leva a uma alteração articular parecida com artrite. Dessa forma, é compreensível que a maioria dos pacientes relatados com deficiência de enzima que cliva a HA (hialuronidase 1 ou *HYAL 1*) tenham sido inicialmente diagnosticados com artrite juvenil idiopática (IMUNDO et al., 2011). Posteriormente essa deficiência de *HYAL 1* foi identificada e classificada como MPS IX ou síndrome de Natowicz. A MPS IX pode ser definida como a mais rara desordem desse tipo, com apenas quatro casos relatados, sendo três pertencentes a uma família consanguínea de origem da Arábia Saudita. Alternativamente, pode ser vista a MPS mais leve, com muitos casos não diagnosticados como MPS IX (KIYKIM et al., 2016; TRIGGS-RAINE; NATOWICZ, 2015).

2. Diagnóstico

A partir da suspeita clínica da doença, o diagnóstico das MPS pode ser estabelecido com a realização de exames bioquímicos específicos e/ou análises moleculares.

Quando não degradados, os GAGs são excretados em quantidades aumentadas na urina. Essa característica permitiu o desenvolvimento dos primeiros testes bioquímicos de diagnóstico da doença, realizado pela quantificação de GAGs urinários (PENNOCK, 1976). Esses testes servem como uma triagem inicial para o diagnóstico e para orientar a direção dos demais exames devido ao grande número de falsos negativos e positivos apresentados (MUENZER, 2011; VAIRO et al., 2015).

A medida da atividade da enzima específica é o diagnóstico bioquímico definitivo do tipo de MPS. O ensaio enzimático pode ser realizado em plasma, leucócitos ou fibroblastos (MUENZER, 2011). A análise molecular é realizada de forma complementar ao diagnóstico bioquímico, auxiliando na predição do fenótipo e no aconselhamento genético. Entretanto, o fluxo de diagnóstico vem se alterando à medida que se obtém maior conhecimento sobre a associação das características moleculares da doença e os custos dos análises decrescem.

As técnicas de sequenciamento de Sanger e de sequenciamento maciço paralelo, mais conhecido como Sequenciamento de Nova Geração (*Next-Generation Sequencing*, NGS) representaram uma mudança de paradigma no estudo de variantes do genoma humano. Reconhece-se hoje a existência de uma grande relação da variação do espectro fenotípico com a heterogeneidade dos defeitos gênicos, como deleções, rearranjos e inúmeras mutações. Entretanto, devido ao aumento na quantidade de dados gerados, cada vez mais aparecem variantes com significado incerto e, assim, sem relação conhecida com o fenótipo. Observa-se também que mutações *missense*, podem causar fenótipos variados (desde grave – se atingirem o sítio ativo – até sem impacto - se ocorrerem em regiões não críticas para o funcionamento da proteína) enquanto mutações *nonsense* e grandes rearranjos, em geral, manifestam-se de forma mais grave (GIUGLIANI, 2012; TERLATO; COX, 2003). Assim, o NGS tornou-se a ferramenta de escolha para estudos de genética humana (GOLDSTEIN et al., 2013).

3. Manejo Clínico

Até o momento, as MPS não possuem cura. Entretanto, existem tratamentos disponíveis que melhoraram a qualidade de vida dos pacientes e retardam o aparecimento de alguns sintomas. Cada tratamento tem suas recomendações, sendo específicas para cada tipo e grau de MPS (BALDO, 2012; HOLLAK; WIJBURG, 2014; PASQUALIM, 2013; VALAYANNOPOULOS; WIJBURG, 2011). Entre as alternativas terapêuticas disponíveis está o transplante de células-tronco hematopoiéticas (TCTH), recomendado para MPS que tenham comprometimento neurológico. Pode ser realizados em crianças de até os dois anos de idade que não demonstrem grave comprometimento cognitivo (DE RU et al., 2011; MUENZER et al., 2009). Entretanto os resultados são limitados e dependem do tipo de MPS (VALAYANNOPOULOS; WIJBURG, 2011).

A terapia de reposição enzimática (TRE) é um tratamento onde é administrada a enzima específica deficiente no paciente, de forma periódica. A TRE não é indicada para tratamento dos sintomas neurológicos, pois a dose recomendada e a forma como é administrada não permitem que a enzima atravesse a barreira hematoencefálica (GIUGLIANI et al., 2010; MUENZER, 2011; MUENZER et al., 2009). Existe a formação de anticorpos em 90% dos pacientes, o que pode levar à diminuição da eficácia da terapia (BALDO, 2012; PASQUALIM, 2013; VALAYANNOPOULOS; WIJBURG, 2011). Uma importante questão da TRE é o custo dessa terapia, podendo custar por um ano de tratamento meio milhão de reais por paciente, obrigando os pacientes e seus familiares a pedir judicialmente o pagamento do tratamento.

As MPS I, MPS II e MPS VI tiveram enzimas aprovadas para tratamento por TRE em 2003, 2006 e 2005 respectivamente. Os resultados apresentados vêm se mostrando conflitantes em relação a real eficiência dessa terapia (VALAYANNOPOULOS; WIJBURG, 2011). Apesar de benefícios terem sido notados, como diminuição nos níveis dos GAGs urinários, aumento de sobrevida dos pacientes, melhoras nas funções respiratórias e motoras, as diferentes idades de início do tratamento e os diferentes fenótipos das MPS geraram níveis variáveis de melhora dos sintomas clínicos. Essa inconsistência nas melhoras por TRE estimulam o desenvolvimento de novas terapias (BALDO, 2012; NOH; LEE, 2014).

Diversas novas técnicas e terapias estão sendo desenvolvidas, principalmente para tratar manifestações neurológicas. As terapias gênicas, terapias de redução de substrato, bem como a combinação das terapias vêm sendo cada vez mais estudadas e apresentam grande potencial para melhorar a qualidade de vida os pacientes (HOLLAK; WIJBURG, 2014; NOH; LEE, 2014; SAWAMOTO et al., 2018). O consenso entre os estudos é a necessidade do início precoce do tratamento, em qualquer modalidade, para garantir a melhor qualidade de vida aos pacientes e a resposta terapêutica. Dessa forma, é importante conhecermos a incidência das MPS para ajudarmos na melhora e no direcionamento das estratégias terapêuticas e do diagnóstico.

4. Frequências

Há poucos dados disponíveis sobre a frequência destas doenças. Por se tratar de um grupo de doenças raras com grande heterogeneidade clínica, acredita-se que a incidência real seja maior que a estimada, principalmente pela dificuldade de identificação dos pacientes com formas mais atenuadas (BAEHNER et al., 2005; GIUGLIANI, 2012). Além da dificuldade do diagnóstico, a falta de obrigatoriedade de registro pelos sistemas de saúde, como no Brasil (GUARANY, 2011), altera consideravelmente os dados de prevalência. Além dessas dificuldades, as variações genéticas entre as populações também contribuem para aumentar a diversidade nos dados de frequência entre os diferentes países no mundo.

Os primeiros estudos estimavam que a incidência em conjunto das MPS fosse aproximadamente 1: 22.000 indivíduos (PJ et al., 1999; POORTHUIS et al., 1999). Contudo, a medida que mais estudos foram realizados, a dificuldade de chegarmos a uma frequência comum aumentou. Moammar e colaboradores (2010) demonstraram que a incidência das MPS em países com altas taxas de consanguinidade, como a Arábia Saudita, chega a 17:100.000. Já um estudo publicado mais recentemente mostrou que a incidência nos Estados Unidos é de 1,1:100.000 (PUCKETT et al., 2017). Esses estudos mostram que os valores de incidência das MPS variam de acordo com a localização que está sendo estudada.

5. Bancos de Dados Genéticos

O desenvolvimento de tecnologias como o sequenciamento completo de exoma (*Whole-Exome Sequencing*, WES) e o sequenciamento completo de genoma (*Whole-Genome Sequencing*, WGS) causou uma mudança no fluxo de informações genéticas disponíveis (PABINGER et al., 2014). O grande número de informações geradas por essas técnicas permitiu a criação de diferentes consórcios que disponibilizam os resultados de análises genômicas em indivíduos saudáveis ou, eventualmente, com patologias definidas. A disponibilidade desses dados ao público possibilita avanços no conhecimento e associação das variantes moleculares as doenças.

Os bancos de mutações contêm variantes descritas em pacientes e se baseiam nas interpretações dos pesquisadores sobre a relação dessas variantes com seus significados clínicos. Esses dados ficam disponíveis em repositórios que podem ser total ou parcialmente pagos como HGMD (*Human Gene Mutation Database*), que possui versão profissional paga (STENSON et al., 2017); ou gratuitos como ClinVar (LANDRUM et al., 2016). Esses bancos têm grande importância clínica, pois ajudam a centralizar os conhecimentos, facilitando o diagnóstico molecular.

Os bancos populacionais reúnem todas as informações geradas por WES ou WGS. Um destes bancos é o *Exome Aggregation Consortium* (ExAC), que reúne dados de exomas de 60.706 indivíduos não relacionados, participantes de projetos de pesquisa de genética de populações e de doenças multifatoriais (LEK et al., 2016). A segunda versão desse banco é o *Genome Aggregation Database* (gnomAD), abrangendo 123.136 exomas e 15.496 genomas de indivíduos não relacionados sequenciados como parte de vários estudos genéticos de doenças e populações específicas. Um fato importante a ser ressaltado é que tanto o ExAC como o gnomAD excluem dados de pacientes com doenças pediátricas graves ou seus familiares. Outro banco de dados importante é o derivado do Projeto 1000 Genomas, que congrega dados de 2504 indivíduos e reporta alterações com frequência maior do que 1% (SUDMANT et al., 2015). Esses bancos possuem dados de diferentes populações. Já o ABraOM (Arquivo Brasileiro Online de Mutações) contém dados de exomas de 609 indivíduos idosos de São Paulo e é o único banco deste tipo disponível para a população brasileira (NASLAVSKY et al., 2017).

5.1 Variantes de Significado Incerto

As VUS (*Variants of Uncertain Significance*) são variantes que apresentam ou discordância ou falta de conhecimento quanto a sua relação com a doença a que é relacionada. A incerteza da relação dessas variantes com a patologia gera grandes problemas na clínica em questões de diagnóstico. Com a popularização das técnicas de NGS, cada vez mais são encontradas variantes que não tem uma relação bem estabelecida com a doença e acabam sendo classificadas como VUS.

5.2 Ferramentas *In Silico*

O grande número de variantes encontradas nos estudos de WES, cerca de 30.000 por indivíduo analisado, estimulou a criação de ferramentas e estratégias específicas para lidar com a priorização de variantes patogênicas (ALEMÁN et al., 2014; JIANG, 2015). Geralmente, os métodos heurísticos de priorização de variantes baseiam-se na suposição de que a variante patogênica altera a sequência codificante da proteína, é rara e possui penetrância completa (STITZIEL; KIEZUN; SUNYAEV, 2011). A abordagem heurística, associada a anotação funcional é uma das principais estratégias utilizadas, sendo capaz de excluir de 50-75% das variantes identificadas (FU et al., 2013; KATSONIS et al., 2014; LI et al., 2013; OHANIAN; OTWAY; FATKIN, 2012).

Em 2015, o *American College of Medical Genetics* e a *Association for Molecular Pathology* publicaram em conjunto diretrizes para a interpretação de variantes genéticas de doenças monogênicas em laboratório clínico, a qual lista uma série de evidências a serem analisadas a fim de se estabelecer o grau de patogenicidade de cada variante (RICHARDS et al., 2015). Dentre essas evidências, encontram-se os programas de predição computacional (*in silico*), que podem auxiliar na interpretação das variantes (PABINGER et al., 2014). Existem diversas ferramentas de predição *in silico* que avaliam o efeito da troca de aminoácidos até alterações de *frameshift* e *indels* (HU; NG, 2012, 2013), alterações de sítio de *splicing* (DESMET et al., 2009; JIAN; BOERWINKLE; LIU, 2014) e alterações sinônimas (BUSKE et al., 2013). Estudos que compararam o desempenho dos preditores e avaliaram sua

habilidade de prever variantes sabidamente patogênicas demonstraram que a acurácia varia entre 65 e 80% (HICKS et al., 2011; THOMPSON et al., 2013).

Aliar a outras informações; como frequência da mutação e ocorrência em homozigose, os preditores e as informações obtidas a partir de bancos de dados populacionais pode auxiliar a determinar a frequência de alterações patogênicas em doenças monogênicas raras (APPADURAI et al., 2015). Essas ferramentas apresentam, então, um grande potencial por facilitar e agilizar a análise de uma grande quantidade de variantes com um bom grau de confiabilidade. Utilizar essas ferramentas nos perfis genéticos populacionais pode contribuir para a melhor fidelidade dos dados de incidência, minimizando o subdiagnóstico das MPS.

6. Justificativa

Estudos sobre a incidência e a prevalência de MPS são de interesse de geneticistas, autoridades da saúde, bem como para os pacientes e suas famílias. Os estudos epidemiológicos são pré-requisitos para cálculos de custo-benefício quando novas e dispendiosas terapias são desenvolvidas. Esses estudos apresentam dificuldades em serem elaborados por ser um grupo de doenças raras e diversos casos não serem diagnosticados pelos sintomas passarem despercebidos, além do diagnóstico exigir métodos laboratoriais não amplamente disponíveis. O uso de dados populacionais que possibilitem a análise de mutações patogênicas nos genes responsáveis por essas doenças pode contribuir para uma estimativa da sua frequência em diferentes regiões do mundo.

7. Objetivos

7.1 Objetivo Geral

Estimar a frequência das diferentes MPS a partir da frequência de variantes consideradas patogênicas, pelo tipo de mutação ou pela análise *in silico*, obtidas em bancos populacionais.

7.2 Objetivo Específico

- 1) Estimar a frequência relativa dos diferentes tipos de MPS no mundo.

ARTIGO

Os materiais e métodos e resultados deste trabalho são apresentados sob a forma de um artigo científico que está em fase de formatação para o envio para *Molecular Genetics and Metabolism*.

Estimating the frequency of Mucopolysaccharidosis from the analysis of genome and exome databases

Matte U, Borges P, Pasqualim G, Leão D, Giugliani R, Vairo F.

Abstract

In this study, the frequency of different types of Mucopolysaccharidoses (MPS) was estimated with the frequency of disease-causing variants. Data from the Exome Aggregation Consortium (ExAC) and The Genome Aggregation Database (gnomAD) were used to identify the potentially disease-causing variants in each gene related to MPS I to IX (but not for the X-linked MPS II). Variants present in homozygosity were excluded. Frameshifts and in-frame insertions or deletions were evaluated with SIFT Indel. Splice related variants were evaluated with Human Splice Finder 3.0 (HSF). Point mutations in the coding region were classified as potentially pathogenic based on type (stop gain, stop loss, start loss) whereas synonymous variants were deemed nonpathogenic. Missense variants were analyzed by *in silico* prediction tools and only those predicted to be damaging in at least 3 out of 5 predictors were considered disease-causing. Also, variants described as pathogenic in HGMD were included regardless of their *in silico* score. Intronic variants, as well as 5' and 3' UTR variants, were excluded. In the end, the frequency of 117 (in *GNS*) to 224 (in *IDUA*) variants were used to calculate disease incidence based on Hardy-Weinberg's equation. The estimated incidence for MPS I was 1:18,289.86; MPS IIIA 1:58,974.05; MPS IIIB 1:101,350.40; MPS IIIC 1:68,947.25; MPS IIID 1:263,641.49; MPS IVA 1:53,397.01; MPS IVB 1:69,648.95; MPS VI 1:117,071.68; MPS VII 1:137,218.05; and MPS IX 1:201,123.12. The estimated incidence for all types of MPS was higher than predicted in the literature. This difference may be due to an underdiagnosis especially of the extreme forms of MPS. An overestimation of the number of disease-causing variants by *in silico* predictors cannot be ruled out. In spite of that, the relative frequencies are similar to what has been detected in diagnosis-based incidence studies.

Introduction

The Mucopolysaccharidosis (MPS) are a group of lysosomal storage diseases characterized by the deficiency of one out of eleven lysosomal enzymes involved in glycosaminoglycans (GAG) degradation. The MPS can be divided into different types according to the deficiency enzyme and accumulated substrate (MPS I, MPS II, MPS III A, B, C, D, MPS IV A, B, MPS VI, MPS VII, and MPS IX). GAGs are constituents of the extracellular matrix and impaired turn over leads to a series of consequences both at the cellular and the physiological levels. Patients usually present coarse features, heart and pulmonary problems, and, depending on the accumulated substrate more pronounced or concomitant bone disorders (dysostosis multiplex) and CNS impairment (behavioral problems or mental retardation). Disease severity is variable and the outcome was greatly improved in recent years for those diseases for which enzyme

replacement therapy has been developed. Even though these are systemic progressive disorders, therapeutic approaches in the form of hematopoietic stem cell transplantation, enzyme replacement therapy or gene therapy are available or under development.

Extensive molecular characterization has been done for the genes encoding the defective enzyme in each of these diseases and both biochemical and molecular diagnosis are available in specialized centers. In spite of that, data on the epidemiology of these diseases are difficult to obtain, as few countries have comprehensive newborn screening programs mainly due to the need of expensive equipment, such as tandem MS, or the lack of treatment for some of these diseases.

In this work, the frequency of disease-causing variants present in population databases was used to estimate the frequency of different types of MPS using Hardy-Weinberg's equation. Data from the Exome Aggregation Consortium (ExAC) and The Genome Aggregation Database (gnomAD), both from the Broad Institute (Lek et al., 2016), were used to identify the potentially disease-causing variants in each gene related to MPS I to IX. Since gnomAD and ExAC give a direct number of alleles (and not subjects), X-linked MPS II could also be calculated. However, due to the excessive low number of retained variants, results for MPS II were not presented.

Methods

Data

Variants from ExAC and gnomAD (Lek et al., 2016) were downloaded separately and merged in order to create a unique database. Table 1 shows the number of variants present in each database and after merging, for each gene analyzed. In the case of variants common to both databases, allele frequencies deposited in gnomAD were used for further analysis, as this database is larger than ExAC. Retained variants were further submitted to a tiered selection.

Table 1. The number of variants for each MPS gene present in ExAC and gnomAD databases. Retained variants represent unique variants after the merge of both databases.

MPS Type	Gene	ExAC variants	gnomAD variants	Common to both	Retained variants
MPS I	<i>IDUA</i>	1246	1439	680	2005
MPS II	<i>IDS</i>	300	920	259	961
MPS III A	<i>SGSH</i>	1188	1400	545	2043
MPS III B	<i>NAGLU</i>	640	805	397	1048
MPS III C	<i>HGSNAT</i>	598	1456	521	1533
MPS III D	<i>GNS</i>	429	1116	404	1141
MPS IV A	<i>GALNS</i>	1390	2254	656	2988
MPS IV B	<i>GLB1</i>	871	1322	564	1629
MPS VI	<i>ARSB</i>	407	1122	370	1159
MPS VII	<i>GUSB</i>	593	891	417	1067
MPS IX	<i>HYAL1</i>	669	700	287	1082

First tier variant selection

Variants located in 5' and 3' UTR, upstream and downstream gene as well as intronic and in non-coding transcript exons, were excluded assuming that no disease-causing mutation in MPS has been described in such positions. In addition, synonymous variants were also excluded, as well as variants in non-canonical transcripts.

Second tier variant selection

At the second tier analysis, missense, nonsense, stop gain and stop loss, frameshift, and splice site variants present in homozygosity (and hemizyosity for *IDS*) were excluded. This was based on the assumption that both ExAC and gnomAD would not include MPS-affected individuals in their database since they exclude samples from patients with severe pediatric diseases and their relatives (Lek et al., 2016). Therefore any homozygous variant should not be pathogenic. Heterozygous stop gain, stop loss and nonsense variants were considered pathogenic, without further analysis.

Third tier variant selection

Alterations in splice site, in-frame insertions and deletions and frameshift were examined in the third tier. Heterozygous splice site variants were further analyzed by Human Splice Finder, using the default algorithm output. In frame insertions and deletions and frameshift mutations were analyzed by SIFT Indel (Hu et al., 2013) also using default parameters.

Fourth tier variant selection

The effect of missense variants was analyzed in the fourth tier. Missense variants were submitted to five *in silico* programs: MutPred (Li et al., 2009), PolyPhen2 (Adzhubei et al., 2010), PROVEAN (Choi et al., 2012), SIFT (Kumar et al., 2009), SNP&Go (Calabrese et al., 2009). For Polyphen2, the only program with more than two categories, results were transformed into binary data considering “possibly pathogenic” as disease-causing. A consensus classification by 3 or more predictors was used to either retain or exclude a variant as disease-causing. A scheme showing all the steps of variant selection (with actual numbers for *IDUA*, as to example) is depicted in figure 1.

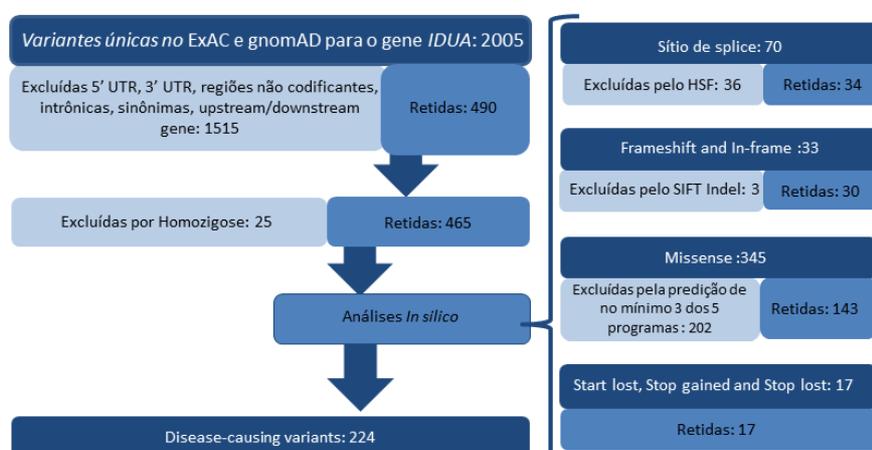
Fifth tier variant selection

All allele frequencies of remained variants were analyzed to ensure that only rare variants were retained. Thereunto all variant that exhibited frequencies above 0.01 were excluded.

Calculation of disease frequency using Hardy-Weinberg equation

The number of chromosomes bearing a variant retained as potentially disease-causing was divided by the total number of chromosomes analyzed at that position since due to a quality filter and other reasons, this number varies in the database. Then the sum of all variant frequencies for each gene was used as the frequency of the recessive allele (q). The incidence was then calculated as q^2 .

Figure 1. Scheme example with all steps of variant selection for *IDUA* gene (MPS I).



Results

Initially, the number of variants for each gene ranged from 961 (*IDS*) to 2988 (*GALNS*), as seen in table 1. After subsequent filtering steps, these numbers were reduced up to 224 (*IDUA*) and 23 (*IDS*) (table 2). A detailed description of the excluded variants can be found at supplementary table 1.

The number of variants excluded due to homozygosity ranged between 3 in *GNS* to 113 in *IDS*. None of stop gain, stop loss or start loss variants were present in homozygosity in any gene. The overall number of heterozygous splice site variants considering all genes was 452. HSF was unable to analyze 11 (supplementary table 3) of them and considered 209 as disease-causing. SIFT Indel analyzed 213 out of 218 frameshift and in-frame insertions and deletions and considered 188 as disease-causing. Variants that were not analyzed by HSF or SIFT Indel were excluded. The total number of variants considered disease-causing in each category can be seen in table 2.

Missense variants were analyzed by 5 different *in silico* programs. From overall 3038 variants, all were analyzed by all 5 programs. Out of these, a pathogenic consensus was reached for 406, whereas 411 were homogenously classified by 4 programs and 430 by 3 programs. For calculating disease frequency, all mutations considered disease-causing by 3 out of 5 programs were included (table 2).

Table 2. The number of variants considered disease-causing at each category for each MPS gene.

Gene	Frameshift	In-frame insertion/deletion	Splice site	Start lost	Stop gained	Stop lost	Missense	Total
<i>IDUA</i>	18	12	34	1	15	1	143	224
<i>IDS</i>	0	1	2	0	0	0	20	23
<i>SGSH</i>	14	7	7	0	14	0	153	195
<i>NAGLU</i>	20	2	10	1	16	0	149	198
<i>HGSNAT</i>	11	4	37	0	9	0	88	149
<i>GNS</i>	5	3	23	0	4	1	81	117
<i>GALNS</i>	11	7	23	1	11	1	163	217
<i>GLB1</i>	13	3	31	1	13	0	142	203
<i>ARSB</i>	12	5	15	0	12	0	118	162
<i>GUSB</i>	13	6	25	2	14	2	79	141
<i>HYAL1</i>	13	8	2	1	9	0	111	144
All genes	130	58	209	7	117	5	1247	1773

The allele frequency of each of these variants for a given gene was summed and considered as the frequency of the disease-causing recessive allele. This number

was then used to calculate disease frequency based on Hardy-Weinberg equation (table 3). Since the number of variants retained for *IDS* was exceedingly low (23), the frequency of MPS II was not estimated. The estimated incidence for MPS IIIB was first calculated as 1: 8,066.47. However, it was observed that one single variant (p.His356Pro) had an exceeding high frequency compared to all other variants (0.007993). When this variant was excluded from the database, the estimated frequency was more similar to that found for other diseases. This was the only gene with any retained variant with a frequency higher than 0.001.

Table 3 Disease frequency based on the sum of allele frequencies of disease-causing variants for each MPS gene.

Gene	Disease-causing variants	Sum of allele frequencies (q)	Disease frequency	In 100,000
<i>IDUA</i>	224	0.007394261	1:18,289.86	5.47
<i>SGSH</i>	195	0.004117841	1:58,974.05	1.69
<i>NAGLU</i>	198	0.00314114	1:101,350.40*	0.99*
<i>HGSNAT</i>	149	0.003800391	1:68,947.25	1.45
<i>GNS</i>	117	0.00194757	1:263,641.49	0.38
<i>GALNS</i>	217	0.004327544	1:53,397.01	1.87
<i>GLB1</i>	203	0.003789158	1:69,648.95	1.43
<i>ARSB</i>	162	0.002922632	1:117,071.68	0.85
<i>GUSB</i>	141	0.002699569	1:137,218.05	0.73
<i>HYAL1</i>	144	0.002229816	1:201,123.12	0.50

* Corrected value. Original value was 1: 8,066.47 or 12.4:100,000

Discussion

In this study, we used public data from the whole exome and whole genome sequencing (WES and WGS) collected from unrelated individuals to estimate the incidence of a group of diseases called Mucopolysaccharidoses. These diseases are a group of debilitating inherited metabolic disorders that manifest themselves in the first decade of life. Therefore, it is highly unlikely that MPS-affected individuals would be part of such databases.

A recent review (Khan et al., 2017) published a comprehensive analysis of estimated frequencies of MPS disorders in different countries based on the number of diagnosed cases in reference centers. The estimated frequency for almost all MPS diseases found in the present study was either above reported data or in the upper limit (table 5), except for MPS VI.

Table 5. Disease frequency estimated in this study compared to incidence (in 100,000) reported by Khan et al. (2017).

MPS Type	Gene	This study	Khan et al. 2017 (max. – min.)
MPS I	<i>IDUA</i>	5.47	3.62 – 0.11
MPS III A	<i>SGSH</i>	1.69	1.62 – 0.08
MPS III B	<i>NAGLU</i>	0.99	0.72 – 0.02
MPS III C	<i>HGSNAT</i>	1.45	0.42 – 0.03
MPS III D	<i>GNS</i>	0.38	0.10 – 0.09
MPS IV A	<i>GALNS</i>	1.87	1.30 – 0.15
MPS IV B	<i>GLB1</i>	1.43	0.14 – 0.01
MPS VI	<i>ARSB</i>	0.85	7.85 – 0.02
MPS VII	<i>GUSB</i>	0.73	0.29 – 0.02
MPS IX	<i>HYAL1</i>	0.50	NA

It is worth noticing that the maximum frequency reported above is for particular countries whereas our data was calculated collectively for the different ethnic backgrounds present in the databases. In that sense, it is likely that we risk overestimating the disease frequency for the general population. Several measures were taken to reduce this risk. Variants were filtered in sequential steps, in order to obtain the most precise data possible. Both homozygous and frequent variants were excluded. Additional steps based on functional annotation were also performed in order to include only those variants more likely to affect protein function. After that, all variants that remained in the analysis were rare (frequencies below 0.001) but a significant number of them was of unknown significance.

We considered using only curated variants reported either on ClinVar or Human Genome Mutation Database (HGMD), however, this would greatly reduce the number of retained variants (for instance, from 224 to 47 for *IDUA*, data not shown). Different *in silico* programs were used to estimate the likelihood of a variant to be disease-causing. However, as no data on the sensitivity and specificity of such programs are available for these specific genes, it is impossible to estimate the number of false-positive results. And that would also increase the estimated disease frequency.

The existence of compound heterozygotes cannot be discarded. In fact, most MPS patients not derived from consanguineous marriage are indeed compound heterozygotes. However, due to the structure of both databases used in this study, it is impossible to determine whether variants are in the same individual or not. Likewise, the occurrence of variants in *cis* cannot be discarded. Both situations would contribute to the overestimation of disease frequency.

Despite these limitations, a similar approach has been used by Appadurai et al. (2015) to estimate the frequency of Cerebrotendinous Xanthomatosis. As in the

present study, the authors suggested an apparent underdiagnosis of this disorder based on the frequency of detected variants. In our case, it is interesting to notice that the discrepancy between molecular-based data and diagnosis-based incidence is more pronounced in the rarest MPS diseases, such as MPS III C, III D, IV B, VII, and IX. For MPS IX it is possible that variants causing enzyme impairment are not clinically recognized as this disorder is characterized by an extremely attenuated phenotype (Kiykim et al., 2016). For MPS VII, on the other hand, severe cases may lead to premature death before the diagnosis is reached or even sought (Sands, 2014).

In the case of MPS IV B, there is an additional complexity that the same gene is involved in a second lysosomal storage disorder with a different accumulated substrate and clinical features, called GM1 Gangliosidosis (Lee et al., 2018). In this study, variants in *GLB1* were considered disease-causing regardless of the associated phenotype. Therefore, the overall frequency of alleles was used to estimate the frequency of MPS IV B whereas in fact only about 13,3% of curated disease-causing variants in this gene are associated with MPS IV B, the rest leading to the three types of GM1 Gangliosidosis (Caciotti et al., 2011).

In sum, we report a simple approach to estimate the disease frequency for MPS disorders that may help to better tailor screening and diagnostic programs for these pathologies.

References

1. Adzhubei, Ivan A. et al. 2010. "A Method and Server for Predicting Damaging Missense Mutations." *Nature Methods* 7(4):248–49.
2. Appadurai, Vivek et al. 2015. "Apparent Underdiagnosis of Cerebrotendinous Xanthomatosis Revealed by Analysis of ~60,000 Human Exomes." *Molecular Genetics and Metabolism* 116(4):298–304.
3. Caciotti, Anna et al. 2011. "GM1 Gangliosidosis and Morquio B Disease: An Update on Genetic Alterations and Clinical Findings." *Human Molecular Genetics* 18(7):782–90.
4. Calabrese, Remo, Emidio Capriotti, Piero Fariselli, Pier Luigi Martelli, and Rita Casadio. 2009. "Functional Annotations Improve the Predictive Score of Human Disease-Related Mutations in Proteins." *Human Mutation* 30(8):1237–44.
5. Choi, Yongwook, Gregory E. Sims, Sean Murphy, Jason R. Miller, and Agnes P. Chan. 2012. "Predicting the Functional Effect of Amino Acid Substitutions and Indels." *PLoS ONE* 7(10).
6. Hu, Jing and Pauline C. Ng. 2013. "SIFT Indel: Predictions for the Functional Effects of Amino Acid Insertions/Deletions in Proteins." edited by A. E. Toland. *PLoS*

One 8(10):e77940.

7. Khan, Shaukat A. et al. 2017. "Epidemiology of Mucopolysaccharidoses." *Molecular Genetics and Metabolism* 121(3):227–40.
8. Kiykim, Ertugrul et al. 2016. "Screening Mucopolysaccharidosis Type IX in Patients with Juvenile Idiopathic Arthritis." edited by Eva MoravaMatthias BaumgartnerMarc PattersonShamima RahmanJohannes ZschockeVerena Peters. *JIMD Reports* 25:21–24.
9. Kumar, Prateek, Steven Henikoff, and Pauline C. Ng. 2009. "Predicting the Effects of Coding Non-Synonymous Variants on Protein Function Using the SIFT Algorithm." *Nature Protocols* 4(7):1073–82.
10. Lee, Jin Sook et al. 2018. "Diagnostic Challenge for the Rare Lysosomal Storage Disease: Late Infantile GM1 Gangliosidosis." *Brain and Development* 1–8.
11. Lek, Monkol et al. 2016. "Analysis of Protein-Coding Genetic Variation in 60,706 Humans." *Nature* 536(7616):285–91.
12. Li, Biao et al. 2009. "Automated Inference of Molecular Mechanisms of Disease from Amino Acid Substitutions." *Bioinformatics* 25(21):2744–50.
13. S Sands, Mark. 2014. *Mucopolysaccharidosis Type VII: A Powerful Experimental System and Therapeutic Challenge*. Vol. 12 Suppl 1.

Sup Table 1. The number of variants excluded at each category for each MPS gene. Bold numbers identify retained variants.

	<i>IDUA</i>	<i>IDS</i>	<i>SGSH</i>	<i>NAGLU</i>	<i>HGSNAT</i>	<i>GNS</i>	<i>GALNS</i>	<i>GLB1</i>	<i>ARSB</i>	<i>GUSB</i>	<i>HYAL1</i>	<i>All genes</i>
Merged variants	2005	961	2043	1048	1533	1141	2988	1629	1159	1067	1082	16656
<i>Excluded in the first tier</i>												
5'UTR	71	65	29	60	0	35	71	42	81	19	124	597
3' UTR	23	299	177	30	298	237	258	29	312	67	41	1771
Upstream gene	12	25	224	11	13	9	374	13	6	3	14	704
Downstream gene	6	1	286	192	7	4	292	8	7	8	367	1178
Intronic	834	197	244	140	375	321	521	770	215	246	94	3957
Non coding transcript exon	204	10	453	0	227	81	700	115	61	191	0	2042
Non-canonical transcript	179	66	132	48	86	31	201	55	14	5	12	829
Synonymous	186	100	94	171	133	124	147	146	122	179	117	1519
Retained variants	490	198	404	396	394	299	424	451	341	349	313	4059
<i>Excluded in second tier</i>												
Homozygosity	25	113	14	7	11	3	19	9	7	4	10	222
Retained variants	465	85	390	389	383	296	405	442	334	345	303	3837
<i>Excluded in third tier</i>												
HSF*	36	8	14	13	31	18	36	30	21	30	6	243
SIFT Indel	3	1	1	3	3	1	4	6	2	2	4	30
Retained variants	426	76	375	373	349	277	365	406	311	313	293	3564
<i>Excluded in the fourth tier</i>												
Consensus for non-pathogenicity	84	29	94	77	120	93	74	85	83	109	81	929
4 out of 5	68	12	36	63	51	38	36	74	39	29	43	489
3 out of 5	50	12	50	34	29	29	38	44	27	34	25	372
Retained variants	224	23	195	197	149	117	217	203	162	141	144	1774
<i>Excluded in the fifth tier</i>												
Allele Frequency	0	0	0	1	0	0	0	0	0	0	0	1
Retained variants	224	23	195	198	149	117	217	203	162	141	144	1773

*HSF: Human Splice Finder 3.0

Sup Table 2. The number of variants excluded out Homozygosity for each MPS gene.

Homozygosity	<i>IDUA</i>	<i>IDS</i>	<i>SGSH</i>	<i>NAGLU</i>	<i>HGSNAT</i>	<i>GNS</i>	<i>GALNS</i>	<i>GLB1</i>	<i>ARSB</i>	<i>GUSB</i>	<i>HYAL1</i>	<i>Total</i>
In-frame insertion/deletion	0	0	0	0	0	0	0	0	1	1	0	2
Splice site	4	9	1	1	1	0	4	1	1	0	0	22
Missense	21	104	13	6	10	3	15	8	5	3	10	198
Total	25	113	14	7	11	3	19	9	7	4	10	222

Sup Table 3. The number of variants excluded by unprocessed for each MPS gene.

Unprocessed	<i>IDUA</i>	<i>IDS</i>	<i>SGSH</i>	<i>NAGLU</i>	<i>HGSNAT</i>	<i>GNS</i>	<i>GALNS</i>	<i>GLB1</i>	<i>ARSB</i>	<i>GUSB</i>	<i>HYAL1</i>	<i>total</i>
Frameshift	0	0	0	0	0	0	2	2	0	0	0	4
In-frame insertion/deletion	0	0	0	0	1	0	0	0	0	0	0	1
Splice site	1	0	0	0	0	0	2	3	1	2	2	11
Total	1	0	0	0	1	0	4	5	1	2	2	16

Discussão

No presente trabalho, foram calculadas as frequências das MPS por uma abordagem não usual, onde utilizamos informações de bancos de dados genômicos. Uma abordagem similar já foi utilizada para erros inatos do metabolismo. Em 2015, Appadurai e colaboradores utilizaram o ExAC para analisar o aparente subdiagnóstico da Xantomatose Cerebrotendinosa e encontraram frequências superiores às previamente descritas. Assim como no estudo da xantomatose, as frequências encontradas para as MPS são maiores que as incidências e prevalências apresentadas nos estudos realizados até o momento.

Dentre os tipos de MPS, a MPS I é o que apresentou maior frequência no nosso estudo. Malm e colaboradores publicaram em 2008 um estudo mostrando a maior prevalência de MPS I encontrada na literatura, de 3,1:100.000 nascimentos, e a segunda maior incidência de 1,85:100.000. A maior incidência é encontrada na Arábia Saudita, com 3,62:100.000 nascimentos (MOAMMAR et al., 2010), o que pode ser explicado pela alta taxa de casamentos consanguíneos no oriente médio. Apesar de serem altos os valores de incidência e prevalência encontrados nesses estudos, eles são menores que a frequência de 5,47:100.000 encontrada no presente estudo.

Em 2014, Beck e colaboradores analisaram dados de registros de MPS I e estimaram a incidência global como 1 a cada 100.000 nascimentos. Esse valor é muito parecido com a incidência de 0,94:100.000 encontrada na Austrália Ocidental (MEIKLE et al., 1999). Porém, esse valor está acima da incidência encontrada em Taiwan e nos Estados Unidos e da prevalência encontrada na Suíça e na Coreia do Sul de 0,11; 0,24; 0,19 e 0,21:100.000 nascimentos respectivamente (CHO; SOHN; JIN, 2014; KHAN et al., 2017; LIN et al., 2009; PUCKETT et al., 2017).

No caso da MPS III, muitos estudos não apresentam os dados específicos de cada subtipo, apenas os valores de frequência global. Como tipo, essa MPS é relatada como mais prevalente nos Países Baixos (1,89:100.000) (POORTHUIS et al., 1999), Suécia (1,63) (MALM et al., 2008), Austrália (1,51) (MEIKLE et al., 1999), República Checa (0,91) (POUPETOVÁ et al., 2010) e Polônia (0,86) (JURECKA et al., 2015). Também é relatada como grupo mais incidente na Austrália Ocidental (1,71) (NELSON

et al., 2003), Alemanha (1,57) (BAEHNER et al., 2005), Tunísia (0,7) (BEN TURKIA et al., 2009) e Estados Unidos (0,38) (PUCKETT et al., 2017). No presente trabalho, calculamos a frequência para cada um dos subtipos da MPS III (A, B, C e D). As frequências foram 1,69; 0,99; 1,45; 0,38:100.000 respectivamente, com uma frequência de 4,51:100.000 para o tipo III em geral, consideravelmente maior do que os valores encontrados na literatura.

As maiores frequências foram encontradas para os subtipos A e C. Apesar da frequência calculada para a MPS IIIA ser similar ao encontrado na literatura, a frequência encontrada para a MPS IIIC está acima do maior valor já descrito. As maiores prevalências descritas para esse subtipo são na Alemanha e na República Checa, com 1.62 e 0.42:100.000 respectivamente (BAEHNER et al., 2005; POUPETOVÁ et al., 2010). Assim como para MPS IIIA, a frequência encontrada para MPS IIIB foi próxima ao maior valor de prevalência relatado, em Portugal, de 0,72:100.000 nascimentos (PINTO et al., 2004). Entretanto, as prevalências relatadas no Japão (KHAN et al., 2017) e no Brasil (GIUGLIANI et al., 2017) e a incidência relatada na Irlanda do Norte (NELSON, 1997) foram consideravelmente menores, de 0,12:100.000 e nenhum caso foi relatado na Estônia (KRABBI et al., 2012).

Apesar de o grupo apresentar uma alta prevalência, a MPS IIID é a doença, dentre as MPS, que obtivemos menor a frequência 0,38:100.000. Esse achado está de acordo com o encontrado na literatura. A maioria dos estudos não apresenta casos relatados desse subtipo (APPLEGARTH; TOONE; LOWRY, 2000; BAEHNER et al., 2005; CHO; SOHN; JIN, 2014; GIUGLIANI et al., 2017; KHAN et al., 2017; KRABBI et al., 2012; NELSON, 1997; PINTO et al., 2004; POUPETOVÁ et al., 2010). Os dois estudos que apresentam algum relato de MPS IIID calculam sua prevalência como 0,0095:100.000 na Austrália (MEIKLE et al., 1999) e 0,1:100.000 nos Países Baixos (POORTHUIS et al., 1999), abaixo do valor calculado.

Diversos estudos apresentam a frequência total da MPS IV e não os valores dos subtipos. Calculamos a frequência da MPS IVA como 1.87:100.000 e da MPS IVB como 1.43. Comparando as nossas frequências com os estudos que apresentam os valores de cada subtipo, notamos que o resultado obtido da MPS IVA está um pouco acima da mais alta incidência descrita, na Irlanda do Norte, de 1,3:100.000

nascimentos (NELSON, 1997). Entretanto, os dados relatados na literatura das incidências da MPS IVB são mais baixos que os encontrados. Os dados mais altos foram relatados nos Países Baixos, com uma prevalência de 0,14:100.000 (POORTHUIS et al., 1999), enquanto diversos países não apresentaram relatos de casos desse subtipo (APPLEGARTH; TOONE; LOWRY, 2000; LIN et al., 2009; MEIKLE et al., 1999; NELSON, 1997; NELSON et al., 2003; PINTO et al., 2004). Essa grande discrepância dos valores de frequência encontrados e relatados na literatura pode ser ocasionada pela presença de variantes patogênicas da Gangliosidose GM1, uma doença causada por mutações no mesmo gene, *GLB1*, que causa uma diferente acumulação de substrato e fenótipo (CACIOTTI et al., 2011). Considerando as variantes presentes no HGMD, a frequência estimada de alelos que causam MPS IVB é 13,3%, as demais são associadas a um dos três tipos de Gangliosidose GM1. Porém, não é possível prever quais variantes são associadas com MPS IVB ou Gangliosidose-GM1. Dessa forma, não foi possível separar as variantes não ligadas a MPS IVB para que fosse realizada uma estimativa mais acurada da frequência.

Em 2010, MOAMMAR e colaboradores publicaram a incidência mais alta já relatada para as MPS: 7,85 casos de MPS VI a cada 100.000 nascimentos na Arábia Saudita. Essa incidência é maior que as demais relatadas e que a frequência encontrada no presente estudo (0,85:100.000). Essa discrepância deve ocorrer pela alta taxa de consanguinidade que ocorre na região, com taxas chegando a 40% entre primos de primeiro grau e até 60% em casamentos entre parentes (MOAMMAR et al., 2010). Os segundos maiores relatos são as prevalências encontradas na Austrália e em Portugal, de 0,43 e 0,42:100.000 respectivamente (MEIKLE et al., 1999; PINTO et al., 2004). O valor de frequência encontrado é consideravelmente maior que as frequências relatadas na literatura, principalmente se compararmos com lugares como a Irlanda do Norte que não relatou casos dessa MPS (NELSON, 1997).

Não existem muitos dados de frequência da MPS VII. A Columbia Britânica no Canadá apresentou a maior incidência da doença, de 0,29:100.000, seguido pela prevalência encontrada nos Países Baixos de 0,24 (APPLEGARTH; TOONE; LOWRY, 2000; POORTHUIS et al., 1999). O Japão, Austrália, República Checa, Suíça e Brasil relataram incidências menores que 0,05:100.000 e os Estados Unidos apresentaram

incidência igual a 0,05 (KHAN et al., 2017; MEIKLE et al., 1999; POUPETOVÁ et al., 2010; PUCKETT et al., 2017). Apesar do cálculo da frequência realizado por nós também colocar a MPS VII como a terceira mais rara, a frequência calculada está acima dos relatos da literatura.

Por só ter quatro casos relatados, os estudos não apresentam valores de incidências ou prevalências para a MPS IX. Pela estratégia utilizada, conseguimos calcular uma frequência de 0.50:100.000. Apesar de ser um alto valor, está abaixo do calculado para MPS IIID, sugerindo uma maior incidência desse tipo de MPS do que o relatado. Isso pode ser verdade, pela grande possibilidade dessa MPS ser diagnosticada como artrite.

Conclusão

No trabalho, foram calculadas as frequências das MPS a partir de dois bancos de dados populacionais disponíveis na *internet*: o gnomAD e o ExAC. As variantes encontradas nesses bancos de dados foram excluídas por tipos, excluindo-se aquelas que foram encontradas em posições menos propensas a causar problemas nas proteínas, por homozigose, baseando-se que ambos os bancos de dados não incluíram pacientes de MPS em seus bancos de dados, e por análise *in silico*. As variantes de alterações sem sentido foram classificadas como patogênicas, as de alteração de sítio de splice foram analisadas pelo HSF, as de frameshift e as indels foram analisadas pelo SIFT Indel e as missense foram analisadas pelo PolyPhen2, MutPred, PROVEAN, SIFT e SNPs&GO.

As frequências calculadas foram, para MPS I, MPS IIIA,B,C e D, MPS IVA e B, MPS VI, MPS VII e MPS IX, respectivamente: 1:18,289.86; 1:58,974.05; 1:101,350.40; 1:68,947.25; 1:263,641.49; 1:53,397.01; 1:69,648.95; 1:117,071.68; 1:137,218.05; 1:201,123.12. A MPS II é a única desse grupo ligado ao X e por isso teve muitas variantes excluídas por homo ou hemizigose, ficando com o menor número de variantes. Assim, a frequência calculada foi muito menor do que o esperado. As demais MPS apresentaram frequências maiores do que as encontradas na literatura, calculadas a partir do número de diagnósticos realizados.

A principal causa da diferença dos valores encontrados no estudo e dos relatados na literatura é a não utilização de dados da doença que partiram do diagnóstico. Isso porque, muitas vezes, o correto diagnóstico é pelo amplo espectro de manifestações fenotípicas encontrado nas MPS. Portanto, o método usado é uma interessante alternativa para futuros estudos de cálculos de frequências de doenças genéticas raras.

Referências Bibliográficas

- ALEMÁN, A. et al. A web-based interactive framework to assist in the prioritization of disease candidate genes in whole-exome sequencing studies. **Nucleic acids research**, [s. l.], v. 42, n. Web Server issue, p. W88-93, 2014. Disponível em: <<http://academic.oup.com/nar/article/42/W1/W88/2437375/A-webbased-interactive-framework-to-assist-in-the>>
- APPADURAI, V. et al. Apparent underdiagnosis of Cerebrotendinous Xanthomatosis revealed by analysis of ~60,000 human exomes. **Molecular Genetics and Metabolism**, [s. l.], v. 116, n. 4, p. 298–304, 2015.
- APPLEGARTH, D. A.; TOONE, J. R.; LOWRY, R. B. Incidence of inborn errors of metabolism in British Columbia, 1969-1996. **Pediatrics**, [s. l.], v. 105, n. 1, p. e10, 2000. Disponível em: <<http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.105.1.e10>>
- BAEHNER, F. et al. Cumulative incidence rates of the mucopolysaccharidoses in Germany. **Journal of Inherited Metabolic Disease**, [s. l.], v. 28, n. 6, p. 1011–1017, 2005. Disponível em: <<http://link.springer.com/10.1007/s10545-005-0112-z>>
- BALDO, G. **Mucopolissacaridoses: mecanismos patogênicos e abordagens terapêuticas baseadas em terapia gênica e reposição enzimática**. 2012. Universidade Federal do Rio Grande do Sul, [s. l.], 2012. Disponível em: <<http://hdl.handle.net/10183/55087>>
- BECK, M. et al. The natural history of MPS I: global perspectives from the MPS I Registry. **Genetics in medicine : official journal of the American College of Medical Genetics**, [s. l.], v. 16, n. 10, p. 759–65, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24675674>>
- BEN TURKIA, H. et al. Incidence of mucopolysaccharidoses in Tunisia. **La Tunisie medicale**, [s. l.], v. 87, n. 11, p. 782–5, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20209839>>
- BUSKE, O. J. et al. Identification of deleterious synonymous variants in human genomes. **Bioinformatics (Oxford, England)**, [s. l.], v. 29, n. 15, p. 1843–50, 2013. Disponível em: <<https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btt308>>
- CACIOTTI, A. et al. GM1 gangliosidosis and Morquio B disease: an update on genetic alterations and clinical findings. [s. l.], v. 1812, n. 7, p. 782–790, 2011.
- CHO, S. Y.; SOHN, Y. B.; JIN, D.-K. An overview of Korean patients with mucopolysaccharidosis and collaboration through the Asia Pacific MPS Network. **Intractable & rare diseases research**, [s. l.], v. 3,

n. 3, p. 79–86, 2014. Disponível em:

<<http://jlc.jst.go.jp/DN/JST.JSTAGE/irdr/2014.01013?lang=en&from=CrossRef&type=abstract>>

COUTINHO, M. F.; LACERDA, L.; ALVES, S. Glycosaminoglycan storage disorders: a review.

Biochemistry research international, [s. l.], v. 2012, p. 471325, 2012. Disponível em:

<<http://www.hindawi.com/journals/bri/2012/471325/>>

DE RU, M. H. et al. Enzyme Replacement Therapy and/or Hematopoietic Stem Cell Transplantation at diagnosis in patients with Mucopolysaccharidosis type I: results of a European consensus procedure.

Orphanet Journal of Rare Diseases, [s. l.], v. 6, n. 1, p. 55, 2011. Disponível em:

<<http://ojrd.biomedcentral.com/articles/10.1186/1750-1172-6-55>>

DESMET, F. O. et al. Human Splicing Finder: An online bioinformatics tool to predict splicing signals.

Nucleic Acids Research, [s. l.], 2009.

DONCEL-PÉREZ, E. et al. Synthesis, physicochemical characterization and biological evaluation of chitosan sulfate as heparan sulfate mimics. **Carbohydrate polymers**, [s. l.], v. 191, p. 225–233, 2018.

Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0144861718302923?via%3Dihub>>.

Acesso em: 30 abr. 2018.

FEDELE, A. Sanfilippo syndrome: causes, consequences, and treatments. **The Application of Clinical**

Genetics, [s. l.], v. 8, p. 269, 2015. Disponível em: <<https://www.dovepress.com/sanfilippo-syndrome-causes-consequences-and-treatments-peer-reviewed-article-TACG>>

<<https://www.dovepress.com/sanfilippo-syndrome-causes-consequences-and-treatments-peer-reviewed-article-TACG>>

FU, W. et al. Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants.

Nature, [s. l.], v. 493, n. 7431, p. 216–20, 2013. Disponível em:

<<http://www.nature.com/doi/10.1038/nature11690>>

GIUGLIANI, R. et al. Mucopolysaccharidosis I, II, and VI: brief review and guidelines for treatment.

Genetics and Molecular Biology, [s. l.], v. 33, n. 4, p. 589–604, 2010. Disponível em: <www.sbg.org.br>

GIUGLIANI, R. Mucopolysaccharidoses: From understanding to treatment, a century of discoveries.

Genetics and molecular biology, [s. l.], v. 35, n. 4 (suppl), p. 924–31, 2012. Disponível em:

<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572012000600006&lng=en&tlng=en>.

Acesso em: 24 abr. 2018.

GIUGLIANI, R. et al. Relative frequency and estimated minimal frequency of Lysosomal Storage

Diseases in Brazil: Report from a Reference Laboratory. **Genetics and molecular biology**, [s. l.], v. 40,

n. 1, p. 31–39, 2017. Disponível em: <[http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572017000100031&lng=en&tlng=en)

<[http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572017000100031&lng=en&tlng=en)

<[http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-47572017000100031&lng=en&tlng=en)

47572017000100031&lng=en&tlng=en>

GOLDSTEIN, D. B. et al. Sequencing studies in human genetics: design and interpretation. **Nature**

reviews. Genetics, [s. l.], v. 14, n. 7, p. 460–70, 2013. Disponível em:

<<http://www.nature.com/articles/nrg3455>>

GUARANY, N. R. **Avaliação do efeito da terapia de reposição enzimática na capacidade funcional**

de pacientes com mucopolissacaridose. 2011. Universidade Federal do Rio Grande do Sul, [s. l.], 2011. Disponível em: <<http://link.springer.com/10.1007/s10545-005-0112-z>>

HICKS, S. et al. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. **Human Mutation**, [s. l.], 2011.

HOLLAK, C. E. M.; WIJBURG, F. A. Treatment of lysosomal storage disorders: successes and challenges. **Journal of inherited metabolic disease**, [s. l.], v. 37, n. 4, p. 587–98, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24820227>>

HORKAY, F. et al. Structure and Properties of Cartilage Proteoglycans. **Macromolecular Symposia**, [s. l.], v. 372, n. 1, p. 43–50, 2017. Disponível em: <<http://doi.wiley.com/10.1002/masy.201700014>>

HU, J.; NG, P. C. Predicting the effects of frameshifting indels. **Genome biology**, [s. l.], v. 13, n. 2, p. R9, 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22322200>>

HU, J.; NG, P. C. SIFT Indel: predictions for the functional effects of amino acid insertions/deletions in proteins. **PloS one**, [s. l.], v. 8, n. 10, p. e77940, 2013. Disponível em: <<http://dx.plos.org/10.1371/journal.pone.0077940>>

IMUNDO, L. et al. A complete deficiency of Hyaluronoglucosaminidase 1 (HYAL1) presenting as familial juvenile idiopathic arthritis. **Journal of inherited metabolic disease**, [s. l.], v. 34, n. 5, p. 1013–22, 2011. Disponível em: <<http://link.springer.com/10.1007/s10545-011-9343-3>>

JIAN, X.; BOERWINKLE, E.; LIU, X. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. **Genetics in Medicine**, [s. l.], v. 16, n. 7, p. 497–503, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24263461>>

JIANG, R. Walking on multiple disease-gene networks to prioritize candidate genes. **Journal of molecular cell biology**, [s. l.], v. 7, n. 3, p. 214–30, 2015. Disponível em: <<https://academic.oup.com/jmcb/article-lookup/doi/10.1093/jmcb/mjv008>>

JURECKA, A. et al. Prevalence rates of mucopolysaccharidoses in Poland. **Journal of Applied Genetics**, [s. l.], v. 56, n. 2, p. 205–210, 2015.

KATSONIS, P. et al. Single nucleotide variations: biological impact and theoretical interpretation. **Protein science : a publication of the Protein Society**, [s. l.], v. 23, n. 12, p. 1650–66, 2014. Disponível em: <<http://doi.wiley.com/10.1002/pro.2552>>

KHAN, S. A. et al. Epidemiology of mucopolysaccharidoses. **Molecular Genetics and Metabolism**, [s. l.], v. 121, n. 3, p. 227–240, 2017. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1096719217302068?via%3Dihub>>. Acesso em: 15 maio. 2018.

KIYKIM, E. et al. Screening Mucopolysaccharidosis Type IX in Patients with Juvenile Idiopathic Arthritis. **JIMD reports**, [s. l.], v. 25, p. 21–24, 2016. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/26122630%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5059191>>

KRABBI, K. et al. The Live-Birth Prevalence of Mucopolysaccharidoses in Estonia. **Genetic Testing and Molecular Biomarkers**, [s. l.], v. 16, n. 8, p. 846–849, 2012. Disponível em: <<http://online.liebertpub.com/doi/abs/10.1089/gtmb.2011.0307>>

LANDRUM, M. J. et al. ClinVar: Public archive of interpretations of clinically relevant variants. **Nucleic Acids Research**, [s. l.], 2016.

LEK, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. **Nature**, [s. l.], v. 536, n. 7616, p. 285–91, 2016. Disponível em: <<http://www.nature.com/articles/nature19057>>

LI, M.-X. et al. Predicting Mendelian Disease-Causing Non-Synonymous Single Nucleotide Variants in Exome Sequencing Studies. **PLoS Genetics**, [s. l.], v. 9, n. 1, p. e1003143, 2013. Disponível em: <<http://dx.plos.org/10.1371/journal.pgen.1003143>>. Acesso em: 24 abr. 2018.

LIN, H. Y. et al. Incidence of the Mucopolysaccharidoses in Taiwan, 1984-2004. **American Journal of Medical Genetics, Part A**, [s. l.], v. 149, n. 5, p. 960–964, 2009.

MALM, G. et al. Mucopolysaccharidoses in the Scandinavian countries: incidence and prevalence. **Acta paediatrica (Oslo, Norway : 1992)**, [s. l.], v. 97, n. 11, p. 1577–81, 2008. Disponível em: <<http://doi.wiley.com/10.1111/j.1651-2227.2008.00965.x>>

MEIKLE, P. J. et al. Prevalence of lysosomal storage disorders. **JAMA**, [s. l.], v. 281, n. 3, p. 249–54, 1999. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9918480>>

MOAMMAR, H. et al. Incidence and patterns of inborn errors of metabolism in the Eastern Province of Saudi Arabia, 1983-2008. **Annals of Saudi Medicine**, India, v. 30, n. 4, p. 271–277, 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2931777/>>

MUENZER, J. et al. Mucopolysaccharidosis I: management and treatment guidelines. **Pediatrics**, [s. l.], v. 123, n. 1, p. 19–29, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19117856>>

MUENZER, J. Overview of the mucopolysaccharidoses. **Rheumatology (Oxford, England)**, [s. l.], v. 50 Suppl 5, n. suppl 5, p. v4-12, 2011. Disponível em: <<https://academic.oup.com/rheumatology/article-lookup/doi/10.1093/rheumatology/ker394>>

NASLAVSKY, M. S. et al. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. **Human mutation**, [s. l.], v. 38, n. 7, p. 751–763, 2017. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/28332257>>

NELSON, J. Incidence of the mucopolysaccharidoses in Northern Ireland. **Human genetics**, [s. l.], v. 101, n. 3, p. 355–8, 1997. Disponível em: <<http://link.springer.com/10.1007/s004390050641>>

NELSON, J. et al. Incidence of the mucopolysaccharidoses in Western Australia. **American journal of medical genetics. Part A**, [s. l.], v. 123A, n. 3, p. 310–313, 2003.

NEUFELD, E. F.; MUENZER, J. The mucopolysaccharidoses. In: GRANT MITCHELL, M. D. V. M. A. L. B. M. B. V. M. K. W. K. P. D. S. E. A. M. D. S. A. B. M. K. M. G. P. D. F. (Ed.). **The Online Metabolic and Molecular Bases of Inherited Disease**. [s.l: s.n.]. p. 3421–3452.

NOH, H.; LEE, J. I. Current and potential therapeutic strategies for mucopolysaccharidoses. **Journal of Clinical Pharmacy and Therapeutics**, [s. l.], v. 39, n. 3, p. 215–224, 2014. Disponível em: <<http://doi.wiley.com/10.1111/jcpt.12136>>

OHANIAN, M.; OTWAY, R.; FATKIN, D. Heuristic methods for finding pathogenic variants in gene coding sequences. **Journal of the American Heart Association**, [s. l.], v. 1, n. 5, p. e002642, 2012. Disponível

em: <<http://jaha.ahajournals.org/cgi/doi/10.1161/JAHA.112.002642>>

PABINGER, S. et al. A survey of tools for variant analysis of next-generation genome sequencing data. **Briefings in bioinformatics**, [s. l.], v. 15, n. 2, p. 256–78, 2014. Disponível em: <<https://academic.oup.com/bib/article-lookup/doi/10.1093/bib/bbs086>>

PASQUALIM, G. **Efeitos da terapia de reposição enzimática com início tardio no modelo miurino de mucopolissacaridose do tipo I**. 2013. Universidade Federal do Rio Grande do Sul, [s. l.], 2013. Disponível em: <<http://hdl.handle.net/10183/72370>>

PENNOCK, C. a. A review and selection of simple laboratory methods used for the study of glycosaminoglycan excretion and the diagnosis of the mucopolysaccharidoses. **Journal of Clinical Pathology**, [s. l.], v. 29, n. 2, p. 111–123, 1976. Disponível em: <<http://jcp.bmj.com/cgi/doi/10.1136/jcp.29.2.111>>

PINTO, L. L. C. **Estudo Sobre a Historia Natural Da Mucopolissacaridose Tipo II (Síndrome De Hunter) Em Pacientes Brasileiros**. 2005. Universidade Federal do Rio Grande do Sul, [s. l.], 2005.

PINTO, R. et al. Prevalence of lysosomal storage diseases in Portugal. **European journal of human genetics : EJHG**, [s. l.], v. 12, n. 2, p. 87–92, 2004. Disponível em: <<http://www.nature.com/articles/5201044>>

PJ, M. et al. Prevalence of lysosomal storage disorders. **JAMA**, [s. l.], v. 281, n. 3, p. 249–254, 1999. Disponível em: <<http://dx.doi.org/10.1001/jama.281.3.249>>

POORTHUIS, B. J. et al. The frequency of lysosomal storage diseases in The Netherlands. **Human genetics**, [s. l.], v. 105, n. 1–2, p. 151–6, 1999. Disponível em: <<http://link.springer.com/10.1007/s004399900075>>

POUPETOVÁ, H. et al. The birth prevalence of lysosomal storage disorders in the Czech Republic: comparison with data in different populations. **Journal of inherited metabolic disease**, [s. l.], v. 33, n. 4, p. 387–96, 2010. Disponível em: <<http://link.springer.com/10.1007/s10545-010-9093-7>>

PUCKETT, Y. et al. Epidemiology of mucopolysaccharidoses (MPS) in the United States: challenges and opportunities. **Molecular Genetics and Metabolism**, [s. l.], v. 120, n. 1–2, p. S111, 2017. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S1096719216307041>>

RICHARDS, S. et al. 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. **Genetics in medicine : official journal of the American College of Medical Genetics**, [s. l.], v. 17, n. 5, p. 405–24, 2015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4544753/pdf/nihms697486.pdf>>

SAWAMOTO, K. et al. Gene therapy for Mucopolysaccharidoses. **Molecular Genetics and Metabolism**, [s. l.], v. 123, n. 2, p. 59–68, 2018. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1096719217306169?via%3Dihub>>. Acesso em: 15 maio. 2018.

SCOTT, H. S. et al. Molecular genetics of mucopolysaccharidosis type I: diagnostic, clinical, and

biological implications. **Human mutation**, [s. l.], v. 6, n. 4, p. 288–302, 1995. Disponível em: <<http://doi.wiley.com/10.1002/humu.1380060403>>

STENSON, P. D. et al. **The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies**, 2017.

STITZIEL, N. O.; KIEZUN, A.; SUNYAEV, S. Computational and statistical approaches to analyzing variants identified by exome sequencing. **Genome biology**, [s. l.], v. 12, n. 9, p. 227, 2011. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21920052>>

SUDMANT, P. H. et al. An integrated map of structural variation in 2,504 human genomes. **Nature**, [s. l.], 2015.

TERLATO, N. J.; COX, G. F. Can mucopolysaccharidosis type I disease severity be predicted based on a patient's genotype? A comprehensive review of the literature. **Genetics in medicine : official journal of the American College of Medical Genetics**, [s. l.], v. 5, n. 4, p. 286–94, 2003. Disponível em: <<http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00125817-200307000-00004>>

THOMPSON, B. A. et al. Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. **Human mutation**, [s. l.], v. 34, n. 1, p. 255–65, 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22949387>>

TRIGGS-RAINE, B.; NATOWICZ, M. R. Biology of hyaluronan: Insights from genetic disorders of hyaluronan metabolism. **World journal of biological chemistry**, [s. l.], v. 6, n. 3, p. 110–20, 2015. Disponível em: <<http://www.wjgnet.com/1949-8454/full/v6/i3/110.htm>>

VAIRO, F. et al. **Diagnostic and treatment strategies in mucopolysaccharidosis VI**, 2015.

VALAYANNOPOULOS, V.; WIJBURG, F. A. Therapy for the mucopolysaccharidoses. **Rheumatology (Oxford, England)**, [s. l.], v. 50 Suppl 5, n. suppl 5, p. v49-59, 2011. Disponível em: <<https://academic.oup.com/rheumatology/article-lookup/doi/10.1093/rheumatology/ker396>>

YOUNG, I. D. et al. A clinical and genetic study of Hunter's syndrome. 2. Differences between the mild and severe forms. **Journal of medical genetics**, [s. l.], v. 19, n. 6, p. 408–11, 1982. Disponível em: <<http://jmg.bmj.com/cgi/doi/10.1136/jmg.19.6.408>>