

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

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

2º Examinador Prof. Dr. Paulo Peitl

3º Examinador Prof. Dr. Carlos Roberto Valêncio

São José do Rio Preto, 21/02/2006.

Storti-Melo, Luciane Moreno.

Mineração de dados em triagem neonatal de hemoglobinopatias / Luciane Moreno
Storti de Melo - São José do Rio Preto: [s.n.], 2006
88 f. : 28.il. ; 30 cm.

Orientador: Claudia Regina Bonini-Domingos

Dissertação (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*. 6. *Visual data mining*. I. Bonini-Domingos, Claudia Regina. II.
Universidade Estadual Paulista. Instituto de Biociências, Letras e Ciências Exatas.
III. Título.

CDU – 616.155

Sumário

Dedicatória.....	i
Lista de Figuras	iii
Lista de Abreviaturas	v
1 Introdução.....	2
1.1 Hemoglobinas humanas	2
1.2 Genética, síntese e ontogenia das Hb normais	3
1.3 Hb Anormais	6
1.3.1 Hb Variantes	7
1.3.2 Talassemias.....	9
1.4 Programas de triagem neonatal	11
1.4.1 Triagem Neonatal de Hemoglobinopatias	11
1.4.2 Implantação do Programa Nacional de Triagem Neonatal (PNTN).....	12
1.5 Bases de dados Eletrônicas	15
1.5.1 Data Mining	16
1.5.1.1 Histórico tecnológico	17
1.5.2 <i>Visual Data Mining (VDM)</i>	18
1.6 Justificativa	21
2 Objetivos	23
2.1 Objetivos específicos	23
3 Material e Métodos	25
3.1 Casuística	25
3.2 Métodos para o diagnóstico de Hemoglobinopatias	26
3.3 Ferramentas <i>Data Mining</i>	36

3.4 Análises Estatísticas	39
4 Resultados	41
4.1 O banco de dados e a ferramenta CLIBIA	41
4.2 Padrões de Hb obtidos na análise pelo <i>FastMapDB</i>	46
4.2.1 Perfil das amostras normais	46
4.2.2 Perfis obtidos para os diferentes fenótipos de Hb	51
5 Discussão	62
6 Conclusões.....	71
6.1 Contribuições para implementação de recursos na ferramenta <i>FastMapDB</i> ..	72
7 Referências.....	74
8 Resumo	83
9 Abstract.....	85
10 Apêndices	87
10.1 Relatório Geral obtido 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 Silvío, 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 Hb.*	2
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.	4
Figura 3: Síntese de cadeias polipeptídicas ao longo do desenvolvimento humano*	6
Figura 4: Mapa representativo da cobertura estadual dos nascidos vivos e da fase de credenciamento dos SRTN. NI = não informado. Fonte: SIA/SUS.	14
Figura 5: Pagina inicial de acesso ao banco de dados LHGDH <i>On-line</i>	41
Figura 6: Página inicial do banco de dados. O círculo em verde destaca o <i>link</i> para a ferramenta CLIBIA.	41
Figura 7: Página inicial da ferramenta CLIBIA. O círculo em verde destaca os <i>links</i> para a seleção das opções consulta ou relatório.	42
Figura 8: Página de seleção do <i>dataset</i> 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 como <i>dataset</i> na opção consulta da ferramenta CLIBIA.	43
Figura 10: Página de seleção do <i>dataset</i> 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 Hb observadas.	45
Figura 12: Projeção da imagem tridimensional para o grupo “Normal”, incluindo os quatro picos como atributos de mapeamento.	47
Figura 13: Projeção da imagem tridimensional para o grupo “Normal” incluindo os picos “F1”, “F” e “A” como atributos de mapeamento. A seta indica o espaço entre os aglomerados.	47
Figura 14: Projeção da imagem tridimensional para o grupo “Normal” incluindo os picos “Fast”, “F” e “A” como atributos de mapeamento.	48
Figura 15: Projeção da imagem tridimensional para o grupo “Normal” incluindo os picos “F” e “A” como atributos de mapeamento.	48
Figura 16: Projeção da imagem tridimensional para o grupo “Normal”, incluindo os quatro atributos de mapeamento e utilizando o pico de “Fast” como classificatório em dois subgrupos sendo o primeiro variando de 0,0 a 5,5 % e o segundo de 5,5% a 10,9%.	50
Figura 17: Projeção da imagem tridimensional para o grupo “Normal”, incluindo os quatro atributos de mapeamento e utilizando o pico “F1” como classificatório em dois subgrupos sendo o primeiro variando de 0,0 a 10,3 % e o segundo de 10,3% a 20,6%.	50
Figura 18: Comparação das imagens projetadas para o grupo “Normal”. A - incluindo os quatro picos como atributos de mapeamento B - utilizando o pico de “F1” dividido em dois subgrupos variando de 0,0 a 10,3% e de 10,3 a 20,6%.	51

Figura 19: Projeção da imagem tridimensional para os grupos “Normal”, “Talassemia” e “Variante”, incluindo os quatro atributos de mapeamento.....	52
Figura 20: Projeção da imagem tridimensional para os grupos “Normal” e “Talassemia” incluindo os quatro atributos de mapeamento.....	52
Figura 21: Projeção da imagem tridimensional para os grupos “Talassemia” e “Normal” utilizando como atributos de mapeamento os picos “F” e “A”.....	54
Figura 22: Projeção da imagem tridimensional para os grupos “Normal”, “Alfa” e “Beta” talassemias, utilizando como atributos de mapeamento os picos “F” e “A”.....	54
Figura 23: Projeção da imagem tridimensional para os grupos “Normal” e “Alfa” talassemia, utilizando como atributos de mapeamento os picos “Fast”, “F” e “A”.....	55
Figura 24: Projeção da imagem tridimensional para os grupos “Normal” e “Alfa” talassemia, utilizando como atributos de mapeamento os picos “Fast” e “A”.....	55
Figura 25: Projeção da imagem tridimensional para os grupos “Normal” e “Variante” utilizando os quatro atributos de mapeamento. Área de sobreposição marcada pelo círculo vermelho.....	57
Figura 26: Projeção da imagem tridimensional para os grupos “Normal”, “Hb AS” e “Hb AC” utilizando os quatro atributos de mapeamento. Área de sobreposição marcada pelo círculo verde.....	57
Figura 27: Projeção da imagem tridimensional para os grupos “Normal” e “Hb S” utilizando os quatro atributos de mapeamento. Área de sobreposição marcada pelo círculo verde.....	58
Figura 28: Projeção da imagem tridimensional para os grupos “Normal” e “Hb C” utilizando os quatro atributos de mapeamento. Área de sobreposição marcada pelo círculo verde.....	58

Lista de Abreviaturas

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 Databases</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>

1 Introdução

1 Introdução

1.1 Hemoglobinas humanas

A hemoglobina (Hb) é uma proteína globular, com peso molecular de 64.458 Daltons, formada por quatro subunidades (Figura 1). Cada subunidade é composta de uma fração protéica, a globina e um grupo prostético heme, que contém o átomo 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 protéica da molécula de Hb é formada por quatro cadeias polipeptídicas totalizando 574 aminoácidos. Duas delas são constituídas por 141 aminoácidos cada e são chamadas tipo alfa. As outras duas possuem 146 aminoácidos, sendo denominadas tipo beta (HONIG; ADAMS III, 1986; WEATHERALL; CLEGG, 2001).

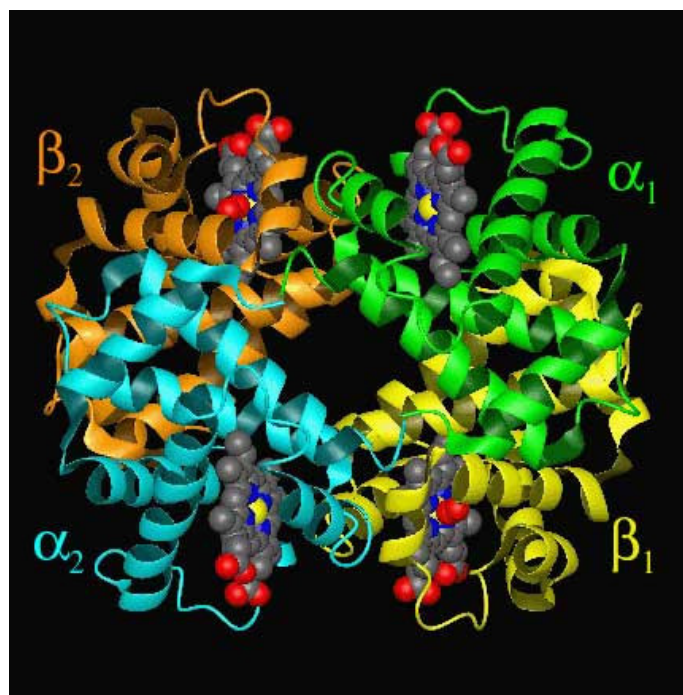


Figura 1: Modelo tridimensional da estrutura quaternária da molécula de Hb.*

(*) Disponível em <http://mercury.bio.uaf.edu/~kevin_mccracken/lab/lab.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 em cromossomos diferentes. Pseudogenes (ψ) são genes com seqüências homólogas aos genes estruturais ativos que acumularam mutações, inibindo sua expressão (LEWIN, 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 1 ($\zeta 2$, $\alpha 2$, $\alpha 1$), dispostos na ordem de sua ativação durante o desenvolvimento, três pseudogenes ($\psi\zeta 1$, $\psi\alpha 2$, $\psi\alpha 1$) 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, mostrado na Figura 2B, está agrupado no braço curto do cromossomo 11 e inclui cinco genes funcionais denominados epsilon, gama alanina, gama glicina, delta e beta (ϵ , γ^A , γ^G , δ e β), dispostos na sua ordem de ativação e um pseudogene ($\psi\beta$) (WEATHERALL; CLEGG, 2001; 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 éxons – e duas regiões não codificadoras, denominados íntrons ou *intervening sequence* – IVS – representadas por seqüências não traduzidas que separam os éxons. O desenvolvimento específico da expressão de cada gene é determinado por interações entre os seus promotores e as respectivas regiões de regulação (WAYE; CHUI, 2001).

Na família dos genes alfa, a regulação é realizada por meio de uma região conhecida como HS-40, localizada a 40Kb acima do grupamento 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

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

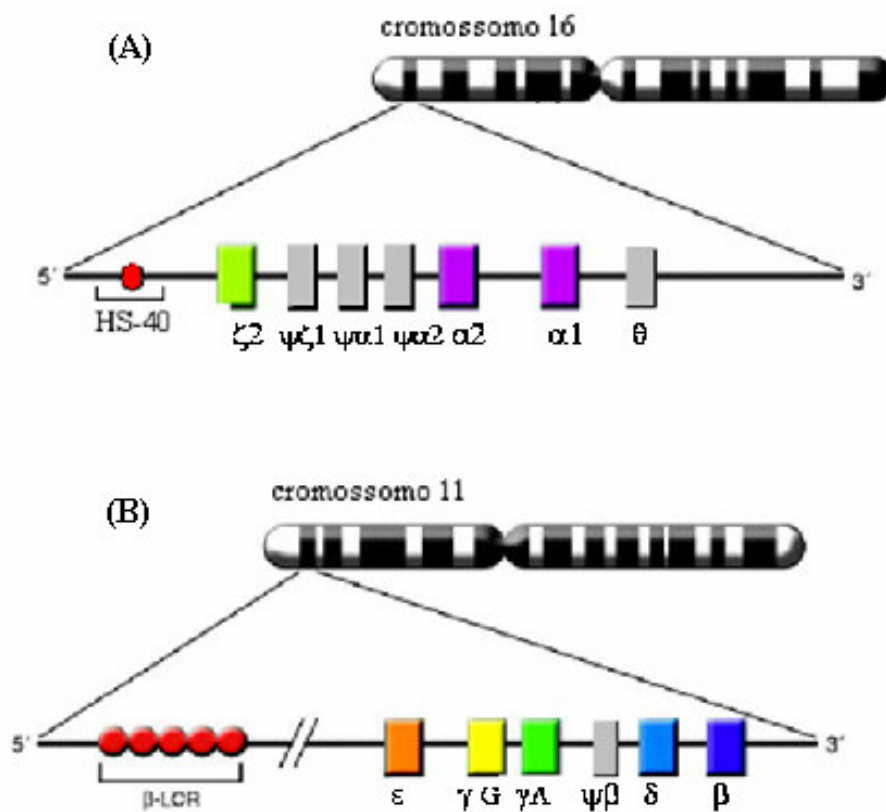


Figura 2: Representação esquemática das famílias gênicas das globinas. (A) Família gênica da globina alfa incluindo a região controladora HS-40, os genes funcionais e pseudogenes. (B) Família gênica da globina beta incluindo a LCR, os genes funcionais e o pseudogene (BANK, 2005, com modificações).

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 globínicas tipo alfa e beta, para a formação das diferentes Hb de forma coordenada (LEWIN, 2000). As Hb diferenciam-se por possuírem características físico-químicas e mobilidades eletroforéticas distintas. Suas funções são marcantes desde os primeiros dias de gestação, adaptando-se ao constante desenvolvimento do embrião e do feto, até estabilizar-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 (WEATHERALL; CLEGG, 2001).

A Hb A ($\alpha_2\beta_2$) começa a ser sintetizada a partir da décima semana de gestação, mantendo-se em concentrações próximas a 10%. Logo após o nascimento, a Hb Fetal e a Hb A apresentam concentrações próximas de 80% e 20%, respectivamente. Suas sínteses se invertem rapidamente até estabilizarem-se, em média, no sexto mês de vida pós-natal. A Hb A₂ ($\alpha_2\delta_2$) tem sua síntese iniciada no final do período fetal e, também, estabiliza-se por volta do sexto mês após o nascimento. As Hb normais encontradas ao nascimento são Hb F (70%-90%), Hb A (0%-10%) e traços de Hb A₂ (0%-1%) (GALACTEROS, 1992, DALMAU et al., 1998). No indivíduo adulto, os valores normais de Hb A são de 96% a 98%, Hb A₂ de 2,5% a 3,5% e Hb F de 0% a 1,0% (BONINI-DOMINGOS, 1993). A Figura 3 ilustra a síntese das globinas em função do período de desenvolvimento.

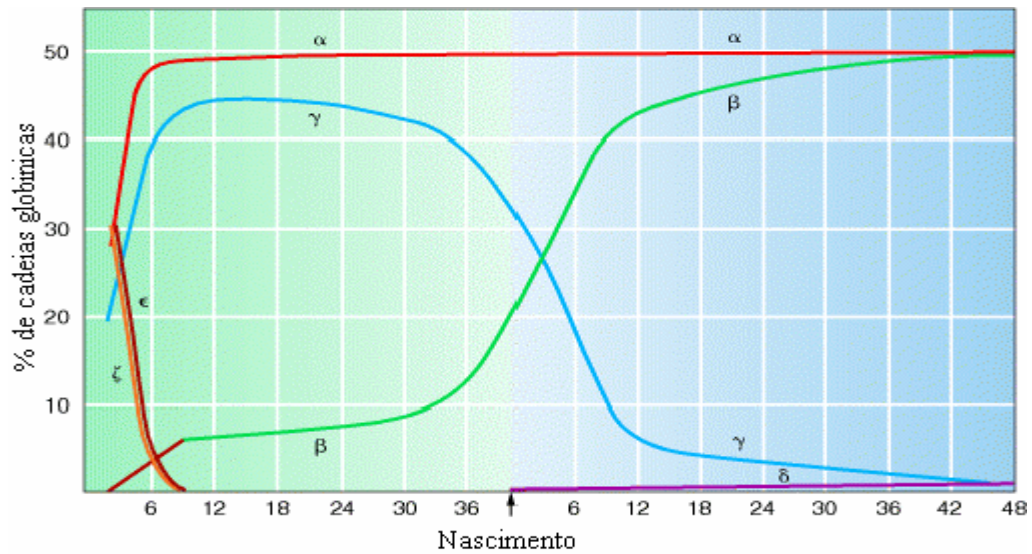


Figura 3: Síntese de cadeias polipeptídicas ao longo do desenvolvimento humano (GRIFFTHS, 1999)

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 promovem a formação de moléculas de hemoglobina com características bioquímicas diferentes das hemoglobinas normais. As talassemias são resultantes de mutações que afetam os genes reguladores promovendo um desequilíbrio no conteúdo quantitativo das cadeias polipeptídicas e conseqüentemente 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 falcização (NAOUM, 1999).

Mais de 1200 alelos mutantes 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 900 variantes de Hb já foram descritas até o momento (HUISMAN, 2005). A maioria delas é originada por simples 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 ocorrem. 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 em 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 destacam-se a Hb S, Hb C e Hb D Los Angeles (CHINELATO-FERNANDES, 2003). A Hb S é uma variante de Hb bem caracterizada, que apresenta prevalência variável nas diferentes regiões do Brasil, dependente dos grupos étnicos formadores de cada região (ZAGO; COSTA, 1985). É causada por uma mutação no gene da globina beta, 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 beta. A mudança de um aminoácido carregado negativamente por um aminoácido apolar gera uma diferença na mobilidade 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,

alterando a morfologia do eritrócito para a forma de foice. A alteração celular, causada pelo processo de falcização, influencia intensamente o fluxo sanguíneo aumentando sua viscosidade. Eventualmente pode ocorrer oclusão de capilares devido à deposição de eritrócitos falciformes na superfície endotelial (SCHNOG et al., 2004).

A Hb C é originada pela alteração no códon GAG para AAG, levando a substituição do ácido glutâmico da cadeia normal pela lisina na posição seis da globina beta. Essa troca de um aminoácido carregado negativamente, por um de carga positiva altera sensivelmente a mobilidade eletroforética da Hb, sendo facilmente diferenciável da Hb A em pH alcalino e ácido. É freqüente em povos oriundos da África Ocidental, onde atinge freqüências de 15% a 30%. Os heterozigotos (Hb AC) são geralmente assintomáticos, não apresentando anemia nem evidências da diminuição do número de eritrócitos, mas a análise morfológica pode apresentar células em alvo. O estado de homozigose para Hb C (Hb CC) é caracterizado por anemia hemolítica de intensidade variável e pode ser equivocadamente diagnosticado (CLARKE; HIGGINS, 2000; CHINELATO-FERNANDES, 2003).

A Hb D foi descoberta em 1951 e ocorre pela troca de uma base no códon GAA por CAA, resultando na substituição do aminoácido ácido glutâmico por uma glutamina na posição 121 da cadeia beta. Essa variante apresenta mobilidade eletroforética idêntica a Hb S em pH alcalino e semelhante a Hb A em pH ácido. Na Índia, região de Punjab, estima-se que 3% da população seja portadora heterozigota dessa alteração (PEREA et al., 1999). A heterozigose para Hb D (Hb AD) é caracterizada pela ausência de sintomatologia clínica, sendo, na maioria dos casos, diagnosticada por rastreamento populacional e estudo familiar. O estado homozigoto (Hb DD) é considerado extremamente raro, podendo apresentar desde parâmetros clínicos e hematológicos normais até uma anemia hemolítica branda com níveis de Hb totais normais. Homozigotos geralmente possuem valores de Hb D acima de 95% e apresentam quantidade normal de Hb A₂. No Brasil, é freqüente sua associação com talassemias e outras variantes. (EL-KALLA; MATHEWS, 1997; CHINELATO-FERNANDES, 2003).

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 em geral resultam em uma anemia microcítica e hipocrômica. A redução da síntese pode ser total ou parcial e são classificadas segundo a cadeia globínica afetada em: 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 a destruição prematura dos eritroblastos na medula óssea e reduzindo a sobrevivência dos eritrócitos na circulação (WEATHERALL; CLEGG, 1981)

As talassemias ocorrem em alta frequência no Mediterrâneo, África, Oriente Médio, Índia e sudoeste da Ásia. Além da importância médica, como o grupo mais comum de patologia monogênica 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 (WEATHERALL, 1997).

As talassemias alfa e beta em 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, 180 defeitos foram relatados, em 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 RNAm 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 podem ser observadas 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 H (β_4). No período fetal, o desaparecimento resulta na formação de tetrâmeros de cadeias gama (γ_4), formando a Hb Bart's. Esses homotetrâmeros, 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 (BANK, 2005). Algumas mutações são específicas de determinados grupos étnicos e mutações predominantes 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 IVS1-6 é a mais encontrada (BONINI-DOMINGOS, 2004). A talassemia beta apresenta heterogeneidade molecular e expressão fenotípica variável. Na forma homozigota, 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 a uma anemia grave (BANK, 2005). A forma heterozigota é caracterizada por uma 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 gama, 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 morbi – mortalidade e melhorando a qualidade de vida dos portadores. Em termos cronológicos pode-se considerar que a Triagem Neonatal (TN) na América Latina deu seus primeiros passos em meados 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 começou 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 1976, quando Benjamin Schmidt criou um laboratório na APAE-São Paulo para 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 mobilização dos setores de saúde na detecção e prevenção dos indivíduos com anemias hereditárias. Estudos preventivos de hemoglobinopatias têm sido realizados em países nos quais a alta incidência constitui um fator de risco para a saúde pública, cada qual enfocando da melhor maneira a sua população e o tipo de hemoglobinopatia que a acomete (W.H.O., 1982). O objetivo fundamental de uma campanha preventiva para hemoglobinopatias é a identificação precoce de crianças com a forma grave da doença (homozigota), 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., 1986; BONINI-DOMINGOS; SIQUEIRA, 2000).

O diagnóstico neonatal de hemoglobinopatias pode ser realizado tanto em sangue de cordão umbilical, quanto em amostras colhidas de punção plantar. Nos programas de triagem em larga escala distribuídos pelo mundo, as amostras são colhidas em papel e nas primeiras semanas de vida (FROST; BELLINGHAM, 1987; HAYASHI ET AL, 1987; EASTMAN, 1999). Para as hemoglobinopatias, a realização de medidas profiláticas e terapêuticas, como o uso de penicilina, vacinação e educação sobre o tratamento da doença, reduz a morbidade e mortalidade devido à septicemia, causada principalmente pelo *Streptococcus pneumoniae*, e os quadros de seqüestro esplênico nas doenças falciformes (WEATHERALL; CLEGG, 2001; BOY, 2002; DAUDT et al., 2002).

Os neonatos diagnosticados como possíveis portadores de hemoglobinopatias devem ser reavaliados laboratorialmente por volta dos três meses de idade, sempre acompanhados da investigação laboratorial dos genitores. O tempo recomendado entre a detecção da anemia falciforme e 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, haverá a inversão na produção das cadeias gama e beta, podendo ser observados os valores definitivos do indivíduo adulto. As frações de Hb anômalas, que aparecem em pequena quantidade no período da triagem, podem ser melhores 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 2001 o Ministério da Saúde instituiu a Portaria GM/MS nº 822 no âmbito do Sistema Único de Saúde (SUS), criando o Programa Nacional de Triagem Neonatal. Devido à diversidade de doenças existentes foram inseridas nesse programa alterações que podem ser detectadas por meio de testes

seguros, confiáveis e com rapidez como é o caso da fenilcetonúria (PKU), hipotireoidismo congênito (HC), anemia falciforme e outras hemoglobinopatias e a fibrose cística (FC). Esse programa permitiu normatizar, definir critérios de inclusão das doenças a serem triadas e cumprir os princípios de equidade, universalidade e integridade da triagem neonatal no país.

As metas de implantação do PNTN foram cobrir 100% dos nascidos vivos no país, definir e ampliar a gama de patologias triadas, implantar uma Rede Nacional de Triagem Neonatal com política baseada em Serviços de Referência em 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 PKU e HC; na fase II, triagem de PKU, HC e a inclusão da anemia falciforme e outras hemoglobinopatias e na fase III os SRTN capacitados à triagem das patologias anteriores com inclusão da FC. Atualmente estão habilitados no PNTN 25 estados com 33 SRTN credenciados. A cobertura nacional para o ano de 2005 é de 75,3% dos nascidos vivos, sendo 16,7% triados apenas na fase I, 58,6% triados na fase II e 19,5% dos nascidos vivos triados também na fase III. A Figura 4 apresenta um mapa ilustrando a cobertura estadual dos nascidos vivos e a fase em que se encontram os SRTN dos estados (CARVALHO, 2005).

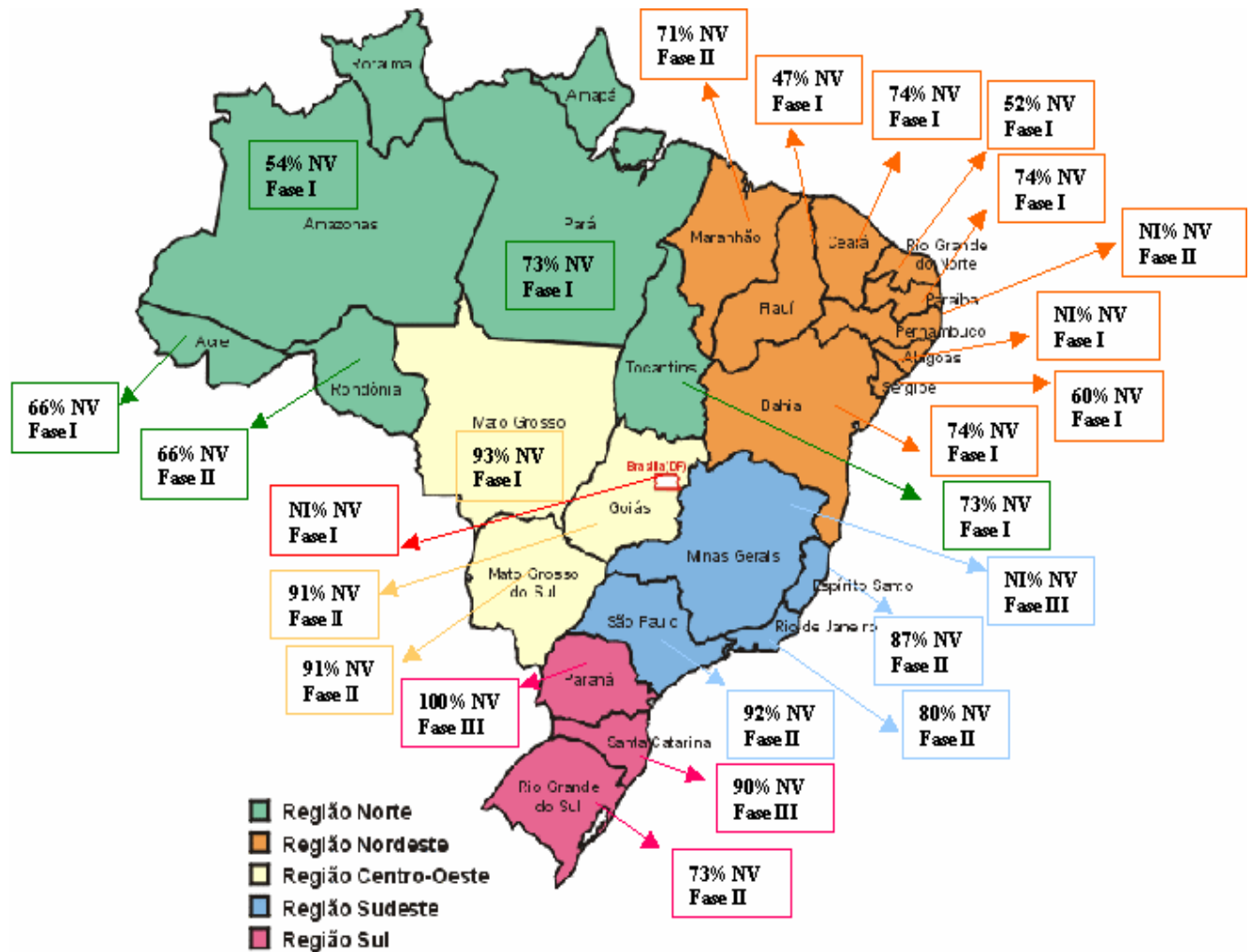


Figura 4: Mapa representativo da cobertura estadual dos nascidos vivos e da fase de credenciamento dos SRTN. NV = nascidos vivos. NI = não informado. Fonte: SIA/SUS, com adaptações.

1.5 Bases de dados Eletrônicas

Com o advento do computador, a necessidade de armazenamento de informações para posterior análise tornou-se fácil e imprescindível. 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 representam as necessidades de um ambiente de trabalho (mini-mundo real). Tal conjunto é estruturado em arquivos de acordo com o ambiente utilizado em seus diferentes desenvolvimentos (KORTH, 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, esta sendo transformada por um crescimento explosivo de dados emergindo dos laboratórios ao redor do mundo. E esta mudança consiste em transformar dados em 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 mesmo conjunto de informação. Por exemplo, uma gravação associada com uma base de dados de uma seqüência de nucleotídeos tipicamente contém informações tais como, nome de contato, a descrição do tipo de molécula, o nome científico do organismo do qual foi isolada e freqüentemente, bibliografia associada com a seqüência. Para a pesquisa em 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 uma questão biológica específica (*).

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 em bases de dados, denominado KDD, do inglês, *Knowledge Discovery in Databases*, ou ainda, Descoberta do Conhecimento em Bases de Dados. KDD é o processo de identificar padrões ou modelos que representem informação válida, inédita, potencialmente útil e essencialmente compreensível em uma coleção de dados (FAYYAD, 1997). Em geral, técnicas de mineração de dados ou *data mining* possuem um papel preponderante nesse processo.

Data mining pode ser definido como um conjunto de técnicas automáticas para a exploração em 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 dificilmente desvendados 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 conseqüente apresentação para o usuário (BARIONE et al., 2001).

(*) Disponível em <<http://ncbi.nlm.nih.gov/About/primer/bioinformatics.html>>. Acesso em 25 de outubro de 2005.

1.5.1.1 Histórico tecnológico

O termo *data mining* é relativamente novo, datando dos anos noventa, porém sua trajetória possui algumas décadas. O princípio da trajetória evolutiva da mineração de dados começou em meados dos 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 se obter 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 em *data mining*, podendo-se dizer que, até os dias atuais, as ferramentas e técnicas *data mining* possuem em seu núcleo, métodos estatísticos (**).

Uma outra vertente de métodos pertencentes à família do *data mining* é 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 demandava 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, porém como demandava uma infra-estrutura pesada, ou seja, computadores com grande capacidade de processamento, tinha sua difusão comercial minimizada (*).

(*) Disponível em: <http://dc.uel.br/~vhmanfredini/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 uma adaptação do aprendizado de máquinas, porém com uma abordagem comercial e cada vez mais tem sido aceito no mundo dos negócios e no científico também (*).

Com o tempo, o *data mining* mostrou algumas deficiências em 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 (**).

1.5.2 Visual Data Mining (VDM)

A visualização gráfica consiste em apresentar graficamente dados nos quais se deseja analisar. É centrada em mecanismos que possibilitam ao usuário a compreensão rápida da informação a ele apresentada (CHITTARO, 2000). O volume de dados a ser analisado é imenso, e na maioria das vezes não se sabe por onde começar o processo de extração da informação. Ou seja, todo o esforço e custo em coletar e armazenar dados pode ser infrutífero se não se souber recuperar dados pertinentes que respondam às consultas desejadas pelos usuários (BARIONE, et al., 2001).

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

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

Devido à natureza das informações armazenadas nestes grandes volumes de dados serem textuais e numéricas, especialmente com muitos atributos, a interpretação destes é realizada com baixa eficiência por seres humanos. Apresentá-los de forma gráfica permite uma melhor percepção dos dados e quais informações podem ser exploradas e extraídas a partir das visualizações criadas (CHITTARO, 2000).

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 um primeiro método para descobrir valores (OLIVEIRA, 2000). 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 (KEIM, 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 mesmos (TRAINA, 2001). 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 (KEIM, 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 em milhões de itens de dados; os itens de dados podem ser arranjados para expressar algum relacionamento entre si (TRAINA, 2001).

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 em grandes bases de dados (KEIM, 1996). Há basicamente 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 exibição dos mesmos, e por outro lado, na capacidade do ser humano em perceber padrões, exceções, tendências e relacionamentos ao analisar a representação visual dos dados de uma maneira gráfica (KEIM, 1996).

As técnicas de VDM atualmente atacam três tópicos, segundo Traina, 2001, 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, 2001).

1.6 Justificativa

Com a implantação do Programa Nacional de Triagem Neonatal através da Portaria do Ministério da Saúde nº 822 as hemoglobinopatias 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 determinada alteração 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 mapeamento 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 1997 a 2002 foi desenvolvido um programa de triagem neonatal de hemoglobinopatias pelo Laboratório de Hemoglobinas e Genética das Doenças Hematológicas – 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 5976 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 em banco de dados, permitem a análise por métodos computacionais na tentativa de estabelecer comportamentos, 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.

2 Objetivos

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, em sistema de visualização 3D - *FastMapDB*, para análise de comportamento dos fenótipos.

3 Material e Métodos

3 Material e Métodos

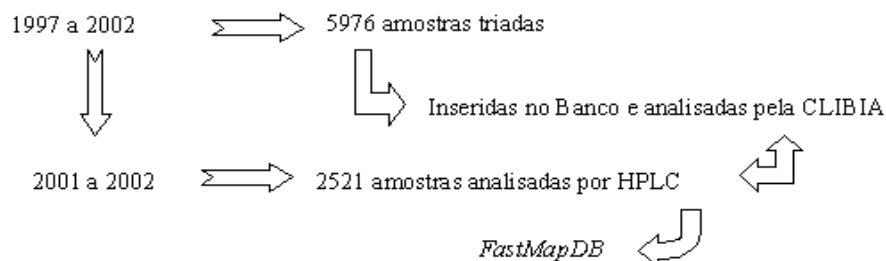
3.1 Casuística

Foram analisados os resultados laboratoriais de 5976 amostras de sangue de cordão umbilical de recém-nascidos no Hospital de Base de São José do Rio Preto no período de 1997 a 2002, de ambos os sexos e distintas origens raciais. As amostras de sangue de cordão umbilical foram colhidas após o parto, em tubos com anticoagulante, identificadas como RN de (nome da mãe), por gravidade e não ordenha do cordão, e acondicionadas em geladeira até o momento das análises, que não excederam uma semana. A coleta das amostras foi realizada por profissionais da enfermagem do Hospital de Base, em plantão 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 àquela amostra somente do conhecimento do responsável pela pesquisa.

Foram 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 ambulatório 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 tem aprovação no CONEP sob o número 3919, processo 25000.034091/2002-88.

Os resultados das 5976 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 2521 amostras triadas nos anos de 2001 e 2002, que incluíram os

resultados quantitativos obtidos por Cromatografia Líquida de Alta Performance (HPLC). Abaixo está representado um organograma 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 a procedimentos eletroforéticos e testes bioquímicos as células foram lisadas para a obtenção da solução de Hb utilizando duas metodologias:

- Hemolisado Rápido - com saponina, (NAOUM, 1990)
- Solução de Hb - com clorofórmio, (NAOUM, 1990)

Hemolisado Rápido: com saponina

Reativo hemolisante:

- Saponina P.A.	1 g
- Água destilada	100 mL

Procedimento:

- Em placa de Kline foi colocado 1 volume de sangue com 1 volume de reativo hemolisante, com posterior homogeneização até a hemólise completa da mistura;

- O hemolisado pôde 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 1mL 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 20 minutos.

- A solução de Hb sobrenadante, 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 10 a 15 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 não é 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 10% - pH 7,4

- NaCl	9,0 g
- Na ₂ HPO ₄	1,36 g
- NaH ₂ PO ₄ .H ₂ O	0,28 g
- Água destilada q.s.p.	100 mL

Solução de trabalho

- NaCl 10%	36 mL
- Água destilada q.s.p.	1000mL

Procedimento:

Em tubo de hemólise colocou-se 2,0 mL de solução de NaCl a 0,36% e 10 µL de sangue total, agitando por inversão, suavemente. A leitura foi feita após 10 minutos.

Interpretação:

O tubo de hemólise com a amostra na solução de NaCl a 0,36% 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, uma vez a resistência aumentada à hemólise do eritrócito tornou a amostra opaca. As linhas foram facilmente visualizadas em amostras com resistência normal à hemólise.

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

Os esfregaços sangüíneos, a fresco, foram analisados ao microscópio óptico, quanto ao tamanho, forma e quantidade de Hb nos 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 Hb normais e grande parte das Hb anormais com mobilidades eletroforéticas diferentes das Hb normais.

Reagentes:

Tampão Tris-EDTA-Borato (TEB), pH 8,6

- Tris hidroximetil aminometano	10,2 g
- Ácido etilenodiaminotetracético	0,6 g
- Ácido Bórico	3,2 g
- Água destilada q.s.p.	1000 mL

Conservado em geladeira

Corante:

Ponceau

- Ponceau S	0,5 g
- Ácido tricloroacético	5,0 g
- Água destilada q.s.p.	100 mL

Solução descorante:

- Ácido acético glacial	100 mL
- Metanol	50 mL
- Água destilada q.s.p.	1000 mL

Procedimento:

- As fitas de acetato de celulose foram embebidas em tampão TEB por 15 minutos no mínimo e 6 horas, no máximo.

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

embeber as fitas, conectando-as com os compartimentos eletrolíticos através de tecido absorvente (*perfex*).

- A solução de Hb foi aplicada a 1,0 cm da extremidade da fita em contato com o pólo negativo.

- As amostras foram submetidas a 300 volts por 30 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 30 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 mL

Solução de Transparentização:

- Ácido acético glacial 7,0 mL

- Metanol 42,5 mL

- Glicerina 0,5 mL

Procedimento:

- A fita de acetato de celulose foi mergulhada no metanol por no máximo 60 segundos.

- Em seguida, ela foi transferida para a solução de ácido acético:metanol:glicerina por 40 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 Hb H e Hb Bart's que apresentam perfil de migração em pH alcalino similar a proteínas plasmáticas.

Reagentes:

Tampão pH neutro

- KH_2PO_4	3,11 g
- Na_2HPO_4	1,66 g
- Água destilada q.s.p.	1000 mL

Conservar em geladeira

Procedimento:

- As fitas de acetato de celulose foram embebidas em tampão neutro por 15 minutos no mínimo e 6 horas, no máximo.

- Após serem secas em folhas de papel absorvente, as fitas foram colocadas na cuba de eletroforese contendo o mesmo tampão utilizado para embeber as fitas, conectando-as com os compartimentos eletrolíticos através de tecido absorvente (*perfex*).

- A solução de Hb foi aplicada a 1,0 cm da extremidade da fita em contato com o pólo negativo.

- As amostras foram submetidas a 300 Volts por 30 minutos.

- As frações foram analisadas sem coloração, seguindo mapa específico de identificação.

3.2.7 Eletroforese em pH ácido (VELLA, 1968)

Princípio:

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

Reagentes:

Tampão Fosfato pH 6,2

Para uso nos compartimentos eletrolíticos e confecção do gel:

- Na ₂ HPO ₄	2,02 g
- NaH ₂ PO ₄ .H ₂ O	7,66 g
- Água destilada q.s.p.	1000 mL

Conservar em geladeira

Gel de Ágar-Fosfato

- Ágar-agar	500 mg
- Tampão fosfato pH 6,2	25 mL

Procedimento:

- Os componentes do gel de ágar-fosfato foram adicionados a um erlenmeyer de 250 mL e levados ao forno microondas até completa dissolução, tomando cuidado para a mistura não ferver.

- Foram pipetados 5,0 mL do gel em lâminas de microscópio que gelificaram à temperatura ambiente.

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

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

- As amostras foram submetidas a 100 volts por 30 minutos.

- As frações foram analisadas sem 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 tetrâmero da Hb. Após coloração esses corpúsculos

apresentam-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 mL

Solução citrato:

- Citrato de sódio	2,2 g
- Água destilada q.s.p.	100 mL

Solução de Azul Cresil Brilhante (ACB):

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

Procedimento:

- Em tubo de ensaio pequeno foram colocados 50 µl de sangue total e adicionados 100 µl de ACB.

- Após agitação suave dos tubos, o material foi incubado a 37°C por 30 e 60 minutos.

- Foram feitos esfregaços finos os quais foram examinados ao microscópio em objetiva de imersão.

Interpretação:

Os corpúsculos de Hb H 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 membrana 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 em um sistema fechado, no qual duas bombas de êmbolo duplo e uma mistura de tampões de diluição, com controles de gradientes pré-programados passam pela coluna detectando as alterações de absorbância a 415 nm. O filtro secundário de 690 nm corrige a linha de base para efeitos provocados pela mistura de tampões com forças iônicas diferentes (EASTMAN et al., 1996). 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 embutido 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 ápice 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 um tubo, foram misturados 5 µL de sangue total com 1,0 mL de água Milique.

Após a hemólise 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 emitidos em modelo próprio que incluiu valores numéricos e perfil cromatográfico. Este Kit identifica

as seguintes frações hemoglobínicas: “Fast”, “F1”, “F” e “A”, sendo as duas primeiras, subfrações de eluição rápida e as duas últimas as Hb normais encontradas em recém-nascidos.

3.3 Ferramentas *Data Mining*

Para analisar os dados por VDM, foi utilizada uma técnica de visualização em três dimensões denominada *FastMapDB*. O *FastMapDB* é uma ferramenta de visualização de dados, como datas, números e textos, onde a apresentação é efetuada através do mapeamento multidimensional dos dados para um espaço tridimensional, promovendo a redução de dimensionalidade. O objetivo da redução de dimensionalidade é apresentar o conjunto de dados com um número menor de atributos, porém preservando 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 minimizar possíveis distorções do mapeamento. Outra característica é 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 em bases relacionais utilizando resultados e algoritmos originalmente desenvolvidos para a redução de dimensionalidade ou de mapeamento de dados métricos para espaços dimensionais. A ferramenta possui recursos para que o usuário crie

interativamente uma função de distância vetorial a partir de qualquer quantidade de atributos de uma tabela em uma 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 mapeia o conjunto de dados para uma representação tridimensional, sendo que cada objeto na tabela é 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.gbdi.icmc.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 2521 amostras de recém-nascidos nos anos de 2001 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 Hb em picos determinados, “Fast”, “F1”, “F” e “A”, estes foram inseridos em planilhas e convertidos para uma base de dados reconhecível pela ferramenta, sendo utilizados como atributos para o mapeamento espacial e, os diferentes fenótipos foram utilizados como atributos classificatórios das amostras. Para estabelecer um padrão de comportamento espacial os resultados foram classificados em grupos segundo delineamento ilustrado no quadro 1.

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

n	Fenótipos de Hb	Grupos de análise para o <i>FastMap</i>
2169	Perfil de Hb normal para idade	“Normal”
55	Hb S em heterozigose	“Variante”
18	Hb C em heterozigose	
21	Beta talassemia heterozigota	“Talassemia”
258	Alfa talassemia heterozigota	

Os resultados laboratoriais das 5976 amostras foram cadastrados no Banco de dados do LHGDH, desenvolvido em colaboração com o 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.locaweb.com.br/banco>. Após o cadastro, as amostras foram analisadas utilizando a ferramenta CLIBIA, especialmente desenvolvida para esse banco de dados. A CLIBIA é uma 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 Hb 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-complementada por Tukey, para comparar três ou mais grupos e Teste t para dois grupos. Quando os dados foram não paramétricos, mesmo após transformação dos valores, foram utilizados os testes não paramétricos de Kruskal-Wallis complementada por Dunn para comparar três ou mais grupos e Mann-Whitney para dois grupos (ZAR, 1999).

4 Resultados

4 Resultados

4.1 O banco de dados e a ferramenta CLIBIA

Os resultados laboratoriais de 5976 amostras analisadas no período de 1997 a 2002 foram inseridos no banco de dados eletrônico, LHGDH *on-line*. Por conterem 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 como 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 7).



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

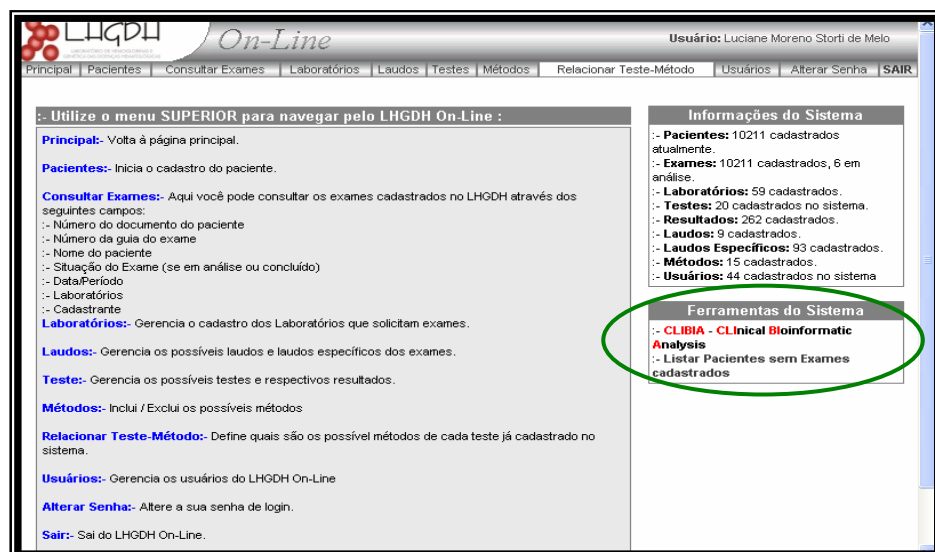


Figura 6: Página inicial do banco de dados. O círculo em verde destaca o *link* para a ferramenta CLIBIA.

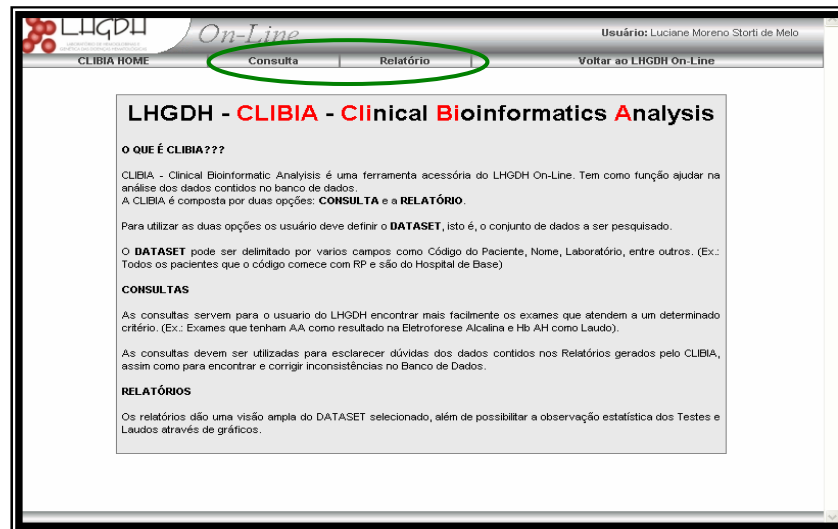


Figura 7: Página inicial da ferramenta CLIBIA. O círculo em verde destaca os links para a seleção das opções consulta ou relatório.

Através do recurso de consulta específica foi possível comparar resultados obtidos em diferentes testes laboratoriais e acessar cada exame individualmente permitindo a observação de todas as informações referentes aquele exame. A Figura 8 ilustra a pagina de seleção do *dataset* para consulta. Nesta página, foi possível buscar as amostras, selecionando-se aqueles exames 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 migrou para uma tela com uma tabela de resultados. Nessa tabela foi possível abrir cada exame 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 em que o nome do paciente tivesse as iniciais RN, com zero anos, e que fossem provenientes do Hospital de Base de São José do Rio Preto.

The screenshot shows the LHGDPH On-Line CLIBIA interface. At the top, there is a header with the logo and the text 'On-Line'. Below the header, there is a navigation bar with buttons for 'CLIBIA HOME', 'Consulta', 'Relatório', and 'Voltar ao LHGDPH On-Line'. The main content area displays the title 'LHGDPH - CLIBIA - CONSULTA : RESULTADOS'. Below the title, there is a table titled 'Items selecionados pelo usuário' with the following data:

Campos	Selecionado pelo usuário
Nome do paciente	RN
Idade entre	0 e 0 anos
Date	1/1/1990 until 9/5/2005
Laboratório	Hospital de Base S. J. do Rio Preto

Below the filter table, there is a summary row: 'TOTAL : 5975'. The main results table has the following columns: N.º, Doc, Nome, Cidade, UF, Sexo, Laboratório, Guia, Inclusão, Laudo, Laudo Específico, and Cadastrante. The first three rows of data are highlighted in red:

N.º	Doc	Nome	Cidade	UF	Sexo	Laboratório	Guia	Inclusão	Laudo	Laudo Específico	Cadastrante
1	116315	RN- [redacted]	São José do Rio Preto	SP	F	Hospital de Base S. J. do Rio Preto	408	16/7/2002	Normal para a idade	FA	[redacted]
2	1163269	RN- [redacted]	São José do Rio Preto	SP	M	Hospital de Base S. J. do Rio Preto	409	16/7/2002	Normal para a idade	FA	[redacted]
3	1163070	RN- [redacted]	São José do Rio Preto	SP	F	Hospital de Base S. J. do Rio Preto	410	16/7/2002	Normal para a idade	FA	[redacted]

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

Através da opção 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 opção relatório da ferramenta CLIBIA, depois de delimitado o *dataset*, foi possível selecionar as configurações do relatório e incluir gráficos de testes, laudos, sexo e relacionamentos. A Figura 10 ilustra a página de seleção do *dataset* na opção relatório.

CLIBIA HOME Consulta Relatório Voltar ao LHGDH On-Line

Usuário: Luciane Moreno Storti de Melo

LHGDH - CLIBIA - RELATÓRIO

Selecionar DATASET

Doc. Paciente: []

Nome do Paciente: RN

Idade entre: 0 e 0

Período de: 01/01/1990 até 01/2006

Laboratórios: Hospital de Base S. J. do Rio Preto/São José do Rio Preto-SP

Cadastrado Por: Todos os usuários...

Configurações

<input checked="" type="checkbox"/> Mostrar relações entre teste, resultado, método e sexo	<input checked="" type="checkbox"/> Mostrar relações laudo, laudo específico e sexo
<input checked="" type="checkbox"/> Mostrar gráfico dos testes	<input checked="" type="checkbox"/> Mostrar gráfico dos resultados
<input checked="" type="checkbox"/> Mostrar gráfico dos laudos	<input checked="" type="checkbox"/> Mostrar gráfico dos laudos específicos
<input checked="" type="checkbox"/> Mostrar gráfico dos sexos	<input type="checkbox"/> Mostrar gráfico das etnias
<input type="checkbox"/> Mostrar gráfico das faixas etárias	<input type="checkbox"/> Mostrar gráfico das regiões

Gerar Relatório

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

Os relatórios gerais contêm uma 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 também foi realizada através de gráficos, coloridos e com legenda que agilizam o processo de análise dos dados. O relatório geral mostrou os seguintes perfis hemoglobínicos para as amostras de neonatos: 81,76% com perfil normal para a idade, 9,91% com resultados sugestivos para alfa talassemia, 1,87% com resultados sugestivos de beta talassemia, 3,38% com hemoglobinas variantes, 0,82% com interações entre hemoglobinas variantes e talassemias e 2,26% amostras com presença de hemoglobinas normais, mas que sugeriram contaminação com o sangue materno. A Figura 11 apresenta o gráfico da distribuição dos laudos dos exames.

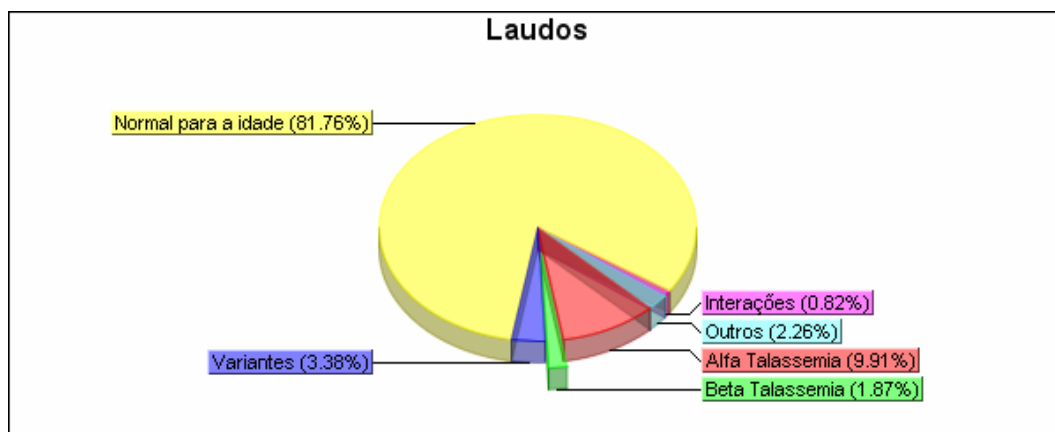


Figura 11: Gráfico obtido pela opção relatório da ferramenta CLIBIA incluindo os laudos dos exames e as frequências fenotípicas das Hb observadas.

O relatório geral completo obtido 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 estabelecer o perfil do fenótipo normal para amostras de sangue de neonatos, foram selecionados os resultados das 2169 amostras com resultados normais para idade, obtidos por HPLC como conjunto amostral. O mapeamento foi realizado utilizando os quatro picos discriminados pelo HPLC, previamente descritos na metodologia.. A visualização tridimensional obtida está ilustrada na Figura 12, e as amostras normais distribuíram-se em dois aglomerados distintos, marcados pelos traçados em 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 mapeamento selecionando os diferentes picos como atributos. A imagem projetada incluindo os picos “F1”, “F” e “A” no mapeamento exibiu uma aproximação dos aglomerados, no entanto, ainda pôde ser observada a presença de dois grupos distintos, como ilustrado na Figura 13.

A Figura 14 representa o mapeamento utilizando os picos “Fast”, “F” e “A”, excluindo-se o pico “F1”, 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 imagem utilizando apenas os picos de “F” e “A”, excluindo as janelas de “Fast” e “F1”. Como observado na Figura 15, nesse mapeamento obteve-se a dispersão linear dos dados.

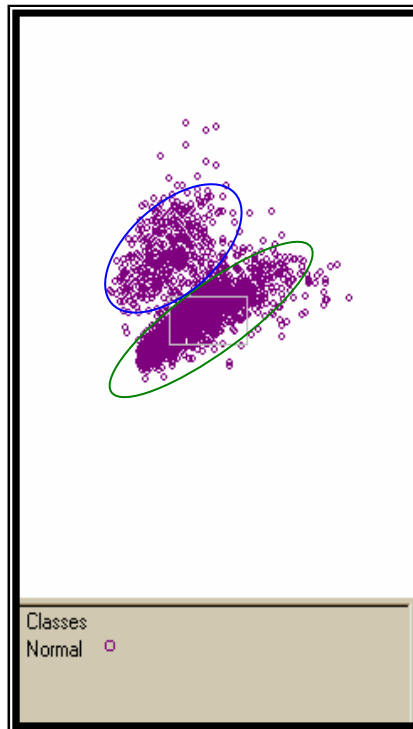


Figura 12: Projeção da imagem tridimensional para o grupo “Normal”, incluindo os quatro picos como atributos de mapeamento.

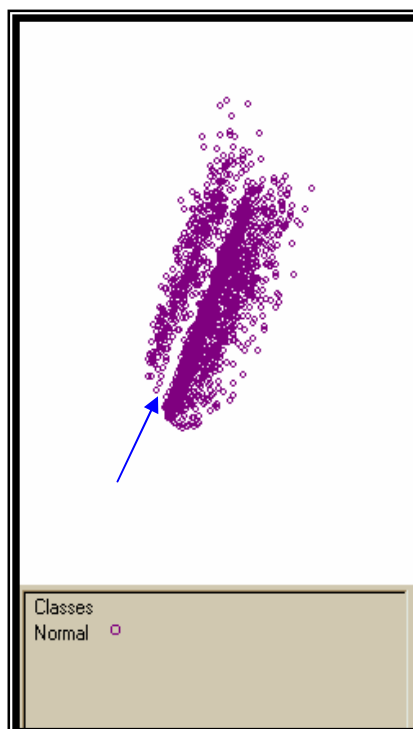


Figura 13: Projeção da imagem tridimensional para o grupo “Normal” incluindo os picos “F1”, “F” e “A” como atributos de mapeamento. A seta indica o espaço entre os aglomerados.

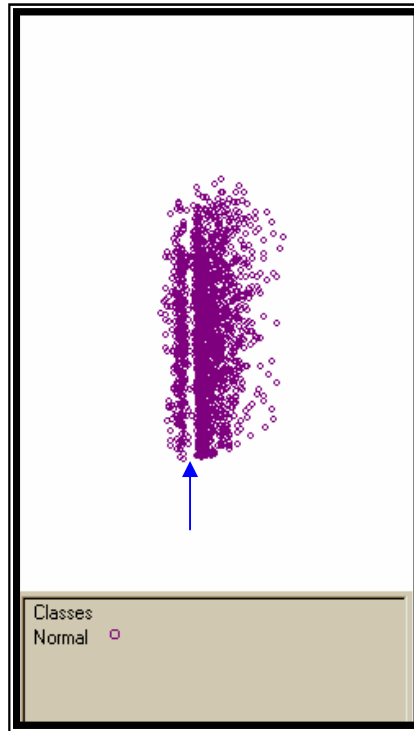


Figura 14: Projeção da imagem tridimensional para o grupo “Normal” incluindo os picos “Fast”, “F” e “A” como atributos de mapeamento.

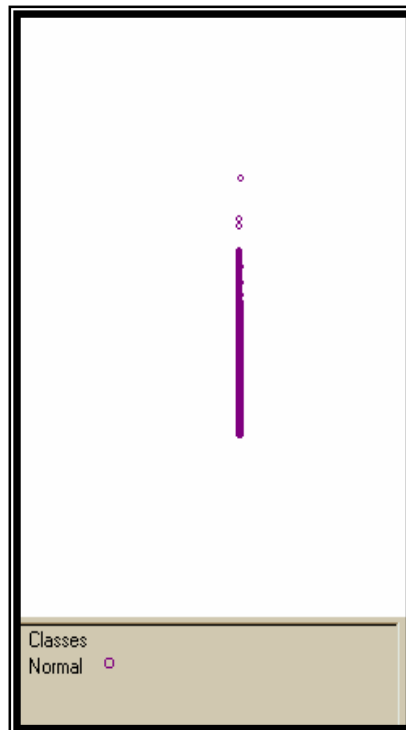


Figura 15: Projeção da imagem tridimensional para o grupo “Normal” incluindo os picos “F” e “A” como atributos de mapeamento.

Para avaliar os valores de “Fast” e ‘F1”, foram geradas imagens utilizando um recurso da ferramenta *FastMapDB*, no qual um atributo pode ser utilizado para dividir o conjunto amostral em dois subgrupos exatamente no valor médio entre os valores mínimo e máximo. A imagem ilustrada na Figura 16 contém os quatro picos de análise e os valores de “Fast” foram utilizados para classificar as amostras em dois subgrupos representados na imagem pela diferença no tamanho dos pontos. Como observado, os pontos marcados pelo círculo azul apresentam valores de “Fast” entre 0,0 e 5,5 % e os pontos dentro do círculo verde correspondem à valores de “Fast” de 5,5 a 10,9%. 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 2040 amostras. O subgrupo variando de 5,5 a 10,9 % foi composto por apenas 129 amostras e a média observada foi 7,35 %. A média percentual obtida no pico de “Fast” para todas as amostras normais foi de 2,64%, 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 “F1” para dividir as amostras no valor médio e a imagem obtida está ilustrada na Figura 17, na qual a distribuição espacial dos dados foi muito semelhante à da Figura 12. As duas nuvens classificadas pelos valores, observadas na Figura 17 não apresentaram sobreposição dos pontos. Para possibilitar a comparação das imagens os mapeamentos das Figuras 12 e 17 foram dispostos lado a lado na Figura 18. Para caracterizar a divisão em duas nuvens amostrais, calculou-se a média percentual para cada subgrupo obtido para o pico de F1. Obteve-se para o subgrupo com valores variando de 0,0 a 10,3% média percentual de 9,45% incluindo 1369 amostras e para o subgrupo com valores variando de 10,3 a 20,6% obteve-se a média de 12,25%, sendo composto por 800 amostras. A média percentual obtida para todas as amostras normais foi de 10,48%, apresentando diferença estatisticamente significativa ($p < 0,05$) das médias obtidas para os dois subgrupos de “F1”.

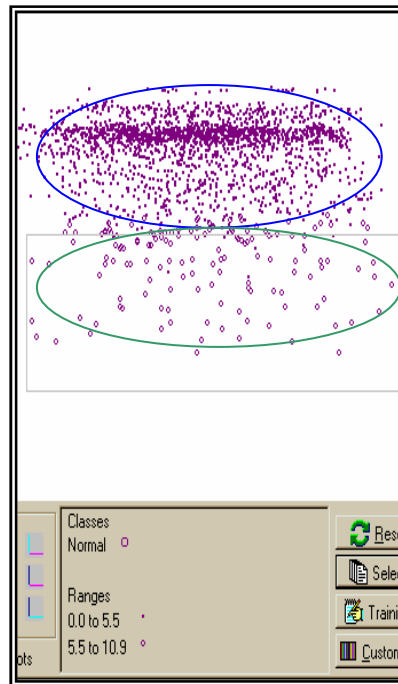


Figura 16: Projeção da imagem tridimensional para o grupo “Normal”, incluindo os quatro atributos de mapeamento e utilizando o pico de “Fast” como classificatório em dois subgrupos, sendo o primeiro variando de 0,0 a 5,5 % e o segundo de 5,5% a 10,9%.

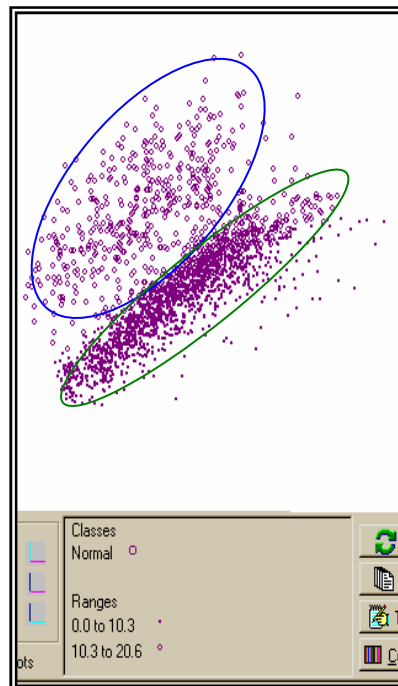


Figura 17: Projeção da imagem tridimensional para o grupo “Normal”, incluindo os quatro atributos de mapeamento e utilizando o pico “F1” como classificatório em dois subgrupos, sendo o primeiro variando de 0,0 a 10,3 % e o segundo de 10,3% a 20,6%.

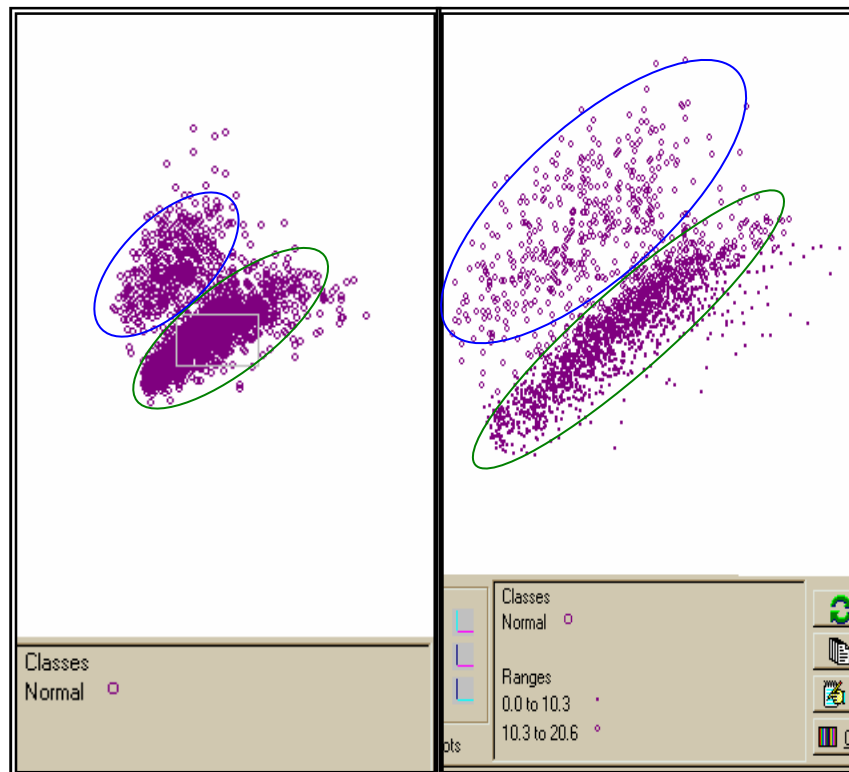


Figura 18: Comparação das imagens projetadas para o grupo “Normal”. A - incluindo os quatro picos como atributos de mapeamento B - utilizando o pico de “F1” dividido em dois subgrupos variando de 0,0 a 10,3% e de 10,3 a 20,6%.

4.2.2 Perfis obtidos para os diferentes fenótipos de Hb

Visando estabelecer um padrão para os diferentes fenótipos observados, foram mapeadas todas as amostras, divididas nos três grupos, “Normal”, “Talassemia” e “Variante”, incluindo os quatro picos como atributos para o mapeamento. Na imagem gerada observaram-se os grupos de análise distribuídos em três nuvens amostrais distintas, como ilustrado na Figura 19, não houve sobreposição entre os grupos. As nuvens correspondentes aos grupos “Talassemia” e “Normal” encontram-se diametralmente opostas.

No mapeamento incluindo as amostras dos grupos “Normal” e “Talassemia” utilizando os quatro atributos observou-se a diferenciação em aglomerados distintos para cada grupo, sem sobreposição espacial como pode ser observado na Figura 20.

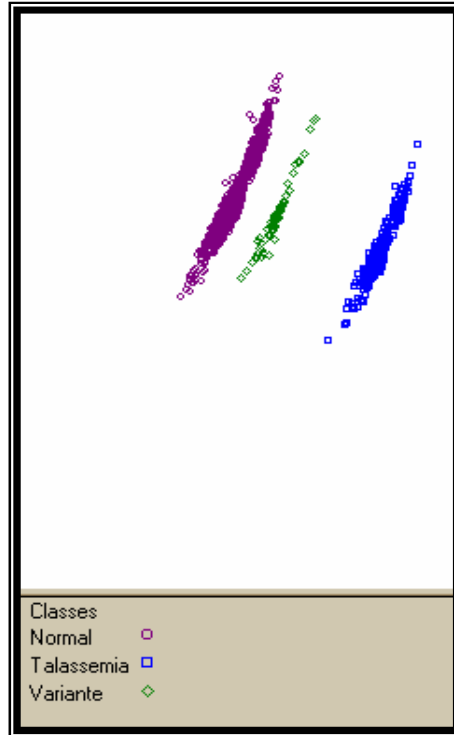


Figura 19: Projeção da imagem tridimensional para os grupos “Normal”, “Talassemia” e “Variante”, incluindo os quatro atributos de mapeamento.

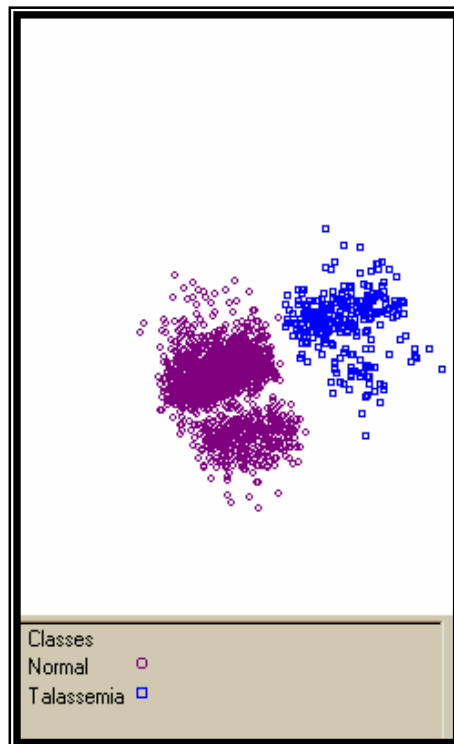


Figura 20: Projeção da imagem tridimensional para os grupos “Normal” e “Talassemia” incluindo os quatro atributos de mapeamento.

Para verificar quais as frações fundamentais na diferenciação de amostras com Hb normais e com alterações talassêmicas, foi gerada uma imagem utilizando apenas os picos “F” e “A” como atributos de mapeamento, por representarem as frações de hemoglobinas em maiores concentrações nesta faixa etária. Na imagem projetada, observou-se a distribuição linear e paralela dos grupos amostrais, como pode ser visualizado na Figura 21.

O grupo “Talassemia” incluiu as amostras com alfa e beta talassemias e apesar de ambas apresentarem alterações nos valores quantitativos de Hb, elas apresentam variações em frações diferentes das Hb, dessa forma realizou-se o mapeamento com esses fenótipos separados e comparados ao grupo “Normal”. A visualização espacial desses grupos incluindo os quatro atributos mostrou sobreposição dos pontos amostrais.

Foram utilizados então, os picos “F” e “A” para o mapeamento e, na visualização obtida observou-se o delineamento dos grupos em três aglomerados representando cada grupo fenotípico, ilustrado na Figura 22. Pôde-se observar que as amostras se distribuíram linearmente no espaço, estando bem definidas em três grupos distintos. A Figura mostrou ainda que, os grupos alfa e beta talassemias estão mais próximos entre si e afastados do grupo normal.

As amostras com alfa talassemia foram projetadas comparadas com as normais e para os mapeamentos realizados com os quatro picos ou mesmos incluindo três picos observou-se sobreposição dos pontos amostrais como visualizado na Figura 23, que incluiu os picos “Fast”, “F” e “A” como atributos de mapeamento. Foram feitos mapeamentos com dois picos incluídos como atributos, e as visualizações obtidas resultaram na concentração dos pontos amostrais em aglomerados distintos para cada fenótipo. A Figura 24 representa a visualização tridimensional do mapeamento feito utilizando os picos “Fast” e “A”, no qual obteve-se a melhor separação amostral.

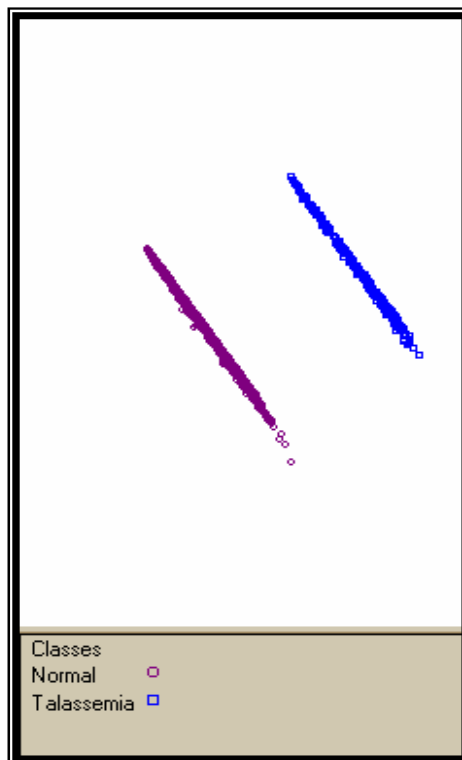


Figura 21: Projeção da imagem tridimensional para os grupos “Talassemia” e “Normal” utilizando como atributos de mapeamento os picos “F” e “A”.

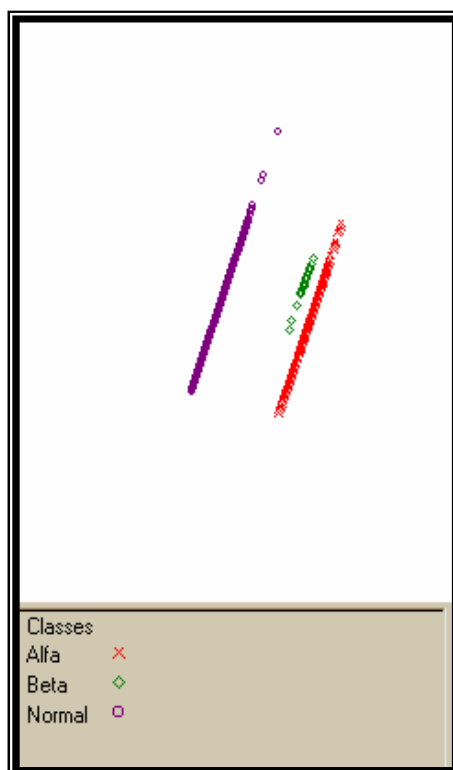


Figura 22: Projeção da imagem tridimensional para os grupos “Normal”, “Alfa” e “Beta” talassemias, utilizando como atributos de mapeamento os picos “F” e “A”.

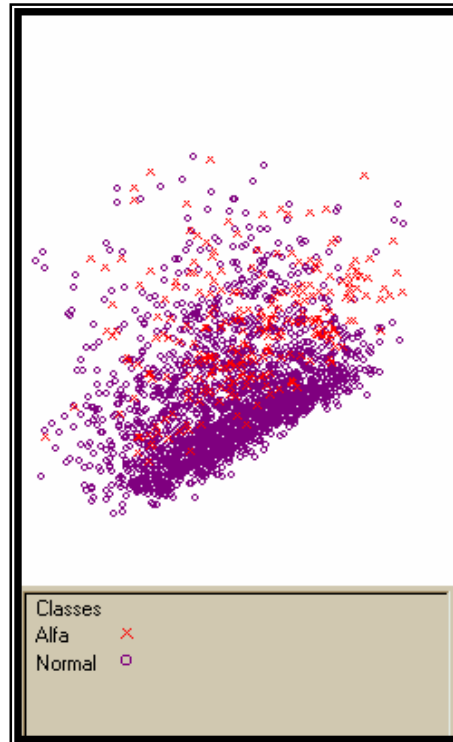


Figura 23: Projeção da imagem tridimensional para os grupos “Normal” e “Alfa” talassemia, utilizando como atributos de mapeamento os picos “Fast”, “F” e “A”.

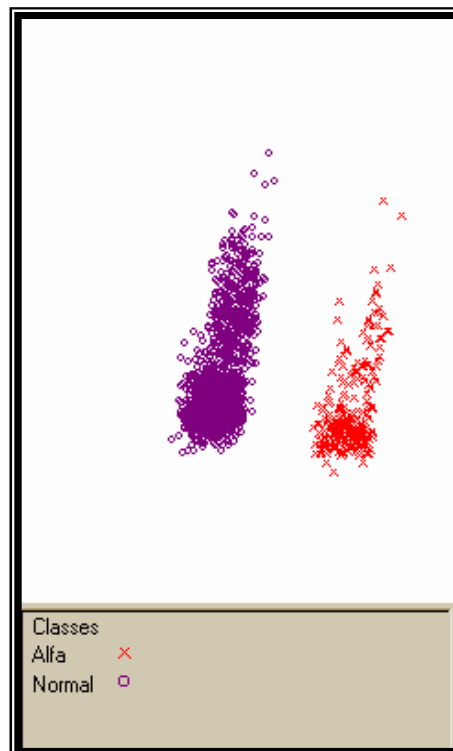


Figura 24: Projeção da imagem tridimensional para os grupos “Normal” e “Alfa” talassemia, utilizando como atributos de mapeamento os picos “Fast” e “A”.

Para as amostras pertencentes ao grupo “Variantes”, observou-se nos mapeamentos 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” é composto pelas amostras contendo Hb AS e Hb AC, para tentar estabelecer um padrão entre essas variantes e diferenciá-las das amostras com Hb normais, realizou-se o mapeamento incluindo o grupo “Normal” com cada uma dessas variantes isoladamente. O padrão de dispersão espacial se manteve, não sendo possível separar as amostras em nuvens amostrais distintas. A Figura 26 representa a visualização tridimensional para o mapeamento dos grupos “Normal”, Hb AS e Hb AC incluindo os quatro atributos de mapeamento. Mesmo analisando as amostras separadamente a área de sobreposição se mantém. As Figuras 27 e 28 ilustram as visualizações espaciais para os mapeamentos das amostras normais com Hb AS e normais com Hb AC respectivamente, incluindo os quatro picos como atributos. Observou-se que mesmo na projeção das amostras de Hb variantes separadamente, a sobreposição espacial se manteve nas visualizações. Deve-se destacar que as frações de Hb variantes são identificadas no HPLC por picos específicos padronizados denominados “S” e “C” que reconhecem o tempo de eluição dessa Hb anormais. Estes picos não foram incluídos como atributos para o mapeamento espacial por estarem ausentes nas amostras com hemoglobinas normais, não sendo possível a comparação.

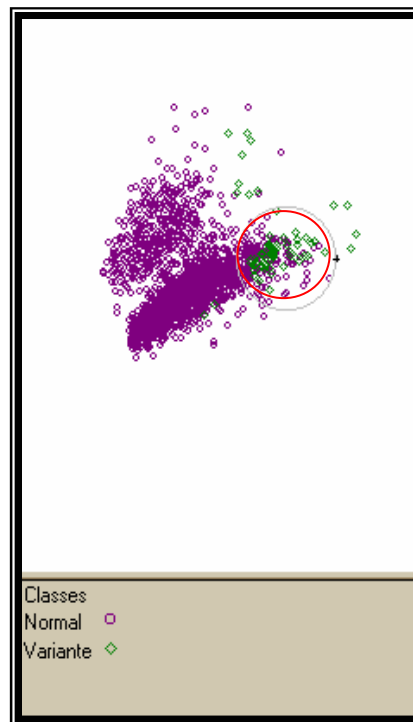


Figura 25: Projeção da imagem tridimensional para os grupos “Normal” e “Variante” utilizando os quatro atributos de mapeamento. Área de sobreposição marcada pelo círculo vermelho.

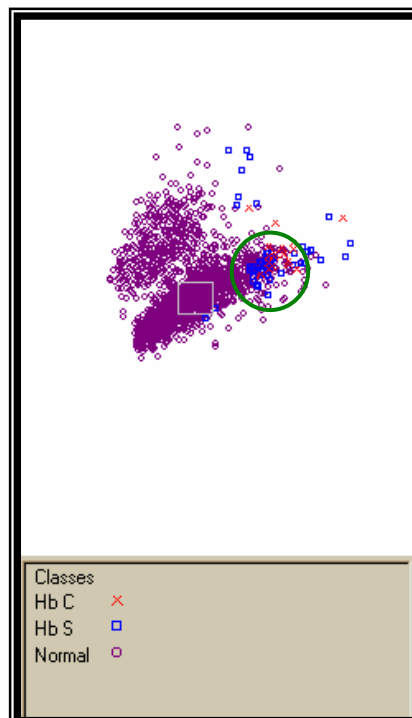


Figura 26: Projeção da imagem tridimensional para os grupos “Normal”, “Hb AS” e “Hb AC” utilizando os quatro atributos de mapeamento. Área de sobreposição marcada pelo círculo verde.

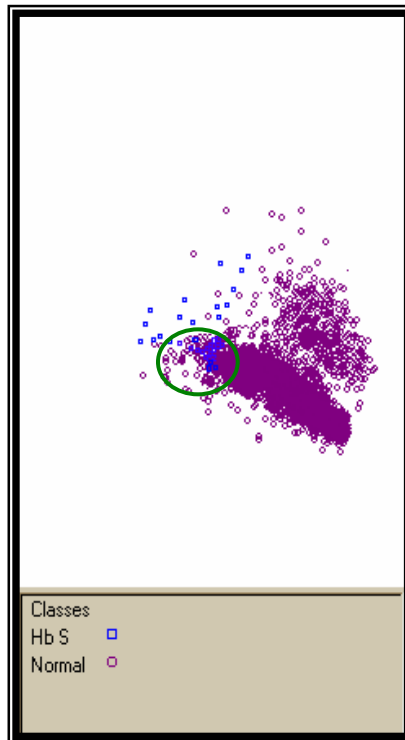


Figura 27: Projeção da imagem tridimensional para os grupos “Normal” e “Hb S” utilizando os quatro atributos de mapeamento. Área de sobreposição marcada pelo círculo verde.

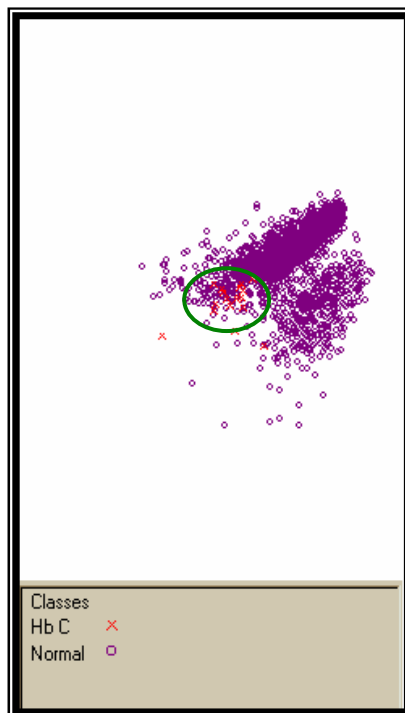


Figura 28: Projeção da imagem tridimensional para os grupos “Normal” e “Hb C” utilizando os quatro atributos de mapeamento. Área de sobreposição marcada pelo círculo verde.

Os valores quantitativos de Hb, obtidos por HPLC, foram submetidos a análise estatística, afim de buscar similaridades ou diferenças que pudessem auxiliar na compreensão dos perfis apresentados pelos diferentes fenótipos nas visualizações tridimensionais. A Tabela 1 contém as médias percentuais dos quatro picos determinados pelo HPLC para todas as alterações de Hb analisadas. As comparações entre as médias fora 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”
Normal	2.169	2,64 ^a	10,48 ^a	73,02 ^a	12,86 ^a
Alfa tal	257	3,46 ^b	10,87 ^b	73,50 ^a	11,02 ^b
Beta tal	21	2,58 ^a	10,86 ^b	80,85 ^b	4,81 ^c
Hb S	55	2,71 ^a	10,13 ^c	72,69 ^a	7,85 ^d
Hb C	18	2,68 ^a	10,55 ^a	76,54 ^c	4,97 ^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 médios dos picos “Fast”, “F1” 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 “F1”, “F” e “A” comparados ao grupo normal. Dessa forma, analisando o grupo de “Talassemias”, observou-se valores médios diferentes estatisticamente para três picos de análise quando comparados às 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, obteve-se $p < 0,05$ para os picos de “F1” e “A” em relação ao grupo normal e, para Hb C, obteve-se diferença significativa nas médias percentuais dos picos “F” e “A” comparados às amostras normais.

5 Discussão

5 Discussão

A utilização do banco de dados para armazenamento dos resultados laboratoriais mostrou-se importante em vários aspectos. A segurança no processo de arquivar as informações, é um ponto a ser destacado. O banco possui acesso restrito, não sendo permitida a todos os usuários a observação dos dados nele 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 aos exames de sua referência.

Além da privacidade das informações, deve-se destacar também as vantagens do arquivo em meio 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 necessitam de grandes arquivos, protegidos de variações de umidade e temperatura e despendem um trabalho laborioso para organizá-los de maneira que seja possível encontrar um dado arquivado no passado com rapidez.

A disponibilização do banco de dados via Internet é mais 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 mais rápido 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 com laudo de alfa talassemia, observou-se que 100% dos exames apresentaram resultado positivo ou sugestivo no teste de Pesquisa intraeritrocitária de Hb H, no entanto, 96% apresentaram a fração de Hb

H ou Bart's em eletroforese pH neutro. Essas observações destacaram a importância da pesquisa de Hb H, e a necessidade de avaliar as amostras para esse teste mesmo quando a eletroforese em pH neutro apresentar resultados negativos. O levantamento manual 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 Globular osmótica revelaram que 24,5% das amostras com laudo normal para idade apresentaram positividade para este teste, enquanto que 21,4% 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 97% dos beta talassêmicos apresentam resistência osmótica aumentada em NaCl a 0,36% (NAOUM, 1997), 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 método em neonatos.

Contabilizar e organizar em tabelas os resultados de análises de aproximadamente 6000 amostras é tarefa demorada 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 é um passo importante, presente na análise de dados da maioria dos trabalhos científicos. Além de despende grande tempo e trabalho, esse processo, de armazenar 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 é cansativo e em geral pouco resolutivo.

A utilização de um banco de dados para armazenar as informações é uma realidade já praticada em alguns laboratórios de rotina em análise. Na área de pesquisa em genética, com os projetos genoma surgiram os grandes bancos de dados implantados para armazenamento de seqüências, sem os quais todo o trabalho de seqüenciar genomas seria improdutivo. Atualmente a possibilidade de

laboratórios de pesquisa, em diversas áreas, utilizarem um banco de dados associado a uma ferramenta *data mining* para processar as informações, abre novos caminhos 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 um sistema computacional pode realizar em segundos. A inserção das informações no banco de dados é uma etapa de responsabilidade do pesquisador, e deve ser momento 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 como Centro de Referência em Diagnóstico de hemoglobinopatias, mas também contribuiu de maneira imensurável o processamento dos dados de pesquisa. Permitiu a observação entre os resultados das diferentes amostras utilizadas em cada projeto, como a comparação entre a eficácia de uma mesma metodologia nas diferentes abordagens e os resultados gerais prontamente utilizados nas publicações e hipóteses de trabalhos subsequentes. Todo o processo de análise de amostras de sangue tornou-se mais dinâmico e menos limitado a um único objetivo, seja este o diagnóstico de uma alteração ou a contribuição científica das análises.

A aplicação de técnicas de VDM é uma abordagem nova nas pesquisas com hemoglobinopatias. Para os dados da triagem neonatal realizada pelo LHGDH as análises pela ferramenta *FastMapDB* apresentaram padrões relevantes. Na projeção das amostras com hemoglobinas normais para a idade observou-se a existência de duas nuvens de distribuição amostral. O mapeamento feito sem a presença dos picos de “Fast” e “F1” mostrou distribuição linear, sugerindo dois padrões de comportamento distintos no grupo com Hb normais. Com o recurso da ferramenta *FastMapDB* os valores do pico de “Fast” foram divididos ao meio separando as amostras em dois subgrupos. Na projeção tridimensional os pontos distribuíram-se em dois grupos sem sobreposição, no entanto, houve um desequilíbrio na distribuição amostral, de forma que o primeiro subgrupo incluiu 94% das amostras, não sendo representativo da distribuição

observada na Figura 12. Esses resultados destacam que o pico de “Fast” não foi o responsável pela formação das duas nuvens amostrais observadas no mapeamento das amostras com hemoglobinas normais para a idade.

No mapeamento feito utilizando a divisão do pico “F1” no valor médio obteve-se dois grupos distribuídos de forma mais igualitária, sendo o primeiro composto por 63,1% das amostras e o segundo incluindo 36,9%. As médias obtidas para esses dois subgrupos apresentaram diferenças estatisticamente significativas ($p < 0,05$), quando comparadas entre si e à média geral obtida para todas as amostras normais. Quando comparadas lado a lado, as visualizações obtidas com os quatro atributos e com o pico de “F1” dividido em dois subgrupos (Figura 18), pôde-se observar semelhança na distribuição dos pontos amostrais. Esses resultados indicam a presença de dois padrões com valores médios diferentes de distribuição para “F1”, devido a grande amplitude de variação observada nos valores deste pico. Deve-se destacar que as frações de Hb degradadas, presentes em amostras acondicionadas por longo tempo, provocam aumento no valor de “F1”. Diante desses achados, deve-se destacar, portanto que a média geral obtida para as 2169 amostras não é valor representativo do pico de “F1” 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 amostrais, não foi possível observar esse comportamento para os valores de “F1” nas amostras de 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 Keim, (2001), o processo de exploração visual dos dados pode ser visto como um processo de geração de hipótese, através das visualizações dos dados, que permitem 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.

A análise global dos resultados, visualizados em projeção espacial ressaltou a existência de um possível padrão obtido por HPLC, para as amostras com talassemias heterozigotas, pois não houve sobreposição espacial dos pontos amostrais, sugerindo diferenças nos perfis quando comparados as amostras 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 beta talassemia. Essa diferença estatística pode explicar a distribuição visual dos pontos amostrais nas imagens geradas para amostras normais e com talassemias, evidenciando a existência de diferenças nesses perfis. Na projeção das amostras normais e com alfa e beta talassemia (Figura 22) utilizando os picos “F” e “A” como atributos de mapeamento observou-se o delineamento dos grupos, estando dispostas em três linhas paralelas, sendo que as amostras com talassemias dispuseram-se próximas entre si e distantes das normais. Esse padrão de distribuição evidencia a potencialidade dessa ferramenta de análise em separar os grupos amostrais segundo suas características.

A projeção da imagem obtida para as amostras normais e com alfa talassemia apresentou sobreposição dos pontos, no entanto, o mapeamento feito utilizando os picos de “Fast” e “A” apresentou distribuição bem definida dos grupos amostrais. Esses picos apresentaram diferenças significativas nos valores médios ($p < 0,05$), como foi ilustrado na Tabela 1. Deve-se destacar também que, as amostras com alfa talassemia, apresentaram aumento estatisticamente significativo ($p < 0,05$) no percentual médio do pico de “Fast” quando comparado aos valores obtidos para esse pico nos demais fenótipos. Para as amostras com beta talassemia houve diferença estatisticamente significativa nos picos de “F”, que apresentou aumento em relação as amostras 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 amostras, pela seleção dos picos para o

mapeamento e pôde-se identificar diferenças posteriormente corroboradas pela análise estatística.

Mendes-Siqueira, (2004), observou aumento estatístico significativo para os picos de “Fast” e “F1” em amostras com alfa talassemia na mesma faixa etária. Para amostras com beta talassemia obteve valores aumentados para o pico de “F” 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 cromatogramas das amostras com talassemias. A HPLC com o sistema Variant e Kit de diagnóstico autorizado pelo Ministério da Saúde no Brasil, não é uma metodologia que permite o diagnóstico das formas talassêmicas nessa faixa etária, com segurança, no entanto, esses resultados sugerem um perfil característico para essas amostras, confirmados por outras metodologias.

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

As Hb S e Hb C são diagnosticadas com precisão pelo sistema de HPLC Variant que identifica com alta sensibilidade as pequenas porcentagens das frações variantes. No entanto o mapeamento espacial das amostras heterozigotas para essas hemoglobinas comparadas ao grupo Normal mostrou sobreposição das nuvens amostrais, não sendo possível diferenciar os grupos. O pico de Hb Fetal, que representa o maior componente hemoglobínico nesta faixa etária, não apresenta diferenças estatisticamente significativas nos valores percentuais entre as amostras normais e com Hb S. Em relação ao pico de Hb A, a diminuição estatisticamente significativa ($p < 0,05$) observada nas amostras com Hb variantes em relação as normais, deve-se ao fato das Hb S e Hb C estarem presentes em heterozigose nas amostras avaliadas. Apesar dessa diferença, para as Hb variantes,

apenas dois picos de análise apresentaram diferença significativa em relação ao grupo Normal, dificultando a separação espacial dos grupos amostrais. 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 amostras normais.

Ondei, (2005), analisando os resultados de HPLC de amostra de sangue de indivíduos adultos por VDM com a ferramenta *FastMapDB*, obteve padrões de distribuição bem definidos para as amostras de Hb S em homozigose (Hb SS) e também para duplos heterozigotos (Hb SC) quando comparadas as amostras normais (Hb AA). Nas imagens obtidas, os grupos Hb SS e Hb SC apresentaram aglomerados em lados opostos da projeção em relação ao aglomerado de amostras normais.

As hemoglobinopatias apresentam diversidade de fenótipos na população brasileira, influenciados por fatores ambientais e genéticos. A observação de características intrínsecas a cada fenótipo 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 amostras, através da separação em nuvens distintas, ressaltando a necessidade de avaliar os resultados cromatográficos em conjunto com outras metodologias.

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. No entanto a análise estatística foi importante para auxiliar na compreensão dos padrões de distribuição. A visualização gráfica permite uma apresentação rápida e global dos dados, e a ferramenta utilizada apresentou-se eficiente na seleção dos parâmetros de análise que pudessem explicar os padrões observados na tela. A utilização desses recursos, ainda é nova no campo das análises com hemoglobinopatias, e a comunidade científica é fortemente dependente dos resultados estatísticos na análise de dados e na inferência de hipótese. Esse trabalho revelou que as análises

estatísticas foram concordantes com os padrões observados nas projeções espaciais. Ondei, (2005), também obteve resultados estatísticos que corroboraram suas visualizações tridimensionais, reforçando a aplicabilidade de grande valor dos métodos de VDM nas análises biológicas. As técnicas *Data mining* e VDM podem contribuir para as análises com hemoglobinopatias, ampliando a exploração dos resultados, corroborados pela análise estatística que auxila na compreensão das imagens.

6 Conclusões

6 Conclusões

1. A utilização de um banco de dados em 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 promovem a mineração dos dados, como a CLIBIA, revelaram a necessidade de reavaliar as metodologias de rotina utilizadas no diagnóstico de hemoglobinopatias em 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 Hb normais, devido a amplitude de variação nos valores obtidos por HPLC para o pico de “F1”.

3. A projeção tridimensional permitiu a separação das com Hb amostras normais e com alfa talassemia ressaltando as diferenças estatísticas nos valores percentuais médios dos picos de “Fast” e “A”. Essa distribuição evidencia a existência de perfis característicos dessa alteração nas análises por HPLC, reforça a necessidade de melhor avaliação dos cromatogramas obtidos e a possibilidade de aplicação dessa metodologia como parte do diagnóstico de alfa talassemia em neonatos.

6.1 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 exatamente 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 amostrais após o mapeamento, 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 um quadrado, círculo, uma mão, etc.

7 Referências

7 Referências

BANK, A. Understanding globin regulation in β -thalassemia: it's as simple as α , β , γ , δ . **The Journal of Clinical Investigation**, v. 115, n. 6, p. 1470-1473, 2005.

BIANCO, I. et al. La prevencione dell, anemia mediterranea nel Lazio: risultati del programa applicato negli ultimi 10 anni. **Prog. Med. Roma**, v.42, p. 521-549, 1986.

BARIONE, M. C. N. et al. Visualização de dados em sistemas de bancos de dados relacionais. Proceeding: XVI Simpósio Brasileiro de Banco de Dados. Rio de Janeiro. p. 95 – 109, 2001.

BONINI-DOMINGOS, C. R. **Hemoglobinopatias no Brasil**: variabilidade genética e metodologia laboratorial. São José do Rio Preto, 1993. 232 f. Tese (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.70-76, 2000.

BONINI-DOMINGOS, C. R. Thalassemia screening in Brazil – Results for 20 years. **Revista Brasileira de Hematologia e Hemoterapia**, v. 26, n. 4, p. 288-289, 2004.

BORRAJO, G. J. C. Situación de la pesquisa neonatal en Latinoamérica. **Rev. Méd. Minas Gerais**. Editora Coopmed, v. 15, n. 2, p. 19-20, 2005.

BOY, R. Rastreamento de doenças genéticas no recém-nascido. **Coletânea de tópicos de erros inatos do metabolismo e triagem neonatal**. Campo Grande-IPD, 2002. 1 CD-ROM

BRASIL. Ministério da Saúde. Portaria da Saúde, nº822, 6 de junho de 2001. Brasília, 2001.

CAO, A; MOI, P. Regulation of the globin genes. **Pediatr. Res.**, v.51, n.4, p.415-421, 2002.

CARVALHO, T. M. Triagem Neonatal no Brasil. **Rev. Méd. Minas Gerais**. Editora Coopmed, v. 15, 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. 227 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. K.; FUCHAROEN, S.; CHAN, V. Hemoglobin H disease: not necessarily a benign disorder. **Blood**, v. 1001, n. 3, p. 791-800, 2003.

CLARKE, G. M.; HIGGINS, T. N. Laboratory Investigation of Hemoglobinopathies and Thalassemias: Review and Update. **Clinical Chemistry**, v. 46, p. 1284-1290, 2000.

CLEGG, J. B. Can the product of the gene θ be a real globin? **Nature**, v. 329, n. 1, p. 465-466, 1987.

DACIE, J. V.; LEWIS, 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.157-60, 1998.

DAUDT, L.E. et al. Triagem neonatal para hemoglobinopatias: um estudo piloto em Porto Alegre, Rio Grande do Sul, Brasil. **Cad. Saúde Pub.**, v.18, n° 3, p. 833-841, 2002.

DEVILLE, M. A. et al. Disruption of the Mechanism of Long Range Activation within the Human α -Globin Complex. **The Journal of Biological Chemistry**, v. 279, n. 21, p. 21793-21801, 2004.

EASTMAN, J. W. et al. Automated HPLC screening of newborns for sickle cell anemia and other hemoglobinopathies. **Clin. Chem.**, v. 42, n. 5, p. 704-710, 1996.

EASTMAN, J. W. et al. Distribution of hemoglobin F, A, S, C, E, and D quantities in 4 million newborn screening specimens. **Clin. Chem**, v. 45, n. 5, p. 683-685, 1999.

EL-KALLA, S.; MATHEWS, A.R. Hb D-Punjab in the United Arab Emirates. **Hemoglobin**, v.21, p.369-375, 1997.

ELMASRI, R. - Fundamentals of Database Systems, 3^a ed, ISBN 0-201-54263-3, Addison-Wesley, 2000.

FALOUTSOS, C.; LIN, K. Fast Map: A Fast algorithm for indexing, Data-Minig and visualization of tradicional and Multimedia Dataset. **ACM SIGMOD**, Zurich,Switzerland, p. 163-174, 1995.

FAYYAD, U. Minig Databases: Towards algorithms for knowledge discovery. **Bulletin of the IEEE Technical Committee on Data Engineering**, v. 21, p. 39-48, 1997.

FROST, B. A.; BELLINGHAM, A. J. Neonatal Haemoglobinopathy screening. **Acta Haemat.**, v. 78, p. 142-143, 1987.

GALACTEROS, F. Diagnostic néonatal des hémoglobinopathies. **Rev. Prat. (Paris)**, v.42, n.15, p.1893-5, 1992.

GRIFFITHS, A. J. H., et al. **Introduction to Genetic Analysis**. 7 ed. New York: W.H> Freeman e Co., 1999.

GTA-TN. **PNTN**: Apresentação sobre o Programa Nacional de Triagem Neonatal. Disponível em: <http://dtr2001.saude.gov.br/sas/dsra/prtn_arquivos/frame.htm> Acesso em: 10 Fevereiro 2004.

HAYASHI, A. et al., Neonatal screening and mass-spectrometric analysis of hemoglobin variants in Japan. **Acta Hemat.** v. 78, p. 114-118, 1987.

HONIG, G. R.; ADAMS III, J. G. **Human hemoglobin genetics**, Wien, N.Y: Springer Verlag, 1986. 452 p.

HUISMAN H. J. et al. HbVar: A Database of Human Hemoglobin Variants and Thalassemias. Summaries of mutation categories. Pennsylvania University USA and McMaster University in Canadá, 1996. Disponível em: <<http://globin.cse.psu.edu/>>. Acesso em 02 de novembro de 2005.

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

JANUÁRIO, J. N. Programa de triagem neonatal apresenta primeiros resultados. **Jornal Hemominas**, v. 8, n. 37, p. 5-6, 1998.

KEIM, D. A.; KRIEGLER, H. P. - Visualization Techniques for Mining Large Databases: A Comparison. **IEEE Transactions on Knowledge and Data Engineering**, v.8, n. 6, p. 923-938, 1996.

KEIM, D. A. - Information Visualization and Visual Data Mining. **IEEE Transactions on Visualization and Computer Graphics**, v.8 , n.1 , 2002.

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

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

MARENGO-ROWE, A. J. Rapid electrophoresis and quantitation of haemoglobin on cellulose acetate. **J. Clin. Path.**, v. 18, p. 790-792, 1965.

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 (doutorado em Genética) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista.

NAOUM, P. C. **Eletroforese, técnicas e diagnósticos**. São Paulo: Santos, 1990. 174 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.113, p.70-98, 1998.

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

OLD, J. M. Screening and genetic diagnosis of haemoglobin disorders. **Blood Reviews**, v. 17, p. 43-53, 2003.

OLIVEIRA, R. B. T.; **O processo de extração de conhecimento de Bases de dados apoiado por agentes de software**. São Carlos, 2000, 104 f. Dissertação (Mestrado em Ciências da Computação) – Instituto de Ciências Matemáticas e de Computação.

ONDEI, L. S. **Perfil eletroforético e cromatográfico das Hb S -“like”**. São José do Rio Preto, 2005. 114 f. Dissertação (Mestrado em 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: Hew publications, 1974.

PEREA, F. J. et al. Hb D-Los Angeles associated with Hb S or β -Thalassemia in four Mexican mestizo families. **Hemoglobin**, v. 23, n. 3, p. 231-237, 1999

RAMALHO, A. S.; MAGNA, L. A.; PAIVA e SILVA, R. B. A portaria MS nº 822/01 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. 62, n. 10, p. 364-374, 2004

SILVESTRONI, E.; BIANCO, I. Screening for microcytemia in Italy: analysis of data collected in the past 30 years. **Am. J. Hum. Genetic**, v. 27, p. 198, 1975.

STAMATOYANNOPOULOS, G. Control of globin gene expression during development and erythroid differentiation. **Experimental Hematology**, v. 33, p. 259-271, 2005.

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, 2001. 104 f. Tese (Livre docente em 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 Minig**. São Jose do Rio Preto, 2003, 56 f.

VELLA, F. Acid agar gel electrophoresis of human hemoglobin. **Am. J. Clin. Path.**, v. 49, n. 3, p. 440-442, 1968.

XIANG-HUI, F.; DE-PEI, L.; CHIH-CHUAN, L. Chromatin Structure and Transcriptional Regulation of the β -Globin Locus. **Experimental Cell Research**, v. 278, p. 1-11, 2002.

WAYE, J.S.; CHUI, D.H. The alpha-globin gene cluster: genetics and disorders. **Clin. Invest Med.**, v.24, n.2, p.103-109, 2001.

W.H.O.- WORLD HEALTH ORGANIZATION. Hereditary anaemias: genetic basis, clinical features, diagnosis and treatment. **Bulletin of the W.H.O.**, v. 60, n. 5. p. 643-59, 1982.

WEATHERALL, D.J.; CLEGG, J.B. **The thalassemia syndromes**. 3.ed Oxford: Blackwell, 1981, 875p.

WEATHERALL, D. J. The thalassaemias. **Brit. Med. J.**, v. 314, p. 1675-8, 1997.

WEATHERALL, D. J.; CLEGG, J. B., Inherited haemoglobin disorders: an increasing global health problem. **Bulletin of the World Health Organization**, v. 79, n. 8, p. 704-712, 2001.

ZAGO, M. A.; COSTA, F. F. Hereditary haemoglobin disorders in Brazil. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 79, p. 385-388, 1985.

ZAR, J.H. **Biostatistical Analysis**. 4.ed. New Jersey: Prentice Hall, 1999. 663p.

8* *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 5976 neonatos nascidos no Hospital de Base de São José do Rio Preto. O montante de dados gerados em todas as fases dos seis anos de programa constitui uma fonte de informações, que armazenadas em banco de dados, permitem a análise por métodos computacionais na tentativa de estabelecer comportamentos, 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, em sistema de visualização 3D - *Fas MapDB*. A utilização de um banco de dados em laboratórios com grande fluxo de amostras mostrou-se imprescindível para a segurança, confiabilidade e agilidade das análises e ferramentas que promovem a mineração dos dados, como a CLIBIA, revelaram 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 Hb normais, devido a amplitude de variação nos valores obtidos por HPLC para o pico de “F1”. Permitiu a separação das amostras com Hb normais e com alfa talassemia, ressaltando as diferenças estatísticas nos valores percentuais médios dos picos de “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.

9Abstract

9 Abstract

Since 1997 until 2002, a neonatal screening program of hemoglobinopathias were carried out in the Hemoglobin and Hematologic Genetic Diseases Laboratory - LHGDH, UNESP, with 5976 newborns 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 data base, 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 become them each time more exact. After an Electronic Database were development and the screening results were inserted was aimed: to evaluate by data mining tool - CLIBIA the methodologies used, test the applicability of a hemoglobinopathies database and to analyze the quantitative hemoglobins 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 datamining, as the CLIBIA, had disclosed the necessity to reevaluate the methodologies used in hemoglobinopathies diagnosis routine in newborns. The three-dimensional visualizations carried out by the *FastMapDB* tool showed the existence of two distribution standards in the samples with normal Hb, due to the variation amplitude into the values obtained by HPLC for the "F1" window. It allowed separating the samples with normal Hb from alpha thalassemia, evidencing the statistical differences in the mean values of the "Fast" and "A" window, emphasizing the necessity of better chromatograms evaluation and the possibility of this methodology application as part of alpha thalassemia diagnosis in newborns.

10 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 Melo¹

Paulo Henrique Mangonaro²

Profa. Dra. Claudia Regina Bonini Domingos¹

1 – Laboratório de Hemoglobinas e Genética das Doenças Hematológicas –
LHGDH - IBILCE/UNESP

Rua Cristóvão Colombo, 2265 – Jd Nazareth, Cep 15054-000 – 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, 2265 – Jd Nazareth,

15054-000 - São José do Rio Preto, SP

Fone: (17) 221-2392. Fax: (17) 221-2390

E-mail: stortilu@yahoo.com.br

Resumo

As hemoglobinopatias foram incluídas no Programa Nacional de Triagem Neonatal a partir de seis de junho de 2001 com a portaria do Ministério da Saúde nº 822. A cromatografia líquida de alta pressão em sistema automatizado foi indicada como uma das metodologias de diagnóstico. A quantidade de informações geradas por esses sistemas é imensa e, nem sempre, um comportamento pode ser observado nas análises individuais. Técnicas de visualização tridimensional podem ser aplicadas para extração dessas informações, como o estabelecimento de padrões, tendências ou relações a partir de resultados armazenados em bases de dados. Objetivou-se nesse trabalho a aplicação de uma ferramenta *datamining* de visualização 3D para detecção de padrões nos resultados de diagnóstico neonatal das hemoglobinopatias por HPLC. Foram utilizados os resultados laboratoriais de 2520 análises de recém-nascidos realizadas nos anos de 2001 e 2002. Os valores quantitativos de hemoglobinas, selecionados como atributos para o mapeamento foram os picos “Fast”, “F1”, “F” e “A” emitidos pelo sistema analítico. Para estabelecer um padrão de comportamento os resultados foram classificados em grupos de estudo segundo o fenótipo de hemoglobina: “Normal” (n = 2169); “Variante” (n = 73) e “Talassemia” (n = 279). A imagem gerada pela análise conjunta dos três grupos, com todos os atributos, mostrou uma separação espacial em nuvens distintas, sem sobreposição. As imagens de “Talassemia” e “Normal” apresentaram-se diametralmente opostas, evidenciando a existência de diferenças nos padrões obtidos por HPLC. A visualização entre os grupos “Normal” e “Variante” apresentou duas nuvens distintas com sobreposição nas extremidades. A área de intersecção nesses grupos pode ser atribuída a heterozigose das amostras com hemoglobina variante. A análise dos resultados permitiu a identificação de um padrão de comportamento diferenciado entre amostras normais e com talassemias, apesar da HPLC não ser utilizada como metodologia diagnóstica para esse tipo de alteração, permitindo inferências nos possíveis diagnósticos finais.

Introdução

As hemoglobinopatias são um grupo de alterações genéticas que representam um problema de saúde pública em 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 hemoglobinas 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 morbi – mortalidade e melhorando a qualidade de vida dos portadores. No Brasil, as hemoglobinopatias foram oficialmente incluídas nos programas de triagem neonatal a partir de 2001 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 hemoglobinopatias 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 ^(5,6,7). 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, é necessária a utilização de técnicas de mapeamento de informação, que têm como primeira ferramenta a apresentação gráfica dos dados ⁽⁹⁾. Essa apresentação consiste em visualizar graficamente dados, os quais se

deseja analisar. É centrada em mecanismos que possibilitam ao usuário a compreensão rápida da informação a ele apresentada. A exploração visual de dados fornece um grau mais alto de confiabilidade nas conclusões da exploração do que uma representação numérica ou textual ^(10.).

O conjunto de técnicas automáticas para a exploração em grandes massas de dados defini-se como *Data Mining*, ou seja, mineração de dados, confeccionadas com o objetivo de encontrar novos padrões, tendências e relações. Através destas técnicas, é possível a extração de informações úteis, que são difíceis de serem desvendadas pelo ser humano a olho nu, sendo melhor interpretadas quando são apresentadas de forma gráfica ⁽¹¹⁾. O *FastMapDB* é uma ferramenta de visualização de dados, como datas, números e textos, onde a apresentação é efetuada através do mapeamento multidimensional dos dados para um espaço tridimensional, promovendo a redução de dimensionalidade. O objetivo da redução de dimensionalidade é apresentar o conjunto de dados com um número menor de atributos, porém preservando as características inerentes a informação armazenada, dessa forma buscando definir quais são os atributos importantes ou que concentram a informação ⁽¹²⁾. Possibilita verificar pontos fora do padrão (*outliers*) e a formação de agrupamentos (*clusters*). Por exemplo, em uma base de dados correspondente a resultados de exames laboratoriais, que no mapeamento desses dados apresente aglomerados de pontos, pode significar que os exames que compõem cada aglomerado têm algo em comum ⁽¹³⁾.

Diante a diversidade de fenótipos de Hb observados na população brasileira e a quantidade de informações geradas em programas neonatais objetivou-se nesse trabalho a aplicação da ferramenta *data mining* de visualização 3D - *FastMapDB* para detecção de padrões de comportamento nos resultados quantitativos obtidos no diagnóstico neonatal de hemoglobinopatias por HPLC.

Casuística e Métodos

Foram utilizados os resultados laboratoriais de Hb obtidos por HPLC de 2521 amostras de sangue de cordão umbilical de recém nascidos da região

noroeste do estado de São Paulo nos anos de 2001 e 2002 sem distinção de sexo e origem étnica. O presente estudo teve aprovação da CONEP e esta em concordância com as normas estabelecidas pela Resolução CNS 196/96.

O diagnóstico final para determinação do fenótipo foi realizado pela combinação de diferentes metodologias laboratoriais, incluindo: Resistência globular osmótica em solução de NaCl a 0,36%⁽¹⁴⁾, um teste seletivo para rastreamento de talassemias do tipo beta; Análise da morfologia eritrocitária⁽¹⁵⁾, para identificar alterações como, hipocromia, microcitose, células em alvo e falcizadas; Eletroforese em acetato de celulose, pH 8,6⁽¹⁶⁾, que permite a separação de todas as hemoglobinas normais e grande parte das anormais; Eletroforese em pH neutro⁽¹⁷⁾, para identificação da hemoglobina Bart's, indicativa de alfa talassemia nesta faixa etária; Eletroforese em ágar-fosfato, pH 6,2⁽¹⁸⁾, que permite separar as hemoglobinas S e C de outras variantes com padrão de migração semelhante em pH alcalino; Pesquisa de corpos de Hb H e Heinz⁽¹⁹⁾, utilizada para a confirmação citológica destes agregados, após coloração vital, na suspeita de alfa talassemia e hemoglobinas instáveis e Cromatografia líquida de alta performance, HPLC⁽²⁰⁾, realizada com o Sistema automatizado VARIANT (BIO-RAD), utilizando os reagentes e o protocolo do fabricante para esta faixa etária.

A ferramenta *FastMapDB* foi desenvolvida para disponibilizar os recursos de visualização para dados armazenados em bases relacionais utilizando resultados e algoritmos originalmente desenvolvidos para a redução de dimensionalidade ou de mapeamento de dados métricos para espaços dimensionais. A ferramenta possui recursos para que o usuário crie interativamente uma função de distância vetorial a partir de qualquer quantidade de atributos de uma tabela em uma 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 mapeia o conjunto de dados para uma representação tridimensional, sendo que cada objeto na tabela é 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 (¹³).

Para aplicação da ferramenta *datamining – FastMapDB*, foram utilizados os resultados quantitativos obtidos por HPLC. Esse sistema identifica os valores percentuais das frações de Hb em picos determinados, “Fast”, “F1”, “F” e “A”, que depois de inseridos em planilhas, foram utilizados como atributos para o mapeamento espacial e, os fenótipos foram utilizados como atributo classificatório das amostras. Para estabelecer um padrão de comportamento espacial os resultados foram classificados em grupos segundo delineamento ilustrado no quadro 1.

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

N	11.1 Fenótipos de Hb	11.2 Grupos de análise para o Fast Map
2169	Perfil de Hb normal para idade	11.3 “Normal”
55	Hb S em heterozigose	“Variante”
18	Hb C em heterozigose	
21	beta talassemia heterozigota	“Talassemia”
258	alfa talassemia heterozigota	

Resultados

Visando estabelecer o perfil do fenótipo normal para amostras de sangue de neonatos, inicialmente, foram geradas imagens utilizando os quatro picos discriminados como atributos de mapeamento para a ferramenta *datamining* (Figura1).

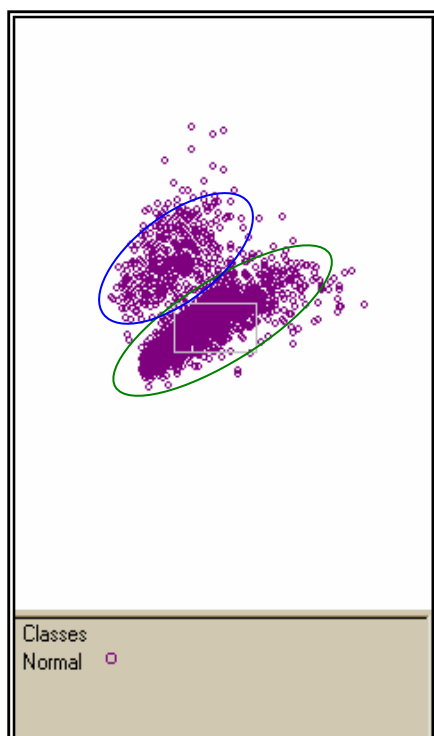


Figura 1. Projeção da imagem para o grupo “Normal”, incluindo os quatro atributos de mapeamento.

O mapeamento das amostras normais apresentou dois aglomerados de pontos marcados pelos círculos na Figura 1. Para avaliar qual o fator responsável pela divisão das amostras 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 pôde ser observada a presença de dois grupos separados. Retirando-se pico de “F1” ocorreu uma aproximação das nuvens de dados, sugerindo que esses valores podem ter determinado a divisão das amostras normais em dois padrões distintos de normalidade.

Associando-se as duas observações, gerou-se uma imagem apenas com os picos de “F” e “A”, retirando-se as janelas de “Fast” e “F1” (Figura 2), e obteve-se a dispersão linear dos dados.

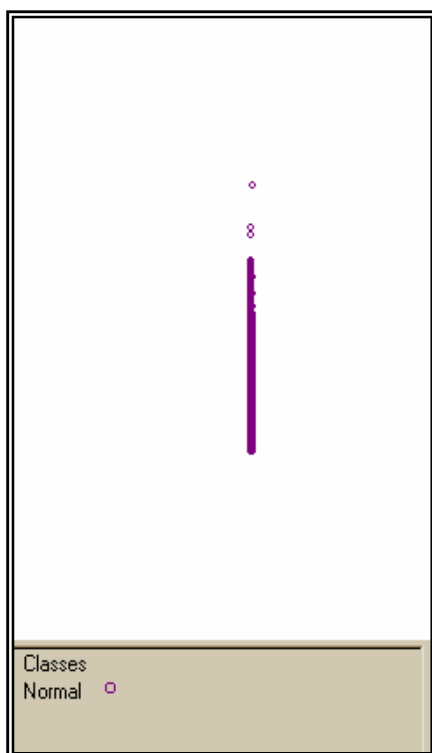


Figura 2. Projeção da imagem para o grupo “Normal”, incluindo os picos “F” e “A” como atributos de mapeamento.

Para avaliar os valores de “Fast” e “F1” foram geradas imagens utilizando um recurso da ferramenta *FastMapDB* que permitiu a divisão exatamente 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.

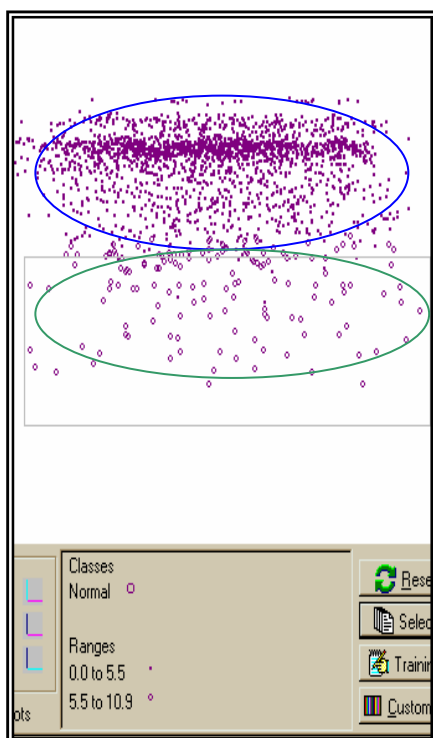


Figura 3. Projeção da imagem para o grupo “Normal”, incluindo os quatro atributos de mapeamento e utilizando o pico de “Fast” como classificatório dividido em dois subgrupos sendo de 0,0 a 5,5 % e de 5,5% a 10,9%.

Como observado, os pontos marcados pelo círculo azul apresentam valores de “Fast” entre 0,0 e 5,5 % e os pontos dentro do círculo verde correspondem à valores de “Fast” de 5,5 a 10,9%. 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 2040 amostras. O subgrupo variando de 5,5 a 10,9 % foi composto por apenas 129 amostras e a média observada foi 7,35 %. A média percentual obtida no pico de “Fast” para todas as amostras normais foi de 2,64%, 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 “F1” para dividir as amostras no valor médio e a imagem obtida está ilustrada na Figura 4, na qual a distribuição espacial dos dados foi muito semelhante à da Figura 1. As duas nuvens classificadas pelos valores, observadas na Figura 4 não apresentaram sobreposição dos pontos. Para possibilitar a comparação das imagens os mapeamentos das Figuras 1 e 4 foram dispostos lado a lado na Figura 5. Para caracterizar a divisão em duas nuvens amostrais, calculou-se a média percentual para cada subgrupo obtido para o pico de F1. Obteve-se para o subgrupo com valores variando de 0,0 a 10,3% média percentual de 9,45% incluindo 1369 amostras e para o subgrupo com valores variando de 10,3 a 20,6% obteve-se a média de 12,25%, sendo composto por 800 amostras. A média percentual obtida para todas as amostras normais foi de 10,48%, apresentando diferença estatisticamente significativa ($p < 0,05$) das médias obtidas para os dois subgrupos de “F1”.

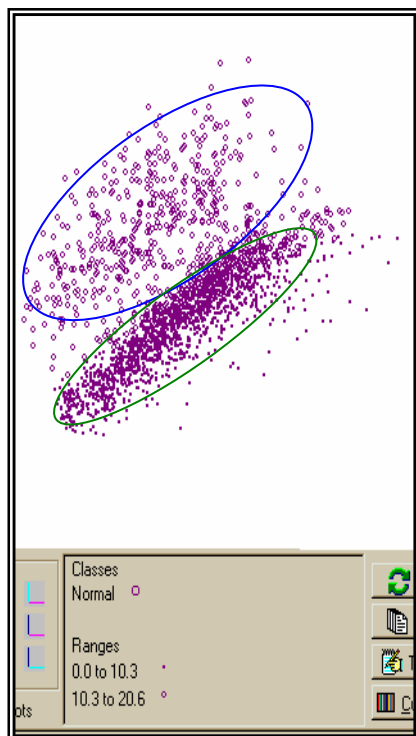


Figura 4. Projeção da imagem para o grupo “Normal”, incluindo os quatro atributos de mapeamento e utilizando o pico de “F1” como classificatório dividido em dois subgrupos sendo de 0,0 a 10,3 % e de 10,3% a 20,6%.

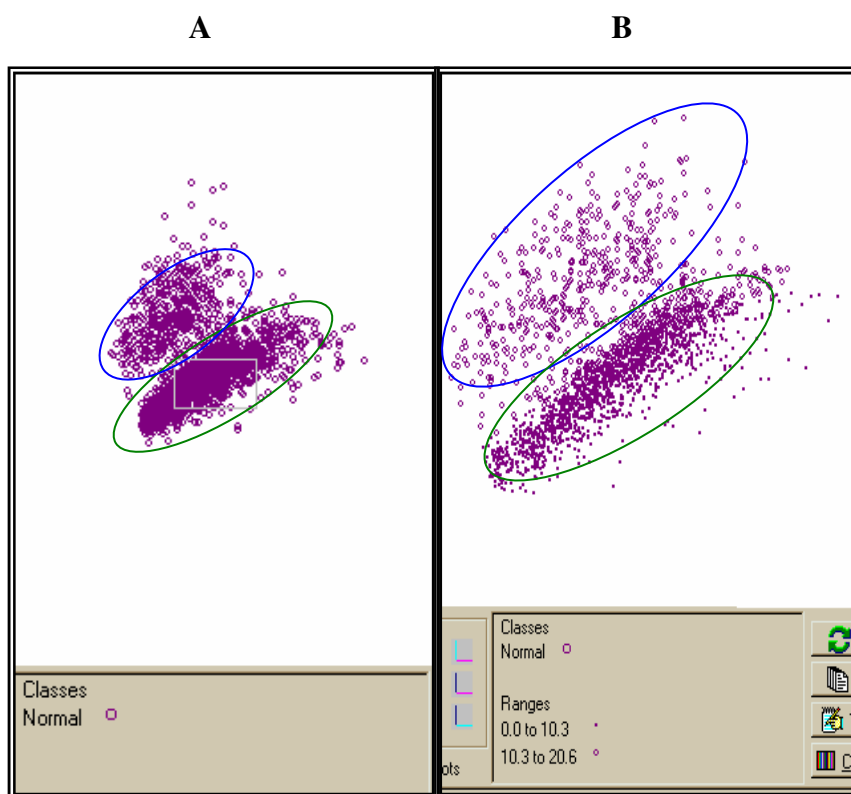


Figura 5. Comparação das imagens projetadas para o grupo “Normal”. A - incluindo os quatro picos como atributos de mapeamento B - utilizando o pico de “F1” dividido em dois subgrupos de 0,0 a 10,3% e de 10,3 a 20,6%.

Visando estabelecer um padrão para os diferentes fenótipos observados foram mapeadas todas as amostras, divididas nos três grupos, “Normal”, “Talassemia” e “Variante”, incluindo os quatro picos como atributos para o mapeamento. Na imagem gerada, os grupos de análise apresentaram-se distribuídos em três nuvens amostrais distintas, como ilustrado na Figura 6, não houve sobreposição entre os pontos. As nuvens correspondentes aos grupos “Talassemia” e “Normal” encontram-se diametralmente opostas.

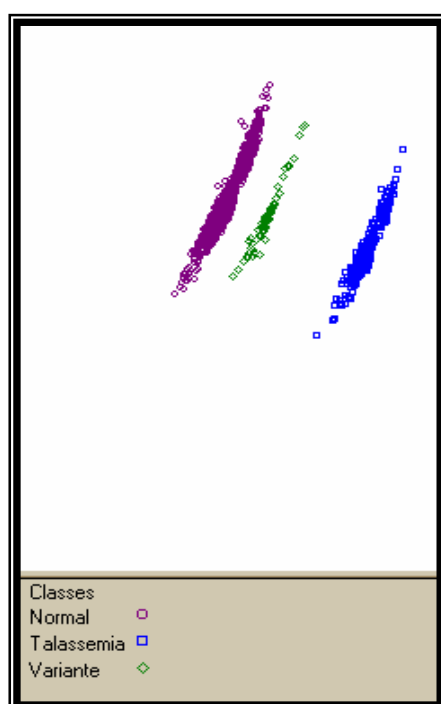


Figura 6. Projeção da imagem tridimensional para os grupos “Normal”, “Talassemia” e “Variante”, incluindo os quatro atributos de mapeamento.

Em seguida foram trabalhadas as amostras do grupo “Normal” e “Talassemia” e na imagem gerada utilizando os quatro atributos de mapeamento observou-se a diferenciação em aglomerados distintos para cada grupo, sem sobreposição espacial. Para verificar quais as frações fundamentais nessa diferenciação, foi gerada uma imagem utilizando apenas os picos “F” e “A” como atributos de mapeamento, por representarem as frações de hemoglobinas em 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 mapeamento espacial, como pode ser visualizado na Figura 7.

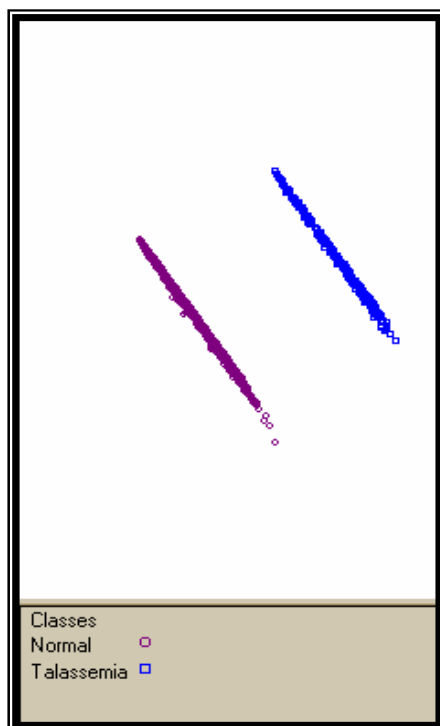


Figura 7. Projeção da imagem tridimensional para os grupos “Talassemia” e “Normal” utilizando como atributos de mapeamento os picos “F” e “A”.

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

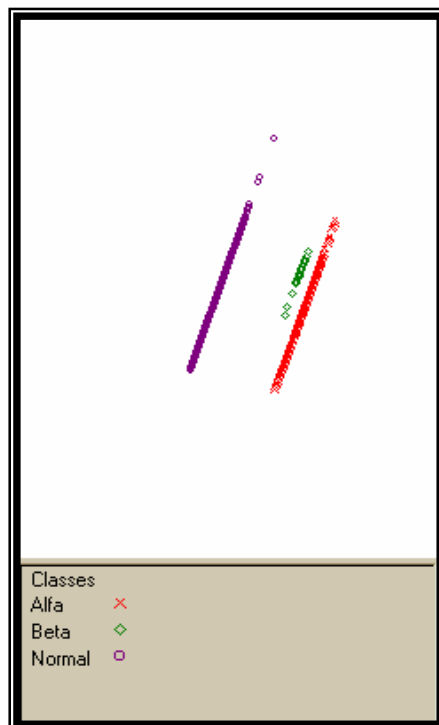


Figura 8. Projeção da imagem tridimensional para os grupos “Normal”, “Alfa” e “Beta” talassemias, utilizando como atributos de mapeamento os picos “F” e “A”.

No mapeamento espacial dos resultados laboratoriais obtidos para as amostras normais e com Hb variantes observou-se a 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 manteve mesmo com a rotação espacial da imagem.

O grupo “Variante” é composto pelas amostras contendo Hb AS e Hb AC, para tentar estabelecer um padrão entre essas variantes e diferenciá-las das amostras com Hb normais, realizou-se o mapeamento incluindo o grupo “Normal” com cada uma dessas variantes isoladamente. O padrão de dispersão espacial se manteve, não sendo possível separar as amostras em nuvens amostrais distintas. A Figura 10 representa o mapeamento incluindo os grupos “Normal”, Hb AS e Hb AC para os quatro atributos de mapeamento. 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 mapeamento espacial por estarem ausentes nas amostras com hemoglobinas normais.

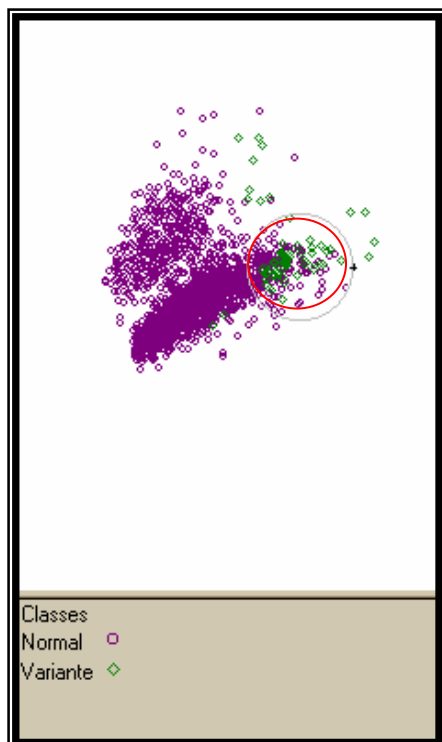


Figura 9. Projeção da imagem tridimensional para os grupos “Normal” e “Variante” utilizando os quatro atributos de mapeamento. Área de sobreposição marcada pelo círculo vermelho.

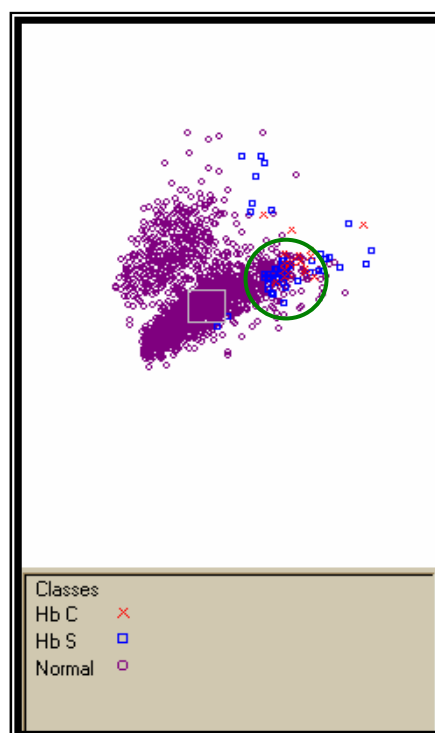


Figura 10. Projeção da imagem tridimensional para os grupos “Normal”, “Hb AS” e “Hb AC” utilizando os quatro atributos de mapeamento. Área de sobreposição marcada pelo círculo verde.

Discussão

A aplicação de técnicas de VDM é uma abordagem nova nas pesquisas com hemoglobinopatias. Para os dados da triagem neonatal realizada pelo LHGDH as análises pela ferramenta *FastMapDB* apresentaram padrões relevantes. Na projeção das amostras com hemoglobinas normais para a idade observou-se a existência de duas nuvens de distribuição amostral. O mapeamento feito sem a presença dos picos de “Fast” e “F1” mostrou a distribuição linear das amostras, sugerindo dois padrões de comportamento distintos nas amostras com Hb normais. Com o recurso da ferramenta *FastMapDB* os valores do pico de “Fast” foram divididos ao meio separando as amostras em dois subgrupos. As amostras distribuíram-se em dois grupos sem sobreposição, no entanto, houve um desequilíbrio na distribuição das amostras, de forma que o primeiro subgrupo

incluiu 94% das amostras, não sendo representativo da distribuição observada na Figura 1. Esses resultados destacam que o pico de “Fast” não foi o responsável pela formação das duas nuvens amostrais observadas no mapeamento das amostras com hemoglobinas normais para a idade.

No mapeamento feito utilizando a divisão do pico “F1” no valor médio obteve-se dois grupos distribuídos de forma mais igualitária, sendo o primeiro composto por 63,1% das amostras e o segundo incluindo 36,9%. As médias obtidas para esses dois subgrupos apresentaram diferenças estatisticamente significativas ($p < 0,05$), quando comparadas entre si e à média geral obtida para todas as amostras normais. Quando comparadas lado a lado, as visualizações obtidas para as amostras normais com os quatro atributos e com o pico de “F1” dividido em dois subgrupos (Figura 4), pôde-se observar semelhança na distribuição dos pontos amostrais. Esses resultados indicam a presença de dois padrões com valores médios diferentes de distribuição para os valores de “F1” nas amostras com Hb normais. Diante desses achados, deve-se destacar, portanto que a média geral obtida para as 2169 amostras não é valor representativo deste pico de Hb para o grupo “Normal”.

A análise global dos resultados, visualizados em projeção espacial através da ferramenta *FastMapDB*, ressaltou a existência de um possível padrão para as amostras talassêmicas por HPLC, pois não houve sobreposição espacial dos pontos amostrais, sugerindo diferenças nos perfis obtidos por HPLC para as amostras com valores normais para a idade e aquelas com formas talassêmicas, alfa ou beta, mesmo que em heterozigose. A HPLC não é uma 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 aumento estatístico significativo para os picos de “Fast” e “F1” em amostras com alfa talassemia. Para amostras com beta talassemia obteve valores aumentados para o pico de “F” e diminuídos para “A”, sendo as diferenças estatisticamente significativas dos valores obtidos para o grupo

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

Apesar da HPLC não poder ser utilizada como método 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 método 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 Hb S e Hb C são diagnosticadas com precisão pela HPLC que identifica com alta sensibilidade as pequenas porcentagens das frações variantes. No entanto o mapeamento espacial das amostras heterozigotas para essas hemoglobinas comparadas ao grupo Normal mostrou sobreposição das nuvens amostrais, não sendo possível diferenciar os grupos. O pico de Hb Fetal, que representa o maior componente hemoglobínico nesta faixa etária, não apresenta alterações nos valores percentuais entre as amostras normais e com Hb variantes, esse fato associado à diferença representativa no número de amostras, pode ser responsável pela sobreposição das nuvens amostrais. Deve-se destacar também que as Hb variantes eluem em picos específicos da cromatografia e estes não foram inseridos no mapeamento por estarem ausentes nas amostras 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 tem 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 milhões 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 observação de características intrínsecas a cada fenótipo 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 amostras normais e talassêmicas através da separação em nuvens distintas, ressaltando a necessidade de avaliar os resultados cromatográficos em 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. 60, n. 5. p. 643-59, 1982.
2. GALACTEROS, F. Neonatal detection of Sickle Cell Disease in metropolitan France. association francaise pour le depistage et la prevention des handicaps de l'enfant (AFDPHE). **Arch. Pediatr.**, v. 3, p. 1026-1031, 1996.
3. BRASIL. Ministério da Saúde. Portaria da Saúde, nº822, 6 de junho de 2001. Brasília, 2001.
4. JANUÁRIO, J. N. Programa de triagem neonatal apresenta primeiros resultados. **Jornal Hemominas**, v. 8, n. 37, p. 5-6, 1998.
5. HAYASHI, A. et al., Neonatal screening and mass-spectrometric analysis of hemoglobin variants in Japan. **Acta Hemat.** v. 78, p. 114-118, 1987.

6. FUCHAROEN, S.; et al., Prenatal and postnatal diagnoses of thalasseмии and hemoglobinopathies by HPLC. **Clinical Chemistry**, v. 44, n. 4, p. 740-748, 1998.
7. OU, C.; ROGNERUD, C. L. Diagnosis of hemoglobinopathies: electrophoresis vs. HPLC. **Clinica Chimica Acta.** , v. 313, p. 187-194, 2001.
8. BRASIL a. 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. KEIM, D. A. Visual exploration of large data sets. **Communications of the ACM**, v. 4, n. 8, p. 39-44, 2001.
11. BARIONE, M. C. N.; et al., Visualização de dados em sistemas de banco de dados relacionais. **Proceeding XVI Simpósio Brasileiro de Banco de Dados**, Rio de Janeiro, p. 95-109, Outubro de 2001.
12. FALOUTASOS, C.; LIN, K. *FastMap*: A fast algorithm for indexing, data mining and visualization of traditional and multimedia datasets, **ACM Sigmod**, Zurich, Switzerland, p. 163-174, 1995.
13. TRONCO, M. N. **Implementação de Recursos de Visualização e Interação em Ferramenta Data Minig.** São Jose do Rio Preto, 2003, 56 f.
14. SILVESTRONI, E.; BIANCO, I. Screening for microcytemia in Italy: analysis of data collected in the past 30 years. **Am. J. Hum. Genetic**, v. 27, p. 198, 1975
15. BONINI-DOMINGOS, C. R. **Hemoglobinopatias no Brasil - Variabilidade genética e metodologia laboratorial.** 1993. 231f. Tese (Doutorado em Ciências) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, 1993.
16. MARENGO-ROWE, A. J. Rapid electrophoresis and quantitation of haemoglobin on cellulose acetate. **J. Clin. Path.**, v. 18, p. 790-792, 1965.
17. DACIE, J. V.; LEWIS, S. M. **Practical haematology.** 8 ed. Livingstone: Churchill, 1995.
18. VELLA, F. Acid agar gel electrophoresis of human hemoglobin. **Am. J. Clin. Path.**, v. 49, n. 3, p. 440-442, 1968.

19. PAPANNOPOULOS, R.; STAMATONYANNOPOULOS, G. **Stains for inclusions bodies.** “In: STANDARDIZATION of laboratory reagents and methods for detection of haemoglobinopathies”. Atlanta: Hew publications, 1974.

20. Instrucion Manual Variant Sickle Cell Short Program, 1994.