

LUCIANE MORENO STORTI DE MELO

***MINERAÇÃO DE DADOS EM TRIAGEM
NEONATAL DE HEMOGLOBINOPATIAS***

**Dissertação apresentada
para obtenção do Título
de Mestre em Genética.**

Orientadora: Profa. Dra. Claudia Regina Bonini Domingos

São José do Rio Preto

2006

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

LUCIANE MORENO STORTI DE MELO

***MINERAÇÃO DE DADOS EM TRIAGEM
NEONATAL DE HEMOGLOBINOPATIAS***

COMISSÃO JULGADORA

DISSERTAÇÃO PARA OBTENÇÃO DO GRAU DE MESTRE

Presidente e Orientador: Profa. Dra. Claudia Regina Bonini Domingos

2. Examinador Prof. Dr. Paulo Peitl

3. Examinador Prof. Dr. Carlos Roberto Valêncio

São José do Rio Preto, 2022.



Storti-Melo, Luciane Moreno.

Mineração de dados e triagem neonatal de hemoglobinopatias / Luciane Moreno Storti de Melo - São José do Rio Preto: [s.n.], 2011.
88 f. : 28.il. ; 3 cm.

Orientador: Claudia Regina Bonini-Domingos

Dissertação de mestrado - Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas

1. Hemoglobinopatia. 2. Triagem neonatal. 3. HPLC. 4. Banco de Dados. 5. Data mining. I. Visual data mining. I. Bonini-Domingos, Claudia Regina. II. Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas. III. Título.

CDU - 55

Sumario

Dedicatória	i
Lista de Figuras	iii
Lista de Abreviaturas	v
1 Introdução	2
. He moglo inas hu anas	2
.2 Gen tica, s ntese e ontogenia das H nor ais	3
.3 H Anor ais	7
.3.1 H Variantes	7
.3.2 Talasse ias.....	9
.4 Progra as de triage neonatal	
.4.1 Triage Neonatal de He oglo inopantias	
.4.2 I planta o do Progra a Nacional de Triage Neonatal PNTN.....	2
.5 Bases de dados Eletronicas	5
.5.1 Data Mining	7
.5.2 Hist rico tecnol gico	7
1.5.2 <i>Visual Data Mining (VDM)</i>	8
.6 Justificativa	2
2 Objetivos	23
2.1 O jetivos espec ficos	23
3 Material e Métodos	25
3.1 Casu stica	25
3.2 M todos para o diagn stico de He oglo inopantias	27
3.3 Ferramentas <i>Data Mining</i>	31

3.4 Análises Estatísticas	39
4 Resultados	41
4.1 O banco de dados e a ferramenta CLIBIA	4
4.2 Padrões de Híbridos na análise pelo <i>FastMapDB</i>	4
4.2.1 Perfil das amostras normais	4
4.2.2 Perfis híbridos para os diferentes fenótipos de H	5
5 Discussão	62
6 Conclusões.....	71
6.1 Contribuições para implementação de recursos na ferramenta <i>FastMapDB</i> ..	27
7 Referências.....	74
8 Resumo	83
9 Abstract.....	85
10 Apêndices	87
10.1 Relatório Geral híbrido pela ferramenta CLIBIA no Banco de Dados LHGDH <i>on-line</i>	87
10.2 Artigo que será submetido à publicação	88

Dedicatória

A minha iluminada filha Maria Eduarda, benção de Deus, por tornar real a minha crença nos verdadeiros valores da vida.

Ao meu esposo Silvio, por me ensinar que ao vencermos, contamos com a ajuda de outros, mas devemos ter a certeza de que foi por nossos próprios méritos.

A minha querida mãe Leontina, por me provar todos os dias que enquanto a palavra Mãe estiver viva, jamais estaremos sozinhos.

Agradecimentos

À Deus, pelo princípio de tudo, pela vida....

Aos meus pais, por me possibilitarem a vida e mais uma vez estar aqui no campo do aprendizado.

À minha mãe, por todos os sacrifícios, por acreditar em mim mais do que eu mesma e por suportar meus defeitos e me ensinar com amor que não há nada de nobre em sermos superiores aos outros, mas sim ao que éramos antes.

Ao meu esposo Silvio, pelo amor, pela amizade, por nossa filha, por não me esperar vencer, mas por decidir trilhar comigo o caminho da construção.

À minha filha Maria Eduarda, por dar vida a minha vida, me ensinar a dividir o tempo e o pensamento e me fazer entender que as coisas simples são as mais extraordinárias.

À minha sobrinha Carol, por ter me ensinado a amar de um jeito que antes dela eu ainda não conhecia e por toda a alegria que traz as nossas vidas.

Ao meu irmão Junior e minha cunhada Graziela, antes de tudo por receber minha filha como deles, por todo apoio, carinho, amizade e pela certeza de sempre poder contar com vocês.

À toda minha família, de longe de perto, direto ou indiretamente, todos que sempre me deram carinho e compartilharam de minha vida.

Aos amigos de antes, de agora, de sempre, os que trilharam comigo a dura jornada da formação, os nomes não são suficientes para expressar o que vai no coração...Obrigada a todos!

"Ninguém conhece tudo, ninguém ignora tudo, aprendemos juntos"

À Profa. Dra. Cláudia Regina Bonini Domingos, pelo exemplo, incentivo, pela confiança e acima de tudo, pela amizade conquistada nesses anos de convivência.

À todos os amigos do Laboratório de Hemoglobinas e Genética das Doenças Hematológicas, aos de hoje, aos que já passaram e muito ensinaram, obrigada pela amizade e pela paciência.

Ao Prof. Dr. Carlos Roberto Valêncio, pelo apoio nas análises com o FastMapDB, e acima de tudo por auxiliar-me a caminhar por caminhos ainda desconhecidos.

Ao Paulo Henrique Mangonaro, pela realização das projeções tridimensionais.

Ao amigo Thiago Yukio Kikuche de Oliveira pela implantação do Banco de Dados e elaboração da Ferramenta CLIBIA, mas principalmente por todas as vezes que iluminou nossas discussões com suas idéias e nos socorreu nas questões computacionais.

À todos os funcionários do IBILCE, que durante esses anos em que aqui passei, tornaram-se velhos conhecidos, muito obrigado pela prestatividade.

Lista de Figuras

Figura 1: Modelo tridimensional da estrutura quaternária da molécula de H₂O^{*}2

Figura 2: Representação esquemática das famílias gênicas das gloínas. A) Família gênica da gloína alfa incluindo a região controladora HS-4, os genes funcionais e pseudogenes. B) Família gênica da gloína beta incluindo a LCR, os genes funcionais e o pseudogene.4

Figura 3: Síntese de cadeias polipeptídicas ao longo do desenvolvimento humano*4

Figura 4: Mapa representativo da cobertura estadual dos nascidos vivos e da fase de credenciamento dos SRTN. NI = não informado. Fonte: SIA/SUS. 4

Figura 5: Página inicial de acesso ao banco de dados LHGDH *On-line*.4

Figura 6: Página inicial do banco de dados. O círculo verde destaca o *link* para a ferramenta CLIBIA.4

Figura 7: Página inicial da ferramenta CLIBIA. O círculo verde destaca os *links* para a seleção das opções consulta ou relatório.42

Figura 8: Página de seleção do *dataset* na opção consulta da ferramenta CLIBIA.42

Figura 9: Página exibindo a tabela de resultados da busca selecionando-se recém-nascidos no Hospital de Base com o *dataset* na opção consulta da ferramenta CLIBIA.43

Figura 10: Página de seleção do *dataset* na opção relatório da ferramenta CLIBIA.44

Figura 11: Gráfico obtido pela opção relatório da ferramenta CLIBIA incluindo os laudos dos exames que refletem as frequências fenotípicas das H₂O servadas.45

Figura 12: Projeto da imagem tridimensional para o grupo "Normal", incluindo os quatro picos com o atributo de amplitude.4

Figura 13: Projeto da imagem tridimensional para o grupo "Normal" incluindo os picos "E", "F" e "A" com o atributo de amplitude. A seta indica o espaço entre os aglomerados.4

Figura 14: Projeto da imagem tridimensional para o grupo "Normal" incluindo os picos "East", "E" e "A" com o atributo de amplitude.48

Figura 15: Projeto da imagem tridimensional para o grupo "Normal" incluindo os picos "E" e "A" com o atributo de amplitude.48

Figura 16: Projeto da imagem tridimensional para o grupo "Normal", incluindo os quatro atributos de amplitude e utilizando o pico de "East" com o classificatório e dois subgrupos sendo o primeiro variando de 5,5% a 5,5% e o segundo de 5,5% a 9%.5

Figura 17: Projeto da imagem tridimensional para o grupo "Normal", incluindo os quatro atributos de amplitude e utilizando o pico "E" com o classificatório e dois subgrupos sendo o primeiro variando de 5,5% a 3% e o segundo de 3% a 2,4%.5

Figura 18: Comparação das imagens projetadas para o grupo "Normal". A - incluindo os quatro picos com o atributo de amplitude B - utilizando o pico de "E" dividido em dois subgrupos variando de 5,5% a 3% e de 3% a 2,4%.5

Figura 19: Projeção da imagem tridimensional para os grupos $\mathbb{N}or$ al”, $\mathbb{T}alasse$ ia” e $\mathbb{V}ariante$ ”, incluindo os quatro atributos de apeamento.....52

Figura 20: Projeção da imagem tridimensional para os grupos $\mathbb{N}or$ al” e $\mathbb{T}alasse$ ia” incluindo os quatro atributos de apeamento.....52

Figura 21: Projeção da imagem tridimensional para os grupos $\mathbb{T}alasse$ ia” e $\mathbb{N}or$ al” utilizando como atributos de apeamento os picos \mathbb{F} ” e \mathbb{A} ”.....54

Figura 22: Projeção da imagem tridimensional para os grupos $\mathbb{N}or$ al”, $\mathbb{A}lfa$ ” e $\mathbb{B}eta$ ” talasseias, utilizando como atributos de apeamento os picos \mathbb{F} ” e \mathbb{A} ”.....54

Figura 23: Projeção da imagem tridimensional para os grupos $\mathbb{N}or$ al” e $\mathbb{A}lfa$ ” talasseia, utilizando como atributos de apeamento os picos $\mathbb{F}ast$ ”, \mathbb{F} ” e \mathbb{A} ”.....55

Figura 24: Projeção da imagem tridimensional para os grupos $\mathbb{N}or$ al” e $\mathbb{A}lfa$ ” talasseia, utilizando como atributos de apeamento os picos $\mathbb{F}ast$ ” e \mathbb{A} ”.....55

Figura 25: Projeção da imagem tridimensional para os grupos $\mathbb{N}or$ al” e $\mathbb{V}ariante$ ” utilizando os quatro atributos de apeamento. rede de substituição arcada pelo círculo vermelho.....5

Figura 26: Projeção da imagem tridimensional para os grupos $\mathbb{N}or$ al”, $\mathbb{H}AS$ ” e $\mathbb{H}AC$ ” utilizando os quatro atributos de apeamento. rede de substituição arcada pelo círculo verde.5

Figura 27: Projeção da imagem tridimensional para os grupos $\mathbb{N}or$ al” e $\mathbb{H}S$ ” utilizando os quatro atributos de apeamento. rede de substituição arcada pelo círculo verde.58

Figura 28: Projeção da imagem tridimensional para os grupos $\mathbb{N}or$ al” e $\mathbb{H}C$ ” utilizando os quatro atributos de apeamento. rede de substituição arcada pelo círculo verde.58

Lista de A reviaturas

ACB	Azul Crezil Brillhante
ASP	<i>Active Server Pages</i>
CLIBIA	<i>Clinical Bioinformatic Analysis</i>
FC	Fibrose Cística
Hb	Hemoglobina
HC	Hipotireoidismo Congênito
HPLC	Cromatografia Líquida de Alta Performance
HS	Sítios Hipersensíveis
IA	Inteligência Artificial
IVS	<i>Intervening sequence</i>
KDD	<i>Knowledge Discovery in Data ases</i>
LCR	Região controladora do <i>Locus</i>
LHGDH	Laboratório de Hemoglobinas e Genética das Doenças Hematológicas
PHHF	Persistência Hereditária de Hemoglobina Fetal
PKU	Fenilcetonúria
PNTN	Programa Nacional de Triagem Neonatal
SGBDR	Sistema Gerenciador de Base de Dado Relacional
SRTN	Serviço de Referência em Triagem Neonatal
SUS	Sistema Único de Saúde
TEB	Tris EDTA Borato
TN	Triagem Neonatal
VDM	<i>Visual Data Mining</i>

Introdu ão

1 Introdução

1.1 Hemoglobinas humanas

A hemoglobina (Hb) é uma proteína globular, com peso molecular de 64.458 Daltons, formada por quatro subunidades. Cada subunidade é composta de uma fração proteica, a globina e um grupo prostético heme, que contém o íon de ferro, o qual se combina com o oxigênio e confere à molécula sua capacidade de transporte de oxigênio dos pulmões para os tecidos e de parte do gás carbônico no sentido inverso. A fração proteica da molécula de Hb é formada por quatro cadeias polipeptídicas totalizando 541 aminoácidos. Duas delas são constituídas por 141 aminoácidos cada e são chamadas tipo alfa. As outras duas possuem 149 aminoácidos, sendo denominadas tipo beta. HONIG; ADAMS III, 1987; EATHERALL; CLEGG, 2007.

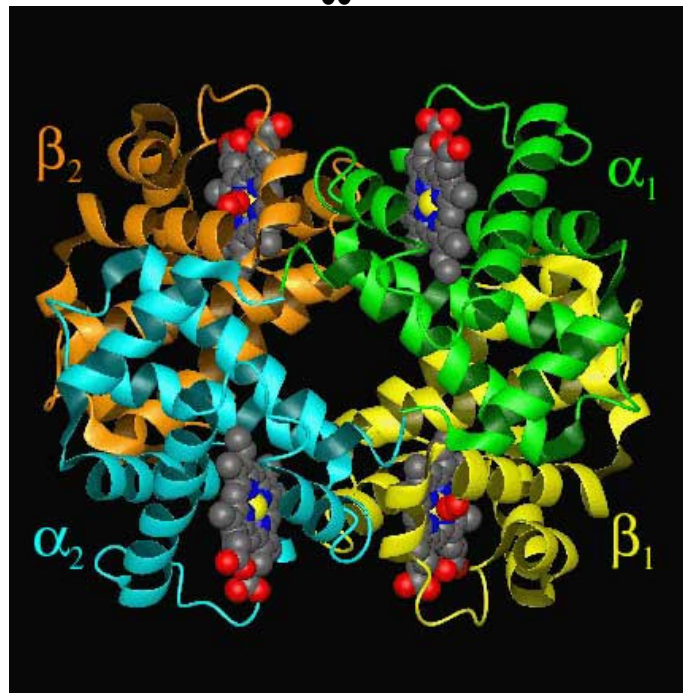


Figura 1: Modelo tridimensional da estrutura quaternária da molécula de Hb.*

* Disponível em <http://mercury.io.uaf.edu/~kevin_ccracken_la_la.html>. Acesso em 22 de outubro de 2005.

1.2 Genética, síntese e ontogenia das Hb normais

As cadeias globínicas são codificadas por genes estruturais pertencentes a *clusters* ou famílias gênicas, que incluem genes funcionais e pseudogenes, localizados e cruzados os diferentes. Pseudogenes (ψ) são genes com seqüências homólogas aos genes estruturais ativos que acumulam mutações, iniciando sua expressão. LEVIN, 2000. O complexo gênico dos genes para as cadeias do tipo alfa, ilustrado na Figura 2A, está localizado no braço curto do cromossomo 16 e compreende três genes funcionais denominados zeta 2, alfa 2 e alfa 2 (ζ_2 , α_2), dispostos na ordem de sua ativação durante o desenvolvimento, três pseudogenes ($\psi\zeta$, $\psi\alpha_2$, $\psi\alpha_1$) e o gene teta (θ), que parece codificar um polipeptídeo que não é incorporado à molécula de Hb, e, por isso, a globina θ pode ser desprovida de qualquer função. CLEGG, 1987; DEVVILE et al., 2004.

O complexo da globina beta, ilustrado na Figura 2B, está agrupado no braço curto do cromossomo 5 e inclui cinco genes funcionais denominados epsilon, gama alfa, gama glicina, delta e beta (ϵ , γ^A , γ^G , δ e β), dispostos na sua ordem de ativação e um pseudogene ($\psi\beta$). EATHERALL; CLEGG, 2000; XIANG-HUI; DE-PEI; CHIH-CHUAN, 2002.

Os genes funcionais das globinas têm a mesma estrutura geral. Apresentam-se divididos em três regiões codificadoras – os exons – e duas regiões não codificadoras, denominados introns ou *intervening sequence* – IVS – representadas por seqüências não traduzidas que separam os exons. O desenvolvimento específico da expressão de cada gene determinado por interações entre os seus promotores e as respectivas regiões de regulação. AYE; CHUI, 2000.

Na família dos genes alfa, a regulação é realizada por meio de uma região conhecida como HS-4, localizada a 4 kb acima do grupo alfa. Essa região atua como sítio de união a fatores de transcrição, cuja integridade é essencial para a expressão dos genes funcionais. DEVVILE et al., 2004. Na família dos genes

eta, a região controladora do loco, LCR, possui cinco sítios hipersensíveis HS, a DNase I e está localizada na porção 5' anterior ao gene ε. A LCR possui papel crítico na expressão da família eta, exercendo atividades de estimulação, inibição ou funções similares complexas, tais como a interação com fatores de transcrição e com a RNA polimerase (CAO; MOI, 2002; STAMATOYANNOPOULOS, 2005).

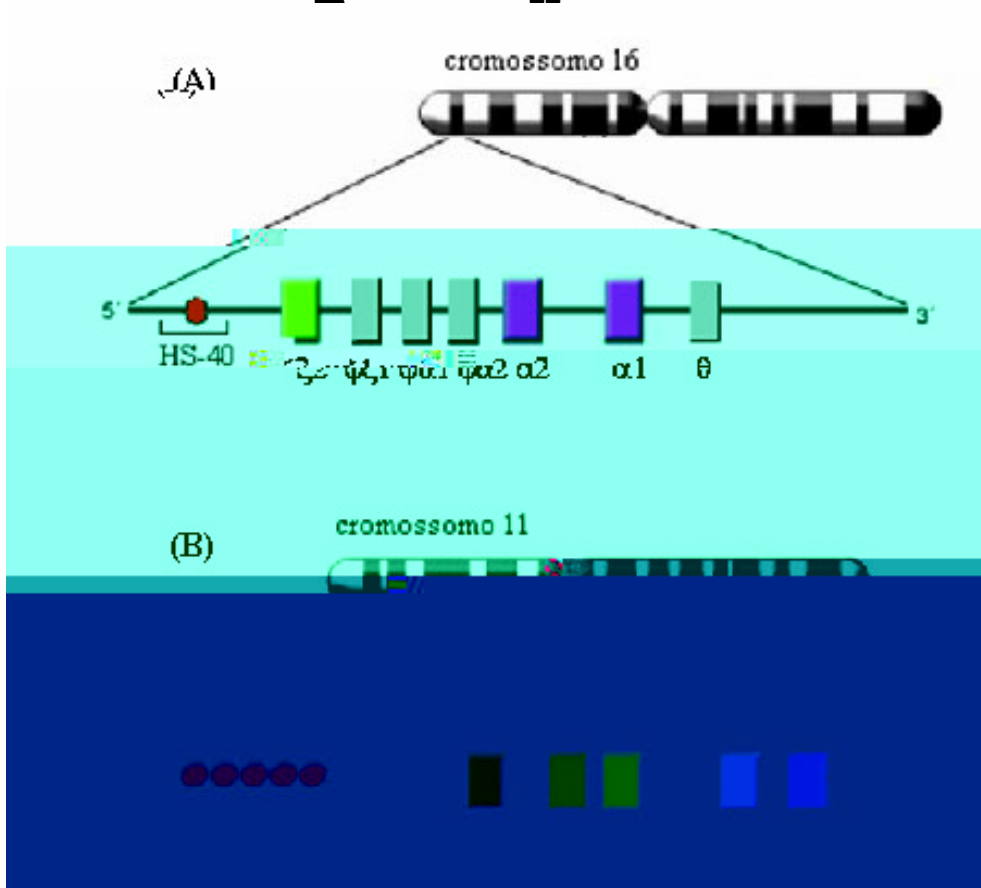


Figura 2: Representação esquemática das famílias gênicas das globinas. A) Família gênica da globina alfa incluindo a região controladora HS-40, os genes funcionais e pseudogenes. B) Família gênica da globina beta incluindo a LCR, os genes funcionais e o pseudogene γ (BANK, 2005, com modificações).

Os genes das famílias alfa e beta expressam-se exclusivamente nos tecidos eritróides, por períodos definidos do desenvolvimento, assegurando o equilíbrio entre a produção de cadeias globulínicas tipo alfa e beta, para a formação das diferentes Hb de forma coordenada (LEVIN, 2000). As Hb diferenciam-se por possuírem características físico-químicas e propriedades eletroforéticas distintas. Suas funções são marcantes desde os primeiros dias de gestação, adaptando-se ao constante desenvolvimento do eritrócito e do feto, atenuando-se por volta dos seis meses após o nascimento. As primeiras Hb sintetizadas são as Hb embrionárias que predominam nos estágios iniciais de gestação. Neste período são encontradas as Hb Gower 1 ($\zeta_2\varepsilon_2$), Gower 2 ($\alpha_2\varepsilon_2$), Hb Portland I ($\zeta_2\gamma_2$) e Hb Portland II ($\zeta_2\beta_2$). Ainda no período embrionário, inicia-se a síntese de Hb Fetal ($\alpha_2\gamma_2$) que substitui gradativamente as Hb embrionárias, atingindo sua plenitude por volta do terceiro mês de gestação (EATHERALL; CLEGG, 2000).

A Hb A ($\alpha_2\beta_2$) começa a ser sintetizada a partir da décima semana de gestação, aumentando-se e concentrando-se progressivamente. Logo após o nascimento, a Hb Fetal e a Hb A apresentam concentrações percentuais de 8% e 2%, respectivamente. Suas sínteses se invertem rapidamente atenuando-se, e a Hb A₂ ($\alpha_2\delta_2$) tem sua síntese iniciada no final do período fetal e, portanto, atenua-se por volta do sexto mês após o nascimento. As Hb normais encontradas ao nascimento são Hb F (90-95%), Hb A (5-10%) e traços de Hb A₂ (1-2%) (GALACTEROS, 1992; DALMAU et al., 1998). No indivíduo adulto, os valores normais de Hb A são de 94% a 98%, Hb A₂ de 2,5% a 3,5% e Hb F de 0% a 1% (BONINI-DOMINGOS, 1993). A Figura 3 ilustra a síntese das globulinas e função do período de desenvolvimento.



Figura 3: Síntese de cadeias polipeptídicas ao longo do desenvolvimento humano (GRIFFITHS, 1999).

1.3 Hb Anormais

Conceitualmente, as Hb anormais apresentam estrutura diferente da Hb A, Hb A2 e Hb Fetal, ou ainda, alterações nos conteúdos quantitativos do componente hemoglobínico normal. Podem ser classificadas em três grandes grupos: as Hb variantes, as talassemias e a persistência hereditária de Hb Fetal (PHHF). As hemoglobinas variantes apresentam alterações envolvendo genes estruturais, que provocam a formação de moléculas de hemoglobina com características bioquímicas diferentes das hemoglobinas normais. As talassemias são resultantes de mutações que afetam os genes reguladores provocando um desequilíbrio no conteúdo quantitativo das cadeias polipeptídicas e consequentemente nos tipos normais de Hb. A PHHF corresponde a alterações nas quais os portadores apresentam quantidades elevadas de Hb Fetal ao longo de sua vida (BONINI-DOMINGOS, 1993). As hemoglobinopatias agrupam as Hb anormais que causam hemólise, policitemia, cianose, ou falciforme (NAOUM, 1999).

Mais de 200 alelos variantes de Hb foram caracterizados molecularmente, incluindo talassemias e variantes. HUISMAN, 2005. Essas mutações são regionalmente específicas, tendo cada país seu espectro de Hb variantes e mutantes talassêmicos. OLD, 2003.

1.3.1 Hb Variantes

Mais de 90 variantes de Hb já foram descritas até o momento. HUISMAN, 2005. A maioria delas originada por substituições de aminoácidos, resultantes de mudanças nas seqüências de nucleotídeos. As alterações estruturais, com conseqüências nas atividades físico-químicas da molécula, dependem da extensão do processo mutacional e dos locais em que ocorre. As substituições de aminoácidos que ocorrem na porção externa da molécula de Hb, não produzem alterações significativas no comportamento funcional da proteína, com exceção da Hb S. No entanto, as mutações nas porções internas, que envolvem a região e torno do grupo heme, causam instabilidade da Hb, geralmente iniciada pela oxidação do grupo heme. BONINI-DOMINGOS, 1993; NAOUM, 1997.

Dentre as Hb variantes mais freqüentes no Brasil destaca-se a Hb S, Hb C e Hb D Los Angeles. CHINELATO-FERNANDES, 2003. A Hb S é uma variante de Hb que é caracterizada, que apresenta prevalência variável nas diferentes regiões do Brasil, dependente dos grupos étnicos formadores de cada região. ZAGO; COSTA, 1985, causada por uma mutação no gene da globina β , na qual ocorre uma substituição da segunda base nitrogenada do códon GAG para GTG, levando à troca do ácido glutâmico pela valina na posição número seis da cadeia β . A mudança de um aminoácido carregado negativamente por um aminoácido apolar gera uma diferença na estabilidade eletroforética da Hb S quando comparada com a Hb A. NAOUM, 1997. A substituição do ácido glutâmico pela valina na porção externa da molécula de Hb favorece a polimerização da Hb S sob condições de baixa concentração de oxigênio,

1.3.2 Talassemias

As talassemias compreendem um grupo de alterações hereditárias caracterizadas pela redução na síntese de uma ou mais cadeias polipeptídicas da globina, que é geralmente resultante de uma aneuploidia cromossômica e hipocromocítica. A redução da síntese pode ser total ou parcial e são classificadas segundo a cadeia globínica afetada e: alfa, beta, delta, delta-beta e gama-delta-beta talassemias (BONINI-DOMINGOS, 1993). Agregados instáveis das cadeias não pareadas precipitam-se, causando efeitos deletérios nos eritrócitos e seus precursores, podendo levar à destruição prematura dos eritrócitos na medula óssea e reduzindo a sobrevivência dos eritrócitos na circulação (EATHERALL; CLEGG, 1998).

As talassemias ocorrem em alta frequência no Mediterrâneo, África, Oriente Médio, Índia e sudoeste da Ásia. Além da importância diagnóstica, como o grupo mais comum de patologia genética na população mundial, proporcionam uma variedade de modelos de ocorrência natural para estudos de regulação da síntese das Hb, do seu desenvolvimento genético e da sua relação antropológica (EATHERALL, 1999).

As talassemias alfa e beta e sua forma heterozigota são as mais frequentes na população brasileira (NAOUM; BONINI-DOMINGOS, 1998). Mais de 80 mutações de talassemia alfa foram descritas, a maioria envolvendo deleções dentro da família alfa. Para as talassemias do tipo beta, 8 defeitos foram relatados, e sua maioria, mutações pontuais (TRAEGER-SYNODINOS et al., 2002).

A talassemia alfa é a doença genética mais frequente no mundo, inclusive na população brasileira, e afeta principalmente grupos de descendência asiática e alguns africanos (BONINI-DOMINGOS, 2004). A gravidade dos fenótipos talassêmicos depende da extensão da deleção e o quanto o gene afetado contribui para a síntese da cadeia alfa. Como o gene $\alpha 2$ produz 2,5 vezes mais RNA que o gene $\alpha 1$, alterações no gene $\alpha 2$ terão efeitos fenotípicos mais graves. Como as globinas alfa são sintetizadas desde o período fetal, as manifestações clínicas

decorrentes da deficiência na sua síntese pode ser observada tanto no período fetal quanto na fase adulta do indivíduo. CHUI; FUCHAROEN; CHAN, 2003.

A deficiência na produção de cadeias alfa resulta na tetramerização de cadeias beta, originando a Hb Hb₄. No período fetal, o desaparecimento resulta na formação de tetrameros de cadeias gamma formando a Hb Bart's. Esses hemotetrameros, que são relativamente instáveis, podem ser oxidados, formando precipitados intracelulares causando destruição na membrana do eritrócito e fagocitose prematura. CHUI; FUCHAROEN; CHAN, 2003. A frequência da talassemia alfa no Brasil foi por muitos anos subestimada devido à dificuldade de diagnóstico dessa alteração. BONINI-DOMINGOS, 2004.

A talassemia beta é causada principalmente por mutações pontuais que provocam a redução ou ausência na síntese da cadeia beta. BANERJEE, 2005. Algumas mutações são específicas de determinados grupos étnicos e mutações predominações nas populações Mediterrâneas, incluindo italianos, espanhóis e portugueses, são frequentemente encontradas na população brasileira. A mutação CD39 apresenta frequência elevada na região sudeste, enquanto na região nordeste do Brasil a mutação IVS-1 é mais encontrada. BONINI-DOMINGOS, 2004. A talassemia beta apresenta heterogeneidade molecular e expressão fenotípica variável. Na forma homocigota, pode ocorrer ausência ou deficiência acentuada na produção de cadeias beta, ocasionando o acúmulo contínuo das cadeias alfa livres que danificam a membrana do eritrócito causando destruição prematura dessas células por hemólise, levando à anemia grave. BANERJEE, 2005. Na forma heterocigota é caracterizada por anemia leve com discreta microcitose e hipocromia e aumento da resistência osmótica dos eritrócitos. As cadeias alfa livres podem combinar-se com cadeias delta e gamma, provocando aumento das Hb A₂ e Hb Fetal, respectivamente. Esses índices alterados são importantes ferramentas para o diagnóstico das talassemias do tipo beta. BONINI-DOMINGOS, 1993.

1.4 Programas de triagem neonatal

Os programas de triagem neonatal têm como objetivo o diagnóstico precoce de alterações genéticas e metabólicas antes do aparecimento dos primeiros sintomas, reduzindo a mortalidade e melhorando a qualidade de vida dos portadores. Entre os cronológicos pode-se considerar que a Triagem Neonatal (TN) na América Latina deu seus primeiros passos e passos dados da década de setenta, a partir de duas iniciativas que abriram espaço para um sistema de prevenção de doenças congênitas. A primeira dessas iniciativas surgiu em 1973 no México, quando Antonio Velásquez concebeu um projeto para a detecção de fenilcetonúria (PKU) e outras aminoacidopatias depois de realizar um treinamento com Robert Guthrie. A segunda destas iniciativas surgiu no Brasil em 1974, quando Benjamin Schmidt criou um laboratório na APAE-São Paulo para a detecção de PKU e outros erros inatos de metabolismo capazes de criar deficiência mental (BORRAJO, 2005).

1.4.1 Triagem Neonatal de Hemoglobinopatias

A Organização Mundial de Saúde, desde 1982, alerta para a utilização dos setores de saúde na detecção e prevenção dos indivíduos com doenças hereditárias. Estudos preventivos de hemoglobinopatias têm sido realizados e países nos quais a alta incidência constitui um fator de risco para a saúde pública, cada qual enfocando de maneira a sua população e o tipo de hemoglobinopatia que acomete. (H.O., 1982). O objetivo fundamental de uma campanha preventiva para hemoglobinopatias é a identificação precoce de crianças com a forma grave da doença hemozigota, seu encaminhamento para tratamento adequado antes do aparecimento dos primeiros sintomas e das complicações decorrentes da alteração genética e a orientação de portadores assintomáticos, heterozigotos, fonte de novos doentes (BIANCO et al., 1984; BONINI-DOMINGOS; SIQUEIRA, 2000).

O diagnóstico neonatal de hemorragias inopatas pode ser realizado tanto e sangue de cordão umbilical, quanto e amostras colhidas de punção plantar. Nos programas de triagem em larga escala distribuídos pelo mundo, as amostras são colhidas e papel e nas primeiras semanas de vida. FROST; BELLINGHAM, 1981; HAYASHI ET AL, 1981; EASTMAN, 1999. Para as hemorragias inopatas, a realização de medidas profiláticas e terapêuticas, como o uso de penicilina, vacinação e educação sobre o tratamento da doença, reduz a mortalidade e mortalidade devido à septicemia, causada principalmente pelo *Streptococcus pneumoniae*, e os quadros de seqüestro espinal nas doenças falciformes. EATHERALL; CLEGG, 2002; BOY, 2002; DAUDT et al., 2002.

Os neonatos diagnosticados como possíveis portadores de hemorragias inopatas devem ser reavaliados laboratorialmente por volta dos três meses de idade, se necessário, com a investigação laboratorial dos genitores. O tempo recomendado entre a detecção da anemia falciforme e o início do tratamento não deve exceder os quatro meses de vida para a adequada prevenção das infecções e outras complicações que podem levar ao óbito. Os componentes hemoglobínicos do neonato apresentam predomínio de cadeias alfa e gama. Após o nascimento e até aproximadamente seis meses de vida, há inversão na produção das cadeias gama e beta, podendo ser observados os valores definitivos do indivíduo adulto. As frações de HbA_{1c}, que aparece em pequena quantidade no período da triagem, podem ser melhor avaliadas após os seis meses. RAMALHO et al., 2002.

1.4.2 Implantação do Programa Nacional de Triagem Neonatal (PNTN)

Em seis de junho de 2000, o Ministério da Saúde instituiu a Portaria GM/MS nº 822 no âmbito do Sistema Nacional de Saúde (SUS), criando o Programa Nacional de Triagem Neonatal. Devido à diversidade de doenças existentes fora inseridas nesse programa alterações que podem ser detectadas por meio de testes

seguros, confiáveis e com rapidez com o caso da fenilcetonúria (P.U.), hipotireoidismo congênito (H.C.), anemia falciforme e outras hemoglobinopatias e a fibrose cística (F.C.). Esse programa permitiu normalizar, definir critérios de inclusão das doenças a serem triadas e cumprir os princípios de equidade, universalidade e integridade da triagem neonatal no país.

As etapas de implantação do PNTN foram: convocar 5% dos nascidos vivos no país, definir e ampliar a gama de patologias triadas, implantar a Rede Nacional de Triagem Neonatal com política baseada em Serviços de Referência e Triagem Neonatal (SRTN). Os fundamentos do PNTN foram os seguintes: a triagem neonatal, a confirmação do diagnóstico e o acompanhamento e tratamento dos doentes. A implantação do programa foi dividida em três fases sendo, na fase I incluídos os SRTN capacitados à triagem neonatal da P.U. e H.C.; na fase II, triagem de P.U., H.C. e a inclusão da anemia falciforme e outras hemoglobinopatias e na fase III os SRTN capacitados à triagem das patologias anteriores com inclusão da F.C. Atualmente estão habilitados no PNTN 25 estados com 33 SRTN credenciados. A cobertura nacional para o ano de 2005 de 5,3% dos nascidos vivos, sendo 1,7% triados apenas na fase I, 58,1% triados na fase II e 9,5% dos nascidos vivos triados também na fase III. A Figura 4 apresenta uma tabela ilustrando a cobertura estadual dos nascidos vivos e a fase em que se encontram os SRTN dos estados. CARVALHO, 2005.

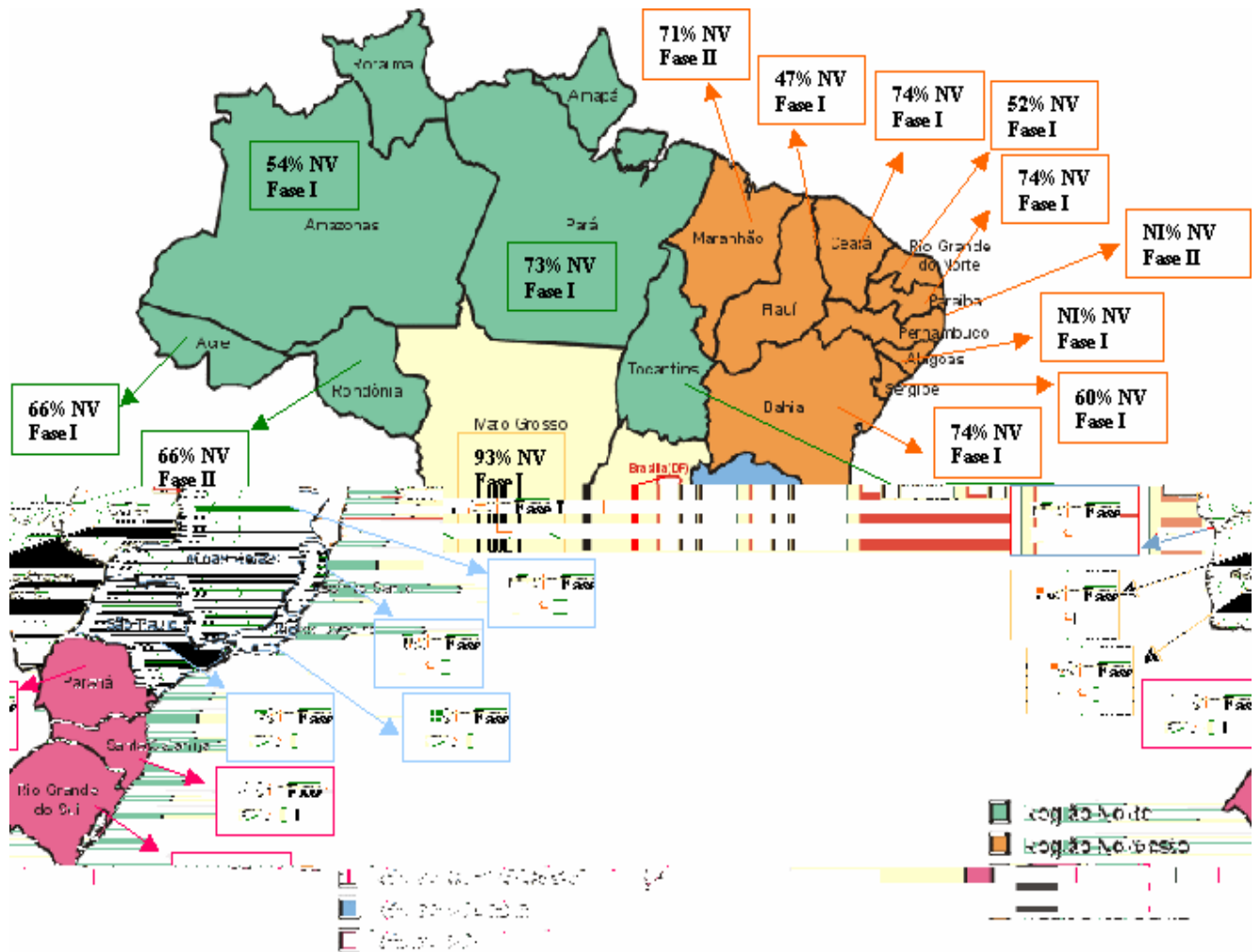


Figura 4: Mapa representativo da cobertura estadual dos nascidos vivos e da fase de credenciamento dos SRTN. NV = nascidos vivos. NI = não informado. Fonte: SIA-SUS, com adaptações.

1.5 Bases de dados Eletrônicas

Com o advento do computador, a necessidade de armazenar e manter informações tornou-se fácil e indispensável. A forma mais eficaz de gerenciamento de informações é realizada através de um Banco de Dados. Além de interligar todo o trabalho de organização, reduz custos, elimina duplicação de tarefas e ajuda na elaboração de estratégias. Banco de Dados ou Base de Dados é um conjunto de informações inter-relacionadas que representa as necessidades de um usuário de trabalho, incluindo real. Tal conjunto estruturado em arquivos de acordo com o ambiente utilizado e seus diferentes desenvolvimentos (ORTH, 1999).

Os antigos bancos de dados eram utilizados apenas para controle interno de empresas e laboratórios, contribuindo muito pouco para agilizar a vida dos usuários. Atualmente, com o auxílio da Internet, bancos de dados podem ser consultados por qualquer pessoa (ELMASRI, 2000). A biologia hoje, está sendo transformada por um crescimento explosivo de dados e surgindo dos laboratórios ao redor do mundo. Esta mudança consiste em transformar dados e conhecimento. Uma base de dados biológica é um grande corpo organizado de dados persistentes, usualmente associado a *softwares* computadorizados projetados para atualizar, pesquisar e recuperar componentes dos dados armazenados dentro do sistema.

Uma base de dados simples pode ser um único arquivo contendo muitas gravações, cada uma das quais incluindo o nome do conjunto de informações. Por exemplo, uma gravação associada com uma base de dados de uma sequência de nucleotídeos típica contém informações tais como, nome de contato, a descrição do tipo de colônia, o nome científico do organismo do qual foi isolada e frequentemente, a filiação associada com a sequência. Para a pesquisa e uma base de dados ser eficiente dois requisitos devem ser observados: fácil acesso à informação e um método para extrair somente a informação necessária para

responder a questãoiol gica especifica *

1.5.1 Data Mining

Nas últimas décadas a capacidade das empresas e laboratórios de gerar e coletar informações aumentou rapidamente. Esta explosão no volume de dados gerou a necessidade do desenvolvimento de novas técnicas e ferramentas que pudessem, além de processar os dados, permitir sua análise, de maneira inteligente e automática, para a descoberta de informações úteis. Isto acarretou o surgimento de um proeminente campo de pesquisa para a extração de informações e bases de dados, denominado *Knowledge Discovery in Databases*, ou ainda, Descoberta do Conhecimento em Bases de Dados. *Knowledge Discovery in Databases* é o processo de identificar padrões ou modelos que representem informações úteis, inéditas, potencialmente essenciais e essenciais para a coleta de dados (FAYYAD, 1999). Geralmente, técnicas de mineração de dados ou *data mining* possuem um papel preponderante nesse processo.

Data mining pode ser definido como o conjunto de técnicas automáticas para a exploração e grandes massas de dados, confeccionadas com o objetivo de encontrar novos padrões, tendências e relações. Através dessas técnicas, são possíveis a extração de informações úteis e o descobrimento de novos conhecimentos, os quais são difícilmente descobertos pelo ser humano, sendo melhor interpretados quando apresentados de forma gráfica. Os principais objetivos do *data mining* são a previsão e a descrição. Para a previsão, variáveis existentes na base de dados são utilizadas para prescrever valores desconhecidos futuros. Já a descrição é voltada para a busca de padrões, descrevendo os dados e a consequente apresentação para o usuário. (BARIONE et al., 2005)

* Disponível em <<http://nci.nih.gov/about/priorities/informatics.html>>. Acesso em 25 de outubro de 2005.

1.5.1.1 Histórico tecnológico

O termo *data mining* relativamente novo, datando dos anos noventa, por sua trajetória possui algumas décadas. O princípio da trajetória evolutiva da mineração de dados começou e se desenvolveu nos anos sessenta, no entanto as tecnologias da época eram extremamente restritas comparadas às atuais. Naquela década, o que se possuía eram coleções de dados gravados em fitas magnéticas e discos, nos quais a distribuição da informação era feita de forma estatística. A IBM era uma grande fornecedora deste tipo de tecnologia da época.*

Na década de setenta foi proposto o modelo relacional para a apresentação dos dados e, a partir de então, surgiram os sistemas gerenciadores de banco de dados relacionais-SGBDR. Até então, toda a tarefa de seleção de informações sobre os dados armazenados era feita através de métodos estatísticos puros como: desvio padrão, distribuição padrão, variância, análise de confiança, entre outros. Sem a estatística nada se teria de *data mining*, podendo-se dizer que, até os dias atuais, as ferramentas e técnicas *data mining* possuem em seu núcleo, todos os estatísticos.**

Uma outra vertente de métodos pertencentes à família do *data mining* a Inteligência Artificial (IA). O objetivo da IA é tentar simular a forma de pensamento humano para processar problemas estatísticos, baseando-se no conceito de heurística. Mas justamente por utilizar essa abordagem, esse processo de análise é um intenso processamento computacional, o que só foi possível nos anos oitenta com a evolução do poder dos computadores aliado ao baixo custo proporcionado pela época. IA teve algumas aplicações governamentais, científicas e militares, por isso o desenvolvimento da infra-estrutura pesada, ou seja, computadores com grande capacidade de processamento, tinha sua difusão comercial iniciada.*

* Disponível em: <http://dc.uel.br/~vh/anfredini/ia/Data_mining.htm>. Acesso em 25 de outubro de 2005.

** Disponível em: <http://data-mining-software.com/data_mining_history.htm> Acesso em 25 de outubro de 2005.

Foi nos anos noventa que o *data mining* tornou-se um forte conceito de aplicação devido a evolução de um outro tipo de abordagem: o aprendizado de máquinas. O aprendizado de máquinas consiste da junção das outras duas abordagens: estatística e IA. *Data Mining* é a adaptação do aprendizado de máquinas, por isso a abordagem comercial e cada vez mais tem sido aceito no mundo dos negócios e no científico ^{*}

Com o tempo, o *data mining* mostrou algumas deficiências e suas formas de representação, que até então era feita de forma textual ou numérica. Com isso, houve a necessidade de se alterar sua forma padrão de exibição para uma mais intuitiva e agradável para o usuário. Surgiram então as técnicas de visualização e de representação visual dos dados. Muitas dessas técnicas evoluíram juntamente com o avanço da computação gráfica e de áreas associadas. A partir de então, recursos no que concerne interação homem-máquina foram sendo adicionados conforme as necessidades exigidas dos usuários ^{**}

.5.2 Visual Data Mining (VDM)

A visualização gráfica consiste em apresentar graficamente dados nos quais se deseja analisar. Centrada em mecanismos que possibilitem ao usuário a compreensão rápida da informação a ele apresentada (CHITTARO, 2000). O volume de dados a ser analisado é menor, e na maioria das vezes não se sabe por onde começar o processo de extração da informação. Ou seja, todo o esforço e custo de coletar e armazenar dados pode ser infrutífero se não se souber recuperar dados pertinentes que respondam às consultas desejadas pelos usuários (BARIONE, et al., 2000)

^{*} Disponível em: <http://data-mining-software.com/data_mining_history.htm>. Acesso em 25 de outubro de 2005.

^{**} Disponível em: <http://dc.uel.br/~vh/anfredini/ia/Data_mining.htm> Acesso em 25 de outubro de 2005.

Devido à natureza das informações apresentadas nestes grandes volumes de dados serem textuais e numéricas, especialmente com muitos atributos, a interpretação destes é realizada com alguma eficiência por seres humanos. Apresentá-los de forma gráfica permite a melhor percepção dos dados e quais informações podem ser exploradas e extraídas a partir das visualizações criadas

CHITTARO, 2003.

A visualização dos dados estimula naturalmente a percepção e a inteligência humana, aumentando a capacidade de entendimento e associação de novos padrões. Dessa forma a visualização utiliza a percepção humana como o primeiro passo para descobrir valores. OLIVEIRA, 2003. A integração do ser humano ao processo de análise de dados pode ser muito proveitosa para explorar grandes volumes de dados. Isso permite combinar a flexibilidade, criatividade e conhecimento geral dos seres humanos com a enorme capacidade de armazenamento e poder computacional dos computadores atuais. EIM, 2002.

Com o crescente aumento no volume de dados, cientistas da área da computação começaram a pesquisar formas de ver e compreender a informação que está presente nesses dados, buscando extrair mais conhecimento dos dados.

TRAINA, 2002. Dessa forma, os pesquisadores reconheceram o potencial das técnicas de visualização para analisar e explorar grandes volumes de dados, e durante esses últimos anos, muitas técnicas para visualização de dados

multidimensionais foram desenvolvidas. EIM, 2002. As técnicas e ferramentas para visualização de dados são instrumentos indispensáveis ao processo de descoberta de conhecimento e têm provado seus valores na exploração e análise de dados. Sendo assim, essas técnicas têm sido utilizadas nos sistemas de *Visual Data Mining* (VDM). Com as técnicas de visualização, grandes volumes de dados podem ser apresentados na tela ao mesmo tempo; cores diferentes permitem ao usuário reconhecer, instantaneamente, similaridade e diferenças entre itens de dados; os itens de dados podem ser arranjados para expressar alguma relação entre si. TRAINA, 2002.

Os sistemas de VDM podem utilizar diversas técnicas de visualização e são de grande valor na exploração e análise de dados, apresentando

um alto potencial para mineração e grandes bases de dados. Hasegawa (1999) classifica entre dois enfoques para sistemas VDM: visualização suportando *data mining*, na qual os resultados são apresentados de forma gráfica; e a realização de *data mining* através de ferramentas visuais. Os sistemas VDM procuram integrar a capacidade de exploração do cérebro humano com os recursos computacionais, de forma a produzir um ambiente para a descoberta de conhecimento. A metodologia baseia-se na funcionalidade característica das estruturas internas dos dados e existência dos nós, e por outro lado, na capacidade do ser humano perceber padrões, exceções, tendências e relacionamentos ao analisar a representação visual dos dados de uma maneira gráfica. Hasegawa (1999).

As técnicas de VDM atualmente atacam três tópicos, segundo Traina (2002), a busca de agrupamentos, classificação de novos objetos e tratamento dos elementos que são exceções no conjunto de dados. A busca por agrupamentos permite indicar quais regiões no conjunto de dados são mais densas, além de apontar os objetos que se encontram isolados. A classificação de novos objetos visa identificar a que classe um certo objeto pertence. E por último, o tratamento de elementos que são exceções no conjunto de dados, permite a identificação de casos raros dentro desse conjunto. TRAINA, 2002.

1.6 Justificativa

Com a implantação do Programa Nacional de Triage Neonatal através da Portaria do Ministério da Saúde nº 822 as hematólogias foram oficialmente incluídas nos programas de triagem neonatal no Brasil. Essa inclusão levou a produção de uma grande quantidade de dados provenientes dos resultados das triagens. Essa quantidade de dados gerada pode conter informações úteis para facilitar o processo de análise das amostras, tais como padrões de resultados que indiquem determinadas alterações de Hb, ou ainda padronização da seqüência metodológica para direcionar o diagnóstico. No entanto, essas informações não têm sido exploradas ainda, pela falta de uma abordagem que permita a análise em larga escala desse montante de resultados.

Para se obter um aproveitamento real das muitas informações contidas nesse grande volume de dados, faz-se necessária a utilização de técnicas especializadas de tratamento de informação. As técnicas *data mining* têm por objetivo a extração de conhecimento capaz de indicar padrões e comportamentos ocultos e de difícil visualização a olho nu. Dentre as técnicas *data mining* utilizadas, destacam-se as denominadas regras de associação, que permitem gerar resultados indicativos do comportamento, por exemplo, de alterações genéticas e seu perfil de herança em diferentes grupos populacionais.

Nos anos de 1999 a 2002 foi desenvolvido um programa de triagem neonatal de hematólogias pelo Laboratório de Hematólogias e Genética das Doenças Hereditárias – LHGDH, do IBILCE, UNESP, com neonatos nascidos no Hospital de Base de São José do Rio Preto, em parceria com o Serviço de Genética da FAMERP. Nesse período foram analisados 59 recém-nascidos. O montante de dados gerados em todas as fases dos seis anos de programa realizado pelo LHGDH constitui uma fonte de informações, que armazenadas e em função de dados, permitem a análise por métodos computacionais na tentativa de estabelecer comportamentos, tendências e padrões, o que, visando dessa forma, contribuir com a Triage Neonatal no país, fornecendo bases para agilizar os diagnósticos e torná-los cada vez mais precisos.

2 O jetivos

2 Objetivos

Analisar os resultados do programa de triagem neonatal de hemoglobinopatias por métodos computacionais, objetivando avaliar as interfaces entre os diferentes fenótipos de hemoglobina e padrões de comportamento nos grupos de indivíduos analisados ao longo dos seis anos de programa neonatal em São José do Rio Preto.

2.1 Objetivos específicos

Após desenvolvimento de uma Base de Dados Eletrônica e inserção dos resultados da triagem objetivou-se:

1. Por meio de ferramenta *data mining* – CLIBIA avaliar as metodologias utilizadas e testar a aplicabilidade de um banco de dados para hemoglobinopatias.

2. Analisar os dados quantitativos de hemoglobinas obtidos por HPLC nos diferentes fenótipos, e sistema de visualização 3D - *FastMapDB*, para análise de comportamento dos fenótipos.

3 Material e Métodos

3 Material e Métodos

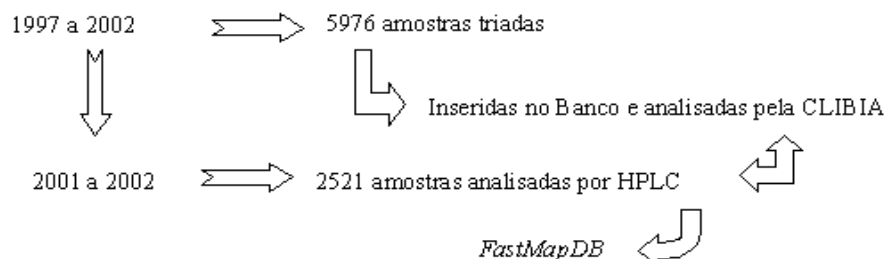
3.1 Casuística

Fora analisados os resultados laboratoriais de 59 amostras de sangue de cordão umbilical de recém-nascidos no Hospital de Base de São José do Rio Preto no período de 1999 a 2002, de ambos os sexos e distintas origens raciais. As amostras de sangue de cordão umbilical foram colhidas após o parto, e tu os co anticoagulante, identificadas com o RN de mãe e da mãe, por gravidade e número ordenha do cordão, e acondicionadas em geladeira até o momento das análises, que não excedera uma semana. A coleta das amostras foi realizada por profissionais da enfermagem do Hospital de Base, e plantada na sala de parto. As informações sobre a pesquisa foram repassadas aos pais ou responsáveis, por profissionais de assistência social da equipe multidisciplinar, e o termo de consentimento livre e esclarecido devidamente preenchido e assinado. Cada amostra ao chegar ao laboratório, foi cadastrada, recebendo um número de código, ficando os dados referentes a aquela amostra sob o conhecimento do responsável pela pesquisa.

Fora utilizados materiais descartáveis, para total segurança dos indivíduos envolvidos, não oferecendo riscos aos seus participantes. Após a realização dos exames, os casos com alterações de hemoglobinas foram encaminhados ao serviço de genética e consultório de hematologia pediátrica da Faculdade de Medicina de São José do Rio Preto, para estudo familiar, aconselhamento genético-educacional e suporte clínico, sendo preservadas a segurança e confiabilidade dos resultados. Esta pesquisa teve aprovação no CONEP sob o número 399, processo 25.349/2002-88.

Os resultados das 59 amostras foram inseridos no Banco de Dados do LHGDH para aplicação de ferramenta *data mining* denominada CLIBIA (*Clinical Bioinformatic Analysis*), especialmente desenvolvida para esse fim. Para aplicação da ferramenta de *Visual data mining*, *FastMapDB*, foram utilizados apenas os resultados das 252 amostras triadas nos anos de 2000 e 2002, que incluam os

resultados quantitativos obtidos por Cromatografia Líquida de Alta Performance (HPLC). A análise está representada no organograma a ilustrando o esquema de análise.



3.2 Métodos para o diagnóstico de Hemoglobinopatias

Após os procedimentos de identificação, as amostras foram submetidas aos seguintes testes de triagem para hemoglobinopatias:

3.2.1 Preparação de hemolisados

Para que as amostras fossem submetidas aos procedimentos eletroforéticos e testes bioquímicos as células foram lisadas para a obtenção da solução de H utilizando duas metodologias:

- Hemolisado Rápido - com saponina, NAOUM, 99
- Solução de H - com clorofórmio, NAOUM, 99

Hemolisado Rápido: com saponina

Reativo hemolisante:

- Saponina P.A.
- água destilada

g
L

Procedimento:

- E placa de Ca^{2+} line foi colocado volume de sangue com volume de reativo heparinizado, com posterior homogeneização até a hemólise completa da célula;

- O hemolisado pode ser utilizado após 5 minutos, e no máximo 4 horas depois da sua preparação.

Solução de Hb: com Clorofórmio

Procedimento:

- Para lavar os eritrócitos, centrifugou-se 1 L de sangue colhido com anticoagulante com solução salina a 0,85%, a 1.500 rpm, durante 5 minutos, descartando o sobrenadante. O processo foi realizado três vezes, no mínimo.

- Ao volume de eritrócitos lavados, adicionou-se outro de água destilada, homogeneizando a solução. A seguir, adicionou-se um volume de clorofórmio idêntico ao do hemolisado formado. A mistura foi agitada vigorosamente e centrifugada a 2.000 rpm, por 2 minutos.

- A solução de hemossobrenadante, ou hemolisado, foi retirada por meio de pipeta Pasteur e transferida para um tubo limpo com identificação da amostra. A concentração do hemolisado, preparado conforme a metodologia apresentada, variou de 0 a 5 g/dL.

3.2.2 Resistência globular osmótica em solução de NaCl a 0,36% (SILVESTRONI; BIANCO, 1975)

Princípio:

Técnica utilizada para detectar talassemias do tipo beta, principalmente na forma heterozigota, pois nesses casos os eritrócitos microcíticos são mais resistentes à hemólise nesta solução. A resistência globular específica para talassemia beta heterozigota, já que resultados positivos são encontrados também em anemias carenciais e outras hemoglobinopatias, como nos heterozigotos para Hb C.

Reagentes:

Solução estoque - NaCl a 0,36% - pH 7,4

- NaCl

- Na₂HPO₄

- NaH₂PO₄.H₂O

- água destilada q.s.p.

9,3 g
0,37 g
0,28 g
1 L

Solução de traço alho

- NaCl 0,3%
- água destilada q.s.p.

3,0 L
 .
 000 L

Procedimento:

Em tubo de hemólise colocou-se 2,0 mL de solução de NaCl a 0,3% e 20 µL de sangue total, agitando por inversão, suavemente. A leitura foi feita após 5 minutos.

Interpretação:

O tubo de hemólise colocado na solução de NaCl a 0,3% foi colocado a 2,0 cm de uma folha branca com linhas negras. O teste foi interpretado como positivo quando as linhas negras não foram visualizadas, ou a vez a resistência auentada à hemólise do eritrócito tornou a amostra opaca. As linhas foram facilmente visualizadas e as amostras com resistência normal à hemólise.

3.2.3 Análise, a fresco, da morfologia eritrocitária (BONINI-DOMINGOS, 2003)

Os esfregaços sanguíneos, a fresco, foram analisados ao microscópio, quanto ao tamanho, forma e quantidade de hemóglóbulos eritrócitos. Os resultados foram divulgados da seguinte maneira, segundo padronização do LHGDH para cada um dos parâmetros avaliados.

- alterações discretas: +)
- alterações moderadas: ++)
- alterações acentuadas: +++)
- células normais: N)

3.2.4 Eletroforese em pH alcalino (MARENGO; ROWE, 1965)

Princípio:

Técnica utilizada para qualificação e quantificação de H⁺ norais e grande parte das H⁺ arorais com mobilidades eletroforéticas diferentes das H⁺ norais.

Reagentes:

Tampão Tris-EDTA-Borato (TEB), pH 8,4

- Tris hidróxi etil a mino etano 1,2 g
 - ácido etilenodiaminotetracético 0,4 g
 - ácido Bórico 3,2 g
 - água destilada q.s.p. 1 L
- Conservado e geladeira

Corante:

Ponceau

- Ponceau S 0,5 g
- ácido tricloroacético 5,0 g
- água destilada q.s.p. 1 L

Solução descorante:

- ácido acético glacial 1 L
- Metanol 5 L
- água destilada q.s.p. 1 L

Procedimento:

- As fitas de acetato de celulose foram desidratadas e tampão TEB por 5 minutos no vácuo e 4 horas, no vácuo.

- Após serem secas e folhas de papel absorvente, as fitas foram colocadas na câmara de eletroforese contendo o tampão utilizado para

e e e as fitas, conectando-as com os dois pontos eletrolíticos através de tecido absorvente (*perfex*).

- A solução de H₂O₂ foi aplicada a cada extremidade da fita e contato com o polo negativo.

- As amostras foram submetidas a 3 volts por 3 minutos.

- As frações foram analisadas, primeiramente, sem coloração e, posteriormente, coradas com Ponceau. Para corá-las, as fitas foram colocadas no corante por 5 minutos, e em solução descorante por 3 minutos, com agitação da vasilha.

3.2.5 Transparentização (BONINI-DOMINGOS, 2003)

Realizada para possibilitar a conservação das fitas de acetato de celulose para registros.

Reagentes:

- Metanol P.A. 25 L

Solução de Transparentização:

- ácido acético glacial 1 L
 - Metanol 42,5 L
 - Glicerina 0,5 L

Procedimento:

- A fita de acetato de celulose foi mergulhada no etanol por no máximo 10 segundos.

- Em seguida, ela foi transferida para a solução de ácido acético: etanol: glicerina por 4 segundos.

- A fita de acetato foi então colocada sobre uma lâmina de vidro e levada à estufa a 60°C para secar por, aproximadamente, 2 minutos.

3.2.6 Eletroforese em pH neutro (DACIE; LEWIS, 1985)

Princípio:

Técnica utilizada para identificação e quantificação das H⁺, H⁻ e H⁰ Bart's que apresenta perfil de migração e pH alcalino similar a proteínas plasmáticas.

Reagentes:

Tampão pH neutro

- H_2PO_4

3, g

- Na_2HPO_4

3, g

- água destilada q.s.p.

1 L

000

Conservar em geladeira

Procedimento:

- As fitas de acetato de celulose foram desidratadas e tampão neutro por 5 minutos no vácuo e 1 hora, no vácuo.

- Após serem secas e folhas de papel absorvente, as fitas foram colocadas na câmara de eletroforese contendo o tampão utilizado para eleger as fitas, conectando-as aos compartimentos eletrolíticos através de tecido absorvente (*perfex*).

- A solução de H⁺ foi aplicada a cada extremidade da fita em contato com o polo negativo.

- As amostras foram submetidas a 3 Volts por 3 minutos.

- As frações foram analisadas se coloração, seguindo o protocolo de identificação.

3.2.7 Eletroforese em pH ácido (VELLA, 1968)

Princípio:

Técnica utilizada para diferenciar alguns tipos de H⁺ que migram e posições semelhantes na eletroforese em pH alcalino e caracterização semi-quantitativa de H⁺ F.

Reagentes:

Tampão Fosfato pH 7,2

Para uso nos componentes eletrolíticos e confecção do gel:

- Na₂HPO₄

2,2 g

- NaH₂PO₄.H₂O

1 g

- água destilada q.s.p.

1 L

Conservar em geladeira

Gel de agar-Fosfato

- agar-agar

5 g

- Tampão fosfato pH 7,2

25 L

Procedimento:

- Os componentes do gel de agar-fosfato foram adicionados a um erlenmeyer de 25 L e levados ao forno microondas até completa dissolução, tomando cuidado para a mistura não ferver.

- Foram pipetados 5 L do gel e lâminas de microscópio que gelificaram à temperatura ambiente.

- As lâminas foram aplicadas na porção dia da lâmina, inserindo o aplicador com cuidado para não partir totalmente o gel.

- Para conexão do gel com os componentes eletrolíticos foi utilizado tecido absorvente (*perfix*).

- As lâminas foram submetidas a 200 volts por 3 minutos.

- As frações foram analisadas se corar e coradas com Ponceau.

3.2.8 Pesquisa de corpúsculos de Heinz e agregados de Hb H (PAPAYANNOPOULOS; STAMATAYANNOPOULOS, 1974)

Princípio:

Os corpúsculos de inclusão de Hb H são formados por cadeias beta oriundas da desnaturação do tetramero da Hb. Após coloração esses corpúsculos

apresenta -se dispostos homogeneamente no interior dos eritrócitos como pequenos pontos azulados.

Reagentes:

Solução salina:

- Cloreto de sódio 0,9 g
- água destilada q.s.p. 100 L

Solução citrato:

- Citrato de sódio 2,2 g
- água destilada q.s.p. 100 L

Solução de Azul Cresil Brilhante (ACB):

- Azul Cresil Brilhante 0,1 g
- Solução salina 100 L
- Solução citrato 25 L

Procedimento:

- Eritrócitos de ensaio pequeno foram colocados 5 µl de sangue total e adicionados 10 µl de ACB.

- Após agitação suave dos tubos, o material foi incubado a 37°C por 30 minutos.

- Foram feitos esfregaços finos os quais foram examinados ao microscópio e ojetiva de imersão.

Interpretação:

Os corpúsculos de Heintz apresentaram-se dispostos homogeneamente no interior dos eritrócitos como fina granulação distribuída homogeneamente, caracterizando um portador de alfa talassemia. Os Corpos de Heinz, provenientes da desnaturação de Hb, apareceram como uma precipitação grosseira junto à periferia dos eritrócitos.

3.2.9 Cromatografia Líquida de Alta Performance (HPLC) (INSTRUCTION MANUAL, 1999)

Princípio:

O equipamento utilizado foi o VARIANT da BIO-RAD com kit de análise Sickle Cell Short Program. A HPLC neste equipamento consiste da cromatografia de troca iônica e sistema fechado, no qual duas colunas de sílica duplo e unidade de tampões de diluição, com controles de gradientes pré-programados passa pela coluna detectando as alterações de absorbância a 450 nm. O filtro secundário de 0,9 µm corrige a linha de base para efeitos provocados pela mistura de tampões com forças iônicas diferentes. EASTMAN et al., 1999. As mudanças na absorbância são monitoradas e exibidas como um cromatograma da absorbância versus tempo. Os dados de análise provenientes do detector são processados por um integrador e utilizados e impressos no relatório da amostra de acordo com o tempo de retenção. O tempo de retenção é o tempo transcorrido entre a injeção da amostra até o pice do pico da Hb. Cada Hb tem um tempo de retenção característico. No final da análise da amostra, uma cópia do cromatograma e os dados do relatório são automaticamente impressos.

Procedimento:

Para kit Sickle Cell Short Program:

Em tubo, foram injetados 5 µL de sangue total com 0,1 mL de água Milique.

Após a análise total, as amostras foram acondicionadas nos recipientes adequados e alojadas no equipamento para realização dos procedimentos de leitura das amostras pré-programados.

Interpretação:

A quantificação das diferentes frações de Hb em uma amostra foi realizada a partir dos valores de porcentagem e tempo de retenção comparados com os valores de calibração específicos fornecidos pelo fabricante e utilizados em modelo prévio que incluiu valores numéricos e perfil cromatográfico. Este kit identifica

as seguintes frações heptagônicas: $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$ e $\frac{1}{5}$, sendo as duas primeiras, suas frações de eluição rápida e as duas últimas as H₂O não encontradas e recém-nascidos.

3.3 Ferramentas *Data Mining*

Para analisar os dados por VDM, foi utilizada uma técnica de visualização e três dimensões denominada *FastMapDB*. O *FastMapDB* é uma ferramenta de visualização de dados, com dados, números e textos, onde a apresentação é efetuada através do mapeamento multidimensional dos dados para um espaço tridimensional, proporcionando a redução de dimensionalidade. O objetivo da redução de dimensionalidade é apresentar o conjunto de dados com um número menor de atributos, por preservar as características inerentes à informação armazenada, dessa forma buscando definir quais são os atributos importantes ou que concentram a informação.

O núcleo da ferramenta *FastMapDB* utiliza o algoritmo denominado *fastmap* (FALOUTSOS; LIN, 1995), que realiza o mapeamento de objetos que podem estar em espaços de diferentes dimensões para o espaço Euclidiano. O mapeamento feito com o *fastmap* procura preservar as distâncias entre os dados e evitar possíveis distorções do mapeamento. Outra característica permitir ao analista compreender como os itens de dados estão relacionados, quais são os padrões envolvidos e detectar agrupamentos e elementos de exceção. A visualização gerada pela ferramenta é simples, e rapidamente manipulada, o que permite que hipóteses sejam formuladas e testadas antes que sejam efetuados cálculos estatísticos e numéricos (FALOUTSOS; LIN, 1995).

A ferramenta *FastMapDB* foi desenvolvida para disponibilizar os recursos de visualização para dados armazenados e bases relacionais utilizando resultados e algoritmos originalmente desenvolvidos para a redução de dimensionalidade ou de mapeamento de dados tridimensionais para espaços bidimensionais. A ferramenta possui recursos para que o usuário crie

interativa entre a função de distância vetorial a partir de qualquer quantidade de atributos de uma tabela e a base de dados. Com a função distância definida o programa escolhe pivôs dentre os pares de tuplas, cada objeto da tabela ou cada amostra, mais distantes entre si. Com a definição dos pivôs, o sistema aplica o conjunto de dados para a representação tridimensional, sendo que cada objeto na tabela representado por um ponto no espaço. Este sistema permite ainda, utilizar um atributo da tabela para classificar as tuplas e a visualização das classes representadas por diferentes cores e formatos. TRONCO, 2003.

A ferramenta *FastMapDB* está disponível para download em <http://www.gdi.icc.usp.br>, foi desenvolvida utilizando a linguagem de programação C++ e a biblioteca gráfica OpenGL. TRONCO, 2003.

Para aplicação da ferramenta *datamining - FastMapDB*, foram utilizados os resultados de 252 amostras de recém-nascidos nos anos de 2000 e 2002 analisados por Cromatografia Líquida de Alta Performance, HPLC, realizada com o Sistema automatizado VARIANT BIO-RAD, utilizando os reagentes e o protocolo do fabricante para esta faixa etária. Esse sistema identifica os valores percentuais das frações de H e picos determinados, F_{fast} , F_1 , F_2 e A , estes foram inseridos em planilhas e convertidos para a base de dados reconhecível pela ferramenta, sendo utilizados como atributos para o planejamento espacial e, os diferentes fenotipos foram utilizados como atributos classificatórios das amostras. Para esta eleição de planejamento espacial os resultados foram classificados em grupos segundo delineamento ilustrado no quadro .

Quadro 1. Divisão e grupos de análise para o *FastMap*, segundo os fenótipos de Hb observados.

n	Fenótipos de Hb	Grupos de análise para o <i>FastMap</i>
29	Perfil de Hb normal para idade	“Normal”
55	H S e heterozigose	“Variante”
8	H C e heterozigose	
2	Beta talassemia heterozigota	“Talassemia”
258	Alfa talassemia heterozigota	

Os resultados laboratoriais das 59 amostras foram cadastrados no Banco de dados do LHGDH, desenvolvido e colocado no Laboratório de Banco de Dados do Departamento de Ciências da Computação e Estatística da Unesp. O cadastro dos pacientes foi realizado via Internet e o endereço de acesso ao banco é <http://www.lhgdh.locawe.com.br>. Após o cadastro, as amostras foram analisadas utilizando a ferramenta CLIBIA, especialmente desenvolvida para esse banco de dados. A CLIBIA é a ferramenta de análise estatística desenvolvida utilizando tecnologia ASP (*Active Server Pages*), para a análise *on-line* do banco de dados do LHGDH. Tem como função ajudar na análise dos dados contidos no banco. A CLIBIA é composta por duas opções: Consulta e Relatório. Para utilizar as duas opções o usuário deve definir o *dataset*, isto é, o conjunto de dados a ser pesquisado. O *dataset* pode ser delimitado por vários campos como Código do Paciente, Nome, Laboratório, Laudo, entre outros.

3.4 Análises Estatísticas

Aos valores quantitativos de H₂O₂ obtidos por HPLC foram aplicados testes estatísticos descritivos e inferenciais. Para as análises inferenciais, os dados foram testados quanto à normalidade e homocedasticidade. Quando as condições de normalidade e homocedasticidade foram satisfeitas aplicaram-se os testes paramétricos, ANOVA-comparação múltipla por Tukey, para comparar três ou mais grupos e Teste t para dois grupos. Quando os dados não foram paramétricos, após a transformação dos valores, foram utilizados os testes não paramétricos de Kruskal-Wallis com comparação múltipla por Dunn para comparar três ou mais grupos e Mann-Whitney para dois grupos. (ZAR, 1999)

Resultados

4 Resultados

4.1 O banco de dados e a ferramenta CLIBIA

Os resultados laboratoriais de 59 amostras analisadas no período de 1999 a 2002 foram inseridos no banco de dados eletrônico, LHGDH *on-line*. Por conter informações de exames os usuários são cadastrados e possuem uma senha de acesso individual. A Figura 5 ilustra a página de acesso ao banco. Na página inicial existe um *link* para a ferramenta CLIBIA que pode ser visualizado no destaque da Figura 6. Esse *link* abre a página principal da ferramenta, na qual selecionou-se a opção consulta ou relatório. Figura

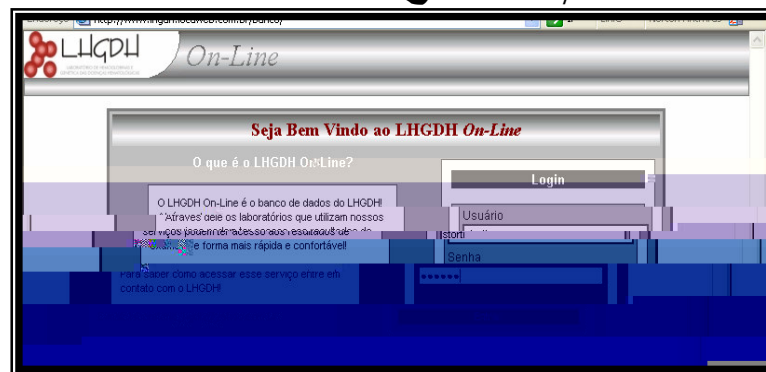


Figura 5: Página inicial de acesso ao banco de dados LHGDH *On-line*.

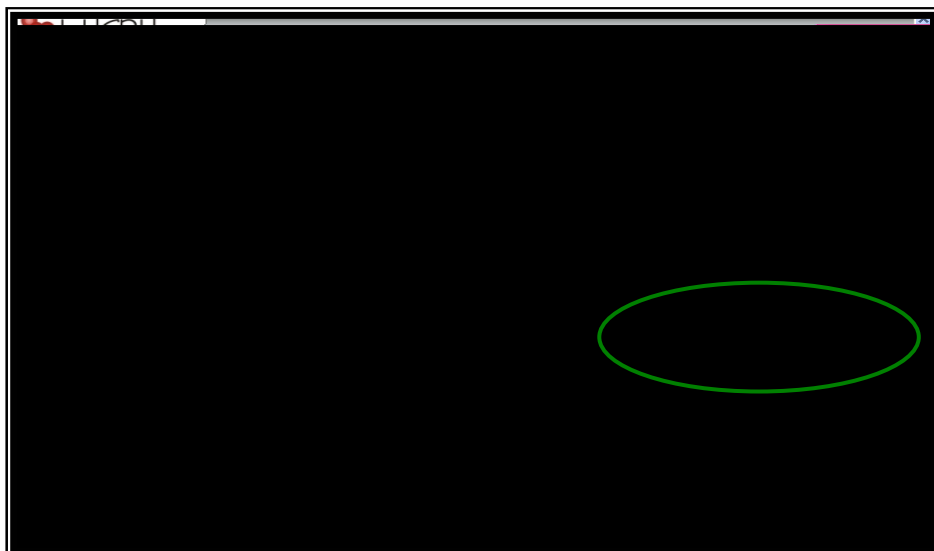


Figura 6: Página inicial do banco de dados. O círculo verde destaca o *link* para a ferramenta CLIBIA.

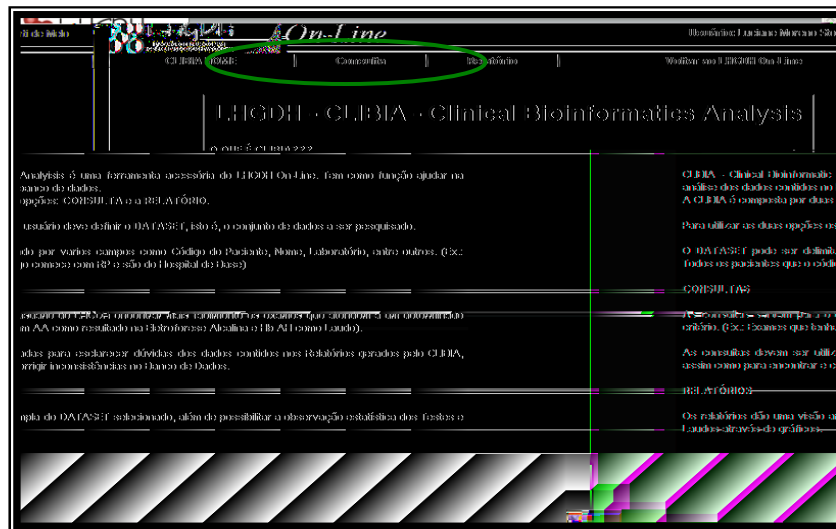


Figura 7: Página inicial da ferramenta CLIBIA. O círculo verde destaca os links para a seleção das opções consulta ou relatório.

Através do recurso de consulta específica foi possível comparar resultados obtidos e diferentes testes laboratoriais e acessar cada uma individualmente permitindo a observação de todas as informações referentes a cada uma. A Figura 8 ilustra a página de seleção do *dataset* para consulta. Nesta página, foi possível usar as alternativas, selecionando-se aquelas que apresentaram resultados específicos, definindo informações dos testes laboratoriais ou dos laudos dos exames.



Figura 8: Página de seleção do *dataset* na opção consulta da ferramenta CLIBIA.

Após a seleção do *dataset* a ferramenta localizou os exames que se enquadraram nos requisitos selecionados e dirigiu para uma tela com uma tabela de resultados. Nessa tela foi possível clicar em cada exame e ao clicar no *link* do nome do paciente. A Figura 9 ilustra os resultados da busca realizada selecionando-se como conjunto de dados todos os exames e que o nome do paciente tivesse as iniciais RN, com zero anos, e que fosse provenientes do Hospital de Base de São José do Rio Preto.

Figura 9: Página exibindo a tabela de resultados da busca selecionando-se recém-nascidos no Hospital de Base como *dataset* na operação consulta da ferramenta CLIBIA.

Através da operação consulta foi possível relacionar os resultados de um teste laboratorial e o laudo do exame, ou relacionar os resultados de diferentes testes, permitindo a avaliação da concordância dos testes entre si e dos testes com os diagnósticos. Dessa forma foi possível identificar discordâncias nas análises, e também testar a eficiência dos testes para identificação das alterações de Hb.

Na operação relatório da ferramenta CLIBIA, depois de delimitado o *dataset*,

The screenshot shows a web application interface for 'LHGDPH On-Line'. The user is logged in as 'Luciane Moreno Storti de Melo'. The main menu includes 'CLIBIA HOME', 'Consulta', 'Relatório', and 'Voltar ao LHGDPH On-Line'. The current page is titled 'LHGDPH - CLIBIA - RELATÓRIO'. The main content area is titled 'Selecionar DATASET' and contains a form with the following fields:

- Doc. Paciente: [Empty text box]
- Nome do paciente: RN
- Idade entre: 0 e 0
- Período de: 01/01/1990 até 01/01/2006
- Laboratórios: Hospital de Base S. J. do Rio Preto/São José do Rio Preto-SP
- Cadastrado Por: Todos os usuários...

Below the form is a 'Configurações' section with several checkboxes:

- Mostrar gráfico dos testes
- Mostrar gráfico dos resultados
- Mostrar gráfico dos laudos
- Mostrar gráfico dos laudos específicos
- Mostrar gráfico dos exames
- Mostrar gráfico das etnias
- Mostrar gráfico das regiões

At the bottom of the configuration section is a button labeled 'Gerar Relatório'.

Figura 10: Página de seleção do *dataset* na operação relatório da ferramenta CLIBIA.

Os relatórios gerais contêm a descrição dos dados, fornecendo rapidamente as proporções entre gêneros, idade, etnia, região de origem, além de fornecerem as frequências dos resultados obtidos nos testes laboratoriais e dos laudos dos exames. A exibição foi realizada através de gráficos, coloridos e com legenda que agiliza o processo de análise dos dados. O relatório geral mostrou os seguintes perfis hemoglobínicos para as amostras de neonatos: 8,9% com perfil normal para a idade, 9,9% com resultados sugestivos para alfa talassemia, 0,8% com resultados sugestivos de beta talassemia, 3,38% com hemoglobinas variantes, 0,82% com interações entre hemoglobinas variantes e talassemias e 2,2% amostras com presença de hemoglobinas normais, as que sugeriram contaminação com o sangue materno. A Figura apresenta o gráfico da distribuição dos laudos dos exames.

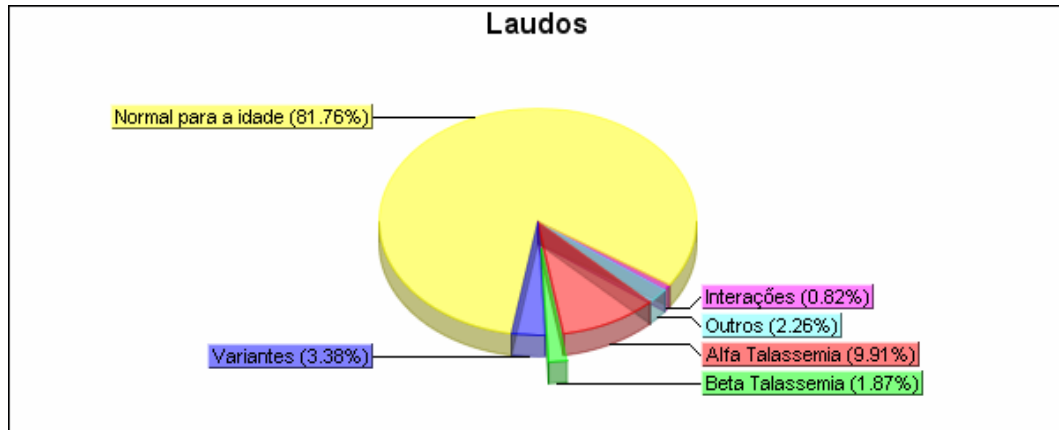


Figura 11: Gráfico de pizza elaborado pela operação relatório da ferramenta CLIBIA incluindo os laudos dos exames e as frequências fenotípicas das Hemoglobinas servadas.

O relatório geral completo elaborado para as amostras de recém-nascidos no Hospital de Base de São José do Rio Preto encontra-se no Apêndice deste trabalho.

4.2 Padrões de Hb obtidos na análise pelo *FastMapDB*

4.2.1 Perfil das amostras normais

Visando esta eleger o perfil do fenótipo normal para as amostras de sangue de neonatos, foram selecionados os resultados das 20 amostras com resultados normais para idade, obtidos por HPLC com o conjunto amostral. O teste então foi realizado utilizando os quatro picos descritos pelo HPLC, previamente descritos na metodologia. A visualização tridimensional obtida está ilustrada na Figura 2, e as amostras normais distribuíram-se em dois aglomerados distintos, marcados pelos traçados e cores azul e verde. A observação da distribuição sugeriu a existência de algum fator adicional dividindo o conjunto amostral. Para avaliar qual seria esse fator, responsável pela divisão das amostras em dois grupos de distribuição espacial, realizou-se o teste selecionando os diferentes picos como atributos. A análise projetada incluindo os picos F_1 , F_2 e A no teste então existiu uma aproximação dos aglomerados, no entanto, ainda pode ser observada a presença de dois grupos distintos, como ilustrado na Figura 3.

A Figura 4 representa o teste utilizando os picos F_{fast} , F_2 e A , excluindo-se o pico F_1 , e pode-se observar uma aproximação maior dos aglomerados indicando grande participação desse atributo na divisão das amostras. Associando-se as duas observações, projetou-se uma análise utilizando apenas os picos de F_2 e A , excluindo as janelas de F_{fast} e F_1 . Com o teste realizado na Figura 5, nesse teste então teve-se a dispersão linear dos dados.

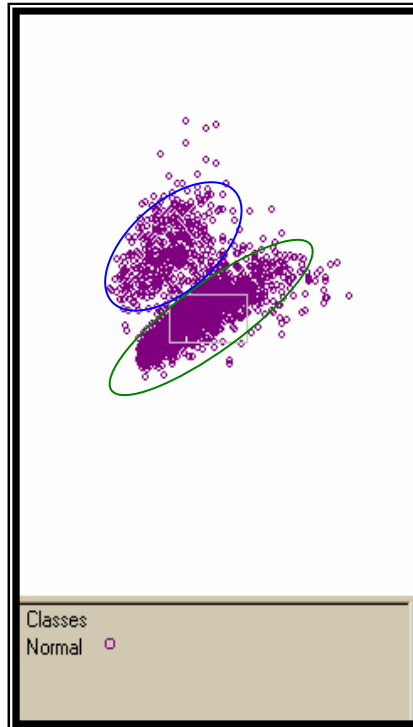


Figura 12: Projeção da imagem tridimensional para o grupo “Normal”, incluindo os quatro picos com o atributo de aparência.

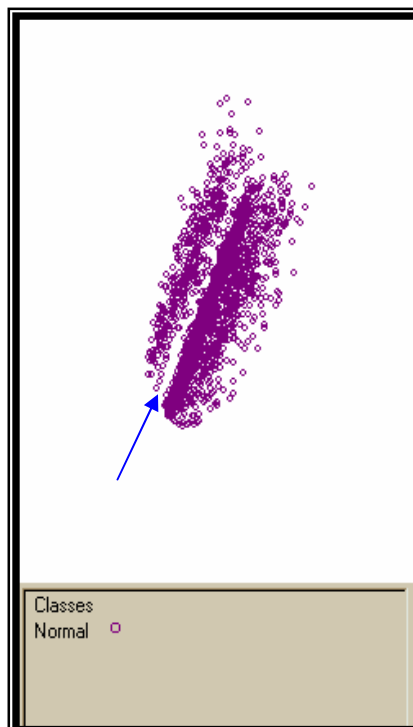


Figura 13: Projeção da imagem tridimensional para o grupo “Normal” incluindo os picos E , E' e A com o atributo de aparência. A seta indica o espaço entre os aglomerados.

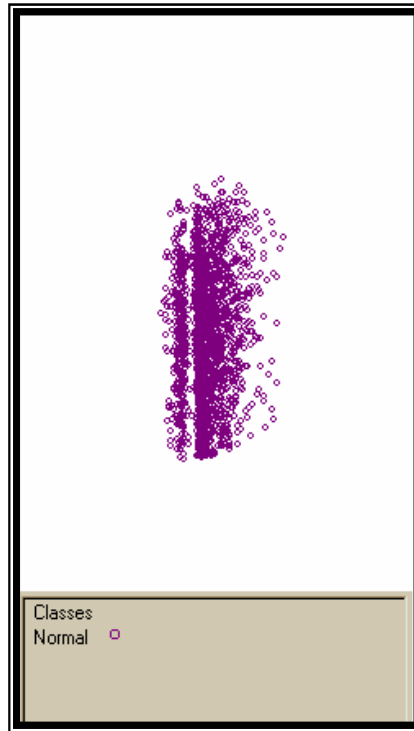


Figura 14: Projeção da imagem tridimensional para o grupo "Normal" incluindo os picos "Fast", "E" e "A" com outros pontos de referência.

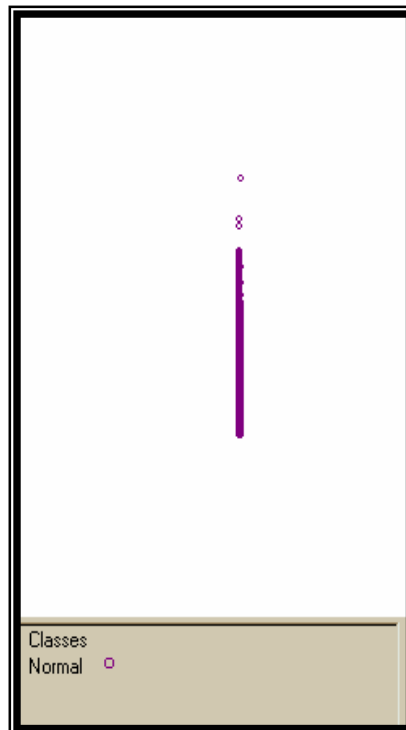


Figura 15: Projeção da imagem tridimensional para o grupo "Normal" incluindo os picos "E" e "A" com outros pontos de referência.

Para avaliar os valores de F_{fast} e F , foram geradas imagens utilizando o recurso da ferramenta *FastMapDB*, no qual o usuário pode ser utilizado para dividir o conjunto aóstral em dois subgrupos exatamente no valor médio entre os valores mínimo e máximo. A imagem ilustrada na Figura 7 contém os quatro picos de análise e os valores de F_{fast} foram utilizados para classificar as aóstras em dois subgrupos representados na imagem pela diferença no tamanho dos pontos. Como observado, os pontos marcados pelo círculo azul apresentam valores de F_{fast} entre 0,0 e 5,5 % e os pontos dentro do círculo verde correspondem a valores de F_{fast} de 5,5 a 9%. Destaca-se que não existe sobreposição dos pontos indicando que os dois subgrupos correspondem a duas nuvens distintas. Analisando estatisticamente esses dados, obteve-se a média de 2,34 % para o subgrupo com valores variando de 0 a 5,5 % e este incluiu um total de 24 aóstras. O subgrupo variando de 5,5 a 9% foi composto por apenas 29 aóstras e a média observada foi 3,5%. A diferença percentual obtida no pico de F_{fast} para todas as aóstras normais foi de 2,44%, apresentando diferença estatisticamente significativa $p < 0,05$ das médias obtidas para os dois subgrupos acima descritos.

Utilizando o mesmo recurso da ferramenta *FastMapDB*, a classificação foi feita utilizando-se o pico de F para dividir as aóstras no valor médio e a imagem obtida está ilustrada na Figura 8, na qual a distribuição espacial dos dados foi muito semelhante à da Figura 2. As duas nuvens classificadas pelos valores observados na Figura 7 não apresentaram sobreposição dos pontos. Para possibilitar a comparação das imagens os pontos das Figuras 2 e 7 foram dispostos lado a lado na Figura 8. Para caracterizar a divisão em duas nuvens aóstrais, calculou-se a diferença percentual para cada subgrupo obtido para o pico de F . Obteve-se para o subgrupo com valores variando de 0,0 a 3,3% uma diferença percentual de 9,45% incluindo 39 aóstras e para o subgrupo com valores variando de 3,3 a 2,9% obteve-se a média de 2,25%, sendo composto por 8 aóstras. A diferença percentual obtida para todas as aóstras normais foi de 0,48%, apresentando diferença estatisticamente significativa $p < 0,05$ das médias obtidas para os dois subgrupos de F .

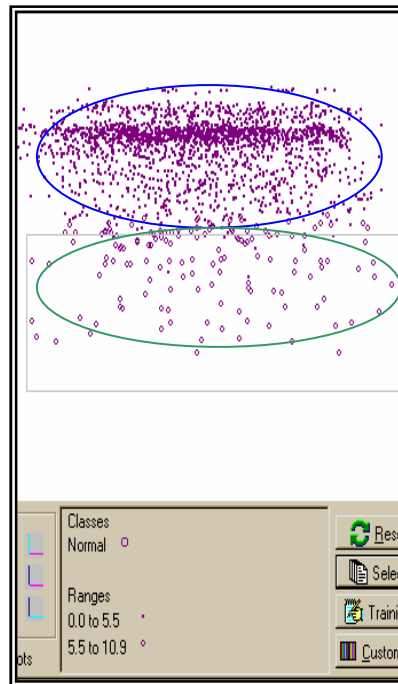


Figura 16: Projeção da idade tridimensional para o grupo “Normal”, incluindo os quatro atributos de desempenho e utilizando o pico de “Fast” com o classificatório e dois subgrupos, sendo o primeiro variando de 0,0 a 5,5% e o segundo de 5,5% a 10,9%.

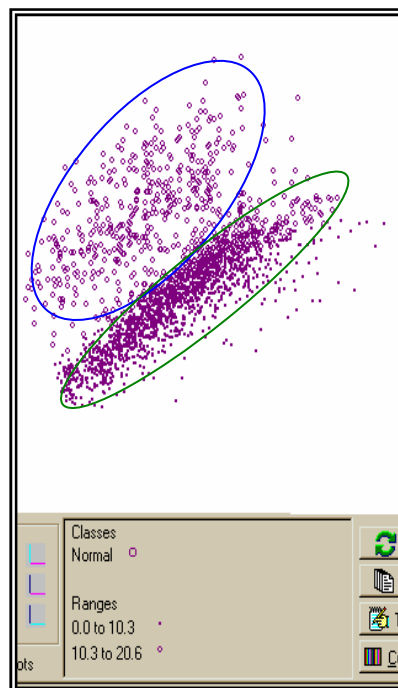


Figura 17: Projeção da idade tridimensional para o grupo “Normal”, incluindo os quatro atributos de desempenho e utilizando o pico “Fast” com o classificatório e dois subgrupos, sendo o primeiro variando de 0,0 a 10,3% e o segundo de 10,3% a 20,6%.

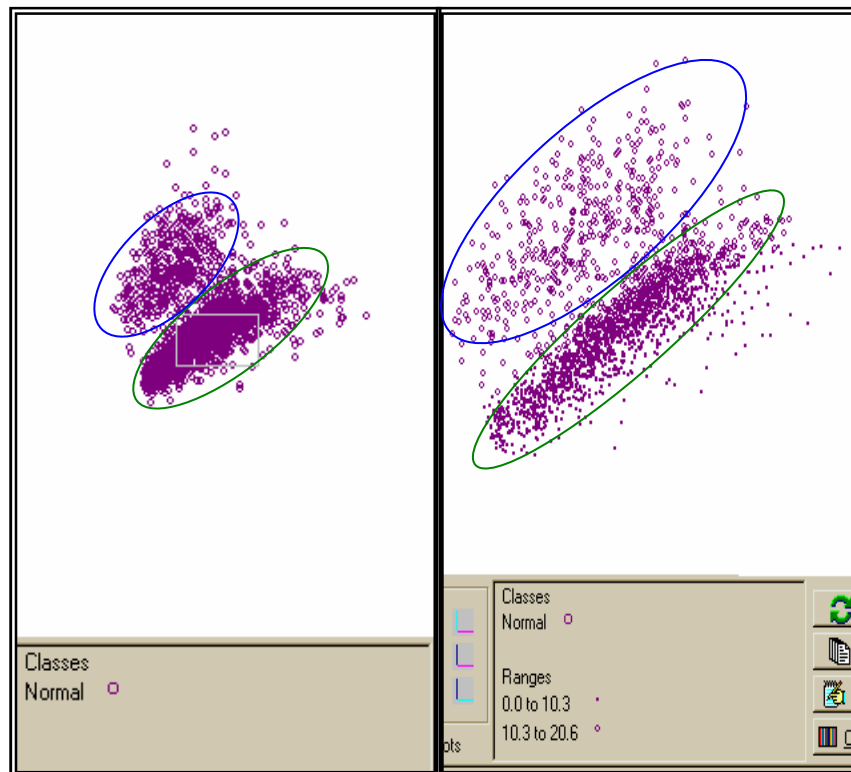


Figura 18: Comparação das imagens projetadas para o grupo “Normal”. A - incluindo os quatro picos com o atributo de aparência B - utilizando o pico de “E” dividido em dois subgrupos variando de 0,3 a 10,3% e de 10,3 a 20,6%.

4.2.2 Perfis obtidos para os diferentes fenótipos de Hb

Visando esta eleger um padrão para os diferentes fenótipos observados, foram analisadas todas as amostras, divididas nos três grupos, “Normal”, “Talassemia” e “Variante”, incluindo os quatro picos com o atributo para o apareamento. Na imagem gerada observaram-se os grupos de análise distribuídos em três nuvens amostrais distintas, como ilustrado na Figura 9, não houve sobreposição entre os grupos. As nuvens correspondentes aos grupos “Talassemia” e “Normal” encontram-se diametralmente opostas.

No apareamento incluindo as amostras dos grupos “Normal” e “Talassemia” utilizando os quatro atributos observou-se a diferenciação e agrupamentos distintos para cada grupo, sem sobreposição espacial como pode ser observado na Figura 2.

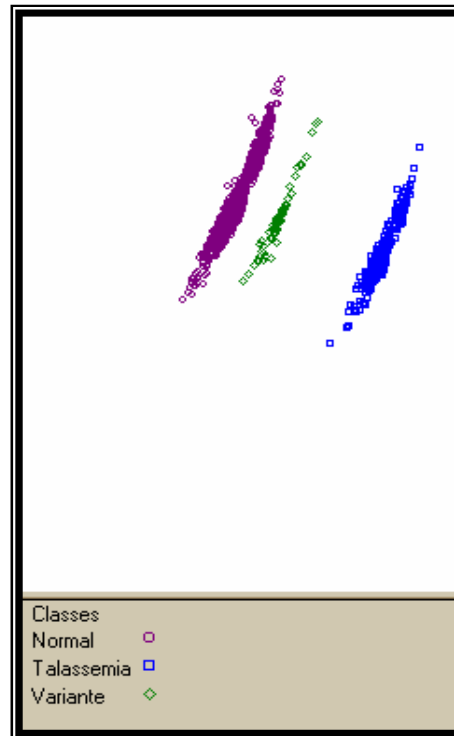


Figura 19: Projeção da dimensão tridimensional para os grupos “Normal”, “Talassemia” e “Variante”, incluindo os quatro atributos de apêndice.

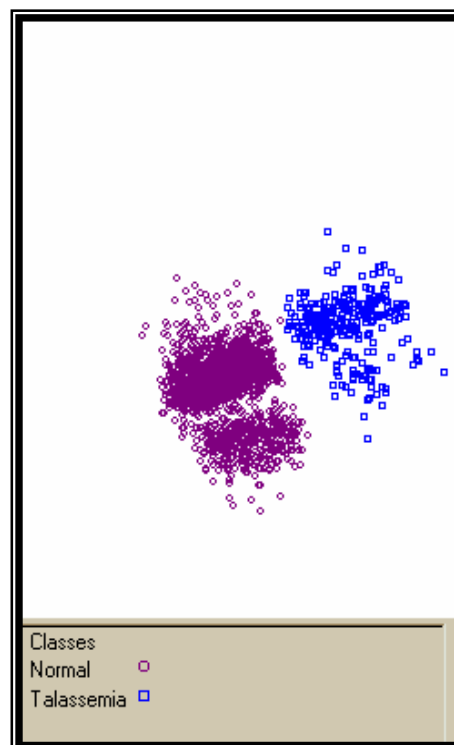


Figura 20: Projeção da dimensão tridimensional para os grupos “Normal” e “Talassemia” incluindo os quatro atributos de apêndice.

Para verificar quais as frações fundamentais na diferenciação de aóstras com H norais e com alterações talassêmicas, foi gerada uma imagem utilizando apenas os picos \underline{F} e \underline{A} com o atributo de apeamento, por representarem as frações de heóglonas e maiores concentrações nesta faixa etária. Na imagem projetada, observou-se a distribuição linear e paralela dos grupos aóstrais, como pode ser visualizado na Figura 2.

O grupo "Talassemia" incluiu as aóstras com alfa e eta talassêmicas e apesar de apresentarem alterações nos valores quantitativos de H, elas apresentam variações e frações diferentes das H, dessa forma realizou-se o apeamento com esses fenótipos separados e comparados ao grupo "Norais". A visualização espacial desses grupos incluindo os quatro atributos mostrou o reposicionamento dos pontos aóstrais.

Foram utilizados então, os picos \underline{F} e \underline{A} para o apeamento e, na visualização obtida observou-se o delineamento dos grupos e três aglomerados representando cada grupo fenotípico, ilustrado na Figura 22. Pode-se observar que as aóstras se distribuíram linearmente no espaço, estando bem definidas e três grupos distintos. A Figura mostrou ainda que, os grupos alfa e eta talassêmicas estão mais próximos entre si e afastados do grupo norais.

As aóstras com alfa talassêmica foram projetadas comparadas com as norais e para os apeamentos realizados com os quatro picos ou com os incluindo três picos observou-se o reposicionamento dos pontos aóstrais como visualizado na Figura 23, que incluiu os picos \underline{F} , \underline{E} e \underline{A} com o atributo de apeamento. Foram feitos apeamentos com dois picos incluídos com o atributo, e as visualizações obtidas resultaram na concentração dos pontos aóstrais e aglomerados distintos para cada fenótipo. A Figura 24 representa a visualização tridimensional do apeamento feito utilizando os picos \underline{F} e \underline{A} , no qual observou-se a melhor separação aóstral.

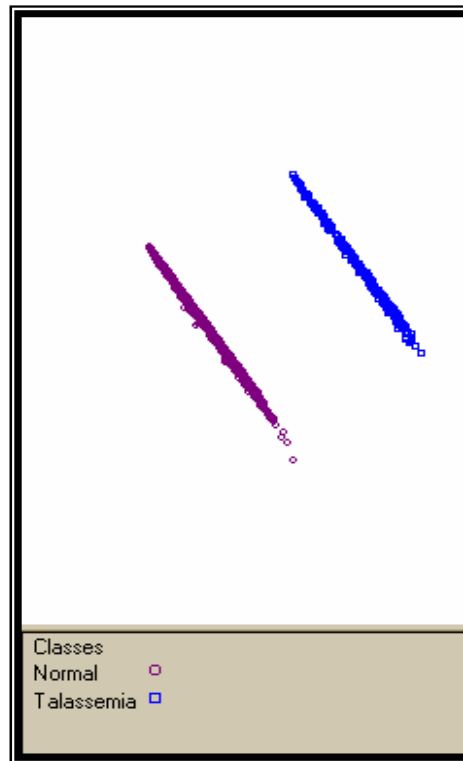


Figura 21: Projeção da imagem tridimensional para os grupos “Talassemia” e “Normal” utilizando como atributos de apeamento os picos F_1 e A_1 .

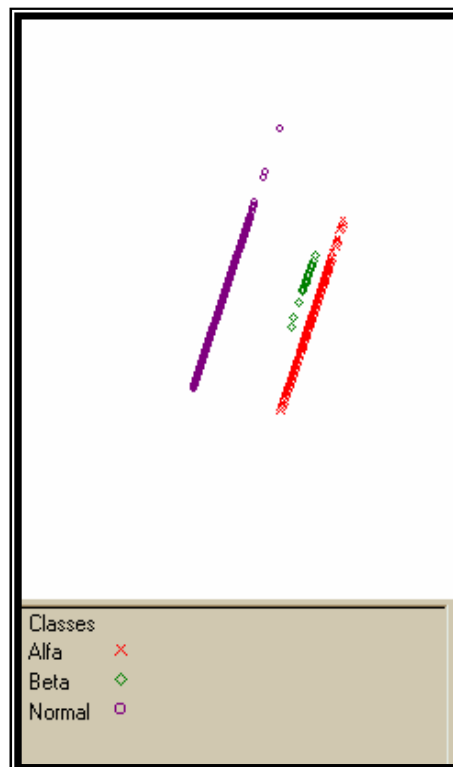


Figura 22: Projeção da imagem tridimensional para os grupos “Normal”, “Alfa” e “Beta” talassemias, utilizando como atributos de apeamento os picos F_1 e A_1 .

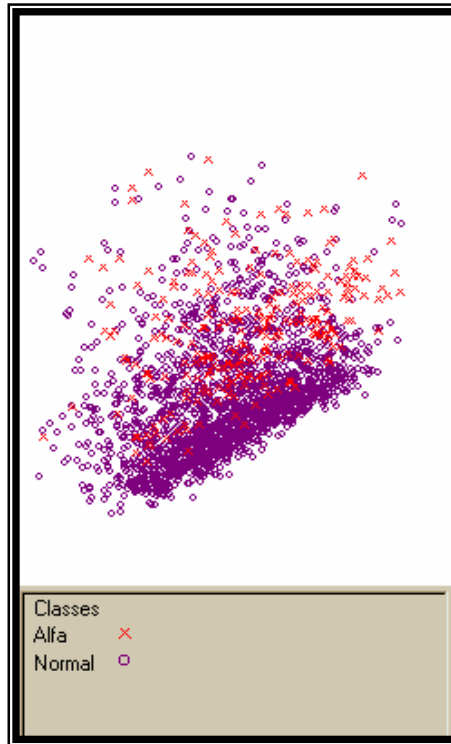


Figura 23: Projeção da imagem tridimensional para os grupos "Normal" e "Alfa" talassêmica, utilizando como atributos de apeamento os picos "Fast", "F" e "A".

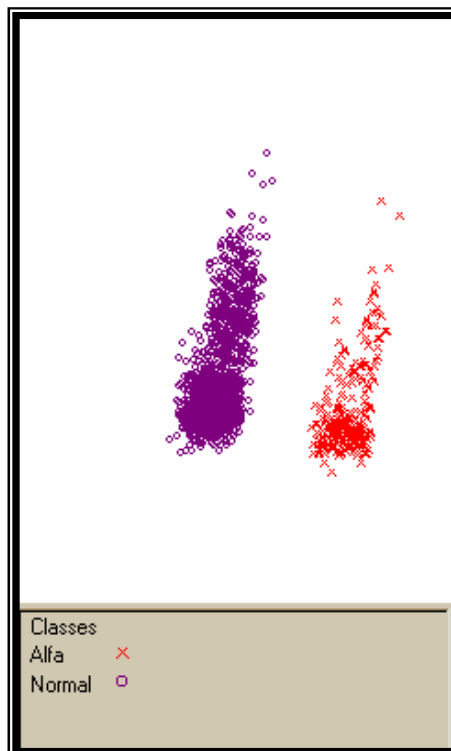


Figura 24: Projeção da imagem tridimensional para os grupos "Normal" e "Alfa" talassêmica, utilizando como atributos de apeamento os picos "Fast" e "A".

Para as amostras pertencentes ao grupo "Variantes", o serviu-se nos experimentos realizados a existência de uma interface entre esse grupo e o grupo de amostras normais, como pode ser visualizado no destaque da Figura 25. O grupo "Variante" foi posto pelas amostras contendo H₂AS e H₂AC, para tentar estabelecer um padrão entre essas variantes e diferenciá-las das amostras com H₂normal, realizou-se o experimento incluindo o grupo "Normal" com cada uma dessas variantes isoladamente. O padrão de dispersão espacial se manteve, não sendo possível separar as amostras e nuvens amostrais distintas. A Figura 26 representa a visualização tridimensional para o experimento dos grupos "Normal", H₂AS e H₂AC incluindo os quatro atributos de experimento. Mesmo analisando as amostras separadamente a readequação espacial se manteve. As Figuras 27 e 28 ilustram as visualizações espaciais para os experimentos das amostras normais com H₂AS e normais com H₂AC respectivamente, incluindo os quatro picos como atributos. O serviu-se que estes o na projeção das amostras de H₂variantes separadamente, a readequação espacial se manteve nas visualizações. Deve-se destacar que as frações de H₂variantes são identificadas no HPLC por picos específicos padronizados denominados "S" e "C" que reconhece o tipo de eluição dessa H₂normal. Estes picos não foram incluídos como atributos para o experimento espacial por estarem ausentes nas amostras com heptóginas normais, não sendo possível a comparação.

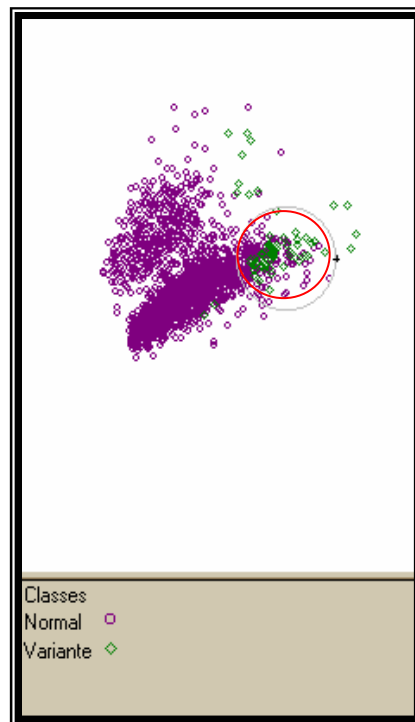


Figura 25: Projeção da imagem tridimensional para os grupos “Normal” e “Variante” utilizando os quatro atributos de apesentamento. Área de interesse ressaltada pelo círculo vermelho.

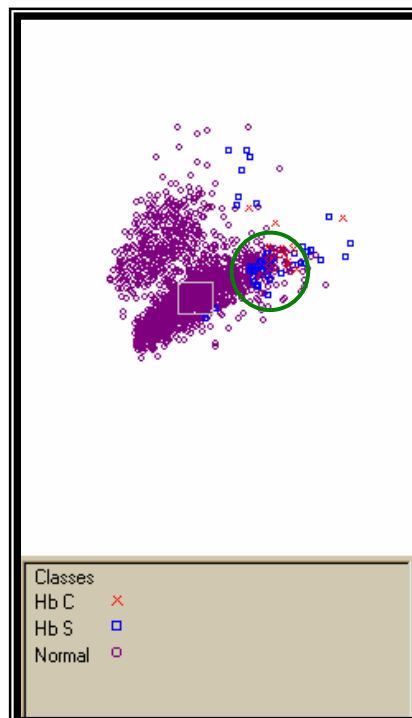


Figura 26: Projeção da imagem tridimensional para os grupos “Normal”, “Hb AS” e “Hb AC” utilizando os quatro atributos de apesentamento. Área de interesse ressaltada pelo círculo verde.

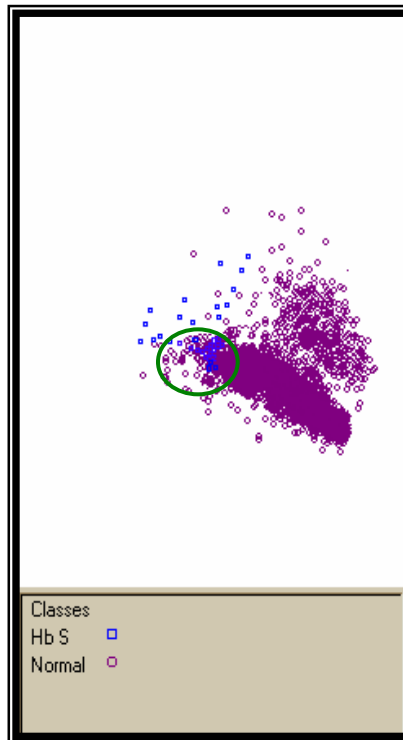


Figura 27: Projeção da imagem tridimensional para os grupos “Normal” e “Hb S” utilizando os quatro atributos deapeamento. área de interesse ressaltada pelo círculo verde.

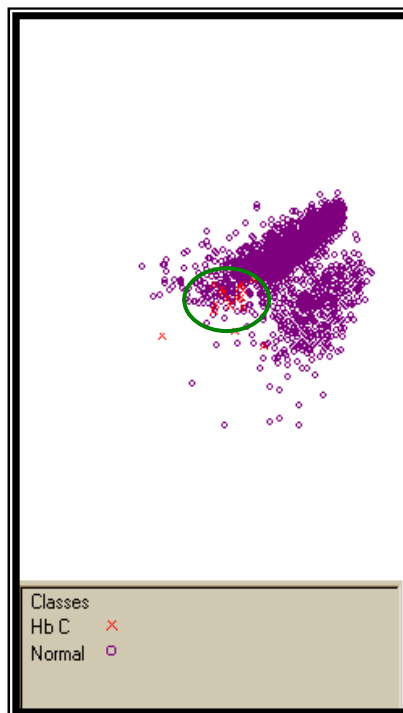


Figura 28: Projeção da imagem tridimensional para os grupos “Normal” e “Hb C” utilizando os quatro atributos deapeamento. área de interesse ressaltada pelo círculo verde.

Os valores quantitativos de Hb, obtidos por HPLC, foram submetidos a análise estatística, a fim de buscar similaridades ou diferenças que pudessem auxiliar na compreensão dos perfis apresentados pelos diferentes fenótipos nas visualizações tridimensionais. A Tabela contém as médias percentuais dos quatro picos determinados pelo HPLC para todas as alterações de Hb analisadas. As comparações entre as médias foram realizadas para cada pico entre os diferentes fenótipos e estão representadas na tabela por letras sobrescritas, sendo cada pico apresentado por uma cor e as letras diferentes indicam que houve diferença estatisticamente significativa ($p < 0,05$).

Tabela 1 - Valores médios percentuais para os picos de Hb obtidos por HPLC nos diferentes fenótipos e suas correlações nas análises estatísticas.

Fenótipos de Hb	N	X %			
		“Fast”	“F1”	“F”	“A”
Nor mal	209	2,4 ^a	3,48 ^a	3,2 ^a	2,8 ^a
Alfa tal	25	3,4 ^b	3,8 ^b	3,5 ^a	3,2 ^b
Beta tal	2	2,58 ^a	3,8 ^b	3,85 ^b	4,8 ^c
H S	55	2,3 ^a	3,3 ^c	2,9 ^a	3,85 ^d
H C	8	2,8 ^a	3,55 ^a	3,54 ^c	4,9 ^c

N= número de amostras. As letras sobrescritas diferentes representam diferença estatisticamente significativa ($p < 0,05$). As cores diferentes representam cada pico de Hb.

Como pode ser observado na Tabela 1, as amostras com alfa talassemia apresentaram diferenças estatisticamente significativas nos valores percentuais dos picos F_{ast} , F_{α} e A em relação ao grupo com amostras normais. Para as amostras com Beta talassemia observaram-se diferenças estatisticamente significativas ($p < 0,05$) para os picos de F_{α} , F_{β} e A em comparação ao grupo normal. Dessa forma, analisando o grupo de "Talasemias", observou-se valores diferentes estatisticamente para três picos de análise quando em comparação com as amostras normais. Para as amostras com Hb variantes as diferenças estatísticas foram significantes apenas em dois picos de análise, sendo que para a Hb S, observou-se $p < 0,05$ para os picos de F_{α} e A em relação ao grupo normal e, para Hb C, observou-se diferença significativa nas áreas percentuais dos picos F_{β} e A em comparação com as amostras normais.

5 Discussão

5 Discussão

A utilização do banco de dados para a análise dos resultados laboratoriais mostrou-se importante em vários aspectos. A segurança no processo de arquivar as informações, um ponto a ser destacado. O banco possui acesso restrito, não sendo permitida a todos os usuários a recuperação dos dados neles armazenados. O livre acesso a todas as informações contidas no banco só foi permitido ao administrador e a responsável pelo LHGDH, os demais usuários, dentre eles, os alunos envolvidos em pesquisas e os colaboradores que enviam amostras ao LHGDH tiveram acesso apenas às análises de sua referência.

Além da privacidade das informações, deve-se destacar também as vantagens do arquivo eletrônico. Muitos laboratórios de rotina ou mesmo laboratórios de pesquisa que trabalham com grande volume de amostras, têm problemas com espaço e organização dos resultados impressos, pois necessita de grandes arquivos, protegidos de variações de umidade e temperatura e depende de um trabalho laborioso para organizá-los de maneira que seja possível encontrar um dado arquivado no passado com rapidez.

A disponibilidade do banco de dados via Internet é um aspecto que deve ser mencionado. Nos dias atuais, a necessidade dos pesquisadores em consultar seus dados de pesquisa, de qualquer lugar e a qualquer momento, torna-se cada vez maior, devido a grande pressão do meio científico em publicar o trabalho possível os seus resultados. Para os laboratórios de rotina, a possibilidade de consultar seus exames pela Internet também se apresenta vantajosa, diante da crescente concorrência nessa área.

A utilização de técnicas *data mining*, como a ferramenta CLIBIA, mostrou-se de grande valor no levantamento dos dados. A localização de um exame foi extremamente rápida e a comparação dos resultados de diferentes testes possibilitou a reavaliação das metodologias laboratoriais utilizadas. Analisando-se, por exemplo, os resultados de exames de colúmbio de alfa talassemia, observou-se que 20% dos exames apresentaram resultado positivo ou sugestivo no teste de Pesquisa intraeritrocitária de HbH, no entanto, 90% apresentaram a fração de Hb

H ou Bart s e eletroforese pH neutro. Essas observações destacaram a importância da pesquisa de H₂H, e a necessidade de avaliar as amostras para esse teste mesmo quando a eletroforese em pH neutro apresentar resultados negativos. O levantamento anual dessas informações dependeria muito tempo, e ainda seria questionável a possibilidade de realizar essas relações entre testes e laudos apenas por tabelas impressas contendo muitas amostras.

Os resultados do teste de Resistência Global os tica revelaram que 24,5% das amostras com laudo normal para idade apresentaram positividade para este teste, enquanto que 2,4% das amostras com laudos de beta talassemia foram positivas para o mesmo teste. Essa metodologia está bem consolidada para rastreamento de beta talassemia em indivíduos adultos, sendo que 9,7% dos beta talassemicos apresentam resistência os tica a NaCl a 0,3% NAOUM, 99 no entanto a análise dos resultados da triagem de recém-nascido por *data mining* sugeriu uma discordância entre os resultados desse teste e o diagnóstico. Diante desses dados deve-se considerar a reavaliação desse todo em neonatos.

Contabilizar e organizar em tabelas os resultados de análises de aproximadamente 1000 amostras tarefa de árdua e trabalhosa. Para aplicar as análises estatísticas, que usualmente são realizadas em dados de pesquisa, necessita-se primeiramente digitalizar os dados em planilhas, convencionalmente utilizadas, como planilhas do Microsoft Excel, esse último passo importante, presente na análise de dados da maioria dos trabalhos científicos. Além de depender grande tempo e trabalho, esse processo, de arquivar e analisar dados, está sujeito à possibilidade de erros em todas as etapas do processo, desde a contagem dos resultados até a digitalização, e na maioria das vezes, o trabalho de conferência desses erros cansativo e é geralmente pouco resolutivo.

A utilização de um banco de dados para arquivar as informações e a realidade já praticada em alguns laboratórios de rotina e análise. Na área de pesquisa e genética, com os projetos genômicos surgiram os grandes bancos de dados implantados para armazenamento de seqüências, sendo os quais todo o trabalho de seqüenciar genomas seria mais produtivo. Atualmente a possibilidade de

laboratórios de pesquisa, e diversas áreas, utilizando o banco de dados associado à ferramenta *data mining* para processar as informações, a renovação para a ciência, aumentando a confiabilidade dos resultados apresentados, permitindo aos pesquisadores investir mais tempo nas questões científicas e menos trabalho contando dados e montando tabelas e gráficos, que o sistema computacional pode realizar em segundos. A inserção das informações no banco de dados é a etapa de responsabilidade do pesquisador, e deve ser objeto de grande atenção, onde inconsistências podem ser corrigidas, as dúvidas verificadas, e dados incorretos reavaliados.

Para o LHGDH a implantação do banco de dados *on-line* agilizou não apenas as atividades de prestação de serviços com o Centro de Referência e Diagnóstico de hepatopatias, mas também contribuiu de maneira insuspeita para o processamento dos dados de pesquisa. Permitiu a interação entre os resultados das diferentes análises utilizadas em cada projeto, com o trabalho entre a eficiência de ambas as metodologias nas diferentes análises e os resultados gerais prontamente utilizados nas publicações e hipóteses de trabalhos subsequentes. Todo o processo de análise de análises de sangue tornou-se mais dinâmico e menos limitado a um único objetivo, seja este diagnóstico de alteração ou a contribuição científica das análises.

A aplicação de técnicas de VDM à análise de dados nas pesquisas com hepatopatias. Para os dados da triagem neonatal realizada pelo LHGDH as análises pela ferramenta *FastMapDB* apresentaram padrões relevantes. Na projeção das análises com hepatopatias normais para a idade observou-se a existência de duas nuvens de distribuição alélica. O que foi feito se a presença dos picos de F_{fast} e F_{norm} mostrou distribuição linear, sugerindo dois padrões de comportamento distintos no grupo controlado. Com o recurso da ferramenta *FastMapDB* os valores do pico de F_{fast} foram divididos ao meio separando as análises em dois subgrupos. Na projeção tridimensional os pontos distribuíram-se em dois grupos separados, no entanto, houve um desequilíbrio na distribuição alélica, de forma que o primeiro subgrupo incluiu 94% das análises, não sendo representativo da distribuição

o servada na Figura 2. Esses resultados destaca que o pico de F_{fast} não foi o responsável pela formação das duas nuvens aóstrais o servadas no apeamento das aóstras com heógloninas normais para a idade.

No apeamento feito utilizando a divisão do pico F_{fast} no valor médio o teve-se dois grupos distribuídos de forma igualmente, sendo o primeiro composto por 3,7% das aóstras e o segundo incluindo 34,9%. As diferenças para esses dois subgrupos apresentaram diferenças estatisticamente significativas $p < 0,05$ quando comparadas entre si e a média geral o tida para todas as aóstras normais. Quando comparadas lado a lado, as visualizações o tidas com os quatro atributos e com o pico de F_{fast} dividido em dois subgrupos

Figura 8, pode-se observar a distribuição dos pontos aóstrais. Esses resultados indicam a presença de dois padrões com valores médios diferentes de distribuição para F_{fast} , devido a grande amplitude de variação o servada nos valores deste pico. Deve-se destacar que as frações de H degradadas, presentes em aóstras acondicionadas por longo tempo, provocam aumento no valor de F_{fast} . Diante desses achados, deve-se destacar, portanto que a média geral o tida para as 249 aóstras não valor representativo do pico de F_{fast} para o grupo "Normal" e ainda que este não é um parâmetro confiável para o diagnóstico.

Destaca-se que apenas através da estatística descritiva, usualmente utilizada quando se pretende determinar valores de referência para dados aóstrais, não foi possível observar esse comportamento para os valores de F_{fast} nas aóstras recém-nascidos com valores normais para a idade. A visualização gráfica dos dados, no entanto, evidenciou já na primeira projeção realizada, a existência desses dois padrões de valores para esse pico. Esses padrões foram facilmente reconhecidos através da capacidade humana de imediatamente reconhecer agrupamentos e, dessa forma, questionar os dados ou inferir hipóteses. Segundo *Leifer* (2002) o processo de exploração visual dos dados pode ser visto como um processo de geração de hipóteses, através das visualizações dos dados, que permite aos usuários obter uma visão deles e surgir com novas hipóteses.

Uma representação visual fornece um grau mais alto de confiabilidade nas conclusões da exploração do que uma representação numérica ou textual.

Análise global dos resultados, visualizados e projeção espacial ressaltou a existência de um possível padrão observado por HPLC, para as aóstras com talassecias heterozigotas, pois não houve o reposicionamento espacial dos pontos aóstrais, sugerindo diferenças nos perfis quando comparadas as aóstras normais para a idade. Deve-se ressaltar que nas análises estatísticas as médias percentuais apresentaram-se significativamente diferentes ($p < 0,05$) para três picos de análise, tanto para alfa, quanto para eta talassecia. Essa diferença estatística pode explicar a distribuição visual dos pontos aóstrais nas imagens geradas para as aóstras normais e com talassecias, evidenciando a existência de diferenças nesses perfis. Na projeção das aóstras normais e com alfa e eta talassecia (Figura 22), utilizando os picos "E" e "A" como atributos de apeamento serviu-se o delineamento dos grupos, estando dispostas em três linhas paralelas, sendo que as aóstras com talassecias dispuseram-se próximas entre si e distantes das normais. Esse padrão de distribuição evidencia a potencialidade dessa ferramenta de análise e separar os grupos aóstrais segundo suas características.

A projeção da idade obtida para as aóstras normais e com alfa talassecia apresentou o reposicionamento dos pontos, no entanto, o apeamento feito utilizando os picos de "E" e "A" apresentou distribuição e definição dos grupos aóstrais. Esses picos apresentaram diferenças significativas nos valores médios ($p < 0,05$) como foi ilustrado na Tabela . Deve-se destacar também que, as aóstras com alfa talassecia, apresentaram aumento estatisticamente significativo ($p < 0,05$) no percentual médio do pico de "E" quando comparado aos valores obtidos para esse pico nos demais fenotipos. Para as aóstras com eta talassecia houve diferença estatisticamente significativa nos picos de "E", que apresentou aumento e relação as aóstras normais, e para o pico de "A", houve diminuição significativa comparada aos valores normais para a idade. A visualização gráfica permitiu nesta comparação, identificar quais os atributos importantes na caracterização das aóstras, pela seleção dos picos para o

apesar disso e pode-se identificar diferenças posteriormente corroboradas pela análise estatística.

Mendes-Siqueira, 2004, serviu ao estudo estatístico significativo para os picos de F_{ast} e F_{c} e para outras cromossomas talassêmicas na faixa etária. Para as cromossomas talassêmicas o teste teve valores aumentados para o pico de F_{c} e diminuídos para A , sendo as diferenças entre os valores obtidos para o grupo normal estatisticamente significativas, sugerindo a existência de um perfil diferenciado para os cromossomas das cromossomas talassêmicas. A HPLC com o sistema Variant é o método diagnóstico autorizado pelo Ministério da Saúde no Brasil, mas é a metodologia que permite o diagnóstico das formas talassêmicas nessa faixa etária, com segurança, no entanto, esses resultados sugerem um perfil característico para essas cromossomas, confirmado por outras metodologias.

Apesar da HPLC não ser utilizada como o método diagnóstico das talassêmicas, a utilização de técnicas de *Visual Data Mining* revelou diferentes perfis nos fenotipos normais e talassêmicos, destacando a necessidade de melhor avaliação dos resultados obtidos na triagem por HPLC, pois diferenças sutis nos cromossomas pode indicar uma possível talassêmica, sendo um teste adicional no diagnóstico dessas alterações e neonatos, contribuindo, desta forma, para o direcionamento das análises, associada a outras metodologias específicas para confirmar esta alteração.

As H_S e H_C são diagnosticadas com precisão pelo sistema de HPLC Variant que identifica com alta sensibilidade as pequenas porcentagens das frações variantes. No entanto o teste é específico das cromossomas heterozigotas para essas hemoglobinas comparadas ao grupo Normal mostrou-se reposição das nuvens cromossomais, não sendo possível diferenciar os grupos. O pico de H_{Fetal}, que representa o maior componente hemoglobínico nesta faixa etária, não apresenta diferenças estatisticamente significativas nos valores percentuais entre as cromossomas normais e com H_S. Em relação ao pico de H_A, a diminuição estatisticamente significativa $p < 0,05$, observada nas cromossomas H_S variantes e em relação às normais, deve-se ao fato das H_S e H_C estarem presentes e heterozigose nas cromossomas avaliadas. Apesar dessa diferença, para as H_S variantes,

apenas dois picos de análise apresentaram diferença significativa e relação ao grupo Normal, dificultando a separação espacial dos grupos aóstrais. Deve-se ressaltar que, os picos característicos de identificação das frações variantes não puderam ser analisados pela projeção visual por não apresentarem contrapartida correspondente no grupo de aóstras normais.

Onde, analisando os resultados de HPLC de aóstra de sangue de indivíduos adultos por VDM I, 2, 4, 1, 2, a, 2, vi, s, f, de 4 or 3-3, a 4 - e 4, i.

estatísticas foram concordantes com os padrões observados nas projeções espaciais. Onde, a partir dos resultados estatísticos que corroboraram suas visualizações tridimensionais, reforçando a aplicabilidade de grande valor dos métodos de VDM nas análises ecológicas. As técnicas *Data mining* e VDM podem contribuir para as análises com heterogeneidades, ampliando a exploração dos resultados, corroborados pela análise estatística que auxilia na compreensão das interações.

6 Conclusões

6 Conclusões

1. A utilização de um banco de dados e laboratórios de análise e pesquisa, que possuem grande fluxo de amostras, mostrou-se imprescindível para a segurança, confiabilidade e agilidade das análises e, ferramentas que proporcionam a mineração dos dados, como a CLIBIA, revelaram a necessidade de reavaliar as metodologias de rotina utilizadas no diagnóstico de hemoglobinopatias e neonatos.

2. A utilização de VDM demonstrou grande aplicabilidade nos dados para hemoglobinopatias. As visualizações tridimensionais realizadas pela ferramenta *FastMapDB* revelaram a existência de dois padrões de distribuição das amostras com hemoglobinas, devido à amplitude de variação nos valores obtidos por HPLC para o pico de β^E .

3. A projeção tridimensional permitiu a separação das com hemoglobinas normais e com alfa talassemia ressaltando as diferenças estatísticas nos valores percentuais relativos dos picos de β^{Fast} e β^A . Essa distribuição evidencia a existência de perfis característicos dessa alteração nas análises por HPLC, reforçando a necessidade de melhor avaliação dos cromatogramas obtidos e a possibilidade de aplicação dessa metodologia como parte do diagnóstico de alfa talassemia e neonatos.

6. Contribuições para implementação de recursos na ferramenta *FastMapDB*

A utilização da ferramenta *FastMapDB* neste trabalho pode contribuir para aprimorar os recursos da ferramenta apresentando sugestões para implementação.

1. Um recurso de grande utilidade na ferramenta é a classificação de um atributo dividido em dois subgrupos exata mente no valor médio. Para melhorar esse recurso sugere-se que o valor para classificação dos subgrupos, nesse recurso, possa ser determinado pelo usuário.

2. A ferramenta possibilita a seleção dos pontos adicionais apenas o clique, no entanto sugere-se a alteração na forma de seleção para facilitar o acesso aos pontos, poderia ser implementado um recurso para seleção com diferentes formatos de cursor como o quadrado, círculo, seta, etc.

Referências

7 Referências

BANERJEE, A. Understanding globin regulation in β -thalassaemia: it's as simple as α , β , γ , δ . **The Journal of Clinical Investigation**, v. 5, n. 4, p. 432-435.

BIANCO, I. et al. La prevenzione dell'anemia mediterranea nel Lazio: risultati del programma applicato negli ultimi 5 anni. **Prog. Med. Roma**, v.42, p. 52-54, 98.

BARIONE, M. C. N. et al. Visualização de dados e sistemas de bancos de dados relacionais. Proceeding: XVI Simpósio Brasileiro de Banco de Dados. Rio de Janeiro. p. 95 - 99.

BONINI-DOMINGOS, C. R. **Hemoglobinopatias no Brasil: variabilidade genética e metodologia laboratorial**. São José do Rio Preto, 1993. 232 f. Tese de Doutorado em Ciências Biológicas - Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista.

BONINI-DOMINGOS, C. R.; SIQUEIRA, F.A.M. Diagnóstico laboratorial das hemoglobinopatias. **NewsLab**, v.42, p. 20-24.

BONINI-DOMINGOS, C. R. Thalassaemia screening in Brazil - Results for 2 years. **Revista Brasileira de Hematologia e Hemoterapia**, v. 24, n. 4, p. 288-289, 294.

BORRAJO, G. J. C. Situação de la pesquisa neonatal en Latinoamérica. **Rev. Méd. Minas Gerais**. Editora Coop. ed, v. 5, n. 2, p. 9-22.

BOY, R. Rastreamento de doenças genéticas no recém-nascido. **Coletânea de tópicos de erros inatos do metabolismo e triagem neonatal**. Campo Grande-IPD, 2002. CD-ROM

BRASIL. Ministério da Saúde. Portaria da Saúde, nº 822, de junho de 2000. Brasília, 2000.

CAO, A; MOI, P. Regulation of the globin genes. **Pediatr. Res.**, v.5, n.4, p.45-42, 2002.

CARVALHO, T. M. Triage Neonatal no Brasil. **Rev. Méd. Minas Gerais**. Editora Coop ed, v. 5, n. 2, p. 20-22, 2005.

CHINELATO-FERNANDES, A. R. **Diferenciação molecular de mutantes de hemoglobinas humanas na população brasileira**. São José do Rio Preto, 2003. 222 f. Tese Doutorado em Genética - Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista.

CHITTARO, L. Information Visualization and its application to medicine. Elsevier. **Artificial Intelligence in Medicine**, 2000.

CHUI, D. H.; FUCHAROEN, S.; CHAN, V. Hemoglobin H disease: not necessarily a benign disorder. **Blood**, v. 80, n. 3, p. 918-920, 2003.

CLARKE, G. M.; HIGGINS, T. N. Laboratory Investigation of Hemoglobinopathies and Thalassemias: Review and Update. **Clinical Chemistry**, v. 40, p. 284-290, 2000.

CLEGG, J. B. Can the product of the gene θ be a real globin? **Nature**, v. 329, n. 6137, p. 445-448, 1998.

DACIE, J. V.; LEIS, S. M. **Practical haematology**. 8 ed. Livingstone: Churchill, 1995.

DALMAU, A.C. et al. Screening neonatal de drepanocitosis en el Consorci Sanitari de Mataró. **An. Esp. Pediatr.**, v.49, p. 5-9, 1998.

DAUDT, L.E. et al. Triage neonatal para he moglo inopatas: u estudo piloto e Porto Alegre, Rio Grande do Sul, Brasil. **Cad. Saúde Pub.**, v. 8, n. 3, p. 833-84, 2002.

DEVILLE, M. A. et al. Disruption of the Mechanis of Long Range Activation within the Human α -Globin Complex. **The Journal of Biological Chemistry**, v. 279, n. 2, p. 293-298, 2004.

EASTMAN, J. et al. Automated HPLC screening of newborns for sickle cell anemia and other hemoglobinopathies. **Clin. Chem.**, v. 42, n. 5, p. 94-99, 1996.

EASTMAN, J. et al. Distribution of hemoglobin in F, A, S, C, E, and D quantities in 4 million newborn screening specimens. **Clin. Chem.**, v. 45, n. 5, p. 83-85, 1999.

ELALLA, S.; MATHE S, A.R. Hemoglobin in the United Arab Emirates. **Hemoglobin**, v.2, p.39-5, 1999.

ELMASRI, R. - Fundamentals of Database Systems, 3 ed, ISBN 0-542-33-3, Addison-Wesley, 2000.

FALOUTSOS, C.; LIN, Y. Fast Map: A Fast algorithm for indexing, Data-Mining and visualization of traditional and Multimedia Dataset. **ACM SIGMOD**, Zurich, Switzerland, p. 3-4, 1995.

FAYYAD, U. Mining Data bases: Towards algorithms for knowledge discovery. **Bulletin of the IEEE Technical Committee on Data Engineering**, v. 2 , p. 39-48, 1997.

FROST, B. A.; BELLINGHAM, A. J. Neonatal Haemoglobinopathy screening. **Acta Haemat.**, v. 8, p. 42-43, 1998.

GALACTEROS, F. Diagnostic neonatal des hemoglobinopathies. **Rev. Prat. (Paris)**, v.42, n. 5, p. 893-5, 1992.

GRIFFITHS, A. J. H., et al. **Introduction to Genetic Analysis**. 2^{ed}. New York: Freeman and Co., 1999.

GTA-TN. PNTN: Apresentação sobre o Programa Nacional de Triage Neonatal. Disponível em : <http://dtr2.saude.gov.br/sas/dsra/prtn_arquivos_fra_e.htm > Acesso em : Fevereiro 2004.

HAYASHI, A. et al., Neonatal screening and mass-spectrometric analysis of hemoglobin variants in Japan. **Acta Haemat.** v. 8, p. 4-8, 1998.

HONIG, G. R.; ADAMS III, J. G. **Human hemoglobin genetics**, 2^{ed}, New York: Springer Verlag, 1997. 452 p.

HUISMAN H. J. et al. HbVar: A Database of Human Hemoglobin Variants and Thalassemias. Subseries of mutation categories. Pennsylvania University USA and McMaster University in Canada, 1997. Disponível em : <<http://globin.cse.psu.edu>>. Acesso em 2 de novembro de 2005.

INSTRUCTION MANUAL OF BIO-RAD, Variant System; Sickle Cell kit, Bio-Rad Laboratories, 1999.

JANUARIO, J. N. Programa de triagem neonatal apresenta primeiros resultados. **Jornal Hemominas**, v. 8, n. 3, p. 5-7, 1998.

KEIM, D. A.; KRIEGER, H. P. - Visualization Techniques for Mining Large Data bases: A Comparison. **IEEE Transactions on Knowledge and Data Engineering**, v.8, n. 4, p. 923-938, 1996.

KEIM, D. A. - Information Visualization and Visual Data Mining. **IEEE Transactions on Visualization and Computer Graphics**, v.8, n. 2, 1998.

KEITH, S. **Sistema de Banco de Dados**. 3. ed., 1999.

LEVIN, B. **Genes VI**. Oxford, International Student Edition, 2000.

MARENGO-ROSE, A. J. Rapid electrophoresis and quantitation of haemoglobin on cellulose acetate. **J. Clin. Path.**, v. 8, p. 91-92, 1955.

MENDES-SIQUEIRA, F. A. **Contribuição para o estudo das alterações moleculares e interferentes na expressão fenotípica das hemoglobinopatias a partir de um programa de diagnóstico neonatal**. São José do Rio Preto, 2004. Tese de doutorado em Genética – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista.

NAOUM, P. C. **Eletroforese, técnicas e diagnósticos**. São Paulo: Santos, 1997. 4 p.

NAOUM, P. C. **Hemoglobinopatias e talassemias**. São Paulo: Sarvier, 1997.

NAOUM, P. C.; BONINI-DOMINGOS, C. R. Talassemias alfa. **Laes; Haes**, v.3, n. 3, p. 97-98, 1998.

NAOUM, P.C. **Eletroforese, técnicas e diagnósticos**. Sao Paulo: Santos, 1999.

OLD, J. M. Screening and genetic diagnosis of haemoglobin disorders. **Blood Reviews**, v. 7, p. 43-53, 2003.

OLIVEIRA, R. B. T.; **O processo de extração de conhecimento de Bases de dados apoiado por agentes de software**. Sao Carlos, 2000, 4 f. Dissertação de Mestrado e Ciências da Computação – Instituto de Ciências Matemáticas e de Computação.

ONDEI, L. S. **Perfil eletroforético e cromatográfico das Hb S -“like”**. Sao Jose do Rio Preto, 2005, 4 f. Dissertação de Mestrado e Genética – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista.

PAPAYANNOPOULOS, R.; STAMATONYANNOPOULOS, G. **Stains for inclusions bodies**. In: STANDARTIZATION of laboratory reagents and methods for detection of haemoglobinopathies”. Atlanta: Hewlett-Packard, 1994.

PEREA, F. J. et al. Hb D-Los Angeles associated with Hb S or β -Thalassaemia in four Mexican mestizo families. **Hemoglobin**, v. 23, n. 3, p. 233-237, 1999.

RAMALHO, A. S.; MAGNA, L. A.; PAIVA e SILVA, R. B. A portaria MS nº 822 e a triagem neonatal das hemoglobinopatias. **Rev. Bras. Hematol. Hemot.**, v. 224, n. 4, p. 244-250, 2002.

SCHNOG, J. B. et al. Sickle cell disease, a general overview. **The Netherlands Journal of Medicine**, v. 42, n. 4, p. 343-347, 2004.

SILVESTRONI, E.; BIANCO, I. Screening for microcytosis in Italy: analysis of data collected in the past 3 years. **Am. J. Hum. Genetic**, v. 23, p. 98, 1975.

STAMATOYANNOPOULOS, G. Control of α globin gene expression during development and erythroid differentiation. **Experimental Hematology**, v. 33, p. 259-265.

TRAEGER-SYNODINOS, J. et al. Best practice guidelines for carrier identification and prenatal diagnosis of haemoglobinopathies. **Eur. Mol. Gen. Qual. Netw.**, 2002.

TRAINA, A. J. M. **Suporte à visualização de consultas por similaridade em imagens médicas através de estrutura de indexação métrica**. São Carlos, 2004. 144 f. Tese Livre docente e Computação – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo.

TRONCO, M. N. **Implementação de Recursos de Visualização e Interação em Ferramenta Data Mining**. São José do Rio Preto, 2003, 54 f.

VELLA, F. Acid agar gel electrophoresis of human haemoglobin. **Am. J. Clin. Path.**, v. 49, n. 3, p. 441-442, 1968.

XIANG-HUI, F.; DE-PEI, L.; CHIH-CHUAN, L. Chromatin Structure and Transcriptional Regulation of the β -Globin Locus. **Experimental Cell Research**, v. 208, p. 1-12, 2002.

AYE, J.S.; CHUI, D.H. The alpha-globin gene cluster: genetics and disorders. **Clin. Invest Med.**, v.24, n.2, p. 13-19, 2001.

W.H.O. - WORLD HEALTH ORGANIZATION. Hereditary anaemias: genetic basis, clinical features, diagnosis and treatment. **Bulletin of the W.H.O.**, v. 15, n. 5, p. 435-59, 1962.

EATHERALL, D.J.; CLEGG, J.B. **The thalassemia syndromes**. 3.ed Oxford: Blackwell, 1985, 85p.

EATHERALL, D. J. The thalassaemias. **Brit. Med. J.**, v. 34, p. 5-8, 1997.

EATHERALL, D. J.; CLEGG, J. B., Inherited haemoglobin disorders: an increasing global health problem. **Bulletin of the World Health Organization**, v. 9, n. 8, p. 422-426.

ZAGO, M. A.; COSTA, F. F. Hereditary haemoglobin disorders in Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 9, p. 385-388, 1985.

ZAR, J.H. **Biostatistical Analysis**. 4.ed. New Jersey: Prentice Hall, 1999. 93p.

Resumo

8 Resumo

Nos anos de 1997 a 2002 foi desenvolvido um programa de triagem de hemoglobinopatias pelo Laboratório de Hemoglobinas e Genética das Doenças Hematológicas – LHGDH, UNESP, com 59 neonatos nascidos no Hospital de Base de São José do Rio Preto. O conjunto de dados gerados em todas as fases dos seis anos de programação constitui a fonte de informações, que arquivadas em banco de dados, permitiu a análise por métodos computacionais na tentativa de estabelecer parâmetros, tendências e padrões, objetivando dessa forma, contribuir com a Triagem Neonatal no país, fornecendo bases para agilizar os diagnósticos e torná-los cada vez mais precisos. Após desenvolvimento de uma Base de Dados Eletrônica e inserção dos resultados da triagem de hemoglobinopatias objetivou-se: avaliar por meio de ferramenta *data mining* – CLIBIA as metodologias utilizadas, testar a aplicabilidade do banco de dados e analisar os dados quantitativos de hemoglobinas obtidos por HPLC nos diferentes fenótipos, e sistema de visualização 3D - *Fast MapDB*. A utilização de um banco de dados e laboratórios com grande fluxo de amostras mostrou-se imprescindível para a segurança, confiabilidade e agilidade das análises e ferramentas que proveem a inserção dos dados, como a CLIBIA, revelando a necessidade de reavaliar as metodologias de rotina utilizadas no diagnóstico de hemoglobinopatias em neonatos. As visualizações tridimensionais revelaram a existência de dois padrões de distribuição das amostras com H normais, devido à amplitude de variação nos valores obtidos por HPLC para o pico de F_{fast} . Permitiu a separação das amostras com H normais e com alfa talassemia, ressaltando as diferenças estatísticas nos valores percentuais dos picos de F_{fast} e A , reforçando a necessidade de melhor avaliação dos cromatogramas obtidos e a possibilidade de aplicação dessa metodologia como parte do diagnóstico de alfa talassemia em neonatos.

9 Abstract

Since 1999 until 2012, a neonatal screening program of hemoglobinopathies were carried out in the Hemoglobin and Hematologic Genetic Diseases Laboratory - LHGDH, UNESP, with 594 new borns from Hospital de Base de São José do Rio Preto. The amount of data obtained during six years of program constitutes a source of information, that stored in database, allows the analysis for computational methods in the attempt to establish behaviors, trends and standards, aiming to contribute with the Neonatal screening in the country and supplied bases to better diagnostic and to reduce the each time more exact. After an Electronic Database were development and the screening results were inserted was aimed: to evaluate the data mining tool - CLIBIA the methodologies used, test the applicability of a hemoglobinopathies database and to analyze the quantitative hemoglobin data obtained by HPLC in the different phenotypes, in 3D visualization system - *FastMapDB*, for to analyze the phenotypes behavior. The use of a database in laboratories with great sample routine were essential for the security, trustworthiness and agility of the analyses and tools that promote the data mining, as the CLIBIA, had disclosed the necessity to reevaluate the methodologies used in hemoglobinopathies diagnosis routine in new borns. The three-dimensional visualizations carried out by the *FastMapDB* tool showed the existence of two distribution standards in the samples with normal H₂, due to the variation amplitude into the values obtained by HPLC for the "F" window. It allowed separating the samples with normal H₂ from alpha thalassemia, evidencing the statistical differences in the mean values of the "Fast" and "A" window, emphasizing the necessity of better chromatogram evaluation and the possibility of this methodology application as part of alpha thalassemia diagnosis in new borns.

9 Apêndices

10 Apêndices

10.1 Relatório Geral obtido pela ferramenta CLIBIA no Banco de Dados LHGDH *on-line*

10.2 Artigo que será submetido à publicação

11 VISUALIZAÇÃO TRIDIMENSIONAL DOS FENÓTIPOS DE HEMOGLOBINAS HUMANAS OBTIDAS POR HPLC

THREE-DIMENSIONAL VISUALIZATION OF THE HUMAN HEMOGLOBIN PHENOTYPES OBTAINED BY HPLC

Luciane Moreno Storti de Melo

Paulo Henrique Mangonaro²

Profa. Dra. Claudia Regina Bonini Domingos

– Laboratório de Hemoglobinas e Genética das Doenças Hematológicas –

LHGDH - IBILCE UNESP

Rua Cristóvão Colombo, 2245 – Jd Nazareth, Cep 13.540-000 – São José do Rio Preto, SP

² – Laboratório de Banco de Dados, Departamento de Ciências da Computação e Estatística – IBILCE UNESP

Trabalho realizado no Laboratório de Hemoglobinas e Genética das Doenças Hematológicas – LHGDH da UNESP campus de São José do Rio Preto

Autor para correspondência: Luciane Moreno Storti de Melo

LHGDH, Departamento de Biologia, IBILCE-UNESP

Rua Cristóvão Colombo, 2245 – Jd Nazareth,

13.540-000 - São José do Rio Preto, SP

Fone: (13) 2242-2392. Fax: (13) 2242-2397

E-mail: stortiluc@yahoo.com.br

Resumo

As he moglio inopatas fora inclu das no Programa Nacional de Triage Neonatal a partir de seis de junho de 2000, com a portaria do Ministerio da Saude nº 822. A cromatografia liquida de alta pressao e sistema automatizado foi indicada como uma das metodologias de diagnostico. A quantidade de informacoes geradas por esses sistemas i ensa e, nesse pre, uco portanto pode ser o servado nas anlises individuais. Tecnicas de visualiza o tridimensional pode ser aplicadas para extra o dessas informacoes, com o o esta eleci o de padroes, tendencias ou rela oes a partir de resultados ar azenados e bases de dados. Ojetivou-se nesse tra alho a aplica o de uma ferramenta *datamining* de visualiza o 3D para detec o de padroes nos resultados de diagnostico neonatal das he moglio inopatas por HPLC. Foram utilizados os resultados laboratoriais de 252 anlises de recém-nascidos realizadas nos anos de 2000 e 2002. Os valores quantitativos de he moglio inas, selecionados como atributos para o apeamento foram os picos "Fast", "F1", "F2" e "A" e itidos pelo sistema analitico. Para esta eleger um padrao de como portanto os resultados foram classificados em grupos de estudo segundo o fenotipo de he moglio ina: "Normal" n = 209, "Variante" n = 37 e "Talassemia" n = 29. A idade gerada pela anlise conjunta dos tres grupos, com todos os atributos, mostrou uma separa o espacial e nuvens distintas, mesmo apos reposi o. As idades de "Talassemia" e "Normal" apresentaram-se dia etral mente opostas, evidenciando a existencia de diferen as nos padroes obtidos por HPLC. A visualiza o entre os grupos "Normal" e "Variante" apresentou duas nuvens distintas com o reposi o nas extremidades. A rea de interseca o nesses grupos pode ser atribuida a heterozigose das a ostras com he moglio ina variante. A anlise dos resultados permitiu a identifica o de um padrao de como portanto diferenciado entre a ostras normais e com talassemias, apesar da HPLC nao ser utilizada como metodologia diagnostica para esse tipo de altera o, permitindo inferencias nos poss veis diagnosticos finais.

Introdução

As herdoenças inopáticas são um grupo de alterações genéticas que representam um problema de saúde pública e países onde a sua incidência é elevada.^{1,2} No Brasil, o processo de colonização teve grande influência na dispersão de genes mutantes, e a distribuição das herdoenças está intimamente relacionada com as etnias que compõem a população.

Os programas de triagem neonatal têm como objetivo o diagnóstico precoce de alterações genéticas e metabólicas antes do aparecimento dos primeiros sintomas, reduzindo a mortalidade e melhorando a qualidade de vida dos portadores. No Brasil, as herdoenças inopáticas foram oficialmente incluídas nos programas de triagem neonatal a partir de 2000 com a implantação do Programa Nacional de Triagem Neonatal através da Portaria do Ministério da Saúde nº 822 visando a identificação de crianças com anemia falciforme, a realização de aconselhamento genético dos casais de risco e acompanhamento clínico adequado dos doentes.³

Estudos para a detecção de herdoenças inopáticas em neonatos foram realizados por diferentes grupos, cada qual enfocando diferentes métodos de coleta e análise.⁴ As metodologias de diagnóstico nesta faixa etária devem ser determinadas por sua eficiência na identificação de frações com pequenas porcentagens e na praticidade para uso em estudos populacionais.^{5,6} A portaria MS nº 822 preconiza a utilização da focalização isoelétrica ou da cromatografia líquida de alta pressão (HPLC), por serem métodos altamente sensíveis e principalmente de alta reprodutibilidade, podendo ser analisados grande número de amostras.⁸ As análises por HPLC fornecem uma grande quantidade de informações geradas pelo sistema automatizado de difícil análise individual quando se objetiva obter padrões de alterações na população.

Para se ter um real aproveitamento das informações contidas em grande volume de dados, necessita-se a utilização de técnicas de tratamento de informações, que têm como primeira ferramenta a apresentação gráfica dos dados.⁹ Essa apresentação consiste em visualizar graficamente dados, os quais se

deseja analisar. centrada e mecanis os que possibilita ao usu rio a co preensao r pida da inform a o a ele apresentada. A explorao visual de dados fornece u grau ais alto de confiabilidade nas conclusoes da explorao do que u a representa o nu rica ou textual¹.

O conjunto de tcnicas auto ticas para a explorao e grandes assas de dados defini-se co o *Data Mining*, ou seja, inera o de dados, confeccionadas co o jectivo de encontrar novos padroes, tendncias e rela oes. Atrav s destas tcnicas, poss vel a extra o de inform oes teis, que so difceis de sere desvendadas pelo ser hu ano a olho nu, sendo elhor interpretadas quando so apresentadas de for a grafica². O *FastMapDB* u a ferramenta de visualiza o de dados, co o datas, n eros e textos, onde a apresenta o efetuada atrav s do apeamento ultidimensional dos dados para u espa o tridimensional, proovendo a redu o de dimensionalidade. O jectivo da redu o de dimensionalidade apresentar o conjunto de dados co u n ero enor de atributos, por preservando as caracter sticas inerentes a inform o ar azenada, dessa for a uscando definir quais so os atributos importantes ou que concentra a inform o². Possi ilita verificar pontos fora do padrao *outliers*, e a for a o de agrupamentos *clusters*. Por exe plo, e u a ase de dados correspondente a resultados de exa es laboratoriais, que no apeamento desses dados apresente aglomerados de pontos, pode significar que os exa es que co poe cada aglomerado fe algo e co u³.

Diante a diversidade de fen tipos de H o servados na popula o brasileira e a quantidade de inform oes geradas e progra as neonatais o jetivou-se nesse tra alho a aplica o da ferramenta *data mining* de visualiza o 3D - *FastMapDB* para detec o de padroes de comporta ento nos resultados quantitativos o tidos no diagn stico neonatal de he oglo inopatias por HPLC.

Casuística e Métodos

Fora utilizados os resultados laboratoriais de H o tidos por HPLC de 252 amostras de sangue de cordao umbilical de recém nascidos da regiao

noroeste do estado de São Paulo nos anos de 2000 e 2002 se distinguiu de sexo e origem étnica. O presente estudo teve aprovação da CONEP e esta é concordância com as normas estabelecidas pela Resolução CNS 96/96.

O diagnóstico final para determinação do fenótipo foi realizado pela combinação de diferentes metodologias laboratoriais, incluindo: Resistência osmótica e solução de NaCl a 0,3%⁴; teste seletivo para rastreamento de talassemias do tipo beta; Análise da morfologia eritrocitária⁵;

per ite ainda, utilizar u atri uto da ta ela para classificar as tuplas e a visualiza o das classes representadas por diferentes cores e for atos ³.

Para aplica o da ferra eta *datamining – FastMapDB*, fora utilizados os resultados quantitativos o tidos por HPLC. Esse siste a identifica os valores percentuais das fra oes de H e picos deter inados, $\underline{F_{ast}}$, $\underline{F_{\alpha}}$, $\underline{F_{\beta}}$ e \underline{A} , que depois de inseridos e planilhas, fora utilizados co o atri utos para o apea ento espacial e, os fen tipos fora utilizados co o atri uto classificat rio das a ostras. Para esta elecer u padrao de co porta ento espacial os resultados fora classificados e grupos segundo delinea ento ilustrado no quadro .

Quadro 1. Divisao e grupos de an lise para o *FastMap*, segundo os fen tipos de H o servados.

N	11.1 Fenótipos de Hb	11.2 Grupos de análise para o Fast Map
2 9	Perfil de H normal para idade	11.3 “Normal”
55	H S e heterozigose	“Variante”
8	H C e heterozigose	
2	eta talasse ia heterozigota	“Talassemia”
258	alfa talasse ia heterozigota	

Resultados

Visando esta eleger o perfil do fen tipo normal para a ostras de sangue de neonatos, inicialmente, foram geradas imagens utilizando os quatro picos discriminados com o atributos de aparência para a ferramenta *datamining*

Figura 2

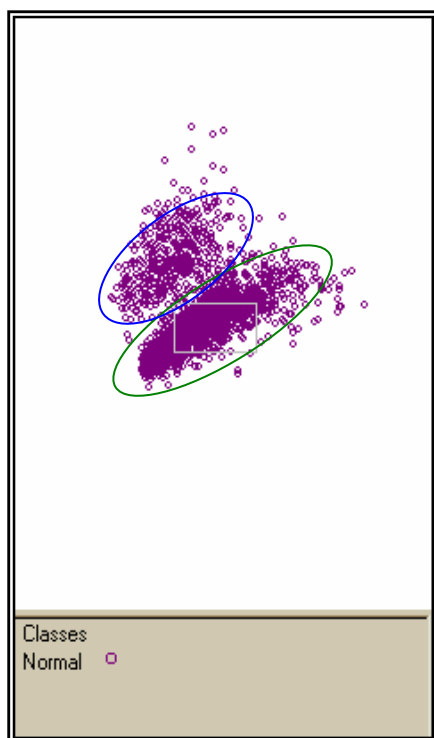


Figura 2. Projeção da imagem para o grupo "Normal", incluindo os quatro atributos de aparência.

O aparecimento das a ostras normais apresentou dois aglomerados de pontos marcados pelos círculos na Figura 2. Para avaliar qual o fator responsável pela divisão das a ostras em dois grupos de distribuição espacial, foram geradas imagens eliminando alguns dos picos, na tentativa de unir os dados em uma única nuvem de concentração. A imagem gerada sem o pico "Fast" mostrou a aproximação das nuvens, no entanto ainda pode ser observada a presença de dois grupos separados. Retirando-se o pico de "E" ocorreu uma aproximação das nuvens de dados, sugerindo que esses valores podem estar determinando a divisão das a ostras normais em dois padrões distintos de normalidade.

Associando-se as duas observações, gerou-se uma imagem apenas com os picos de "E" e "A", retirando-se as janelas de "Fast" e "E". Figura 2, e observou-se a dispersão linear dos dados.

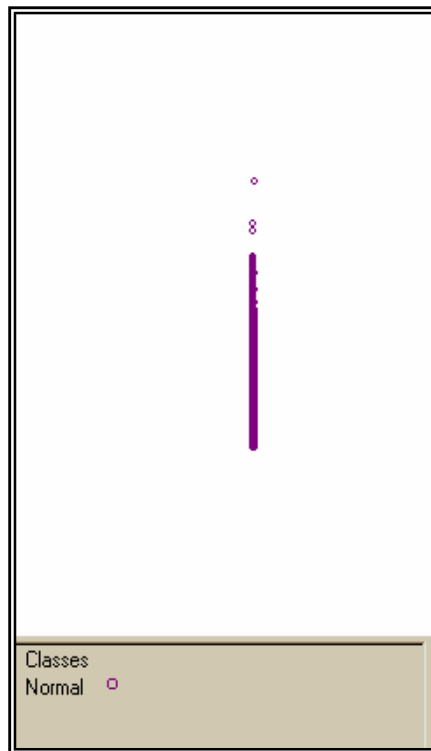


Figura 2. Projeção da imagem para o grupo "Normal", incluindo os picos "E" e "A" com o atributo de aparência.

Para avaliar os valores de "Fast" e "E" foram geradas imagens utilizando um recurso da ferramenta *FastMapDB* que permitiu a divisão exata no valor médio, constituindo dois grupos de classificação para cada pico de análise. A imagem ilustrada na Figura 3 contém os quatro picos de análise e os valores de "Fast" divididos e utilizados na classificação, representada na imagem pela diferença no tamanho dos pontos.

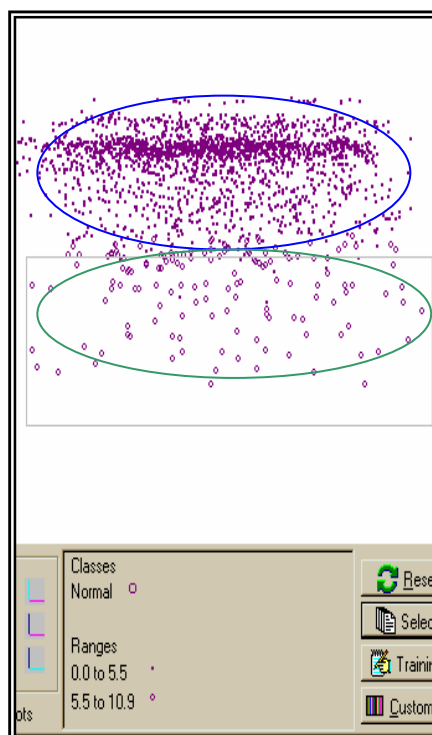


Figura 3. Projeção da imagem para o grupo "Normal", incluindo os quatro atributos de aparência e utilizando o pico de "Fast" como classificatório dividido e dois subgrupos sendo de 0,0 a 5,5% e de 5,5% a 10,9%.

Como o observado, os pontos marcados pelo círculo azul apresentam valores de "Fast" entre 0,5 e 5,5 % e os pontos dentro do círculo verde correspondem a valores de "Fast" de 5,5 a 9,9%. Destaca-se que não existe sobreposição dos pontos indicando que os dois subgrupos correspondem a duas nuvens distintas. Analisando estatisticamente esses dados, obteve-se a média de 2,34 % para o subgrupo com valores variando de 0,5 a 5,5 % e este incluiu um total de 24 amostras. O subgrupo variando de 5,5 a 9,9 % foi composto por apenas 29 amostras e a média observada foi de 3,5%. A diferença percentual obtida no pico de "Fast" para todas as amostras normais foi de 2,94%, apresentando diferença estatisticamente significativa $p < 0,05$ das médias obtidas para os dois subgrupos acima descritos.

Utilizando o mesmo recurso da ferramenta *FastMapDB*, a classificação foi feita utilizando-se o pico de "F" para dividir as amostras no valor médio e a distribuição obtida está ilustrada na Figura 4, na qual a distribuição espacial dos dados foi muito semelhante à da Figura 3. As duas nuvens classificadas pelos valores observados na Figura 4 não apresentaram sobreposição dos pontos. Para possibilitar a comparação das imagens os gráficos das Figuras 3 e 4 foram dispostos lado a lado na Figura 5. Para caracterizar a divisão em duas nuvens amostrais, calculou-se a diferença percentual para cada subgrupo obtido para o pico de F. Obteve-se para o subgrupo com valores variando de 0,5 a 5,3% uma diferença percentual de 9,45% incluindo 39 amostras e para o subgrupo com valores variando de 5,3 a 9,9% obteve-se a média de 2,25%, sendo composto por 8 amostras. A diferença percentual obtida para todas as amostras normais foi de 0,48%, apresentando diferença estatisticamente significativa $p < 0,05$ das médias obtidas para os dois subgrupos de "F".

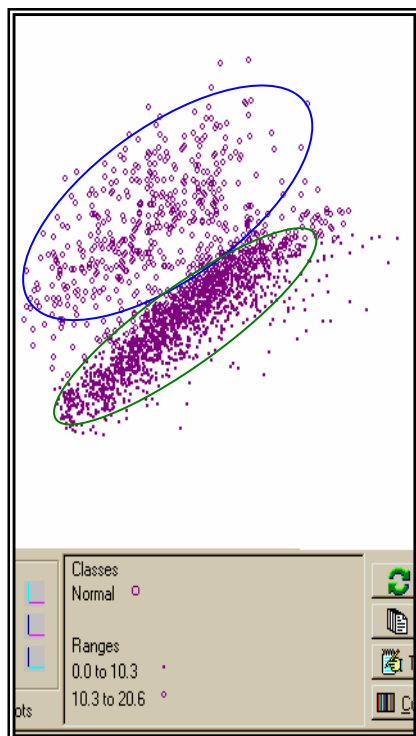


Figura 4. Projeção da idade para o grupo “Normal”, incluindo os quatro atributos de apeamento e utilizando o pico de “E” com o classificatório dividido em dois subgrupos sendo de 2,3 a 2,3% e de 2,3% a 2,4%.

A

B

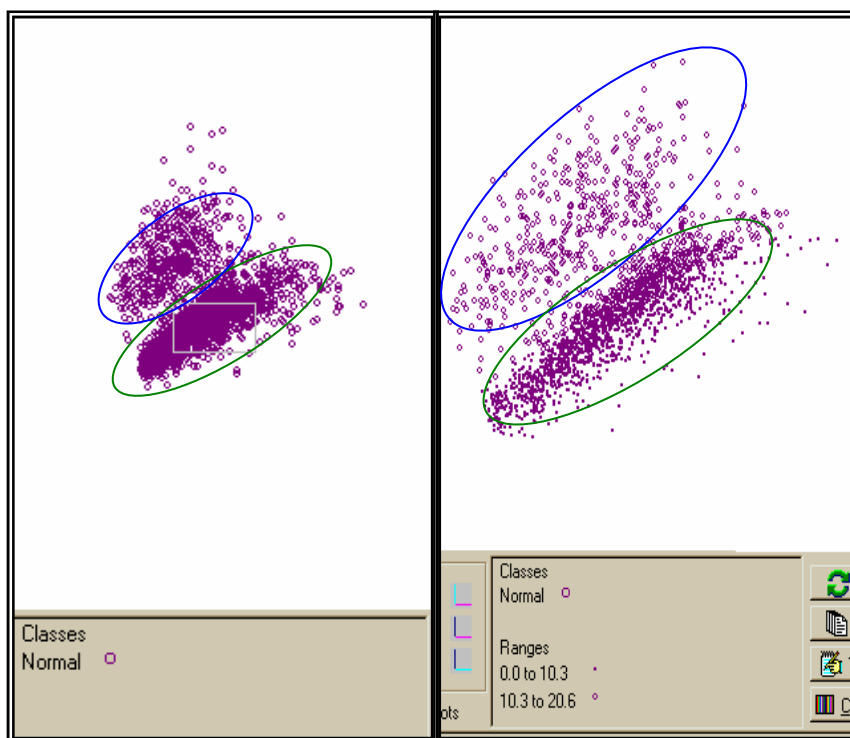


Figura 5. Comparação das idades projetadas para o grupo “Normal”. A - incluindo os quatro picos com os atributos de apeamento B - utilizando o pico de “E” dividido em dois subgrupos de 2,3 a 2,3% e de 2,3 a 2,4%.

Visando esta eleger um padrão para os diferentes fenótipos observados foram traçadas todas as amostras, divididas nos três grupos, "Normal", "Talassemia" e "Variante", incluindo os quatro picos com o atributo para o eixo Y. Na imagem gerada, os grupos de análise apresentaram-se distribuídos em três nuvens amostrais distintas, como ilustrado na Figura 4, não houve sobreposição entre os pontos. As nuvens correspondentes aos grupos "Talassemia" e "Normal" encontram-se diametralmente opostas.

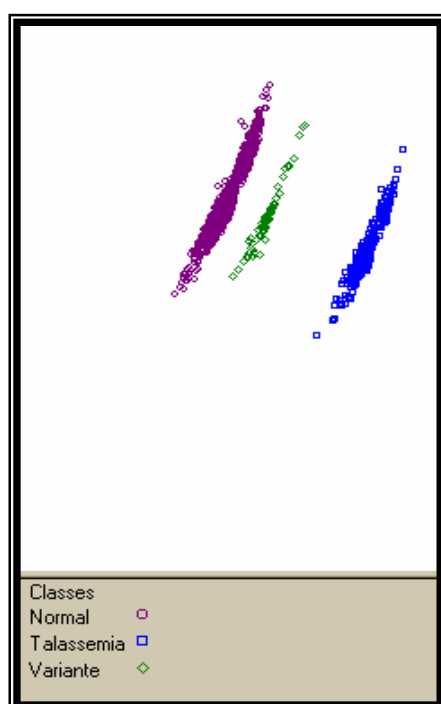


Figura 4. Projeção da análise tridimensional para os grupos "Normal", "Talassemia" e "Variante", incluindo os quatro atributos de eixo Y.

Em seguida foram traçadas as amostras do grupo "Normal" e "Talassemia" e na imagem gerada utilizando os quatro atributos de eixo Y observou-se a diferença nos eixos gerados distintos para cada grupo, sem sobreposição espacial. Para verificar quais as frações fundamentais nessa diferença, foi gerada uma imagem utilizando apenas os picos "F" e "A" com o atributo de eixo Y, por representarem as frações de hemoglobinas e maiores concentrações nesta faixa etária. Na imagem gerada, observou-se a distribuição linear e paralela dos grupos amostrais, destacando que os valores percentuais das frações "F" e "A", obtidos por HPLC, apresentam diferenças significativas para o eixo espacial, como pode ser visualizado na Figura 5.

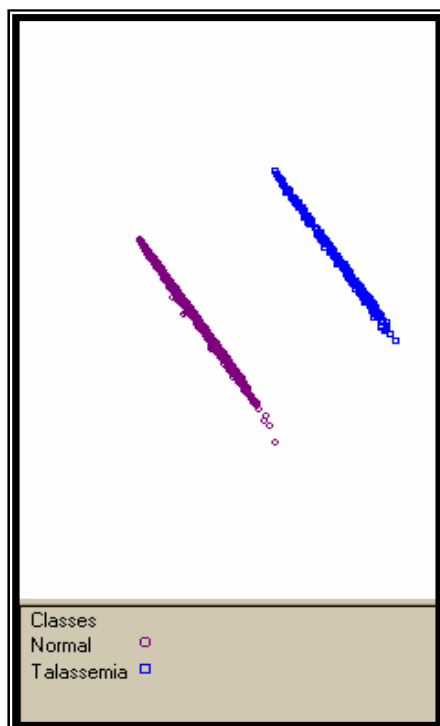


Figura 8 – Projeção da análise tridimensional para os grupos “Talassemia” e “Normal” utilizando como atributos de apeamento os picos F_2 e A_2 .

Para análise mais precisa, foram geradas imagens com os fenotipos alfa e beta talassemia separados e comparados com o normal. O apeamento desses grupos incluindo os quatro atributos mostrou o reposicionamento dos pontos aóstrais. Dessa forma, esses grupos foram apeados utilizando dois picos da HPLC combinados dois a dois e foi possível verificar nas imagens geradas o delineamento dos grupos e três aglomerados distintos representando cada grupo fenotípico. A Figura 8 ilustra o apeamento incluindo os picos F_2 e A_2 , no qual pode-se observar que as aóstras se distribuíram linearmente no espaço, estando bem definidas e três conjuntos aóstrais distintos. A imagem mostrou ainda que, os grupos alfa e beta talassemias estão mais próximos entre si e afastados do grupo normal.

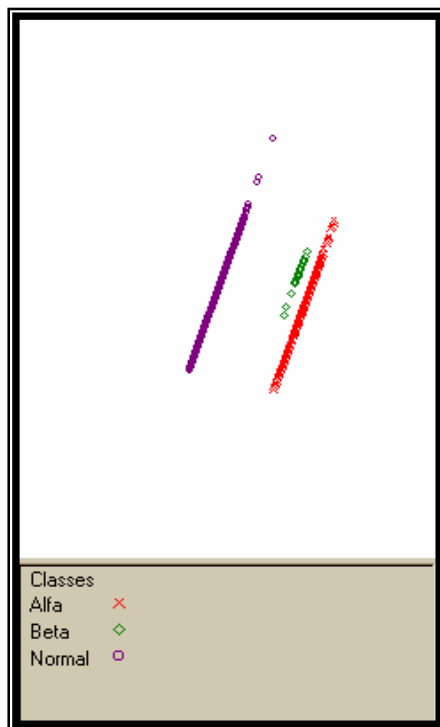


Figura 8. Projeção da idade tridimensional para os grupos Normal”, Alfa” e Beta” talassemias, utilizando como atributos de dependência os picos F” e A”.

No planejamento espacial dos resultados laboratoriais obtidos para as amostras normais e com as variantes, serviu-se da existência de uma interface entre os grupos, como pode ser visualizado no destaque da Figura 9. As amostras se distribuíram apresentando uma área de sobreposição na extremidade, que se antevia ser o caminho espacial da idade.

O grupo Variante” foi composto pelas amostras contendo H_{AS} e H_{AC}, para tentar estabelecer um padrão entre essas variantes e diferenciá-las das amostras com as normais, realizou-se o planejamento incluindo o grupo Normal” com cada uma dessas variantes isoladamente. O padrão de dispersão espacial se antevia, não sendo possível separar as amostras e nuvens amostrais distintas. A Figura 10 representa o planejamento incluindo os grupos Normal”, H_{AS} e H_{AC} para os quatro atributos de planejamento. Mesmo analisando as amostras separadamente a área de sobreposição se mantém. Deve-se destacar que as frações variantes estão representadas no HPLC por picos específicos denominados S” e C”, e estes não foram incluídos como atributos para o planejamento espacial por estarem ausentes nas amostras com hemoglobinas normais.

Discussão

A aplicação de técnicas de VDM à a ordenação nova nas pesquisas com o hecoglo inopáticas. Para os dados da triagem neonatal realizada pelo LHGDH as análises pela ferramenta *FastMapDB* apresentaram padrões relevantes. Na projeção das aóstras com o hecoglo inas normais para a idade observou-se a existência de duas nuvens de distribuição aóstral. O apeamento feito se a presença dos picos de *Fast* e *F.* mostrou a distr

incluiu 94% das amostras, não sendo representativo da distribuição observada na Figura 3. Esses resultados destacam que o pico de "Fast" não foi o responsável pela formação das duas nuvens amostrais observadas no alinhamento das amostras com o genoma humano normal para a idade.

No alinhamento feito utilizando a divisão do pico "Fast" no valor médio, teve-se dois grupos distribuídos de forma quase igualitária, sendo o primeiro composto por 43,7% das amostras e o segundo incluindo 34,9%. As diferenças para esses dois subgrupos apresentaram diferenças estatisticamente significativas ($p < 0,05$) quando comparadas entre si e com a média geral observada para todas as amostras normais. Quando comparadas lado a lado, as visualizações observadas para as amostras normais com os quatro atributos e com o pico de "Fast" divididos em dois subgrupos (Figura 4), pode-se observar a existência de dois pontos amostrais. Esses resultados indicam a presença de dois padrões com valores médios diferentes de distribuição para os valores de "Fast" nas amostras com H normal. Diante desses achados, deve-se destacar, portanto, que a média geral observada para as 209 amostras não é valor representativo deste pico de H para o grupo "Normal".

Analisando globalmente os resultados, visualizados e projetados espacialmente através da ferramenta *FastMapDB*, ressaltou a existência de um possível padrão para as amostras talassêmicas por HPLC, pois não houve o reposicionamento espacial dos pontos amostrais, sugerindo diferenças nos perfis observados por HPLC para as amostras com valores normais para a idade e aquelas com formas talassêmicas, alfa ou beta, o que é heterozigose. A HPLC não é a metodologia precisa para o diagnóstico das formas talassêmicas nessa faixa etária, no entanto os resultados isolados sugeriram um perfil característico para as amostras que apresentaram talassemias, diagnosticadas por outras metodologias. Mendes-Siqueira, 2004 observou um aumento estatisticamente significativo para os picos de "Fast" e "Fast" e amostras com alfa talassemia. Para as amostras com beta talassemia observou-se valores aumentados para o pico de "Fast" e diminuídos para "A", sendo as diferenças estatisticamente significativas dos valores observados para o grupo

normal, sugerindo a existência de um perfil diferenciado para os cromatogramas com as outras talassêmicas.

Apesar da HPLC não poder ser utilizada como todo diagnóstico das talassemias, os diferentes perfis observados na projeção espacial destacam a necessidade de melhor avaliar os resultados obtidos na triagem por HPLC, pois diferenças sutis nos cromatogramas podem indicar uma provável talassemia, sendo um todo adicional no diagnóstico dessas alterações em neonatos, de forma a contribuir para o direcionamento das análises associado a outras metodologias específicas para confirmar esta alteração.

As H_S e H_C são diagnosticadas com precisão pela HPLC que identifica com alta sensibilidade as pequenas porcentagens das frações variantes. No entanto o aparecimento espacial das talassemias heterozigotas para essas hemoglobinas comparadas ao grupo Normal mostrou-se reposição das nuvens alóstrais, não sendo possível diferenciar os grupos. O pico de H_{Fetal}, que representa o maior componente hemoglobínico nesta faixa etária, não apresenta alterações nos valores percentuais entre as talassemias normais e com H_{variantes}, esse fato associado à diferença representativa no número de talassemias, pode ser responsável pela substituição das nuvens alóstrais. Deve-se destacar também que as H_{variantes} exibem picos específicos da cromatografia e estes não foram inseridos no aparecimento por estarem ausentes nas talassemias normais, não sendo possível a comparação.

As técnicas e ferramentas para visualização de dados são instrumentos indispensáveis ao processo de descoberta do conhecimento e têm provado seus valores na exploração e análise de dados, sendo assim, essas técnicas têm sido utilizadas nos sistemas de *Visual Data Mining* (VDM), ou seja, mineração visual de dados. Com as técnicas de visualização grandes volumes de dados podem ser apresentados na tela ao mesmo tempo; cores diferentes permitem ao usuário reconhecer, instantaneamente, similaridades e diferenças de clusters de itens, e os itens podem ser arranjados para expressar algum relacionamento entre si, se este existir.

Conclusões

As hemoglobinopatias apresentam diversidade de fenótipos na população brasileira, influenciados pelos fatores ambientais e genéticos. A diversidade de características intrínsecas a cada fenotipo pode ser destacada individualmente pelos resultados laboratoriais. No entanto, a análise global dos grupos fica prejudicada quando se pretende avaliar um grande número de amostras. A projeção espacial dos resultados permitiu a identificação de comportamentos diferenciados entre as amostras normais e talassêmicas através da separação em nuvens distintas, ressaltando a necessidade de avaliar os resultados cromatográficos e conjunto com outras metodologias, para o diagnóstico das talassemias. O *FastMapDB* permitiu o agrupamento das informações e a projeção tridimensional dos dados no espaço, facilitando a visualização dos perfis dos diferentes fenótipos e a interface entre grupos teoricamente distintos.

Referências

1. W.H.O.- WORLD HEALTH ORGANIZATION. Hereditary anaemias: genetic basis, clinical features, diagnosis and treatment. **Bulletin of the W.H.O.**, v. 4, n. 5, p. 43-59, 1962.
2. GALACTEROS, F. Neonatal detection of Sickle Cell Disease in metropolitan France. association française pour le dépistage et la prevention des handicaps de l'enfant AFDPE, **Arch. Pediatr.**, v. 3, p. 24-3, 1994.
3. BRASIL. Ministério da Saúde. Portaria da Saúde, nº 822, de junho de 2000. Brasília, 2000.
4. JANUARIO, J. N. Programa de triagem neonatal apresenta primeiros resultados. **Jornal Hemominas**, v. 8, n. 3, p. 5-7, 1998.
5. HAYASHI, A. et al., Neonatal screening and mass-spectrometric analysis of hemoglobin variants in Japan. **Acta Hemat.** v. 8, p. 4-8, 1998.

7. FUCHAROEN, S.; et al., Prenatal and postnatal diagnoses of thalassaemias and haemoglobinopathies by HPLC. **Clinical Chemistry**, v. 44, n. 4, p. 44-48, 1998.

8. LOU, C.; ROGNERUD, C. L. Diagnosis of haemoglobinopathies: electrophoresis vs. HPLC. **Clinica Chimica Acta.**, v. 333, p. 89-94, 2000.

8. BRASIL. Ministério da Saúde. **Manual de normas técnicas e rotinas operacionais do programa de triagem neonatal**. Anvisa. Brasília, 2002.

9. CHITTARO, L. Information visualization and its application to medicine, artificial Intelligence in Medicine, Elsevier, 2000.

10. EIM, D. A. Visual exploration of large data sets. **Communications of the ACM**, v. 4, n. 8, p. 39-44, 2000.

11. BARIONE, M. C. N.; et al., Visualização de dados e sistemas de banco de dados relacionais. **Proceeding XVI Simpósio Brasileiro de Banco de Dados**, Rio de Janeiro, p. 95-99, Outubro de 2000.

12. FALOUTASOS, C.; LIN, C. *FastMap*: A fast algorithm for indexing, data mining and visualization of traditional and multimedia datasets, **ACM Sigmod**, Zurich, Switzerland, p. 3-4, 1995.

13. TRONCO, M. N. **Implementação de Recursos de Visualização e Interação em Ferramenta Data Minig**. São Jose do Rio Preto, 2003, 50 f.

14. SILVESTRONI, E.; BIANCO, I. Screening for sickle cell anemia in Italy: analysis of data collected in the past 3 years. **Am. J. Hum. Genetic**, v. 28, p. 98, 1957.

15. BONINI-DOMINGOS, C. R. **Hemoglobinopatias no Brasil - Variabilidade genética e metodologia laboratorial**. 993. 23 f. Tese Doutorado em Ciências, – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, 1993.

16. MARENGO-ROE, A. J. Rapid electrophoresis and quantitation of haemoglobin on cellulose acetate. **J. Clin. Path.**, v. 8, p. 9-12, 1955.

17. DACIE, J. V.; LEISS, S. M. **Practical haematology**. 8 ed. Livingstone: Churchill, 1995.

18. VELLA, F. Acid agar gel electrophoresis of human haemoglobin. **Am. J. Clin. Path.**, v. 49, n. 3, p. 44-442, 1968.

9. PAPAYANNOPOULOS, R.; STAMATONYANNOPOULOS, G. **Stains for inclusions bodies.** In: STANDARTIZATION of laboratory reagents and methods for detection of haemoglobinopathies". Atlanta: Hew publications, 1994.

2. Instrucion Manual Variant Sickle Cell Short Progra , 1994.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)