

ANA FLÁVIA BRUGNEROTTO

**ANÁLISE DA EXPRESSÃO GÊNICA EM PACIENTES
TALASSÊMICOS HOMOZIGOTOS PARA
MUTAÇÃO BETA 39 COM EVOLUÇÕES CLÍNICAS
DISTINTAS, MAIOR E INTERMEDIÁRIA, POR
*SERIAL ANALYSIS OF GENE EXPRESSION (SAGE)***



UNICAMP

CAMPINAS

2010

ANA FLÁVIA BRUGNEROTTO

**ANÁLISE DA EXPRESSÃO GÊNICA EM PACIENTES
TALASSÊMICOS HOMOZIGOTOS PARA
MUTAÇÃO BETA 39 COM EVOLUÇÕES CLÍNICAS
DISTINTAS, MAIOR E INTERMEDIÁRIA, POR
*SERIAL ANALYSIS OF GENE EXPRESSION (SAGE)***

*Dissertação de Mestrado apresentada ao Programa de
Pós-Graduação em Clínica Médica da Universidade
Estadual de Campinas para a Obtenção do Título de
Mestre em Clínica Médica, área de concentração em
Ciências Básicas*

Orientador: Dr. Fernando Ferreira Costa

Coorientador: Dr. Anderson Ferreira da Cunha

CAMPINAS

2010

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA FACULDADE DE CIÊNCIAS MÉDICAS DA UNICAMP**

Bibliotecário: Sandra Lúcia Pereira – CRB-8ª / 6044

Brugnerotto, Ana Flávia
B833a Análise da expressão gênica em pacientes talassêmicos homozigotos para mutação beta 39 com evoluções clínicas distintas, maior e intermediária, por Serial Analysis of Gene Expression (SAGE) / Ana Flávia Brugnerotto. Campinas, SP : [s.n.], 2010.

Orientadores: Fernando Ferreira Costa, Anderson Ferreira da Cunha
Dissertação (Mestrado) Universidade Estadual de Campinas. Faculdade de Ciências Médicas.

1. Talassemia beta. 2. Análise serial da expressão gênica. I. Costa, Fernando Ferreira. II. Cunha, Anderson Ferreira da. III. Universidade Estadual de Campinas. Faculdade de Ciências Médicas. IV. Título.

Título em inglês: “Analysis of gene expression from patients with beta thalassemia, carrying the same mutation (Cd39) but with different phenotypes (major and intermedia), using the Serial Analysis of Gene Expression (SAGE)”

Keywords: • Beta thalassemia
• SAGE

Titulação: Mestre em Clínica Médica
Área de concentração: Ciências Básicas

Banca examinadora:

Prof. Dr. Fernando Ferreira Costa
Profa. Dra. Maria de Lourdes Rios Barjas Castro
Profa. Dra. Maria Stella Figueiredo

Data da defesa: 31-08-2010

Banca examinadora da Dissertação de Mestrado

Ana Flávia Brugnerotto

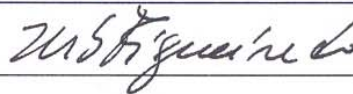
Orientador: Prof. Dr. Fernando Ferreira Costa

Membros:

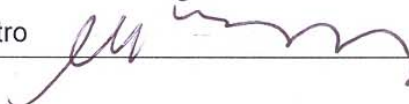
1. Prof. Dr. Fernando Ferreira Costa



2. Prof^a. Dr^a. Maria Stella Figueiredo



3. Prof^a. Dr^a. Maria de Lourdes Rios Barjas Castro



Curso de Pós-Graduação em Clínica Médica da Faculdade de Ciências Médicas da Universidade Estadual de Campinas.

Data: 31/08/2010

DEDICATÓRIA

Aos meus pais, Vitório e Ana Maria

Ao meu marido e companheiro, Rogério

À minha irmã e amiga, Vanessa

AGRADECIMENTOS

Agradeço a Deus por ter me direcionado até aqui e daqui para adiante.

Ao meu orientador, Fernando Ferreira Costa, pela sua bondade, experiência e por ter me recebido como sua aluna.

Ao Anderson, que me recebeu como a uma filha, me iniciou, me ensinou, cuidou sempre para que tudo desse certo. Todas as palavras do mundo não seriam suficientes para expressar minha gratidão.

Aos meus pais, por estarem sempre prontos e presentes, mesmo que por pensamento, em todos os momentos, bons ou ruins da minha vida.

À minha irmã, companheira e amiga, por estar sempre por perto.

Ao Rogério, pela compreensão, amor e companheirismo.

À Laura e ao Paulo, por cuidarem de mim.

À Dulcinéia, pela sua disposição em ajudar sempre, seja na bancada ou dando conselhos, valeu Dul.

Agradeço a Lena e a Simone, que sempre faz os dias serem mais legais.

Agradeço a todos do laboratório de Biologia Molecular do Hemocentro da Unicamp, com carinho especial a turma do Osmar, que inclui Regiane, Camila, Flávia Pallis, Fernanda e Rose, além da Sheley, Flávia Boquinha, Cíntia, Diana, Emília, Carol, Carla, Andréia, Nic e Taís, pessoas que com certeza me ajudaram e apoiaram em momentos especiais. Agradeço também a Venina, Lediane, Dani, Tati, Vanessa, Adriana, Patrícia, Renatinha, Andrey, Kléber, Marcos André, Liciane e Gabriela.

Agradeço também a amiga da vida inteira, Renata, que diariamente me aguenta e me entende. Muito obrigada por fazer parte da minha vida.

Agradeço a CAPES, FAPESP e CnPQ, pelo apoio financeiro.

Enfim, agradeço a todos que participaram direta ou indiretamente da realização desse trabalho. Obrigada.

RESUMO

As síndromes talassêmicas compreendem um grupo heterogêneo de doenças hereditárias em que existe uma redução no ritmo de síntese de uma ou mais cadeias polipeptídicas da hemoglobina. Nas talassemias β ocorre a supressão total ou parcial da produção de cadeias β . O estado homozigótico da maioria das variantes genéticas da talassemia β produz o quadro clínico da talassemia maior. Esses pacientes apresentam acentuada anemia e necessitam de transfusões sanguíneas regulares para sobreviverem. Os indivíduos heterozigotos para a talassemia β apresentam, com raras exceções, apenas discreta anemia. Existem ainda alguns quadros clínicos não tão graves quanto à forma homozigótica clássica, geralmente não dependem de transfusões, que são denominados de talassemia intermediária. Os genes responsáveis pela síntese das cadeias da β globina estão organizados em um *cluster* localizado no braço curto do cromossomo 11. Quase 200 alelos de talassemia β já foram caracterizados. Destes, 125 representam mutações pontuais em regiões funcionalmente importantes do gene, uma de particular interesse ao nosso estudo, é a mutação Cd 39 (C $\rightarrow$ T). Este trabalho teve como objetivo analisar o perfil global de expressão gênica de células CD 34⁺ de pacientes com talassemia β , portadores da mesma mutação genética (Cd 39), mas com evoluções clínicas distintas, maior e intermediária, pelo método de SAGE (*Serial Analysis of Gene Expression*). Foram gerados 2718 transcritos únicos para o perfil de talassemia β intermediária e 3052 para o perfil de talassemia β maior, os quais foram classificados como genes identificados, *no matches*, ESTs e outras sequências preditas e anotadas. As expressões de 14 genes foram quantificadas pela reação em cadeia da polimerase em tempo real nas amostras de células CD34⁺ dos pacientes talassêmicos intermediário e maior, com intuito de validar os resultados obtidos pelo SAGE. As expressões foram concordantes em 57,14% dos genes (*ABCB10*, *APEX1*, *APOC1*, *EYA3*, *HMBS*, *OAZ1*, *SRGN*, *TAGLN2*) e discordantes nos demais 42,85% (*EIF5a*, *GRIN2C*, *HMGB1*, *NAE1*, *PCBP2*, *RAD23B*). Quando ambos os perfis foram comparados entre si, 42 transcritos foram ditos como diferencialmente expresso ($p < 0,05$). A análise funcional comparativa dos 42 transcritos diferencialmente expressos foi realizada de acordo com o *Gene Ontology Consortium* afim de obtermos a classificação funcional destes transcritos. Em conjunto, os resultados podem colaborar na

identificação de transcritos importantes, que possam auxiliar na melhor compreensão da fisiopatologia desta doença e desempenhar papéis moduladores no fenótipo em talassemia β .

ABSTRACT

Thalassemia syndromes are a heterogeneous group of hereditary diseases in which there is a reduction in the synthesis of one or more hemoglobin chains. In β -thalassemia there is a reduction or total suppression of the expression of β globin genes. The homozygous state of most β -thalassemia genetic variants produces the clinical evolution of thalassemia major. These patients present severe anemia and require regular blood transfusions in order to survive. Heterozygous individuals present, with exceptions, discrete anemia. There are some clinical evolutions that are not as severe as the classic homozygous form, which are often blood transfusion independent, named β -thalassemia intermedia. The gene responsible for the synthesis of β the globin chain is arranged in a cluster located on the short arm of chromosome 11. Nearly 200 β -thalassemia alleles have been characterized. Of these, 125 represent mutations in functionally-important regions of the gene. A nonsense mutation, ie base substitution, which introduces a premature stop codon, destroying the normal reading and interfering in mRNA translation, is of particular interest to our study, especially the Cd 39 mutations (C $\rightarrow$ T). The aim of this study was to evaluate the global gene expression pattern of CD34+ culture cells from patients with β -thalassemia, carrying the same genetic mutation (Cd 39) but with different phenotypes (major and intermedia), using Serial Analysis of Gene Expression (SAGE). Two SAGE profiles were generated: INT and MAIOR. Comparison of the 2718 and 3052 distinct tags from the INT and MAIOR profiles, represented by identified transcripts and novel tags, 42 tags demonstrated with a differential expression at a statistically significant level ($p < 0,05$) and corresponded to known genes, ESTs or no matches. The expression of 14 genes was further investigated by the real-time polymerase chain reaction in cells cultured from patients with β -thalassemia intermedia and major, with the purpose of validating the results obtained by the SAGE method. Similar expressions were seen in 57.14% (*ABCB10*, *APEX1*, *APOC1*, *EYA3*, *HMBS*, *OAZ1*, *SRGN*, *TAGLN2*) and discordant expressions were seen in 42.85% (*EIF5a*, *GRIN2C*, *HMGB1*, *NAE1*, *PCBP2*, *RAD23B*). The functional classification was performed according to the Gene Ontology Consortium and a total of 42 transcripts were submitted to functional classification. Together, our results may contribute to identify important transcripts that

may play roles in modulating phenotype in β -thalassemia and to better understand the pathophysiology of this disease

LISTA DE ABREVIACÕES

µg	Micrograma
µL	Microlitro
µM	Micromolar
BAC	Gene constitutivo beta actina
BSA	Albumina de soro bovino
cDNA	DNA complementar
CGAP	<i>Cancer Genome Anatomy Project</i>
DEPC	Dietilpirocarbonato
DNA	Ácido desoxirribonucléico
DNase	Desoxirribonuclease
dNTP	Deoxiribonucleotídeo
EST	Sequência de gene não classificada ou <i>expressed gene tag</i>
FITC	Isotiocianato de fluoresceína
GAPDH	Gene constitutivo gliceraldeído-3-fosfato-desidrogenase
M	Molar
mL	Mililitro
mM	Milimolar
mRNA	RNA mensageiro
ng	Nanograma
No match	Sequência gênica não descrita
ORF	Sequência aberta de leitura ou <i>open reading frame tag</i>
Pb	Pares de bases
PBS	Solução salina fosfato tamponada
PCR	Reação em cadeia da polimerase
PE	Ficoeritrina
Pmol	Picomol
qRT-PCR	Reação em cadeia da polimerase em tempo real
RNA	Ácido ribonucléico
rpm	Rotações por minuto
SAGE	Análise seriada da expressão gênica
Tm	Temperatura de dissociação
U	Unidade

LISTA DE TABELAS

Tabela 1 – 10 tags mais detectadas na biblioteca de paciente β talassêmico intermediário. Para cada tag estão indicados a sequência de nucleotídeo, símbolo, descrição e número no <i>UniGene</i> do gene correspondente, e a frequência com que foi detectada.....	82
Tabela 2 – 10 tags mais detectadas na biblioteca de paciente β talassêmico maior. Para cada tag estão indicados a sequência de nucleotídeo, símbolo, descrição e número no <i>UniGene</i> do gene correspondente, e a frequência com que foi detectada.	83
Tabela 3 – Padronização das reações em cadeia da polimerase quantitativas em tempo real. Genes selecionados para análise, sequências dos primers desenhados para sua amplificação, tamanho e temperatura de desnaturação dos produtos, predita pelo programa <i>GeneAmp</i> ® 5700 SDS (Applied Biosystems).....	85
Tabela 4 – Concentração de <i>primers</i> utilizadas na amplificação dos genes de estudo e eficiência da amplificação obtida. As concentrações foram definidas pela eficiência de amplificação gerada nas condições testadas.	87
Tabela 5 – Identificação das <i>tags</i> diferencialmente expressas entre células de cultura de um paciente portador de talassemia β maior e intermediária. <i>Tags</i> foram classificadas como: <i>no match</i> (não apresentam nenhuma identificação no <i>UniGene</i>), sequências preditas (correspondem à ESTs, ORFs, clones de cDNA ou proteínas hipotéticas) ou genes conhecidos.....	92
Tabela 6 – Ontologia dos genes diferencialmente expressos. Os genes foram classificados segundo o <i>Gene Ontology</i> quanto às funções e processos da qual participam.	94

LISTA DE FIGURAS

- Figura 1 - Desenvolvimento ontogenético normal das cadeias globínicas** - Na figura 1A podemos observar a composição e a disposição gênica dos cluster α e β . De acordo com a composição dos clusters temos a formação de vários tipos de hemoglobinas: as de predomínio embrionário, as presentes no período fetal e as que encontramos durante toda a vida adulta. O evento visualizado na figura 1B representa a troca sincronizada do predomínio do tipo de hemoglobina, fenômeno denominado “switching” (WEATHERALL, 2001).38
- Figura 2 - Representação esquemática da fisiopatologia da talassemia β** – As cadeias α globinas livres precipitam no organismo danificando a parede de precursores eritróides levando à eritropoese ineficaz e diminuição da vida média da hemácia (parte direita da figura). Este evento leva à produção de progenitores celulares, que ocasiona expansão da medula óssea e aumento na absorção de ferro. Em casos de anemia muito grave necessita-se de transfusão, situação na qual aumentam ainda mais os níveis de ferro no organismo e consequentemente sua absorção, levando à clínicas relacionadas com as talassemias β que são geralmente a causa dos óbitos dos pacientes. Em casos onde se observa aumento da produção de hemoglobina fetal (HbF) (lado esquerdo da figura), o paciente apresenta um quadro clínico melhor (WEATHERALL et al., 2001).41
- Figura 3 - Diferentes mecanismos moleculares no *cluster* da beta globina que causam talassemia**.....43
- Figura 4 – Gel desnaturante de agarose.** As amostras de RNA foram submetidas à eletroforese em gel desnaturante de agarose 1,2%, onde se observa as subunidades 18S e 28S do RNA ribossomal. A presença dessas subunidades é indicativa do grau de integridade da amostra.....60
- Figura 5 – Síntese de DNA complementar (cDNA).** A) Representação esquemática do RNA mensageiro com cauda poli-A ligada à esfera magnética e síntese de DNA complementar dupla-fita B) Representação esquemática da reação em cadeia polimerase para verificação da síntese do cDNA, utilizando primers para os genes *EF1* e *GAPDH* C) Gel de agarose 1% para verificação da síntese de cDNA: 1 - λ -Hind 2 – amplificação de 350pb do gene *EF* 3 – amplificação de 540pb do gene *GAPDH*.....62
- Figura 6 – Digestão com *Nla* III.** A) Representação esquemática da digestão enzimática do DNA complementar (cDNA) com *Nla* III B) Representação esquemática da reação em cadeia da polimerase para verificação da digestão com *Nla* III, utilizando *primers* para os genes *EF* e *GAPDH* C) Gel de agarose 1% para verificação da digestão com *Nla* III: 1 – λ -Hind 2 – amplificação de 540pb do gene *GAPDH* em amostra de cDNA não digerida com *Nla* III 3 – após a digestão, o sítio reconhecido pelo *primer* do gene *GAPDH* é perdido e por isso não ocorre amplificação 4 – amplificação de 350pb do gene *EF* em amostra de cDNA não digerida com *Nla*

III 5 – mesmo após a digestão o sítio reconhecido pelo *primer* do gene *EF* é mantido e por isso ocorre amplificação de 350pb63

Figura 7 – Ligação aos adaptadores. A) Representação esquemática da ligação dos adaptadores A ou B ao DNA complementar (cDNA) B) Representação esquemática da reação em cadeia da polimerase para verificação da ligação dos adaptadores, utilizando *primers antisense EF* e *sense DTP-1* ou *DTP-2* C) Gel de agarose 1% para verificação da ligação dos adaptadores: 1 – λ -Hind 2 – amplificação de 350pb utilizando-se os *primers antisense EF* e *sense DTP-1* em amostra de cDNA ligada ao adaptador A 3 – utilizando-se os *primers antisense EF* e *sense DTP-1* em amostra de cDNA ligada ao adaptador B não ocorre amplificação 4 - utilizando-se os *primers antisense EF* e *sense DTP-2* em amostra de cDNA ligada ao adaptador A não ocorre amplificação 5 - amplificação de 350pb utilizando-se os *primers antisense EF* e *sense DTP-2* em amostra de cDNA ligada ao adaptador B.....64

Figura 8 – Formação de tags – digestão com *BsmF I*. A) Representação esquemática da digestão do DNA complementar (cDNA) ligado aos adaptadores com a enzima *BsmF I* B) Representação esquemática da reação em cadeia da polimerase para verificação da digestão do cDNA, utilizando *primers antisense EF* e *sense DTP-1* ou *DTP-2* C) Gel de agarose 1% para verificação da digestão: 1 – amplificação de 350pb utilizando-se os *primers antisense EF* e *sense DTP-1* em amostra de cDNA ligada ao adaptador A, antes da digestão 2 - amplificação de 350pb utilizando-se os *primers antisense EF* e *sense DTP-2* em amostra de cDNA ligada ao adaptador B, antes da digestão 3 – após a digestão, utilizando-se os *primers antisense EF* e *sense DTP-1* em amostra de cDNA ligada ao adaptador A não ocorre amplificação 4- após a digestão, utilizando-se os *primers antisense EF* e *sense DTP-2* em amostra de cDNA ligada ao adaptador B não ocorre amplificação 5 - λ -Hind.....66

Figura 9 – Formação de *ditags*. Representação esquemática da reação de preenchimento de extremidades com a enzima *Klenow* polimerase e posterior ligação das *tags* para formar *ditags*.....67

Figura 10 – Amplificação e isolamento de *ditags*. A) Representação esquemática da amplificação das *ditags* pela reação em cadeia da polimerase (PCR) B) Gel de poliacrilamida 12% para teste da concentração de *ditags* a ser usada na PCR em larga escala. Como molde foram utilizados: 1 – água DEPC (controle negativo da PCR) 2 – controle negativo da reação de ligação 5 – amostra de *ditags* na diluição 1:50 6 – amostra de *ditags* na diluição 1:100 7 – amostra de *ditags* na diluição 1:200 8 – controle positivo I-SAGE; como marcadores de peso molecular foram utilizados: 3 – *Ladder* 25pb 4 – *Ladder* 50pb C) Géis poliacrilamida 12%: 2 – 10 – produtos da amplificação em larga escala das *ditags* e isolamento do fragmento de 100pb 1- *Ladder* 100pb.....68

Figura 11 – Clivagem dos adaptadores. A) Representação esquemática da digestão das *ditags* com *Nla III* B) Géis de poliacrilamida 12%: 1 – *Ladder* 25pb 2 – 4 – fragmentos de 100pb,

correspondente a *ditags* não digeridas; 60pb, resultantes de digestão parcial; 40pb, correspondentes aos adaptadores; e 26pb, correspondente as *ditags* clivadas, que foram isoladas do gel.....69

Figura 12 - Formação de concatâmeros. A) Representação esquemática da ligação das *ditags* para formação de concatâmeros B) Géis de poliacrilamida 8%: 3 – concatâmeros de tamanhos variados e isolamento daqueles de tamanhos de 300 a 500pb, de 500 a 800pb e de 800 a 1000pb; os marcadores de peso molecular 1 – *Ladder* 100pb e 2 – λ -*Hind* foram utilizados para orientar no isolamento dos fragmentos70

Figura 13 – Verificação de clonagem. Gel de agarose 1% para verificação da clonagem: 1 – λ -*Hind* 2 – 16 – amplificações obtidas por reação em cadeia da polimerase utilizando *primers M13 forward* e *reverse* e como molde, colônias resistentes ao antibiótico Zeocin™.....72

Figura 14 – Sequenciamento automático das bibliotecas. A) Representação esquemática de uma placa de cultura para a proliferação de bactérias contendo insertos de diferentes tamanhos B) Produtos da amplificação direta dos concatâmeros clonados no vetor pZero pela reação em cadeia da polimerase (PCR), onde 1- marcador de peso molecular λ -*Hind* 2 – 12 - fragmentos amplificados C) eletroferogramas gerados pela análise no *MEGA BACE 1000 DNA Analysis System*, a partir da amplificação dos fragmentos de DNA pela PCR73

Figura 16 – Temperatura de dissociação do produto amplificado. Curva de dissociação do produto amplificado do gene *SRGN*, gerada pelo programa *7500 Software v2.03* (Applied Biosystems, USA).86

Figura 17 – Curva padrão para a reação de amplificação. Curva padrão do produto amplificado do gene *SRGN*, gerada pelo programa *7500 Software v2.03* (Applied Biosystems, USA). 88

Figura 18 – Comparação de resultados obtidos por SAGE e por qRT-PCR. Representação gráfica de 8 genes com expressão gênica concordantes pelo método SAGE e pela reação em cadeia da polimerase em tempo real, em amostras de cultura de células CD34⁺ do paciente talassêmico β intermediário e maior, a partir da qual foi construída a biblioteca SAGE.89

Figura 19 – Comparação de resultados obtidos por SAGE e por qRT-PCR. Representação gráfica de 6 genes com expressão gênica discordantes pelo método SAGE e pela reação em cadeia da polimerase em tempo real, em amostras de cultura de células CD34⁺ do paciente talassêmico β intermediário e maior, a partir da qual foi construída a biblioteca SAGE.90

Figura 20 - Análise de cluster dos genes diferencialmente expressos. Diagrama de cores: branco, baixa expressão; vermelho, alta expressão. A intensidade das cores reflete a intensidade da expressão dos genes.....91

SUMÁRIO

DEDICATÓRIA	iv
AGRADECIMENTOS.....	ix
RESUMO	xi
ABSTRACT	xv
LISTA DE ABREVIACÕES	xix
LISTA DE TABELAS	xxi
SUMÁRIO	xxix
INTRODUÇÃO	35
CASUÍSTICA E ASPECTOS ÉTICOS	51
2.1 Pacientes talassêmicos β	53
2.2 Aspectos éticos da Pesquisa.....	53
MÉTODOS	55
1 - Culturas de células CD 34 ⁺ hematopoéticas de sangue periférico	57
1.1 Separação de células mononucleares	57
1.2 – Isolamento de células CD 34 ⁺ de sangue periférico.....	57
1.3 – Cultivo das células CD 34 ⁺	57
1.4 - Citometria de fluxo	58
1.5 – Extração de RNA	60
2- Análise da expressão gênica	61
2.1 – Serial Analysis of gene expression (SAGE)	61
2.1.1 - Separação do RNA mensageiro (mRNA).....	61
2.1.2 - Síntese de DNA complementar (cDNA)	61
2.1.3 - Digestão com Nla III	62
2.1.4 - Ligação aos adaptadores	63

2.1.5 - Formação de tags – digestão com BsmF I	65
2.1.6 - Formação de ditags	66
2.1.7 - Clivagem dos adaptadores	68
2.1.8 - Formação dos concatâmeros	69
2.1.9 - Clonagem em plasmídeo pZERO®-1	70
2.1.10 - Transformação de bactérias	71
2.1.11 - Sequenciamento automático das bibliotecas	72
2.1.12 - Análise de sequências	73
2.1.13 - Extração e contagem de tags	74
2.1.14 - Identificação das tags	74
2.1.15 - Critérios para identificação de tags diferencialmente expressas	74
2.1.16 - Perfil funcional dos transcritos	74
3 - PCR quantitativo em tempo real (qRT-PCR)	74
3.1 – Síntese de DNA complementar (cDNA)	74
3.2 – Princípios da qRT-PCR	75
3.3 – Padronizações	76
3.3.1 - Desenho dos primers	76
3.3.2 - Concentração de primer	76
3.3.3 - Eficiência de reação	77
3.3.4 - Especificidade da reação	77
3.4 – Reações	77
3.5 – Análise dos dados	78
RESULTADOS	79
1 – Bibliotecas SAGE	81
1.1 - Biblioteca SAGE INTERMEDIÁRIA	81
1.2 - Biblioteca SAGE MAIOR	82

2- Validação dos resultados de SAGE por qRT-PCR.....	83
2.1 - Síntese de DNA complementar (cDNA).....	83
2.2 - Padronizações da qRT-PCR	84
2.3 - Reações.....	88
3 - Expressão gênica diferencial: INT X MAIOR.....	90
DISCUSSÃO.....	96
CONCLUSÃO	104
REFERÊNCIAS BIBLIOGRÁFICAS	109
ANEXOS.....	117

INTRODUÇÃO

Talassemias

O termo hemoglobina foi descrito pela primeira vez em 1674 como sendo um pigmento de eritrócitos que apresenta a propriedade de se ligar à molécula de oxigênio. A molécula de hemoglobina é constituída de 4 sub-unidades, cada uma contendo um grupo heme, um derivado porfirínico contendo radical Fe^{+2} , ligado a uma cadeia globínica (STEINBERG *et al.*, 2001).

Nos pulmões onde a pressão de oxigênio (pO_2) é alta, cada molécula de hemoglobina se combina com 4 moléculas de O_2 , sendo uma molécula de O_2 para cada grupo heme, formando a oxihemoglobina. Esta combinação é reversível e o O_2 transportado é transferido para os tecidos, onde a pO_2 é baixa. A combinação de hemoglobina com CO_2 normalmente produzidos nos tecidos também é reversível e constitui a carboxihemoglobina. Este evento de ligação com O_2 e/ou CO_2 ocorre em células específicas denominadas hemácias, eritrócitos ou glóbulos vermelhos (HOFFMAN *et al.*, 1991).

Os genes das globinas localizam-se em 2 grupamentos (*cluster*): agrupamento α , com aproximadamente 30kb de DNA localizados no cromossomo 16 e agrupamento β , com aproximadamente 50kb de DNA localizados no cromossomo 11. A disposição gênica nestes agrupamentos apresenta grande importância na transcrição gênica, como podem ser visualizados na Figura 1:

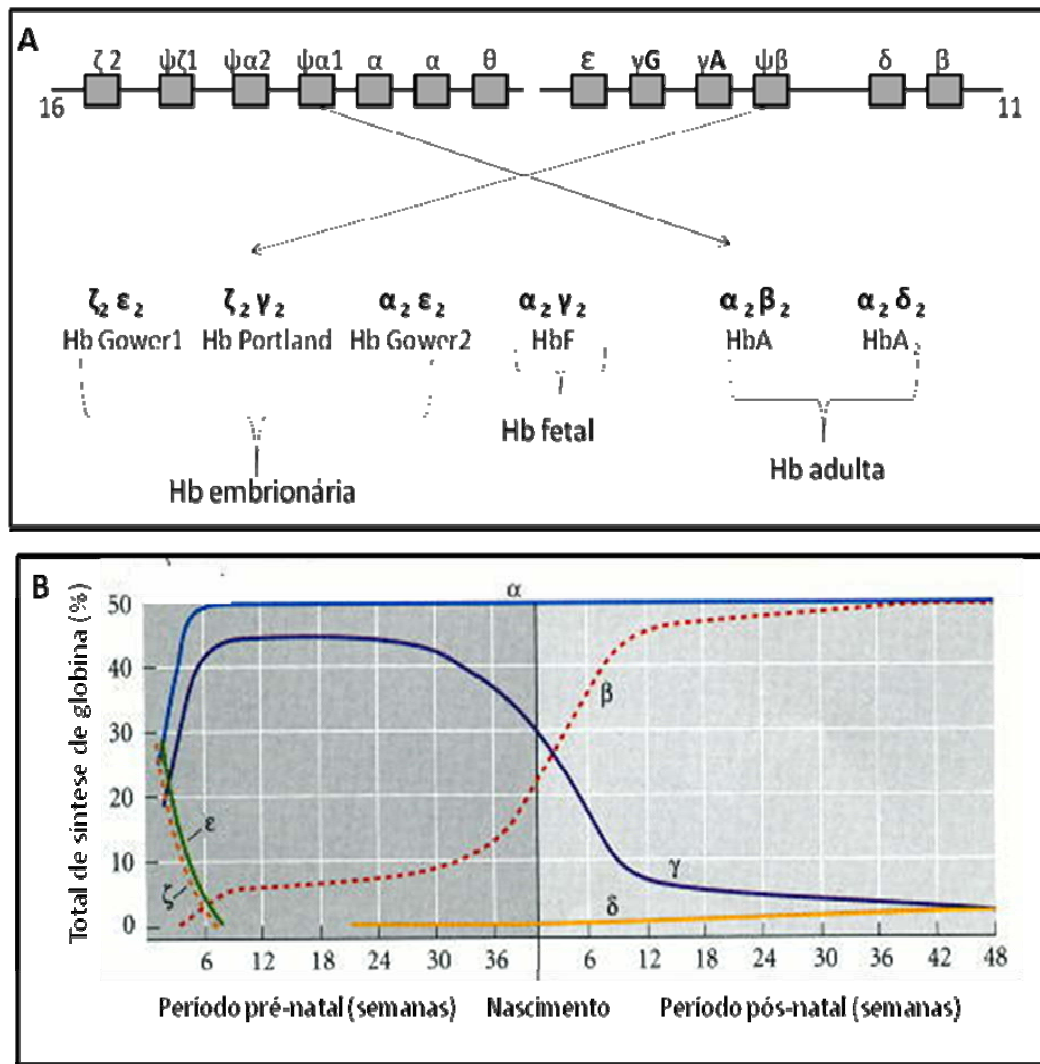


Figura 1 - Desenvolvimento ontogenético normal das cadeias globínicas - Na figura 1A podemos observar a composição e a disposição gênica dos cluster α e β . De acordo com a composição dos clusters temos a formação de vários tipos de hemoglobinas: as de predomínio embrionário, as presentes no período fetal e as que encontramos durante toda a vida adulta. O evento visualizado na figura 1B representa a troca sincronizada do predomínio do tipo de hemoglobina, fenômeno denominado “switching” (WEATHEDRALL *et al.*, 2001).

Alterações hereditárias na molécula de hemoglobina causam doenças denominadas hemoglobinopatias, que podem ser classificadas em: hemoglobinopatias estruturais, nas quais ocorrem mutações em genes de cadeias globínicas resultando na produção de proteínas anômalas ou talassemias, nas quais há redução ou ausência de síntese de um ou mais tipos de cadeias globínicas.

As talassemias compreendem um grupo heterogêneo de doenças hereditárias caracterizadas pela redução no ritmo de síntese de uma ou mais cadeias polipeptídicas da hemoglobina. Esta anormalidade, de natureza complexa, ocasiona a supressão total ou parcial da produção de cadeias α nas α -talassemias e de cadeias β nas β -talassemias (THEIN *et al.*, 1998), que são as mais conhecidas pela frequência e manifestações clínicas de considerável importância nos portadores, já que as cadeias α e β formam a hemoglobina A (BUNN e FORGET, 1986; HIGGS, 1993).

Daí advém a hemoglobinizacão deficiente e, conseqüentemente, microcitose e hipocromia características dessas síndromes. Além disso, a síntese de cadeias não afetada permanece inalterada provocando acúmulo e formação de agregados instáveis das cadeias não pareadas. A precipitação desses agregados provoca numerosos efeitos deletérios sobre a hemácia ou seus precursores (HOFFMAN *et al.*, 1991; STEINBERG, 2001). São essas lesões as responsáveis pela destruição prematura dos eritroblastos na medula óssea, caracterizando a eritropoese ineficaz comum a essas síndromes, e também pela reduzida sobrevivência dos eritrócitos na circulação, determinante do quadro hemolítico comum a essas síndromes (BANK, 1978; WEATHERALL *et al.*, 1972).

As β -talassemias são decorrentes de um grupo de alterações moleculares causadas por redução parcial ou completa da síntese de um ou mais cadeias de globina β , levando a uma menor produção de hemoglobina, originando dessa forma anemia de graus variados. Mutações no gene da globina β em heterozigose levam, com raras exceções, a uma anemia leve e assintomática, denominada β -talassemia menor. Mutações em homozigose, dependendo do grau de impedimento da síntese de globina, podem provocar anemia sintomática moderada (β -talassemia intermediária), que pode exigir transfusões de eritrócitos apenas esporadicamente ou formas graves (β -talassemia maior), que são sintomáticas e exigem regime transfusional mensal, na maioria dos casos. Mutações no gene da globina β que suprimem totalmente a síntese são chamadas de Beta Zero (β^0).

Mutações em que existe alguma síntese de cadeia de globina β são chamadas de Beta Mais (β^+) (THEIN, 1998).

Entretanto, a síntese normal de globinas α em pacientes com talassemia β resulta em desequilíbrio na síntese de cadeias globínicas. As cadeias de globina α livres são extremamente instáveis e precipitam formando corpos de inclusão. Estas inclusões de cadeias α são responsáveis pela destruição intramedular de precursores de eritrócitos denominada de eritropoese ineficaz, que caracterizam as talassemias β (WICKRAMASINGLE e HUGHES, 1980).

Os produtos da degradação das cadeias α em excesso, que são globina, heme, hemina (forma oxidada do heme) e ferro livre, provavelmente são responsáveis por grande parte dos danos à membrana de eritrócitos nestas síndromes. Células vermelhas de indivíduos portadores de talassemia β apresentam redução em sua vida-média, que são o reflexo das múltiplas lesões causadas pelo excesso de cadeias α (Figura 2).

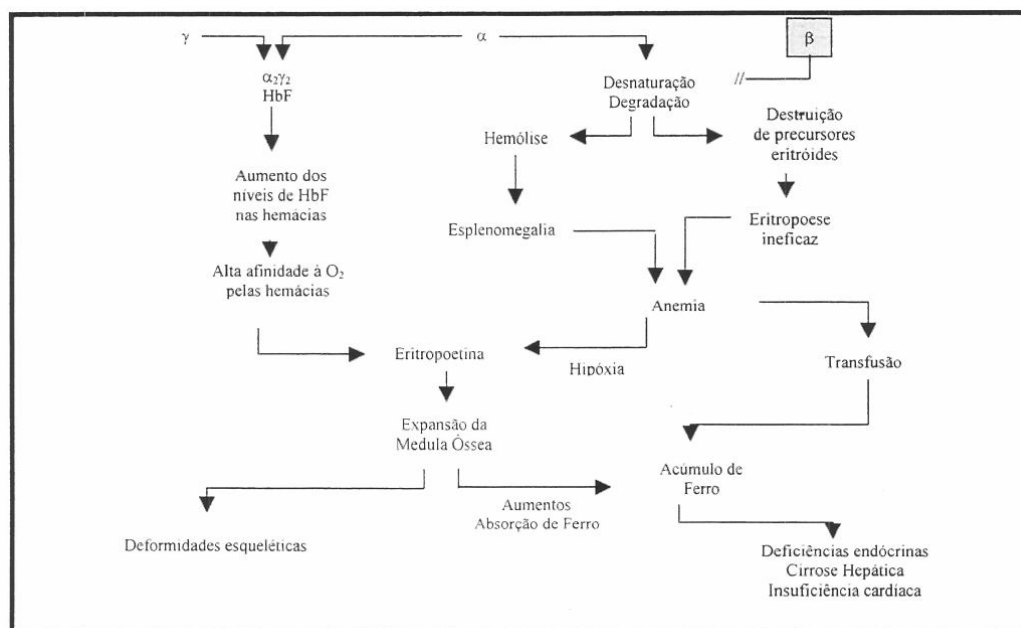


Figura 2 - Esquema sobre a fisiopatologia da talassemia β – As cadeias α globinas livres precipitam no organismo danificando a parede de precursores eritróides levando à eritropoese ineficaz e diminuição da vida média da hemácia (parte direita da figura). Este evento leva à produção de progenitores celulares, que ocasiona expansão da medula óssea e aumento na absorção de ferro. Em casos de anemia muito grave necessita-se de transfusão, situação na qual aumentam ainda mais os níveis de ferro no organismo e consequentemente sua absorção, levando à clínicas relacionadas com as talassemias β que são geralmente a causa dos óbitos dos pacientes. Em casos onde se observa aumento da produção de hemoglobina fetal (HbF) (lado esquerdo da figura), o paciente apresenta um quadro clínico melhor (WEATHERALL et al., 2001).

Formas clínicas da talassemia beta

O diagnóstico da talassemia β é realizado a partir de dados clínicos, laboratoriais e moleculares. É embasado fundamentalmente na gravidade da anemia que se apresenta microcítica e hipocrômica nas três formas de talassemias β : menor, intermediária e maior (CAO et al., 1997).

Talassemia β menor: também denominado traço talassêmico, são heterozigotos clinicamente assintomáticos e habitualmente não necessitam de tratamento apesar de apresentar microcitose, hipocromia e níveis de hemoglobina normal ou discretamente diminuídos. Embora o defeito possa ser detectado por exames laboratoriais específicos, é frequentemente confundida com anemia por carência de ferro (ZAGO, 2004).

Talassemia β intermediária: essa denominação reserva-se para casos sintomáticos que espontaneamente mantém níveis de hemoglobina de 6-9 g/dL, que não dependem de transfusões regulares (ZAGO, 2004).

Talassemia β maior: é resultante de homozigose ou heterozigose composta para as mutações nos genes da globina β , e corresponde à forma mais grave da doença, dependente de transfusão, com quadro clínico composto de anemia intensa, esplenomegalia, deformidades ósseas e graves alterações no crescimento, no desenvolvimento e na reprodução. As manifestações surgem durante o primeiro ano de vida com menor aumento de peso, episódios de febre, diarreia, apatia, irritabilidade e palidez (WEATHERALL e CLEGG, 1981). A maioria dos pacientes necessita de transfusões regulares, a cada 3 ou 4 semanas, seguidas de um regime de quelação do excesso de ferro. Este regime de transfusão/quelação deve ser iniciado o mais precocemente possível, para evitar deformidades ósseas e danos ao pâncreas, fígado, miocárdio de outros órgãos (CAPELLINI e COMINO, 2000).

No passado, o quadro de talassemia maior era um reflexo direto da falta de um regime transfusional regular, como também de uma quelação regular do excesso de ferro. Na ausência de tratamento, o quadro clínico se agrava progressivamente e a morte ocorre geralmente na primeira década de vida. Há anemia intensa ($Hb < 6,0\text{g/dL}$), esplenomegalia volumosa, atraso do crescimento, redução da massa muscular e alterações craniofaciais características (CAPELLINI e COMINO, 2000).

A diferenciação entre talassemia maior e intermediária é fundamental na prática clínica, pois define o tipo de tratamento que o paciente receberá. No entanto, essa distinção nem sempre é fácil, uma vez que o quadro clínico pode ser resultante de um grande número de combinações moleculares (WEATHERALL, 1990).

Bases moleculares da talassemia β

Os genes responsáveis pela síntese de cadeias da globina β e similares estão organizados num agrupamento ou *cluster*, localizado no braço curto do cromossomo 11 que contém 5 genes funcionais (5' $-\epsilon-\gamma^G-\gamma^A-\phi\beta-\delta-\beta-3'$), organizados na ordem de suas expressões durante o desenvolvimento do embrião e feto (THEIN, 1998).

Mais de 200 alelos da talassemia β foram caracterizados. Destes, 125 representam mutações pontuais em regiões funcionalmente importantes do gene, causadas por uma variedade de mecanismos (Figura 3) ou em sequências flanqueadoras do gene (THEIN, 1998).

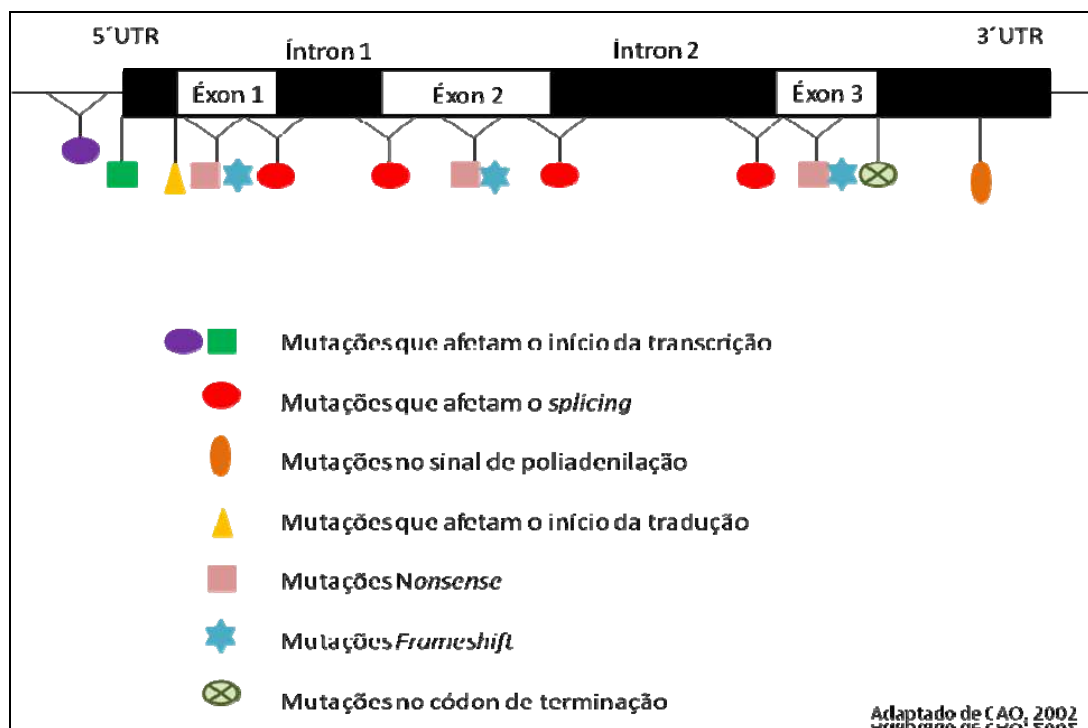


Figura 3 - Diferentes mecanismos moleculares no *cluster* da beta globina que causam talassemia.

Mutações nas regiões reguladoras (*CAT box* e *TATA box*) que antecedem aos genes, diminuem a eficiência da transcrição do RNA mensageiro (RNAm) por diminuir a ligação a fatores de transcrição, originando assim talassemias do tipo β^+ .

Mutações no sinal de poliadenilação dificultam a adição da cauda poli A ao RNAm. Esses transcritos são instáveis, mas a maior parte dos transcritos produzidos são normais.

Essas mutações geralmente levam a um fenótipo de talassemia β^+ leve (FORGET *et al.*, 2001; THEIN, 1998).

Aproximadamente a metade dos alelos beta talassêmicos produz talassemias β^0 . Muitos são *frameshifts*, ou seja, alteração na região codificante do gene beta. Essas mutações destroem a leitura normal e interferem na tradução do RNAm pela introdução de um códon de terminação prematura. Outras são mutações chamadas *nonsense*, isto é, substituição de bases que causam o aparecimento de um códon de terminação prematuro na produção de proteína, formando um polipeptídeo não funcional, como é o caso da troca de C→T no códon 39 da beta globina originando hemoglobinas muito instáveis e que são rapidamente degradadas, originando o fenótipo de talassemia β^0 . Cadeias globínicas truncadas podem ser produzidas como resultado de tais mutações, mas elas não são usualmente detectadas, presumivelmente porque são muito instáveis e rapidamente degradadas, originando fenótipo de talassemia β^0 (THEIN, 1998; FORGET *et al.*, 2001).

As mutações que afetam o processamento do RNA talvez sejam as de maior interesse funcional. O RNA inicialmente transcrito contém éxons e íntrons. A perda dos íntrons é essencial para formar um RNA funcional. Mutações nas uniões éxon-íntron (ou próximas a elas) impedem ou dificultam a retirada do íntron, originando talassemia β^0 ou β^+ , respectivamente. De fato, a análise desses genes β talassêmicos tem proporcionado minuciosa reflexão acerca das sequências de RNA requeridas para um *splicing* de RNA específico e eficiente. Mutações que afetam os invariáveis dinucleotídeos GT na extremidade 5' (doadora) e na AG na extremidade 3' (receptora), suprimem o *splicing* normal e produzem talassemia do tipo β^0 , onde nenhum RNA é encontrado. Como exemplo, podemos citar a mutação mediterrânea IVS-I-1 resultante da troca G→A no primeiro nucleotídeo do primeiro íntron que impede o processamento do RNA para retirar o íntron, impedindo a síntese de cadeias β (ZAGO, 2004; FORGET *et al.*, 2001).

Ao lado desses invariáveis dinucleotídeos existem sequências que estão bem conservadas e uma certa leitura consensual pode ser reconhecida. Mutações que ocorrem nessas sequências reduzem a eficácia do *splicing* em vários graus e produzem fenótipos de talassemia que variam de leves a graves. Por exemplo, a mutação na posição 5 do íntron I (IVS-I-5, G→C, T ou A) produz uma talassemia β^+ grave, enquanto que a substituição T→C no nucleotídeo seguinte, IVS-I posição 6, produz uma talassemia β^+ leve

(TREISMAN *et al.*, 1983). Esta última é geralmente citada como a forma portuguesa de talassemia (TAMAGNINI *et al.*, 1983) e faz parte da maioria das mutações leves encontradas em talassêmicos do Mediterrâneo (THEIN, 1998). Os indivíduos homozigotos IVS-I-6/IVS-I-6, em geral, tem quadro clínico, ou expressão fenotípica, compatível com a talassemia intermediária (COSTA *et al.*, 1991).

Algumas mutações no íntron ou na região codificante podem trazer um efeito inverso, criando um sítio alternativo de *splicing*. Cada molécula de RNA poderá então ser processada pela via normal (RNAm funcional) ou alternativamente pela via anômala (RNAm não funcional). Estas mutações são causadoras de talassemia β^+ , e a quantidade de cadeias β produzidas dependerá da proporção de moléculas de RNAm processadas pela via normal.

Um exemplo desse tipo de mutação é a substituição de A→G na posição 110 do íntron 1, que produz uma forma grave de talassemia β^+ , pois apenas 10-20% do RNAm são processados pela via normal. Essa mutação é uma das maiores causas de talassemia β^+ no Mediterrâneo (ZAGO, 2004; FORGET *et al.*, 2001).

Talassemia β Intermediária

Uma forma de talassemia β é clinicamente definida como intermediária, quando se revela como uma anemia crônica moderada com níveis de hemoglobina entre 6 e 9g/dL, que se mantém espontaneamente sem necessidade transfusional contínua e na presença de esplenomegalia. Geralmente, o início dos sintomas e o reconhecimento da anemia ocorrem mais tardiamente, em geral, após o segundo ano de vida (WEATHERALL, 1990). Portanto, a definição de talassemia intermediária subentende um diagnóstico clínico que envolve um amplo espectro de fenótipos em dois extremos, em que se distingue desde formas leves até formas graves. Formas leves podem permanecer assintomáticas e bem toleradas até a idade adulta, geralmente, se apresentam com marcada hepatoesplenomegalia e sinais de hiperesplenismo (HO *et al.*, 1998).

Entre os dois extremos da talassemia intermediária coloca-se um espectro de quadros clínicos de gravidade variável, que parecem depender basicamente de um mecanismo principal que é a eritropoese ineficaz, que por sua vez, é fruto do desequilíbrio da síntese de globinas α e β (HO *et al.*, 1998).

Baseado nesse presuposto fisiopatológico há uma tendência imediata em se pensar que um fenótipo intermediário seja ligado a vários genótipos talassêmicos. Nos últimos anos, com aquisição de conhecimentos moleculares, foram feitas várias tentativas de relacionar manifestações clínicas com mutações e outros arranjos moleculares que pudessem, de alguma forma, traduzirem-se na expressão do fenótipo (HO *et al.*, 1998). É o caso de certas mutações leves no gene β globínico (β^+ ou β^{++}) que se caracterizam por uma menor redução na síntese de globina β levando, conseqüentemente, a uma menor quantidade de cadeias α livres. Os exemplos com maior frequência estão relacionados com quadros de talassemia intermediária são as mutações -101 (C→T), -87 (C→T), -88 (C→T), -29 (A→G) e a IVS-I-6 (T→C) (FORGET, 2001).

A caracterização clínico-laboratorial da talassemia β intermediária é frequentemente dificultada pela grande variabilidade de sintomologia e dos dados hematológicos dos pacientes. A correlação entre a apresentação clínica e as mutações tem sido uma dificuldade para vários pesquisadores (CAO *et al.*, 1994). Por outro lado, a talassemia intermediária é uma forma clínica de menor gravidade que a talassemia maior, que resulta geralmente da combinação dos seguintes defeitos genéticos:

- Homozigose para genes β^+ talassêmicos de menor gravidade (como IVS-I-6);
- Combinação do gene β talassêmico grave com talassemia β^+ particularmente benigna (como β talassemia “silenciosa”);
- Associação de $\delta\beta$ com β^+ talassemia;
- Presença de um defeito adicional que reduz o excesso de cadeias α (como a co-herança de talassemia α) ou que aumenta a produção de cadeias γ (PHHF);
- Heterozigose para o gene talassêmico β particularmente grave. Em geral, o heterozigoto β talassêmico é assintomático, mas raramente o defeito é suficientemente grave para determinar manifestações clínicas;
- Heterozigose para gene talassêmico β com genes α extras;
- Talassemia β dominante (variantes de cadeia β hiperinstável).

Zago (2004), Thein (2005) e outros observaram considerável variação nos fenótipos clínicos resultantes dessas interações moleculares.

Fisiopatologia da talassemia β intermediária

As complicações clínicas da talassemia intermediária são decorrentes de três processos principais: eritropoese ineficaz, anemia crônica e sobrecarga de ferro. A gravidade clínica depende primariamente do defeito molecular no gene β . A diminuição da síntese de cadeia β leva a um desequilíbrio com as cadeias α . Excessos de cadeias α são extremamente instáveis e precipitam nos precursores eritróides da medula óssea, formando corpos de inclusão, que causam danos à membrana e morte celular, levando à eritropoese ineficaz (OLIVIERI, 1999). É importante enfatizar, que, quanto menor for o desequilíbrio da síntese das globinas, menor será a taxa de globinas α livres que precipitam dentro da hemácia e, conseqüentemente, menor será a taxa de eritropoese ineficaz, o grau de anemia e suas complicações clínicas (CAPPELLINI e COMINO, 2000). A eritropoese leva a uma tal expansão medular, que ocupa a matriz óssea e provoca osteoporose e deformidades ósseas. O estresse eritropoético a que a medula é constantemente imposta pode determinar compressão por massas eritropoéticas em vários locais, prevalentemente intravertebrais, mas também no fígado ou intra-raqueanas (CAMASHELLA e CAPPELLINI, 1995). O nível de eritropoese ineficaz é o determinante primário no desenvolvimento da anemia, levando a um aumento da absorção gastro-intestinal de ferro, resultando em sobrecarga de ferro, que pode ser a causa de várias complicações sérias, incluindo cardiopatia e anormalidades endócrinas, tais como diabetes mérito e hipogonadismo (TAHER *et al.*, 2006).

Fatores moduladores de gravidade na talassemia β

A notável diversidade fenotípica das talasemias β é protótipo de como o amplo espectro da gravidade da doença pode ser gerado, e está diretamente relacionada ao desequilíbrio na produção das cadeias α e β . Muitas das variações podem ser explicadas pelo heterogeneidade das lesões moleculares que afetam o gene da globina β , e que podem ser influenciadas por outros determinantes genéticos ligados ou não aos genes globínicos (THEIN, 2004).

A capacidade de prever o fenótipo a partir do genótipo tem importantes implicações clínicas, na triagem de portadores de talassemia β , no aconselhamento genético e diagnóstico pré-natal, e no planejamento de um regime de tratamento apropriado. A análise

do genótipo torna-se importante para se estabelecer o diagnóstico precoce de uma talassemia β^+ leve. O conhecimento da diversidade fenotípica da talassemia β tem avançado com análises das bases moleculares das formas de talassemia β e análises da relação genótipo/fenótipo na talassemia intermediária (THEIN, 2004). Entretanto, prever fenótipo pelo genótipo na talassemia intermediária é ainda difícil devido a fatores moduladores genéticos e ambientais (WEATHERALL, 2001).

Os modificadores genéticos primários são os diferentes alelos de talassemia β que causam completa ou marcada redução na síntese de cadeias β . Modificadores secundários são aqueles que têm um efeito direto no excedente de cadeias α , tais como: genótipo α globínico e determinantes genéticos que aumentam a produção de cadeias γ . Modificadores terciários são polimorfismos que ocorrem fora do *cluster* α e β , que envolvem osso, ferro, metabolismo da bilirrubina, entre outros e que podem influenciar na expressão clínica da doença. Os fatores ambientais incluem: condições sociais, nutrição e a disponibilidade de cuidados médicos (THEIN, 2004; FORGET *et al.*, 2001; CAO, 2002; TAHER *et al.*, 2006, GALANELLO e CAO, 1998).

Mutação β talassêmica

Com exceção de poucas deleções, a grande maioria das talassemias β é causada por mutações de ponto no gene ou em suas regiões flangeadoras, e podem afetar qualquer nível de regulação gênica (THEIN, 2004). O fator genotípico mais consistente que pode ser usado na predição do fenótipo é o tipo de alelo β talassêmico (HO *et al.*, 1998; WEATHERALL, 2001; CAO, 2002; THEIN, 2005).

Método da análise seriada da expressão gênica (SAGE)

O método SAGE, descrito por VELCULESCU *et al.* (1995), é baseado nos quatro princípios apresentados a seguir.

A obtenção de pequenas sequências de DNA complementar (cDNA), denominadas *tags*, com 9 a 21 pares de bases de uma região definida e específica de cada um do total de genes expressos no tipo celular avaliado.

A ligação das *tags* entre si, formando uma longa molécula, denominada concatêmero, que é clonada e sequenciada possibilitando, assim, a identificação de cada uma e suas *tags*

como pertencentes a genes previamente descritos em bancos de dados, bem como a sua caracterização como pertencentes a genes ainda não descritos, denominados *no matches*.

A quantificação da expressão dos genes, avaliada por meio do número de vezes que uma determinada *tag* é encontrada em um tipo celular obtido do caso a ser investigado comparado ao número de vezes em que é encontrado no mesmo tipo celular do controle. Assim, se um número significativamente maior de *tags* for observado em células do paciente em comparação ao número obtido em células do controle, é dito que o gene por ela representado tem expressão aumentada na doença. Se o número de *tags* for significativamente menor no paciente do que no encontrado no controle, é dito que o gene por ela representado tem expressão diminuída na doença. Ainda, tem expressão normal na doença o gene que apresentar um número de *tags* similar ao encontrado no controle. A identificação, a quantificação e a análise comparativa da expressão dos diferentes genes nos tipos celulares avaliados, anormais e normais, são realizadas por sistemas computadorizados próprios.

E, finalmente, para que os resultados da expressão gênica obtida no SAGE sejam aceitos como verdadeiros, a confirmação ou validação por outro método é necessária. Para tal, a expressão de um grupo de genes escolhidos aleatoriamente é também quantificada em células anormais e normais por qRT-PCR, método considerado como “*gold standard*” para a quantificação da expressão gênica.

Objetivo

Analisar a expressão gênica global de pacientes talassêmicos homozigotos para a mutação Cd 39 (C → T) com evoluções clínicas distintas (maior e intermediária) pelo método da análise seriada da expressão gênica (SAGE), buscando identificar genes diferencialmente expressos que possam estar relacionados ao fenótipo diferenciado entre os dois pacientes.

CASUÍSTICA E
ASPECTOS ÉTICOS

2.1 Pacientes talassêmicos β

Para a construção da biblioteca SAGE de talassemia β intermediária, foi selecionado um paciente atendido no ambulatório do Hemocentro da UNICAMP, diagnosticado como β talassêmico no primeiro ano de vida, portador da mutação Cd 39. Esse paciente não é dependente de transfusão. A amostra foi obtida na ocasião da coleta para exames de rotina e não fazia uso de medicamentos há pelo menos 2 meses.

Para a construção da biblioteca SAGE de talassemia β maior, foi selecionado um paciente portador da mesma mutação Cd 39, diagnosticado como talassêmico β no primeiro ano de vida, atendido no ambulatório do Hemocentro da UNICAMP e com dependência transfusional a cada 15 dias. A amostra foi obtida na ocasião da coleta para exames de rotina e antes do procedimento de transfusão de sangue.

Polimorfismos associados aos genes das globinas alfa e gama, bem como a existência de outros fatores associados foram avaliados, não mostrando diferença.

2.2 Aspectos éticos da Pesquisa

Este estudo foi submetido ao Comitê de Ética em Pesquisa da Faculdade de Ciências Médicas da UNICAMP (CEP) e à Comissão Nacional de Ética em Pesquisa (CONEP). Aprovado pelo CEP em 16 de março de 2004 durante a III Reunião Ordinária de CEP/FCM (registro CEP: 612/2003).

Todos os indivíduos foram informados sobre os procedimentos da coleta e seus riscos e assinaram um Termo de Consentimento Livre e Esclarecido (TCLE) para a autorização da mesma.

MÉTODOS

1 - Culturas de células CD 34⁺ hematopoéticas de sangue periférico

1.1 Separação de células mononucleares

Amostras de sangue periférico dos pacientes beta-talassêmicos foram coletadas em tudo contendo heparina lítica e processadas imediatamente após a coleta. As amostras foram diluídas 1:2 em solução PBS+ACD+BSA e em seguida colocadas em gradiente de Ficoll-Paque (Amersham Bioscience, Uppala Sweden) por 30min a 317g em temperatura ambiente. A camada formada na interface (*buffy coat*) foi coletada e as células lavadas em PBS+ACD+BSA. Em seguida, as células foram lisadas em tampão de lise para hemácias por 15min em gelo e centrifugadas por 5min a 361g. Novamente as células foram lavadas em PBS e ressuspensas para contagem em câmara de Neubauer.

1.2 – Isolamento de células CD 34⁺ de sangue periférico

Após a contagem das células, estas foram incubadas em *FcR Blocking Reagent* e *Hapten-Antibody* por 15min e após lavagem, incubadas com *Anti-Hapten MicroBeads*, segundo as orientações do fabricante (Miltenyi Biotec GmbH, Bergish Gladbach, Germany). Após o tempo de incubação, as células foram passadas por uma coluna magnética LS (Miltenyi Biotec GmbH, Bergish Gladbach, Germany) fixada em um suporte magnético. As células de interesse para o nosso estudo ficaram presas à coluna, enquanto as contaminantes (monócitos e linfócitos) foram coletadas em tubo falcon. Em seguida, a coluna foi retirada do suporte magnético e uma leve pressão foi realizada através de um êmbolo para fazer com que as células CD 34⁺ se soltassem da coluna e pudessem ser coletadas em um novo tubo. As células CD 34⁺ foram então lavadas com PBS+ACD+BSA por 10min a 203g e ressuspensas em 1mL de PBS+ACD+BSA para contagem em câmara de Neubauer. 1.10⁴ células de CD 34⁺ dos pacientes beta-talassêmicos foram marcada com fluoróforo RPE (*anti-mouse imunoglobulins RPE, Rabbit F (ab') 2*) para citometria de fluxo, para verificar a pureza das células obtidas. Lâminas de *cytospin* coradas com Leishman foram preparadas para verificação da morfologia celular. A viabilidade celular foi analisada por meio da técnica de azul de trypan, sendo superior a 98% em todos os pontos coletados.

1.3 – Cultivo das células CD 34⁺

As células CD 34⁺ foram cultivadas em metil celulose *Methocult™ H-4230* (Stem Cell Technologies Inc., Vancouver, British Columbia, Canadá). Para cada 1 mL de metil celulose

foram cultivadas 3.10^3 células de CD 34⁺, acrescidas de 50ng de *Stem Cell factor* (SCF; R&D Systems, Inc., Minneapolis, MN), 3 unidades de eritropoetina (Epo; Vetter Pharma Fertigung GmbH, Ravensburg, Alemanha) e 30 UI de interleucina 3 (IL-3; Pepro Tech México, S.A. de C.V.). A metil celulose foi pipetada em placa de 6 poços na concentração de 1mL por poço em atmosfera úmida no interior da placa. A placa foi incubada por 7 dias em estufa de CO₂ – 5% a 37°C (Fase I). Após sete dias, as colônias formadas por unidades formadoras de colônias eritróides (CFU-E) e proeritroblastos foram coletadas, lavadas, ressuspensas em meio *Minimum Essential Medium Alpha* (Alpha MEM, GIBCO™ Grand Island, N.Y., USA) e contadas em câmara de Neubauer. Nesse momento foram coletadas células para extração de RNA total, citometria de fluxo e citospin. O restante das células foi colocado novamente em cultivo, agora em fase líquida, tendo como base o meio Alpha MEM (Fase II). Na Fase II, utilizamos 30% de soro fetal bovino (SFB), 8% de albumina sérica bovina (BSA), 300mg/mL de holotransferrina humana (Calbiochem, USA), 10⁻⁵M de β-mercaptoetanol (Sigma Chemical CO, St. Louis, USA) e 2 unidades de Epo. As células foram cultivadas novamente em 5% CO₂ a 37°C (CROISILLE *et al.*, 1999).

Além do dia 10, foram coletadas células no dia 7 e 13 da Fase II para extração de RNA total. A diferenciação eritróide foi acompanhada por citometria de fluxo e morfologia por *citospin* corado com Leishman. A viabilidade celular da Fase I e II foi analisada por meio da técnica de azul de trypan.

1.4 - Citometria de fluxo

A técnica de citometria de fluxo consiste no reconhecimento de proteínas *in situ*, com um anticorpo marcado com um fluorocromo, comumente ficoeritrina (PE) ou fluoresceína isotiocianato (FITC). Estes fluorocromos ao serem estimulados por um feixe de laser emitem um fóton que é captado por sensores no aparelho. O citômetro de fluxo faz análises qualitativas (quantas células expressam esse produto) e quantitativas (quanto de certo produto é expresso por célula) da mesma amostra (OWENS *et al.*, 1995).

O processo de diferenciação eritróide foi acompanhado através da técnica de citometria de fluxo utilizando os seguintes anticorpos de superfície celular (Caltag Laboratories, Burlingame, CA):

- Anti-receptor de Transferrina – CD 71 (conjugado com FITC): marcador de células eritróides jovens;
- Anti-glicoforina A (conjugado com PE): marcador de células eritróides maduras.

Como controle negativo, foi utilizado um isotipo controle não relacionado (IgG, Caltag Laboratories, Burlingame, CA).

Foi utilizado um anticorpo de marcação citoplasmática: Anti-hemoglobina fetal (conjugado com FITC) para acompanhar a produção de hemoglobina fetal durante o processo de diferenciação celular. Antes da marcação com o anticorpo Anti-hemoglobina fetal foi realizada a permeabilização das membranas citoplasmáticas das células eritróides com o *Kit Fix & Perm* (Caltag Laboratories, Burlingame, CA) e em seguida marcação com anticorpo.

Outros anticorpos de marcação de superfície celular (Dako Cytomation Denmark A/S, Denmark) foram utilizados para verificar a presença de contaminação por outras linhagens celulares (mielóide e linfóide) durante a diferenciação, são eles:

- Anti-CD 3: anti linfócito T (clone UCHT1);
- Anti-CD 19: anti linfócito B (clone HD37);
- Anti-CD 14: anti-precursor de monócito (clone TUK4);
- Anti-CD 15: anti-precursor de granulócito (clone C3D-1).

Para todos os anticorpos de superfície celular citados anteriormente, a concentração de células utilizadas variou de 1.10^4 a 1.10^5 células por tubo num volume final de 100µL em PBS. As células foram então incubadas com 3-5µL dos anticorpos por 30min a 4°C, na ausência de luz. Em seguida, estas células foram lavadas com 500µL de PBS, centrifugadas a 203g por 5min. O sobrenadante foi descartado e o precipitado ressuspensão em 500µL de PBS. As células foram conservadas a 4°C, ao abrigo da luz até o momento da aquisição pelo citômetro. Já as células marcadas com anticorpo anti-hemoglobina fetal tiveram primeiramente suas membranas fixadas por 15min pela Solução A e após lavagem com PBS, foi adicionado Solução B para permeabilização das mesmas. Depois disso, o anticorpo foi adicionado num volume de 3µL, incubado por 30min a 4°C, na ausência de luz. As células foram lavadas com PBS e o pellet ressuspensionado em 500µL. As células foram conservadas a 4°C, ao abrigo da luz até o momento da aquisição pelo citômetro.

As células foram analisadas a 488nm em um citômetro *FACScalibur Becton-Dickinson*. Gráficos tipo *Dot plots* foram utilizados para análise dos dados obtidos. A intensidade da

fluorescência de cada célula foi comparada com as células incubadas com um controle isotipo (IgG).

1.5 – Extração de RNA

As amostras de RNA foram extraídas com kit comercial – *Rneasy Mini* e *Macro Kit* (Qiagen GmbH, Hilden) segundo as instruções do fabricante. Optamos pelo uso do kit, uma vez que a quantidade de células disponíveis não era abundante. Além disso, a pureza e a qualidade do RNA eram fundamentais para o sucesso da transcrição destas amostras.

A integridade das amostras foi verificada por eletroforese em gel desnaturante de agarose 1,2%. As amostras íntegras apresentaram duas subunidades ribossomais: 18S e 28S (Figura 4). Após a eletroforese, as amostras de RNA foram armazenadas em freezer -80°C.

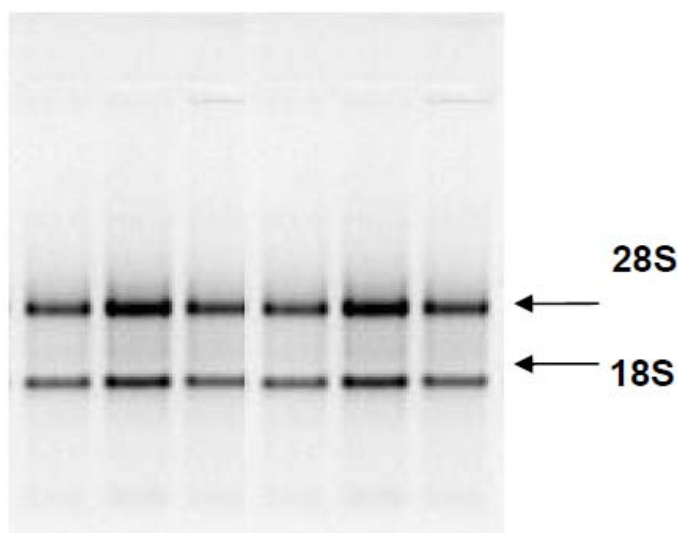


Figura 4 – Gel desnaturante de agarose. As amostras de RNA foram submetidas à eletroforese em gel desnaturante de agarose 1,2%, onde se observa as subunidades 18S e 28S do RNA ribossomal. A presença dessas subunidades é indicativa do grau de integridade da amostra

2- Análise da expressão gênica

2.1 – Serial Analysis of gene expression (SAGE)

Técnica utilizada para análises globais de expressão gênica. O protocolo foi realizado segundo recomendações do fabricante (I-SAGE™ Kit – Invitrogen, Life Technologies). Informações também estão disponíveis nos sites www.invitrogen.com, www.sagenet.org, www.ncbi.nlm.nih.gov/SAGE/.

2.1.1 - Separação do RNA mensageiro (mRNA)

Foi iniciado com a separação do mRNA a partir do RNA total do extraído do 10º dia de cultura de células dos dois pacientes portadores de beta-talassemia, feita por meio de interação de suas caudas poli-A a oligos dT adaptados a esferas magnéticas. Utilizando uma estante magnética, o mRNA ligado às esferas foi atraído para a parede do tubo, sendo o restante da solução descartado.

2.1.2 - Síntese de DNA complementar (cDNA)

Para a síntese da primeira fita de cDNA, as esferas foram ressuspensas na seguinte reação: 1X *First Strand Buffer* (50mM Tris-HCl, pH 8,3; 75mM KCl; 3mM MgCl₂), 10mM DTT, 500µM dNTPs, 40U de RNase OUT RT™ (Invitrogen, Life Technologies, USA) e 600U de Superscript II RT™ (Invitrogen, Life Technologies, USA), em volume final de 90µL. A reação foi realizada por 1h a 37°C.

Para a síntese da segunda fita de cDNA, foi realizada a seguinte reação: 90µL do produto da reação anterior, 1X *Second Strand Buffer* (20mM Tris-HCl, pH 6,9; 90mM KCl; 23mM MgCl₂; 13uM β-NAD⁺; 10mM sulfato de amônia), 200µM dNTPs, 50U de *E. coli* DNA ligase, 200U de *E. coli* DNA polimerase e 10U de *E. coli* RNase H, em volume final de 750µL (Figura 5). A reação foi realizada por 2h a 16°C, sendo paralisada com 28,3mM de EDTA.

A verificação da síntese de cDNA foi feita por meio de PCR, com a utilização de *primers* para os genes *EF* e *GAPDH* (Figura 5). A reação realizada foi: 1X *BV Buffer* (33mM sulfato de amônia; 134mM Tris-HCl, pH 8,8; 15,4mM MgCl₂; 200mM β-mercaptoetanol, 3µL de DMSO, 1mM dNTPs, 200ng de *primers antisense GAPDH ou E*, 5U de *Taq* polimerase e 0,5µL de amostra de cDNA em volume final de 50µL. O programa foi iniciado por 1min/95°C, seguido de 30 ciclos: 30s/95°C – 30s/55°C – 2min/72°C, finalizando com 5min/72°C. Foram geradas

amplificações de 350 a 540 pb, respectivamente para os genes *EF* e *GAPDH*, visualizadas em gel de agarose 1% (Figura 5).

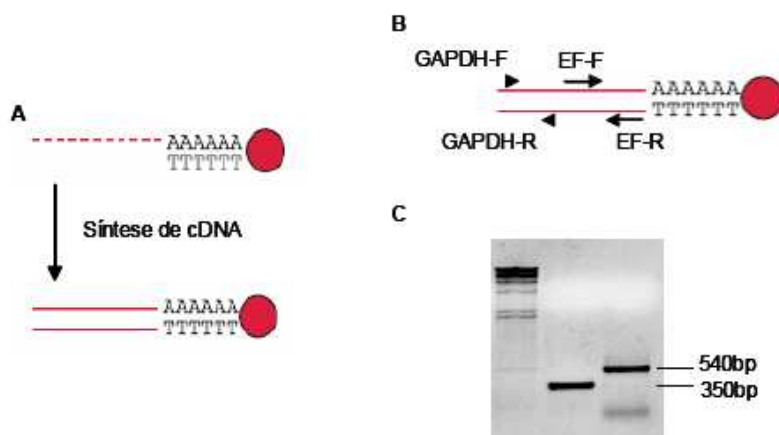


Figura 5 – Síntese de DNA complementar (cDNA). A) Representação esquemática do RNA mensageiro com cauda poli-A ligada à esfera magnética e síntese de DNA complementar dupla-fita B) Representação esquemática da reação em cadeia polimerase para verificação da síntese do cDNA, utilizando *primers* para os genes *EF* e *GAPDH* C) Gel de agarose 1% para verificação da síntese de cDNA: 1 - λ -Hind 2 – amplificação de 350pb do gene *EF* 3 – amplificação de 540pb do gene *GAPDH*

2.1.3 - Digestão com *Nla* III

A digestão do cDNA com enzima *Nla* III, que reconhece a sequência repetitiva CATG, normalizou o tamanho das moléculas de cDNA presentes na amostra. Após a digestão, estas apresentaram cerca de 256 pb e permaneceram ligadas às esferas magnéticas (Figura 6). Para a digestão com *Nla* III, as esferas foram ressuspensas na seguinte reação: 172 μ L de LoTE (3mM Tris-HCl, pH 7,5; 0,2mM EDTA, pH 7,5), 1X BSA (100 μ g/mL), 1X *Buffer* 4 (20mM Tris-acetato, pH 7,9; 10mM acetato de magnésio; 50mM acetato de potássio; 1mM DTT) e 60U de *Nla* III, em volume final de 200 μ L. A reação foi realizada por 1h a 37°C.

A digestão com *Nla* III pode ser verificada por PCR (reação idêntica à utilizada para verificação da síntese de cDNA) pela perda do sítio reconhecido pelo *primer* do gene *GAPDH*, mas não do gene *EF*. (Figura 6)

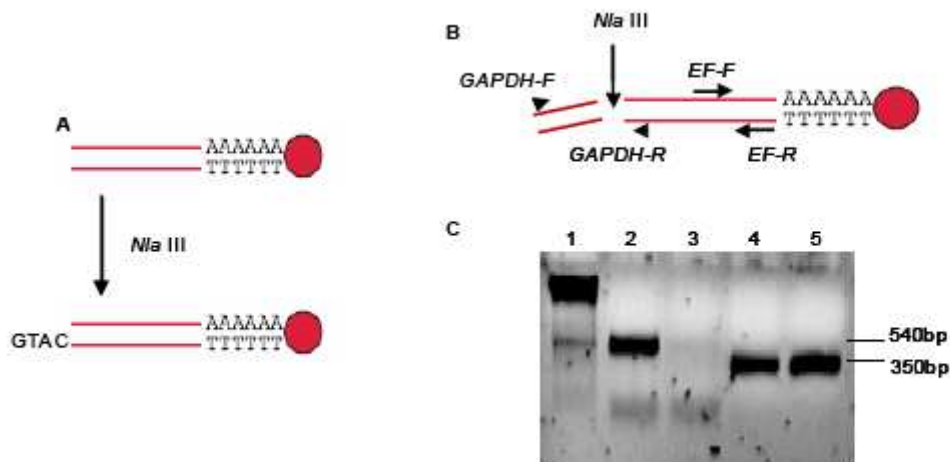


Figura 6 – Digestão com *Nla* III. A) Representação esquemática da digestão enzimática do DNA complementar (cDNA) com *Nla* III B) Representação esquemática da reação em cadeia da polimerase para verificação da digestão com *Nla* III, utilizando *primers* para os genes *EF* e *GAPDH* C) Gel de agarose 1% para verificação da digestão com *Nla* III: 1 – λ -Hind 2 – amplificação de 540pb do gene *GAPDH* em amostra de cDNA não digerida com *Nla* III 3 – após a digestão, o sítio reconhecido pelo *primer* do gene *GAPDH* é perdido e por isso não ocorre amplificação 4 – amplificação de 350pb do gene *EF* em amostra de cDNA não digerida com *Nla* III 5 – mesmo após a digestão o sítio reconhecido pelo *primer* do gene *EF* é mantido e por isso ocorre amplificação de 350pb

2.1.4 - Ligação aos adaptadores

A amostra foi, a seguir, dividida em duas, para ligação aos adaptadores A ou B, que apresentam extremidades coesivas às geradas no cDNA pela digestão com *Nla* III (Figura 7). Cada uma das amostras de esferas magnéticas foi ressuspensa na seguinte reação: 14 μ L LoTE, 60ng de adaptador A ou B, 1X *Ligase Buffer* (6mM Tris-HCl, pH 7,5; 6mM MgCl₂; 5mM NaCl, 100 μ g/mL BSA; 7mM β -mercaptoetanol; 10 μ M ATP; 2mM DTT; 1mM *spermidin*) e 10 unidades *Weiss* de T4 DNA ligase, em volume final de 20 μ L. A reação foi realizada por 2h a 16°C.

A verificação da ligação do cDNA aos adaptadores foi feita por PCR, utilizando os *primers antisense EF* e *sense DTP-1*, que reconhece sítio localizado no adaptador A ou *sense DTP-2*, que reconhece sítio localizado no adaptador B (Figura 7). A reação realizada foi: 1X *BV Buffer* (33mM sulfato de amônia; 134mM Tris-HCl, pH 8,8; 15,4mM MgCl₂; 200mM β-mercaptoetanol, 3μL de DMSO, 1mM dNTPs, 200ng de *primers antisense EF* e 350ng de *primer sense DTP-1* ou *DTP2*, 5U de *Taq* polimerase e 0,5μL de amostra de cDNA em volume final de 50μL. O programa de PCR utilizado foi o mesmo descrito anteriormente.

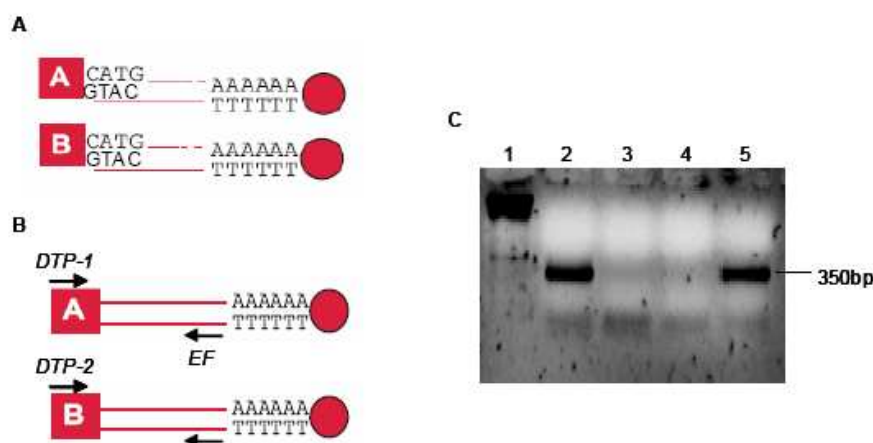


Figura 7 – Ligação aos adaptadores. A) Representação esquemática da ligação dos adaptadores A ou B ao DNA complementar (cDNA) B) Representação esquemática da reação em cadeia da polimerase para verificação da ligação dos adaptadores, utilizando *primers antisense EF* e *sense DTP-1* ou *DTP-2* C) Gel de agarose 1% para verificação da ligação dos adaptadores: 1 – λ-Hind 2 – amplificação de 350pb utilizando-se os *primers antisense EF* e *sense DTP-1* em amostra de cDNA ligada ao adaptador A 3 – utilizando-se os *primers antisense EF* e *sense DTP-1* em amostra de cDNA ligada ao adaptador B não ocorre amplificação 4 - utilizando-se os *primers antisense EF* e *sense DTP-2* em amostra de cDNA ligada ao adaptador A não ocorre amplificação 5 - amplificação de 350pb utilizando-se os *primers antisense EF* e *sense DTP-2* em amostra de cDNA ligada ao adaptador B

2.1.5 - Formação de tags – digestão com *BsmF I*

As duas amostras de cDNA, ligadas aos adaptadores A ou B, foram digeridas com enzima *BsmF I*. Esta enzima reconhece sítio localizado nos adaptadores e digeriu a fita de cDNA em região distante de 10 a 14 pares de bases do sítio de reconhecimento. Assim, foram obtidos fragmentos de 10-14pb específicos para cada molécula de cDNA, denominados *tags*, ligados ao adaptador A ou B (Figura 8). Para tanto, as esferas ligadas ao adaptador A ou B foram ressuspensas na seguinte reação: 174µL de LoTE, 1X *Buffer 4*, 2X BSA e 4U de *BsmF I*, em volume final de 200µL. A reação foi realizada por 1h a 65°C. Neste ponto, as esferas magnéticas foram descartadas e os sobrenadantes, contendo as *tags* ligadas aos adaptadores, foram recuperados. Para purificação, adicionamos aos sobrenadantes um volume igual ao de fenol-clorofórmio. Após centrifugação (14000rpm, 2min, temperatura ambiente), a fase aquosa foi transferida para *ependorf* novo. Para precipitação do cDNA, foram adicionados 2/3 volumes de 7,5M acetato de amônio, 5µL de glicogênio e 1mL de etanol absoluto, a amostra foi incubada por 20min em banho de gelo seco com etanol, seguido de centrifugação (14000rpm, 40min, 4°C). Após duas lavagens com etanol 75%, o cDNA precipitado foi eluído em 10µL de LoTE.

A verificação da digestão com *BsmF I* pode ser feita por PCR (reação idêntica à utilizada para verificação da ligação dos adaptadores), pela perda do sítio reconhecido pelo *primer* do *EF* (Figura 8).

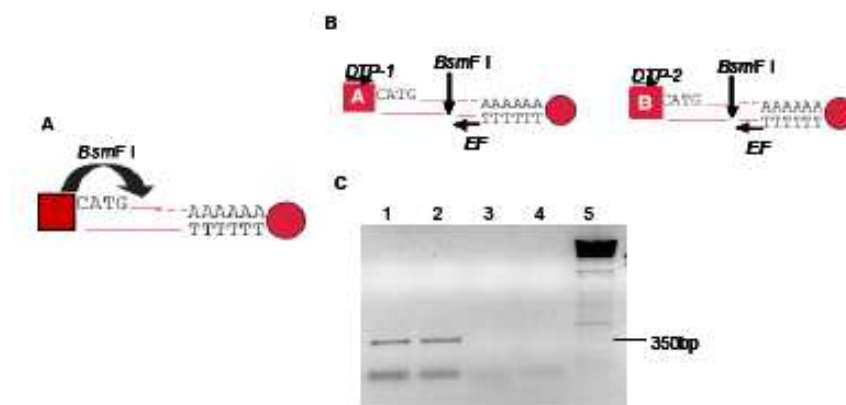


Figura 8 – Formação de tags – digestão com *BsmF I*. A) Representação esquemática da digestão do DNA complementar (cDNA) ligado aos adaptadores com a enzima *BsmF I* B) Representação esquemática da reação em cadeia da polimerase para verificação da digestão do cDNA, utilizando *primers antisense EF* e *sense DTP-1* ou *DTP-2* C) Gel de agarose 1% para verificação da digestão: 1 – amplificação de 350pb utilizando-se os *primers antisense EF* e *sense DTP-1* em amostra de cDNA ligada ao adaptador A, antes da digestão 2 - amplificação de 350pb utilizando-se os *primers antisense EF* e *sense DTP-2* em amostra de cDNA ligada ao adaptador B, antes da digestão 3 – após a digestão, utilizando-se os *primers antisense EF* e *sense DTP-1* em amostra de cDNA ligada ao adaptador A não ocorre amplificação 4- após a digestão, utilizando-se os *primers antisense EF* e *sense DTP-2* em amostra de cDNA ligada ao adaptador B não ocorre amplificação 5 - λ -Hind

2.1.6 - Formação de ditags

As extremidades deixadas pela digestão com *BsmF I* foram preenchidas por ação da enzima *Klenow* polimerase (Figura 9) na seguinte reação: 10 μ L da amostra de cDNA ligada ao adaptador A ou B, 1X *Klenow Buffer* (50mM Tris-HCl, pH 8,0; 10mM MgCl₂; 50mM NaCl), 0,5X BSA, 500 μ M dNTPs e 3-9U de *Klenow* polimerase, em volume final de 50 μ L, realizada por 30min, 37°C.

A seguir, as alíquotas contendo as *tags* ligada aos adaptadores A e ligada aos adaptadores B foram combinadas. Após purificação e precipitação, apenas 2/3 da amostra foram utilizados para formar a chamada *ditag*, por ação da enzima T4 DNA ligase (Figura 9). Para tanto, 1,5µL do seguinte *mix* de reação foi adicionado: 1mM Tris-HCl, pH 7,5, 2X *Ligase Buffer* e 4U *Weiss* de T4 DNA ligase, em 3,75µL de volume final. Como controle negativo da ligação, 1/3 de amostra restante foi adicionado a 1,5µL do *mix*: 600µM Tris-HCl, pH 7,5, 2X *Ligase Buffer*, em 3,75µL de volume final. A reação foi realizada por 16h a 16°C.



Figura 9 – Formação de *ditags*. Representação esquemática da reação de preenchimento de extremidades com a enzima *Klenow* polimerase e posterior ligação das *tags* para formar *ditags*

As *ditags* obtidas na reação de ligação foram submetidas à amplificação por PCR em larga escala (Figura 10). Para isto, foi verificada anteriormente, qual diluição de amostra seria adequada para ser utilizada com molde, tendo como padrão a amplificação de controle positivo I-SAGE. As reações realizadas foram: 1X *BV Buffer*, 3µL DMSO, 1,5mM dNTPs, 350ng de *primer DTP-1*, 350ng de *primer DTP-2*, 2,5U *Platinum® Taq* polimerase e 1µL das *ditags* nas diluições 1:50 ou 1:100 ou 1:200, ou 1µL do controle positivo I-SAGE, ou 1µL do controle negativo da ligação, ou 1µL de água DEPC (controle negativo da PCR), em volume final de 50µL. Os produtos foram aplicados em gel de poliacrilamida 12% (Figura 10). A diluição adequada encontrada foi aquela cujo produto amplificado mais se aproximou em intensidade ao controle positivo I-SAGE. O programa foi iniciado por 2min/95°C, seguido de 27 ciclos: 30s/95°C – 1min/55°C – 1min/70°C, finalizando com 5min/70°C.

Após a amplificação das *ditags* por PCR em larga escala (300 reações), os produtos foram purificados utilizando fenol-clorofórmio seguido de precipitação com 7,5M acetato

de amônio, glicogênio e etanol absoluto. Ressuspensão em LoTE, tais produtos foram aplicados em gel de poliacrilamida 12% e submetidos à eletroforese. O fragmento de 100pb, correspondente a *ditag* ladeada pelos adaptadores A e B, foi isolado do gel (Figura 10). A seguir, foi eluído em solução 5V LoTE, 1V 7,5M acetato de amônio, purificados em colunas S.N.A.P.TM, seguido de fenol-clorofórmio, sendo ao final precipitado com etanol, 7,5M acetato de amônio e glicogênio e eluído em 126μL de LoTE.

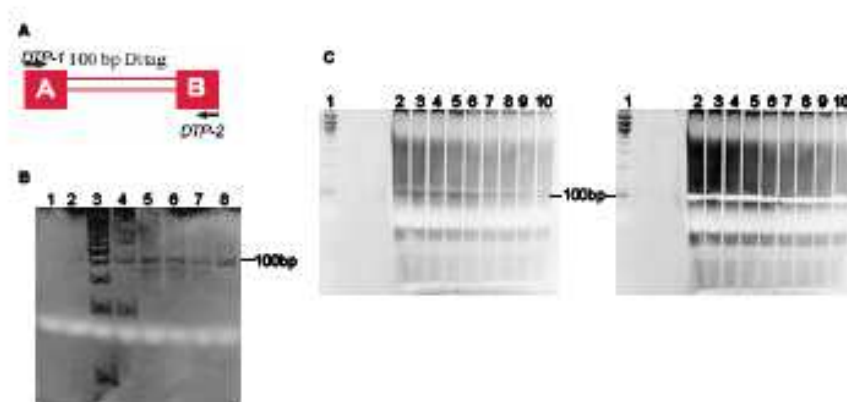


Figura 10 – Amplificação e isolamento de *ditags*. A) Representação esquemática da amplificação das *ditags* pela reação em cadeia da polimerase (PCR) B) Gel de poliacrilamida 12% para teste da concentração de *ditags* a ser usada na PCR em larga escala. Como molde foram utilizados: 1 – água DEPC (controle negativo da PCR) 2 – controle negativo da reação de ligação 5 – amostra de *ditags* na diluição 1:50 6 – amostra de *ditags* na diluição 1:100 7 – amostra de *ditags* na diluição 1:200 8 – controle positivo I-SAGE; como marcadores de peso molecular foram utilizados: 3 – Ladder 25pb 4 – Ladder 50pb C) Géis poliacrilamida 12%: 2 – 10 – produtos da amplificação em larga escala das *ditags* e isolamento do fragmento de 100pb 1- Ladder 100pb

2.1.7 - Clivagem dos adaptadores

A enzima *Nla* III foi utilizada para clivagem dos adaptadores, gerando *ditags* de 26pb, ladeadas pela sequência de reconhecimento desta enzima, CATG (Figura 11). Para realizar a reação foram divididos os 126μL de amostra de cDNA (100pb) em 3 aliquotas de 42μL. A reação realizada foi: 42μL de cDNA, 1X Buffer 4, 1,3X BSA e 120U de *Nla* III,

em volume final de 200 μ L, por 2h a 37°C. A seguir, a amostra foi purificada com fenol-clorofórmio e o cDNA, precipitado com descrito anteriormente.

Os produtos da digestão foram aplicados em gel de poliacrilamida 12% e submetidos à eletroforese. Puderam ser visualizados fragmentos de 100pb, correspondente a *ditags* não digeridas, 60pb, resultante de digestão parcial, 40pb, correspondente aos adaptadores, e 26pb, correspondente às *ditags* clivadas, que foram então isolados (Figura 11). A eluição, purificação e precipitação do cDNA foram realizadas como descrito anteriormente, sendo a amostra ressuspensa em 7,75 μ L de LoTE.

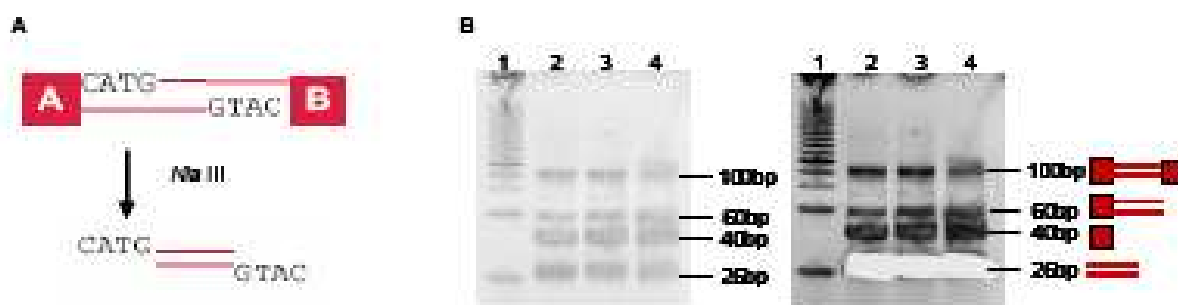


Figura 11 – Clivagem dos adaptadores. A) Representação esquemática da digestão das *ditags* com *Nla* III B) Géis de poliacrilamida 12%: 1 – Ladder 25pb 2 – 4 – fragmentos de 100pb, correspondente a *ditags* não digeridas; 60pb, resultantes de digestão parcial; 40pb, correspondentes aos adaptadores; e 26pb, correspondente as *ditags* clivadas, que foram isoladas do gel.

2.1.8 - Formação dos concatâmeros

As *ditags* purificadas foram ligadas umas as outras com a enzima T4 DNA ligase para formação de concatâmeros (Figura 12). A reação realizada foi: 7,75 μ L de cDNA (26pb), 1X *Ligase Buffer* e 5 unidades *Weiss* de T4 DNA ligase, em volume final de 10 μ L, sendo realizada por 3h a 16°C.

Em seguida, a amostra foi submetida à eletroforese em gel de poliacrilamida 8%. Foi feita então uma seleção por tamanho, isolando do gel fragmentos de 300 a 500pb, de 500 a 800pb e de 800 a 1000pb (Figura 12). A eluição, purificação e precipitação do cDNA foram realizadas como descrito anteriormente, sendo as amostras ressuspensas em 6 μ L de LoTE.

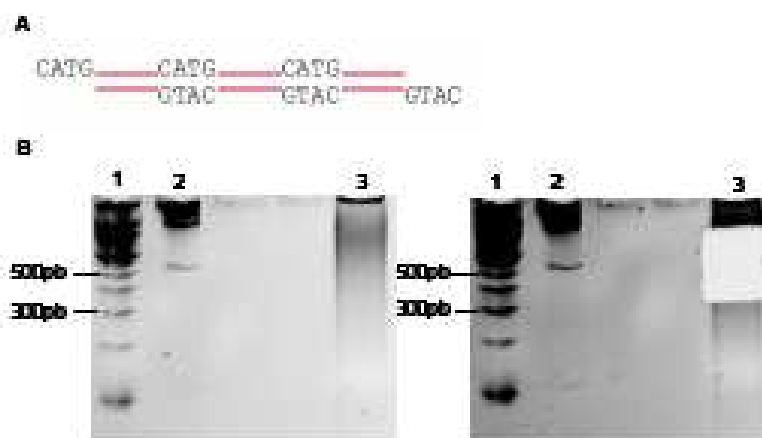


Figura 12 - Formação de concatâmeros. A) Representação esquemática da ligação das *ditags* para formação de concatâmeros B) Géis de poliacrilamida 8%: 3 – concatâmeros de tamanhos variados e isolamento daqueles de tamanhos de 300 a 500pb, de 500 a 800pb e de 800 a 1000pb; os marcadores de peso molecular 1 – *Ladder* 100pb e 2 – λ -*Hind* foram utilizados para orientar no isolamento dos fragmentos

2.1.9 - Clonagem em plasmídeo pZERO®-1

Para clonagem dos concatâmeros, o plasmídeo pZERO®-1 foi previamente linearizado utilizando a enzima *Sph* I. A reação utilizada foi: 2µg de pZERO®-1, 1X *Buffer* 2 (10mM Tris-HCl, pH 7,9; 1mM MgCl₂; 5mM NaCl; 100µM DTT) e 7U de *Sph* I, em volume final de 25µL, por 25min a 37°C. A seguir foi feita a purificação com fenol-clorofórmio e precipitação de cDNA conforme descrito anteriormente, sendo ressuspensão em 60µL de LoTE.

Para clonagem, a reação utilizada foi: 6µL de cDNA (concatâmeros), 1µL de pZERO®-1 digerido com *Sph* I, 1X *Ligase Buffer* e 8 unidades *Weiss* de T4 DNA ligase, em volume final de 10µL. Como controles negativos, foram feitas reações sem DNA (os 6µL de amostra foram substituídos por água estéril em reação idêntica a descrita) e sem ligase (os volumes de amostra e de enzima foram substituídos por água estéril em reação idêntica a descrita). Após a purificação e precipitação as amostras foram ressuspensas em LoTE.

2.1.10 - Transformação de bactérias

A transformação de bactérias *One Shot® TOP10 Eletrocomp™ E. coli* foi feita por eletroporação com a utilização de cubetas de 0,1cm e eletroporador *Gene Pulser II* (Bio-Rad Laboratories, Hercules, CA, USA), ajustado para 1,8 kV, 25mF de capacitância e 200 ohms de resistência, utilizando 2µL de amostra da biblioteca construída por alíquota de 40µL de bactérias.

Imediatamente após a eletroporação, as bactérias foram incubadas em meio líquido SOB (20g de triptona/L, 5g de extrato de levedura/L, 0,5g NaCl/L, 2,5mM KCl, 10mM MgCl₂, pH 7,0) a 37°C por 1h, sendo a seguir plaqueadas em meio LB *Low Salt* (10g triptona/L, 5g extrato de levedura/L, 5g NaCl/L), 2% ágar, contendo 50µg de antibiótico Zeocin™/mL e incubadas por 16h a 37°C.

Uma característica importante do vetor pZERO®-1 é a existência do gene “letal” *ccdB*. Quando ocorre a religação do vetor sem inserto, este gene é transcrito e a proteína traduzida, interferindo no mecanismo de ação da enzima bacteriana essencial DNA girase, levando à quebra e morte celular. Assim, apenas bactérias contendo plasmídeos com insertos crescem nas placas de cultura.

Para verificação da clonagem, foi realizada a seguinte PCR: 1X tampão *Taq* polimerase, 1,66mM MgCl₂, 125 µM dNTPs, 0,12pmol *primer M13 forward* (5' CGCCAGGGTTTCCCAGTCACGAC 3'), 0,12pmol *primer M13 reverse* (5' TTTCACACAGGAAACAGCTATGAC 3') e 1U de *Taq* polimerase, em volume final de 13µL. Como molde da PCR, utilizou-se diretamente a colônia resistente a Zeocin™, homogeneizada no volume da reação. O programa foi iniciado por 2min/94°C, seguido de 35 ciclos: 20s/94°C – 15s/58°C – 1min/72°C. Os produtos obtidos foram verificados em gel de agarose 1% (Figura 13).

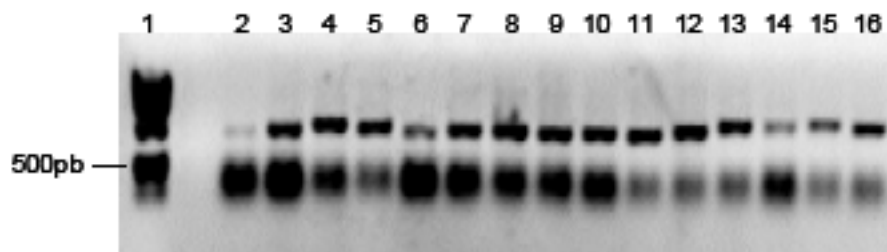


Figura 13 – Verificação de clonagem. Gel de agarose 1% para verificação da clonagem: 1 – λ -Hind 2 – 16 – amplificações obtidas por reação em cadeia da polimerase utilizando *primers M13 forward* e *reverse* e como molde, colônias resistentes ao antibiótico Zeocin™

2.1.11 - Sequenciamento automático das bibliotecas

As colônias obtidas foram inoculadas em placas de cultura com 96 poços contendo meio líquido SOB com 50 μ h de antibiótico Zeocin™/mL e incubadas a 130rpm, por 16h a 37°C. A seguir, 2 μ L desta cultura foram usados com molde em PCR idêntica a descrita para verificação da clonagem, sendo o programa utilizado também o mesmo. Para verificação da reação, as amplificações de diferentes tamanhos foram visualizadas em gel de agarose 1%.

Para sequenciamento, 1 μ L desta PCR foi usado como molde na seguinte reação: 0,3pmol *primer M13 forward* e 4 μ L de *Dyenamic ET Terminator reagent premix*, em volume final de 15 μ L. O programa utilizado foi o mesmo descrito para PCR, sendo a temperatura de anelamento de 57°C. As amplificações foram precipitadas, utilizando 7,5M acetato de amônia e etanol absoluto e analisadas em sequenciador *MEGA BACE 1000 DNA Analysis System* (Figura 14).

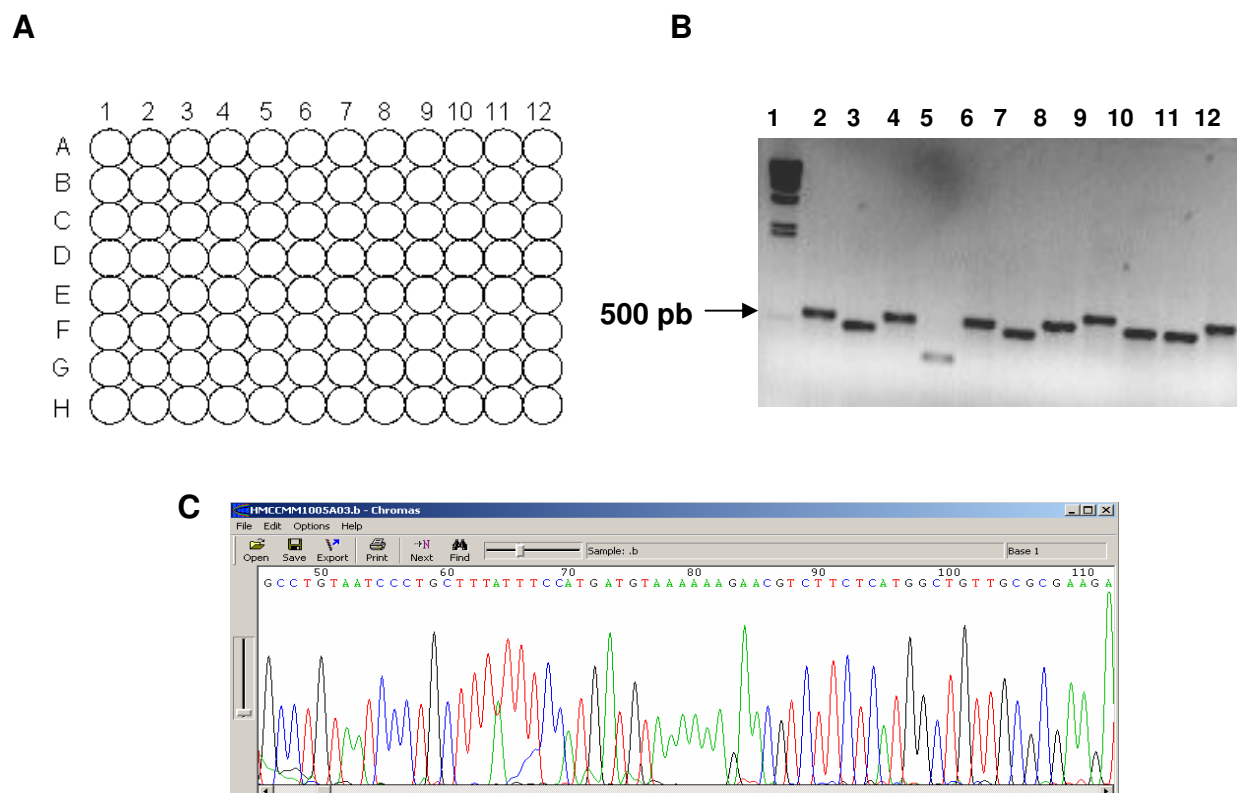


Figura 14 – Sequenciamento automático das bibliotecas. **A)** Representação esquemática de uma placa de cultura para a proliferação de bactérias contendo insertos de diferentes tamanhos **B)** Produtos da amplificação direta dos concatâmeros clonados no vetor pZero pela reação em cadeia da polimerase (PCR), onde 1- marcador de peso molecular λ -Hind 2 – 12 - fragmentos amplificados **C)** Eletroferogramas gerados pela análise no *MEGA BACE 1000 DNA Analysis System*, a partir da amplificação dos fragmentos de DNA pela PCR

2.1.12 - Análise de sequências

As sequências obtidas foram analisadas por meio dos programas *MEGA BACE Sequence Analyser* (Amersham – Life Science), *Phred* e *Phrap* (<http://genome.washington.edu/phrap.docs/phred.html>; <http://genome.washington.edu/phrap.docs/phrap.html>), sendo excluídas bases de baixa qualidade e sequências correspondentes a vetores.

2.1.13 - Extração e contagem de tags

A extração e contagem das *tags* foi realizada por meio do programa eSAGE Software v1.2 (MARGULIES EH e INNIS JW, 2000).

2.1.14 - Identificação das tags

Para identificação das *tags* foi utilizado o banco de dados *Best Gene for a tag*, *SAGEGenie* – *CGAP* (BOON K *et al.*, 2002) (banco *CGAP*), (<http://cgap.nci.nih.gov/SAGE>), que relacionou *tags* a números de *cluster Unigene* do banco de dados *Homo sapiens (Unigene Cluster Number, Hs.)* (PONTIUS JU *et al.*, 2003) de maneira não redundante. Ainda, por sua identificação no *Unigene*, as *tags* foram classificadas de acordo com critérios estabelecidos em nosso laboratório, nas seguintes categorias: *no match*; sequências denominadas preditas ou anotadas, as quais incluem *Expressed Sequence Tags* (ESTs), *Open Reading Frames* (ORFs), clones de cDNA e proteínas hipotéticas; e ainda a categoria genes conhecidos, que são aqueles bem caracterizados, com diversas sequências depositadas no *Unigene*.

2.1.15 - Critérios para identificação de tags diferencialmente expressas

Convencionamos considerar como genes diferencialmente expressos aqueles cujas expressões tiveram um valor de $p < 0,05$.

2.1.16 - Perfil funcional dos transcritos

Para análise do perfil funcional, transcritos observados foram classificados segundo o sistema do *GeneOntology* (ASHBURNER M *et al.*, 2000).

3 - PCR quantitativo em tempo real (qRT-PCR)

Técnica utilizada para validação dos resultados obtidos por SAGE.

3.1 – Síntese de DNA complementar (cDNA)

As amostras de RNA foram submetidas à síntese de DNA complementar (cDNA) utilizando-se o kit *RevertAid™ H Minus First Strand cDNA Synthesis* (Fermentas Life Sciences): 2µg de RNA foram tratados com DNase 1, RNase-free (Fermentas Life

Science), na seguinte reação: 1U DNase I, 1X DNase *Reaction buffer with MgCl₂* (100mM Tris-HCl, pH7,5; 25mM MgCl₂, 1mM CaCl₂), em volume final de 10μL. A reação foi realizada por 15min a temperatura ambiente e paralisada com 2,27mM EDTA, incubando-se por 5min a 65°C.

Para síntese do cDNA, foram adicionados 1μL de 100mM *Oligo(dt) Primer* (volume de 12μL) e incubado por 5min a 65°C. Adicionou-se então, 20X *Reaction Buffer* (250mM Tris-HCl, pH 8,3; 250mM KCl; 20mM MgCl₂; 50mM DTT), 2μL de 10mM dNTP Mix, 200U *RevertAid™ H Minus M-MuLV Reverse Transcriptase* e 200U *RiboLock™ RNase Inhibitor* e incubado por 1h/42°C – 5min/70°C.

A verificação da síntese de cDNA foi feita por meio de PCR para amplificação do gene beta-actina (BAC). As reações realizadas foram: 1X tampão *Taq* polimerase, 1mM MgCl₂, 200μM dNTPs, 200μM *primer BAC forward* (5' – AAGAGATGGCCACGGCTGCT – 3'), 200μM *primer BAC reverse* (5' – TCGCTCCAACCGACTGCTGT – 3'), 2,5U de *Taq* polimerase e aproximadamente 50ng de cDNA, em volume final de 50μL. O programa foi iniciado por 2min/95°C, seguido de 30 ciclos: 30s/95°C – 45s/50°C – 1min/72°C, sendo finalizado por 7min/72°C. Os produtos obtidos foram submetidos à eletroforese em gel de agarose 1% para a verificação da amplificação de 640pb.

3.2 – Princípios da qRT-PCR

Após a quantificação do cDNA, realizada no espectrofotômetro *NanoDrop ND-100* (NanoDrop Technologies), alíquotas de cDNA foram utilizadas como molde em qRT-PCR. A técnica consiste no monitoramento óptico da fluorescência emitida durante a reação de PCR, através da ligação de uma sonda específica ou um corante na fita recém sintetizada.

As reações, feitas sempre em duplicata, foram realizadas utilizando o reagente *SYBRGreen PCR Master Mix®* (Invitrogen), que além de conter todos os reagentes necessários para a PCR (dNTPs, MgCl₂, tampão, *Taq* Ampli-Gold), contém o corante SYBRGreen, componente intercalante de dupla fita. Além disso, utilizamos também amostra de cDNA e *primers* específicos para o gene analisado. Em todos os casos foram feitos controles negativos, contendo água estéril em substituição à amostra. As reações foram preparadas em placas de 96 poços (Applied Biosystems, USA) com adesivos

plásticos que permitem a passagem da luz. O programa foi iniciado por 10min/95°C, seguido de 45 ciclos: 15s/95°C – 1min/60°C.

A detecção de amplificação em tempo real foi realizada no equipamento *7500 Fast Real-time PCR System®* (Applied Biosystems, USA) e os resultados foram apresentados pelo programa *7500 Software v2.03* (Applied Biosystems, USA) em gráficos de fluorescência *versus* número de ciclos. O ciclo na qual se detecta fluorescência acima do limite basal estabelecido (*threshold*) é denominado ciclo de *threshold* ou Ct. Quanto maior a expressão de um gene, ou seja, quanto mais cópias existirem no início da reação, mais precocemente ocorre a amplificação e, conseqüentemente, menor é o Ct.

3.3 – Padronizações

De acordo com o gene escolhido para análise, foram necessárias padronizações para a realização da reação, descritas a seguir:

3.3.1 - Desenho dos *primers*

Os *primers* utilizados foram desenhados com o uso do site www.idt.com, analisados no programa *Blast* (www.ncbi.nlm.nih.gov/bblast) para confirmar homologia com o gene de interesse. A formação de estruturas como *hairpins* e *dimers* também foram avaliadas com o uso do programa *GeneRunner* (Hastings Softwares). Os *primers forward* e *reverse* foram desenhados em éxons separados, de forma a se detectar uma eventual contaminação com DNA genômico (amplificação do íntron entre os éxons aumentaria o tamanho do produto amplificado).

3.3.2 - Concentração de *primer*

A seguir, foi padronizada a concentração ótima de *primer* a ser utilizada, que deve ser a mínima suficiente para permitir a duplicação de todas as cópias do gene presentes na amostra. Utilizando a mesma quantidade de amostra foram feitas reações contendo cada um dos *primers (forward e reverse)* nas concentrações de 70nM, 150nM e 300nM. A concentração ideal foi aquela em que o gene de interesse obteve o menor valor de Ct (ciclo *threshold*), ou seja, foi a concentração em que o gene foi amplificado mais precocemente.

3.3.3 - Eficiência de reação

A reação de PCR em tempo real precisa ser confiável e reproduzível. Utilizando a concentração ótima de *primer*, foi determinada a eficiência da reação, onde as amplificações apresentem 100% de eficiência a cada ciclo de PCR. Para tanto, são feitas reações com quantidades conhecidas de amostra em escala logarítmica. Os resultados foram utilizados para construção de uma curva padrão Ct *versus* quantidade de amostra. A eficiência de amplificação (E) foi obtida a partir da fórmula $E = 10^{(-1/\text{slope})}$, onde *slope* corresponde ao coeficiente de inclinação da reta (PFAFFL MW, 2001). Observando-se tal linearidade, qualquer quantidade de amostra dentro da faixa analisada pode ser utilizada nas reações.

3.3.4 - Especificidade da reação

Na reação de PCR em tempo real, a especificidade pode ser verificada pela temperatura de desnaturação do produto amplificado, que depende do seu tamanho e constituição dos nucleotídeos. A temperatura de desnaturação foi determinada ao final do programa de amplificação, em protocolo de dissociação, que consistiu em aumento gradativo de temperatura de 60°C a 95°C. À medida que os produtos gerados por PCR desnaturam com o aumento de temperatura, cai o sinal fluorescente do *SYBR Green*. O gráfico resultante permite verificar se há um ou mais produtos de PCR presentes em cada reação, devido a diferentes temperaturas de desnaturação.

3.4 – Reações

Tendo sido feitas as devidas padronizações, as reações realizadas continham 6µL do reagente *SYBR Green Master Mix*® (Invitrogen), 5ng de amostra de cDNA e a concentração ótima de *primer* determinada, perfazendo um volume final de 12µL.

Para a validação das bibliotecas SAGE, foi utilizado como molde as amostras de cDNA do 10º dia da cultura dos pacientes portadores de beta-talassemia.

3.5 – Análise dos dados

A expressão dos genes de interesse foi determinada de forma relativa, sendo normalizada com relação a um gene calibrador (GAPDH). Esse é um gene cuja expressão é dita constitutiva, ou seja, apresenta pouca variação entre diversas condições.

Foram calculadas as médias entre os valores de Ct obtidos das reações realizadas em duplicata, para cada amostra. A derivação aritmética ΔCt foi a diferença entre a média do Ct de cada gene entre a média do Ct do gene controle. A seguir, foi aplicada a fórmula $\Delta\Delta Ct$, que foi a diferença entre o ΔCt de cada gene e o ΔCt da amostra calibradora, que foi uma amostra de cultura controle. Após esse passo, foi aplicada a fórmula aritmética $2^{-\Delta\Delta Ct}$ e gerado o valor da expressão de cada gene.

RESULTADOS

1 – Bibliotecas SAGE

1.1 - Biblioteca SAGE INTERMEDIÁRIA

5µg de RNA de cultura de células CD34+ do paciente talassêmico β intermediário, portador da mutação Cd 39 foram utilizados para construção da biblioteca SAGE, denominada SAGE INT. Essa biblioteca foi construída conforme descrito e então submetida a sequenciamento automático.

Após tratamento das sequências, contou-se para a biblioteca SAGE INT, um total de 4967 *tags* geradas, representando 2718 *tags* únicas. No site www.lge.ibi.unicamp.br/reticulocito, pode-se encontrar a distribuição da frequência, ou seja, o número absoluto de vezes que uma determinada *tag* foi detectada, em relação ao número de *tags* totais e de *tags* únicas.

Realizando a identificação das 2718 *tags* únicas da biblioteca INT com o banco de dados CGAP, observou-se cerca de 17% de *no matches*, 11% de *tags* correspondentes à sequências preditas, 3% de *tags* ORFs e 69% das *tags* correspondentes a genes conhecidos.

As *tags* mais expressas em INT correspondem aos genes: *HBA₂*, *HBG₁*, *BLVRB* e *GRIN2C* (Tabela 1).

Tabela 1 – 10 tags mais detectadas na biblioteca de paciente β talassêmico intermediário. Para cada tag estão indicados a sequência de nucleotídeo, símbolo, descrição e número no *UniGene* do gene correspondente, e a frequência com que foi detectada.

<i>Tag</i>	Símbolo	Gene	<i>Unigene</i>	Freq
CCCAACGCGC	HBA ₂	Hemoglobin, alpha 2	HS.654744	792
TGGATCCTGA	HBG ₁	Hemoglobin, gamma A	HS.705371	363
CCCATCGTCC		Transcribed locus, moderately similar to NP_536846.1 cytochrome C oxidase subunit II [HOMO SAPIENS]	HS.559716	43
TCCCTATTAA		Transcribed locus	HS.615061	40
GTGACCACGG	GRIN2C	Glutamate receptor, ionotropic, N-methyl D-aspartate 2C	HS.436980	39
AGGAGCAAAG	BLVRB	Biliverdin reductase B (flavin reductase (NADPH))	HS.515785	39
CACCTAATTG	<i>NO MATCH</i>		HS.336431	38
TCCCCGTACA		Transcribed locus	HS.536923	37
GAGGGAGTTT	RPL27A	Ribosomal protein L27A	HS.523463	29
GCATAATAGG	RPL21	Ribosomal protein L21	HS.381123	28

1.2 - Biblioteca SAGE MAIOR

5 μ g de RNA de cultura de células CD34⁺ do paciente talassêmico β maior, portador da mutação Cd 39 foram utilizados para construção da biblioteca SAGE, denominada SAGE MAIOR. Essa biblioteca foi construída conforme descrito e então submetida a sequenciamento automático.

Após tratamento das sequências, contou-se para a biblioteca SAGE MAIOR, um total de 6353 *tags* geradas, representando 3052 *tags* únicas. No site www.lge.ibi.ubicamp.br/reticulocito, pode-se consultar o número absoluto de vezes que uma determinada *tag* foi detectada, em relação ao número de *tags* totais e de *tags* únicas.

Realizando a identificação das 3052 *tags* únicas da biblioteca MAIOR com o banco de dados CGAP, observou-se cerca de 18% de *no matches*, 9% de *tags* correspondentes à sequências preditas, 3% *tags* ORFs e 70% das *tags* correspondentes a genes conhecidos.

As *tags* mais expressas em MAIOR correspondem aos genes: *HBA₂*, *HBG₁*, *BLVRB* (Tabela 2).

Tabela 2 – 10 tags mais detectadas na biblioteca de paciente β talassêmico maior.

Para cada tag estão indicados a sequência de nucleotídeo, símbolo, descrição e número no UniGene do gene correspondente, e a frequência com que foi detectada.

<i>Tag</i>	Símbolo	Gene	<i>Unigene</i>	Freq
CCCAACGCGC	HBA ₂	Hemoglobin, alpha 2	HS.654744	227
TGGATCCTGA	HBG ₁	Hemoglobin, gamma A	HS.705371	157
CCCATCGTCC		Transcribed locus, moderately similar to NP_536846.1 cytochrome C oxidase subunit II [HOMO SAPIENS]	HS.559716	53
TCCCTATTAA		Transcribed locus	HS.615061	46
GCATAATAGG	RPL21	Ribossomal protein L21	HS.381123	32
AGGAGCAAAG	BLVRB	Biliverdin reductase B (flavin reductase (NADPH))	HS.515785	32
GGATTGGCC	RPLP2	Ribossomal protein, large, P2	HS.437594	27
ACTTTTCAA		CDNA clone IMAGE:4328048	HS.349570	26
CACCTAATTG	<i>NO MATCH</i>			26
GAAAAATGGT	RPSA	Ribossomal protein SA	HS.449909	23

2- Validação dos resultados de SAGE por qRT-PCR

2.1 - Síntese de DNA complementar (cDNA)

A síntese de cDNA foi realizada conforme descrito, a partir das amostras de RNA dos pacientes β talassêmicos maior e intermediário estudados por SAGE. A verificação da síntese foi feita por PCR através da amplificação de um fragmento do gene da beta actina que é um gene constitutivo e expresso em diferentes tecidos e tipos celulares (Figura 15).

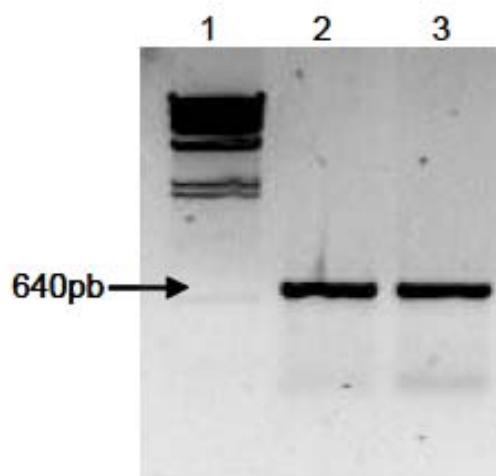


Figura 15 – Síntese de DNA complementar. Gel de agarose 1% para verificação da síntese de cDNA por reação de PCR para amplificação do gene *BAC*: 1 λ -*Hind* 2-3 amplificações de 640pb obtidas utilizando-se como molde: (2) cDNA de paciente portador de β talassemia intermediária (3) cDNA do paciente β talassêmico maior.

2.2 - Padronizações da qRT-PCR

Dentre os dados obtidos por SAGE, foram selecionados arbitrariamente, os genes *ABCB10*, *APEX1*, *APOC1*, *EIF5A*, *EYA3*, *GRIN2C*, *HMBS*, *HMGB1*, *NAE1*, *OAZ1*, *PCBP2*, *RAD23B*, *SRGN* e *TAGLN2*. Os *primers* para sua amplificação foram desenhados por meio do programa no site www.idt.com e analisados no programa GeneAmp® 5700 SDS (Applied Biosystems), de onde se obteve também tamanho e temperatura de desnaturação previstas do produto amplificado (Tabela 3).

Tabela 3 – Padronização das reações em cadeia da polimerase quantitativas em tempo real. Genes selecionados para análise, sequências dos primers desenhados para sua amplificação, tamanho e temperatura de desnaturação dos produtos, predita pelo programa *GeneAmp® 5700 SDS* (Applied Biosystems)

Gene	Sequência dos <i>primers</i> desenhados	Temperatura de desnaturação	Tamanho
ABCB10	F - 5' AGTTAGCAAGGAAAGAGGCAG 3' R - 5' AGGACAGAAAGCACGATCAG 3'	82 °C	85pb
APEX1	F - 5' GGTA AAAATATTGCTTCGGTGGG 3' R - 5' GTCTTACTCTTCTTGGCCTCTG 3'	83 °C	144pb
APOC1	F - 5' TGGTTCTGTCGATCGTCTTG 3' R - 5' GTTCACTCTGTTTGATGCGG 3'	83 °C	144pb
EIF5A	F - 5' GGAGATTGAGCAGAAGTACGAC 3' R - 5' GGAGCCAGTTATTTTGCCATG 3'	81 °C	111pb
EYA3	F - 5' CAAGAGGAAAGCTGATGCCAC 3' R - 5' CATCCAAGTCCACAGAAATACC 3'	82 °C	120pb
GRIN2C	F - 5' AACATCCGCAGTAACTACCG 3' R - 5' CTGCCATGTAGTTGAGGACAG 3'	83 °C	142pb
HMBS	F - 5' ATTGAAAGCCTCGTACCCTG 3' R - 5' TTCTCCAGGGCATGTTCAAG 3'	79 °C	137pb
HMGB1	F - 5' GATATGGCAAAGCGGACAAG 3' R - 5' GGCGATACTCAGAGCAGAAG 3'	79 °C	148pb
NAE1	F - 5' ATGGAAATCAGGTCAGCGG 3' R - 5' CAAAACCTCCAGAGACATCGC 3'	77 °C	132pb
OAZ1	F - 5' AGGGAATAGTCAGAGGGATCAC 3' R - 5' CAGTTAATGCGTTTGGCGTC 3'	79 °C	147pb
PCBP2	F - 5' TTACCATCACTGGATCTGCTGCC 3' R - 5' AGATGGATCATGGGTGGTGGTG 3'	82 °C	114pb
RAD23B	F - 5' ACGGAATCAGCCTCAGTTTC 3' R - 5' GAATAAAATGCTCCTGGTGTGG 3'	77 °C	140pb
SRGN	F - 5' CCTCATCCTGGTTCTGGAATC 3' R - 5' TGTTGGATTACCTGGAAGTAG 3'	81 °C	149pb
TAGLN2	F - 5' GATCCCAACTGGTTCCTAAG 3' R - 5' CATCCCGTAGCCAGTCATG 3'	82°C	144pb

A especificidade do produto amplificado foi verificada experimentalmente por meio de qRT-PCR, pela temperatura de desnaturação específica do produto amplificado. Na figura 16, está ilustrada a curva de dissociação (derivativa da fluorescência *versus* temperatura) do produto amplificado do gene *SRGN*, gerada pelo programa 7500 Software v2.03 (Applied Biosystems, USA). Observou-se desnaturação em concordância com o esperado.

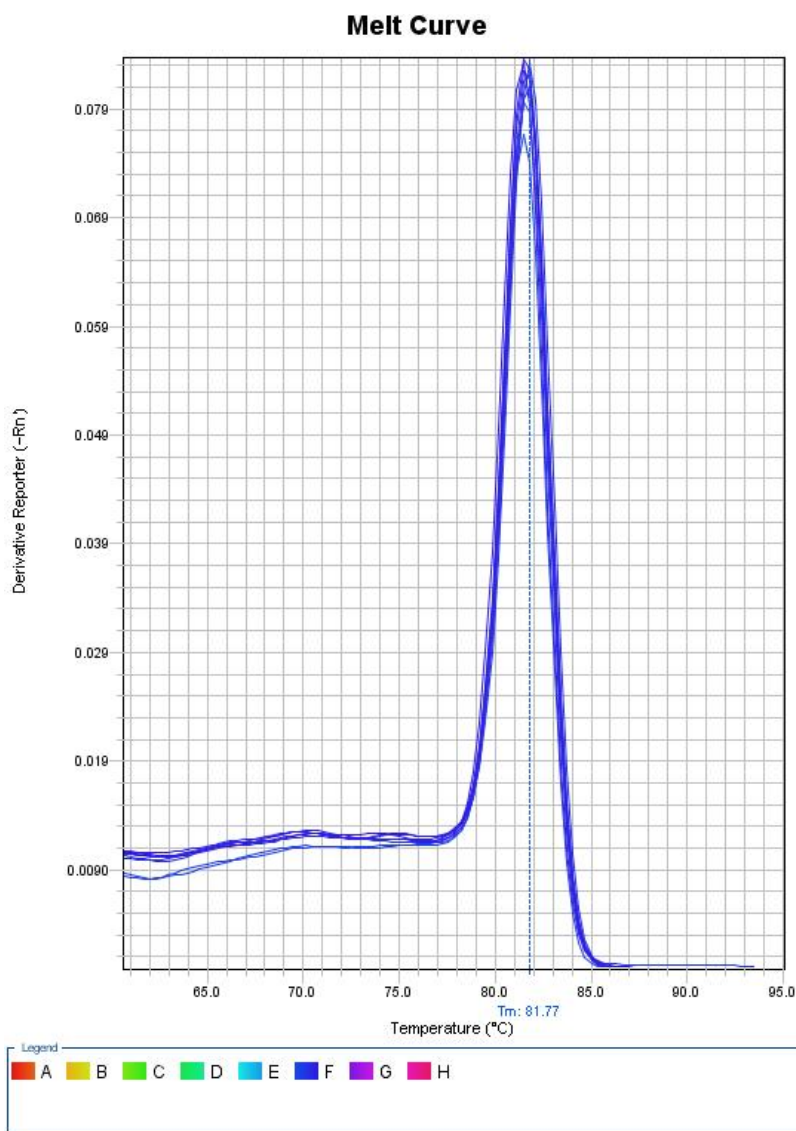


Figura 16 – Temperatura de dissociação do produto amplificado. Curva de dissociação do produto amplificado do gene *SRGN*, gerada pelo programa 7500 Software v2.03 (Applied Biosystems, USA).

Foram então realizadas as reações para determinação da concentração ótima de *primer* a ser utilizada e da eficiência da reação (Tabela 4). Na figura 17, está ilustrada a curva padrão da reação de amplificação do gene *SRGN*, gerada pelo programa 7500 Software v2.03 (Applied Biosystems, USA). O coeficiente de inclinação da reta, ou *slope*, e o valor da eficiência da reação foi dado pelo programa 7500 Software v2.03 (Applied Biosystems, USA).

Tabela 4 – Concentração de *primers* utilizadas na amplificação dos genes de estudo e eficiência da amplificação obtida. As concentrações foram definidas pela eficiência de amplificação gerada nas condições testadas.

Primer	Concentração utilizada	Eficiência do <i>primer</i>
ABCB10	150 nM	99%
APEX1	150 nM	100%
APOC1	150 nM	100%
EIF5A	150 nM	100%
EYA3	150 nM	100%
GRIN2C	70 nM	99%
HMBS	150 nM	98,5%
HMGB1	150 nM	100%
NAE1	150 nM	98,8%
OAZ1	150 nM	100%
PCBP2	150 nM	100%
RAD23B	150 nM	99,9%
SRGN	150 nM	100%
TAGLN2	150 nM	100%

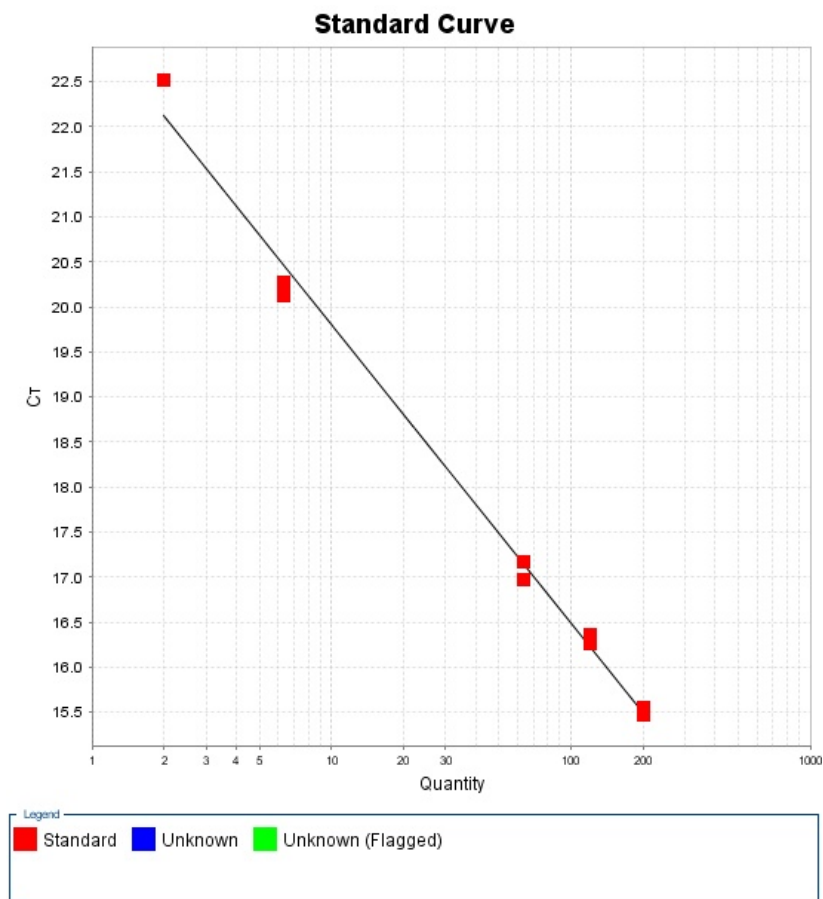


Figura 17 – Curva padrão para a reação de amplificação. Curva padrão do produto amplificado do gene *SRGN*, gerada pelo programa *7500 Software v2.03* (Applied Biosystems, USA).

2.3 - Reações

Para validação dos resultados de SAGE, foi selecionado como gene calibrador o gene GAPDH e os valores de expressão gênica normalizados foram calculados utilizando a fórmula aritmética $2^{-\Delta\Delta Ct}$.

Expressões gênicas concordantes, medidas por SAGE e qRT-PCR, foram identificadas em 8 dos 14 genes avaliados (57,14%).

Os genes *EYA3*, *TAGLN2*, *HMBS1*, *ABCB10* e *OAZ1* apresentaram-se mais expressos no paciente com talassemia β intermediária do que no paciente com talassemia β maior (Figura 18). Já a expressão dos genes *APEX1*, *SRGN* e *APOC1* mostrou-se maior na biblioteca de paciente talassêmico β maior do que no paciente talassêmico β intermediário.

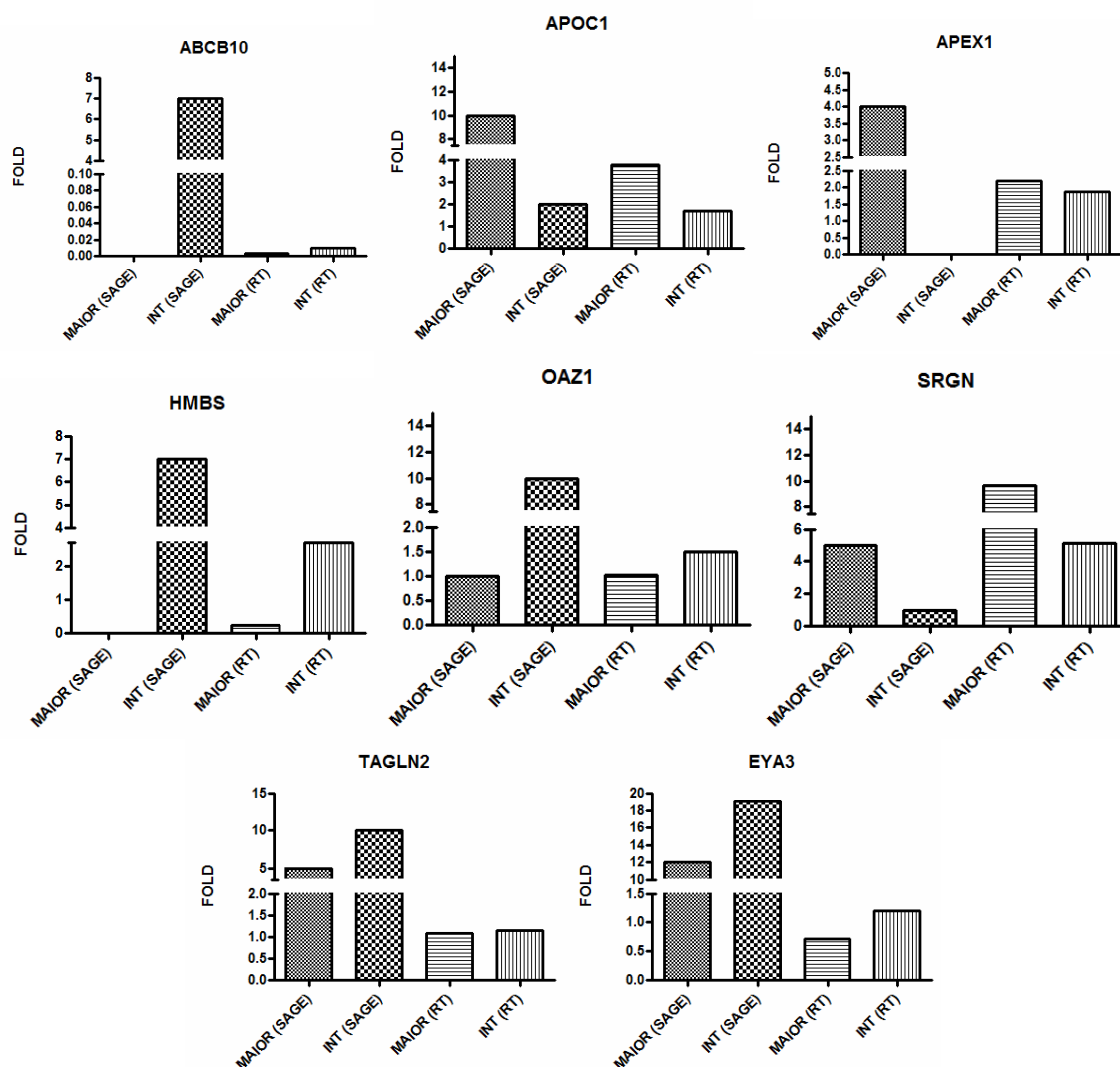


Figura 18 – Comparação de resultados obtidos por SAGE e por qRT-PCR. Representação gráfica de 8 genes com expressão gênica concordantes pelo método SAGE e pela reação em cadeia da polimerase em tempo real, em amostras de cultura de células $CD34^+$ do paciente talassêmico β intermediário e maior, a partir da qual foi construída a biblioteca SAGE.

Expressões gênicas discordantes, medidas por SAGE e qRT-PCR, foram identificadas em 6 genes (Figura 19).

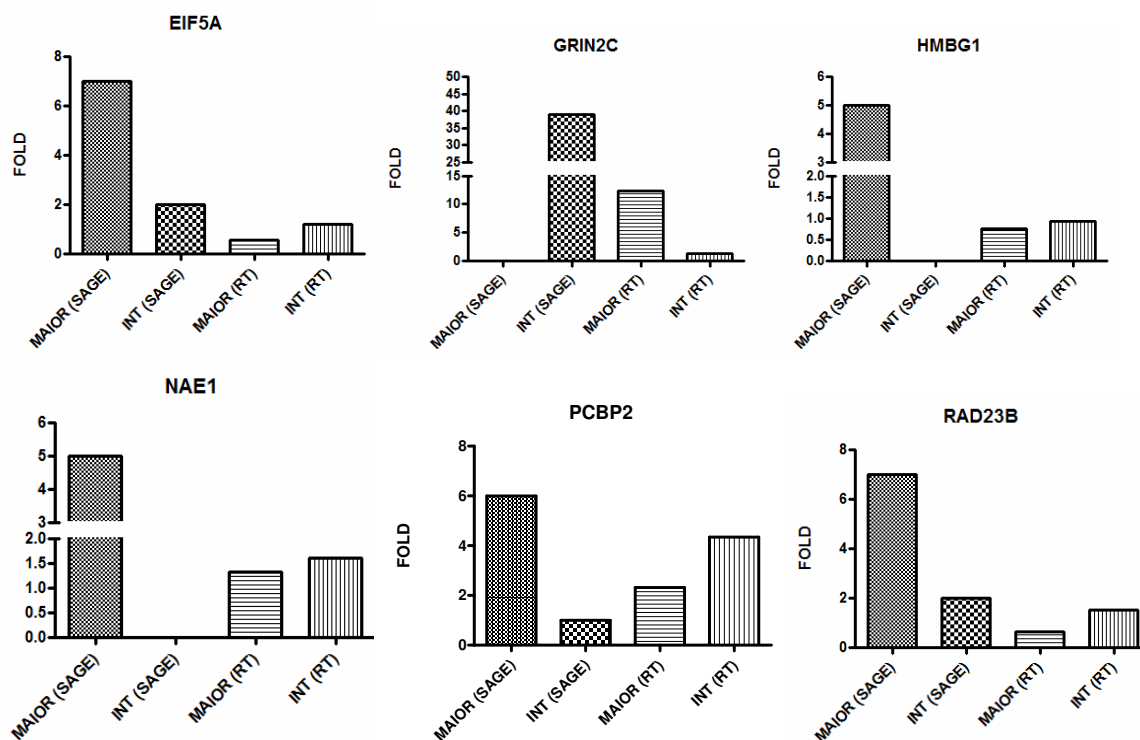


Figura 19 – Comparação de resultados obtidos por SAGE e por qRT-PCR. Representação gráfica de 6 genes com expressão gênica disconcordantes pelo método SAGE e pela reação em cadeia da polimerase em tempo real, em amostras de cultura de células CD34⁺ do paciente talassêmico β intermediário e maior, a partir da qual foi construída a biblioteca SAGE.

3 - Expressão gênica diferencial: INT X MAIOR

Tendo sido caracterizados e validados os perfis globais de expressão gênica em talassêmico β intermediário e maior, estes foram comparados entre si, no intuito de se identificarem alterações que pudessem estar relacionadas à doença.

Usando os critérios descritos, foram encontradas 42 *tags* diferencialmente expressas entre as bibliotecas SAGE (Tabela 5).

A figura 20 representa a clusterização dos genes diferencialmente expressos e a identificação das *tags* encontra-se na Tabela 5.

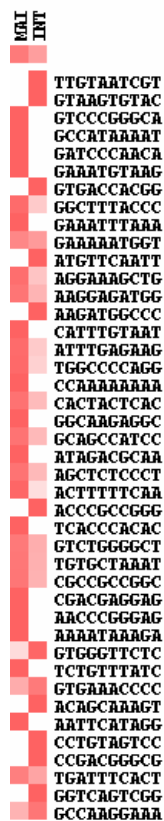


Figura 20 - Análise de cluster dos genes diferencialmente expressos. Diagrama de cores: branco, baixa expressão; vermelho, alta expressão. A intensidade das cores reflete a intensidade da expressão dos genes.

Tabela 5 – Identificação das *tags* diferencialmente expressas entre células de cultura de um paciente portador de talassemia β maior e intermediária. *Tags* foram classificadas como: *no match* (não apresentam nenhuma identificação no *Unigene*), sequências preditas (correspondem à ESTs, ORFs, clones de cDNA ou proteínas hipotéticas) ou genes conhecidos.

<i>Tag</i>	INT	MAIOR	PVALOR	(-)LOG PVALOR	FOLD CHANGE	SYMBOL	TITLE	FUNÇÃO
ACTTTTCAA	2	26	5,93E-08	24,00848	-16,6275	--	CDNA CLONE IMAGE:4328048	Unknown
TGATTTCAC	9	22	0,002537	8,622698	-3,12655	--	CDNA CLONE IMAGE:5265638	Unknown
CCGACGGGCG	9	0	0,006199	7,33364	0,001417	--	TRANSCRIBED LOCUS, STRONGLY SIMILAR TO NP_001090497.1 SOLUTE CARRIER FAMILY 25 (MITOCHONDRIAL CARRIER; OXOGLUTARATE CARRIER), MEMBER 11 [XENOPUS LAEVIS]	Unknown
GGTCAGTCGG	9	0	0,006199	7,33364	0,001417	--	TRANSCRIBED LOCUS, STRONGLY SIMILAR TO NP_001041438.1 HYPOTHETICAL PROTEIN LOC503325 [RATTUS NORVEGICUS]	Unknown
CACTACTCAC	3	10	0,018227	5,777783	-4,26347	--	TRANSCRIBED LOCUS, STRONGLY SIMILAR TO NP_536855.1 CYTOCHROME B [HOMO SAPIENS]	Unknown
AATTCATAGG	0	4	0,032529	4,942137	-0,00081	--	TRANSCRIBED LOCUS	Unknown
CATTGTGAAT	0	4	0,032529	4,942137	-0,00081	--	CDNA FLJ44518 FIS, CLONE UTERU3002701	Unknown
GCCAAGGAAA	18	5	0,033443	4,902142	2,814607	--	RED CELL ANION EXCHANGER (EPB3, AE1, BAND 3) 3' NON-CODING REGION	
ATGTTCAATT	7	0	0,019683	5,666912	0,001102	ABCB10	ATP-BINDING CASSETTE, SUB-FAMILY B (MDR/TAP), MEMBER 10	
GTCCCGGGCA	0	5	0,014273	6,130564	-0,00101	ACSS1	ACYL-COA SYNTHETASE SHORT-CHAIN FAMILY MEMBER 1	
AAAATAAAGA	0	4	0,032529	4,942137	-0,00081	APEX1	APEX NUCLEASE (MULTIFUNCTIONAL DNA REPAIR ENZYME) 1	
TGGCCCCAGG	2	10	0,006491	7,267345	-6,39521	APOC1	APOLIPOPROTEIN C-I	lipid metabolic process- it is activated when monocytes differentiate into macrophages
GATCCCAACA	0	4	0,032529	4,942137	-0,00081	ATP5B	ATP SYNTHASE, H ⁺ TRANSPORTING, MITOCHONDRIAL F1 COMPLEX, BETA POLYPEPTIDE	
AACCCGGGAG	0	6	0,006263	7,318991	-0,00121	BMS1P5	BMS1 PSEUDOGENE 5	
GGCTTTACCC	2	7	0,046244	4,434579	-4,47665	EIF5A	EUKARYOTIC TRANSLATION INITIATION FACTOR 5A	
CGACGAGGAG	0	6	0,006263	7,318991	-0,00121	EMP3	EPITHELIAL MEMBRANE PROTEIN 3	
GTGGGTTCTC	19	2	0,000965	10,01747	7,427436	EYA3	EYES ABSENT HOMOLOG 3 (DROSOPHILA)	transcription activator
GTGACCACGG	39	0	1,85E-10	32,33456	0,006139	GRIN2C	GLUTAMATE RECEPTOR, IONOTROPIC, N-METHYL D- ASPARTATE 2C	

AAGATGGCCC	7	0	0,019683	5,666912	0,001102	HMBS	HYDROXYMETHYLBILANE SYNTHASE	
GAAATTTAAA	0	5	0,014273	6,130564	-0,00101	HMGB1	HIGH-MOBILITY GROUP BOX 1	
CCTGTAGTCC	6	0	0,035072	4,833548	0,000944	MAFF	V-MAF MUSCULOAPONEUROTIC FIBROSARCOMA ONCOGENE HOMOLOG F (AVIAN)	
GTGAAACCCC	28	8	0,008268	6,918177	2,736424	MOG	MYELIN OLIGODENDROCYTE GLYCOPROTEIN	
ATAGACGCAA	0	6	0,006263	7,318991	-0,00121	MORF4L1	MORTALITY FACTOR 4 LIKE 1	
GGCAAGAGGC	0	5	0,014273	6,130564	-0,00101	NAE1	NEDD8 ACTIVATING ENZYME E1 SUBUNIT 1	
GTAAGTGTAC	13	1	0,004393	7,83058	10,16386	NO MATCH	NO MATCH	
ACCCGCCGGG	9	0	0,006199	7,33364	0,001417	NO MATCH	NO MATCH	
TCACCCACAC	0	5	0,014273	6,130564	-0,00101	NO MATCH	NO MATCH	
TTGTAATCGT	10	1	0,020272	5,624353	7,818354	OAZ1	ORNITHINE DECARBOXYLASE ANTIZYME 1	Protein binding
CCAAAAA	0	4	0,032529	4,942137	-0,00081	OCIAD1	OCIA DOMAIN CONTAINING 1	
GAAATGTAAG	1	6	0,030866	5,017833	-7,67425	PCBP2	POLY(RC) BINDING PROTEIN 2	DNA/RNA binding- The encoded protein is also suggested to play a part in formation of a sequence-specific alpha-globin mRNP complex which is associated with alpha-globin mRNA stability
ATTTGAGAAG	2	7	0,046244	4,434579	-4,47665	RAD23B	RAD23 HOMOLOG B (S. CEREVISIAE)	damaged DNA binding/protein binding
ACAGCAAAGT	9	0	0,006199	7,33364	0,001417	RHCE	RH BLOOD GROUP, CCEE ANTIGENS	
AGCTCTCCCT	3	9	0,033238	4,911035	-3,83713	RPL17	RIBOSOMAL PROTEIN L17	ribosomal
GCAGCCATCC	3	9	0,033238	4,911035	-3,83713	RPL28	RIBOSOMAL PROTEIN L28	ribosomal
AAGGAGATGG	4	13	0,007394	7,079473	-4,15689	RPL31	RIBOSOMAL PROTEIN L31	ribosomal
TGTGCTAAAT	4	11	0,023637	5,402784	-3,51736	RPL34	RIBOSOMAL PROTEIN L34	ribosomal
CGCCGCCGGC	4	12	0,013316	6,230689	-3,83713	RPL35	RIBOSOMAL PROTEIN L35	ribosomal
AGGAAAGCTG	3	15	0,000744	10,39238	-6,39521	RPL36	RIBOSOMAL PROTEIN L36	ribosomal
GAAAAATGGT	14	23	0,026116	5,258898	-2,10128	RPSA	RIBOSOMAL PROTEIN SA	ribosomal
TCTGTTTATC	0	4	0,032529	4,942137	-0,00081	SRP14	SIGNAL RECOGNITION PARTICLE 14KDA (HOMOLOGOUS ALU RNA BINDING PROTEIN)	
GCCATAAAAT	1	5	0,0623348	4,0038174	6,3952084	SRGN	SERGLYCIN	Protein binding/apoptosis
GTCTGGGGCT	4	10	0,041279	4,598431	-3,1976	TAGLN2	TRANSGELIN 2	

A fim de se obter a ontologia dos 42 genes diferencialmente expressos, submetemos os genes à análise no site www.ncbi.nlm.gov/gene. Na tabela 6, está demonstrado os dados referentes à classificação dos transcritos.

Tabela 6 – Ontologia dos genes diferencialmente expressos. Os genes foram classificados segundo o *Gene Ontology* quanto às funções e processos da qual participam.

<i>Símbolo</i>	<i>Descrição</i>	<i>HS</i>	<i>Função</i>
	CDNA CLONE IMAGE:4328048	349570	Unknown
	CDNA CLONE IMAGE:5265638	129283	Unknown
	TRANSCRIBED LOCUS, STRONGLY SIMILAR TO NP_001090497.1 SOLUTE CARRIER FAMILY 25 (MITOCHONDRIAL CARRIER; OXOGLUTARATE CARRIER), MEMBER 11 [XENOPUS LAEVIS]	586920	Unknown
	TRANSCRIBED LOCUS, STRONGLY SIMILAR TO NP_001041438.1 HYPOTHETICAL PROTEIN LOC503325 [RATTUS NORVEGICUS]	631428	Unknown
	TRANSCRIBED LOCUS, STRONGLY SIMILAR TO NP_536855.1 CYTOCHROME B [HOMO SAPIENS]	631491	Unknown
	TRANSCRIBED LOCUS	625475	Unknown
	CDNA FLJ44518 FIS, CLONE UTERU3002701	689535	Unknown
	RED CELL ANION EXCHANGER (EPB3, AE1, BAND 3) 3' NON-CODING REGION	654584	Unknown
ABCB10	ATP-BINDING CASSETTE, SUB-FAMILY B (MDR/TAP), MEMBER 10	17614	ATP binding, ATPase/transporter activity
ACSS1	ACYL-COA SYNTHETASE SHORT-CHAIN FAMILY MEMBER 1	529353	ATP/nucleotide/protein binding, ligase activity, metabolic process
APEX1	APEX NUCLEASE (MULTIFUNCTIONAL DNA REPAIR ENZYME) 1	73722	DNA/ metal ion binding, endonuclease activity, base excision repair
APOC1	APOLIPOPROTEIN C-I	110675	Lipase inhibitor activity, cholesterol metabolic process, lipid metabolic process
ATP5B	ATP SYNTHASE, H ⁺ TRANSPORTING, MITOCHONDRIAL F1 COMPLEX, BETA POLYPEPTIDE	406510	Hydrolase/lipoprotein receptor activity, transmembrane transporter activity, angiogenesis
BMS1P5	BMS1 PSEUDOGENE 5	647203	
EIF5A	EUKARYOTIC TRANSLATION INITIATION FACTOR 5A	534314	RNA binding, protein transport, transmembrane transport, mRNA export from nucleus, negative regulation of apoptosis
EMP3	EPITHELIAL MEMBRANE PROTEIN 3	9999	Cell growth, negative regulation of cell proliferation
EYA3	EYES ABSENT HOMOLOG 3 (DROSOPHILA)	706831	Chromatin binding, chromatin modification metabolic process, transcription activator
GRIN2C	GLUTAMATE RECEPTOR, IONOTROPIC, N-METHYL D-ASPARTATE 2C	436980	Cation channel activity, receptor and transporter activity

HMBS	HYDROXYMETHYLBILANE SYNTHASE	82609	Hydroxymethylbilane synthase activity, coenzyme binding, heme biosynthetic process
HMGB1	HIGH-MOBILITY GROUP BOX 1	434102	Transcription factor activity/binding, damage DNA binding, DNA ligation involved in DNA repair
MAFF	V-MAF MUSCULOAPONEUROTIC FIBROSARCOMA ONCOGENE HOMOLOG F (AVIAN)	517617	Sequence specific DNA binding, transcription factor activity
MOG	MYELIN OLIGODENDROCYTE GLYCOPROTEIN	141308	Cell adhesion, central nervous system development
MORF4L1	MORTALITY FACTOR 4 LIKE 1	374503	DNA repair, chromatin modification, protein binding, regulation of growth/transcription
NAE1	NEDD8 ACTIVATING ENZYME E1 SUBUNIT 1	460978	DNA replication, regulation of apoptosis, cell cycle, metabolic process, catalytic activity
NO MATCH		NO MATCH	
NO MATCH		NO MATCH	
NO MATCH		NO MATCH	
OAZ1	ORNITHINE DECARBOXYLASE ANTIZYME 1	446427	Ornithine decarboxylase inhibitor activity, protein binding
OCIAD1	OCIA DOMAIN CONTAINING 1	518750	Endosome component
PCBP2	POLY(RC) BINDING PROTEIN 2	546271	DNA/RNA enzyme protein binding, RNA splicing, mRNA metabolic process
RAD23B	RAD23 HOMOLOG B (S. CEREVISIAE)	521640	Damage DNA/protein, single stranded DNA binding, nucleotide excision repair
RHCE	RH BLOOD GROUP, CCEE ANTIGENS	449968	Transmembrane transport
RPL17	RIBOSOMAL PROTEIN L17	374588	Translation elongation
RPL28	RIBOSOMAL PROTEIN L28	652114	RNA/protein binding, translation elongation
RPL31	RIBOSOMAL PROTEIN L31	469473	RNA/protein binding, translation elongation
RPL34	RIBOSOMAL PROTEIN L34	438227	RNA/protein binding, translation elongation
RPL35	RIBOSOMAL PROTEIN L35	182825	RNA/protein binding, translation elongation
RPL36	RIBOSOMAL PROTEIN L36	408018	RNA/protein binding, translation elongation
RPSA	RIBOSOMAL PROTEIN SA	449909	Receptor activity, ribosome binding, cell adhesion
SRP14	SIGNAL RECOGNITION PARTICLE 14KDA (HOMOLOGOUS ALU RNA BINDING PROTEIN)	533732	RNA/protein binding, response to drug
SRGN	SERGLYCIN	1908	Collagen/protein binding, apoptosis
TAGLN2	TRANSGELIN 2	517168	Protein binding

DISCUSSÃO

As síndromes talassêmicas compreendem um grupo heterogêneo de doenças hereditárias em que existe uma redução no ritmo de síntese de uma ou mais cadeias polipeptídicas da hemoglobina. Nas talassemias β ocorre a supressão total ou parcial da produção de cadeias β .

Aproximadamente a metade dos alelos beta talassêmicos produz talassemias β^0 . Muitos são mutações *nonsense*, isto é, substituição de bases que introduzem também um códon de terminação prematura. Como exemplo desse tipo de mutação, podemos citar a troca de C→T no códon 39 da beta globina que produz uma proteína anormal e não funcional que é degradada.

Em nosso estudo, incluímos 2 pacientes talassêmicos β , portadores da mesma mutação genética (Cd 39) mas com quadros clínicos distintos, maior e intermediária com o objetivo de avaliar possíveis moduladores genéticos, diferentes dos já descritos, e que podem estar envolvidos na gravidade da doença.

Foram utilizadas para construção dos perfis INT e MAIOR, células CD34⁺ isoladas do sangue periférico dos pacientes e cultivadas durante 13 dias.

Após a identificação dos 2718 transcritos únicos encontrados no perfil INT e 3052 transcritos únicos no perfil MAIOR, esses transcritos foram classificados de acordo com o banco de dados do *The Cancer Genome Anatomy (CGAP)* e os perfis foram comparados entre si a fim de se obter os transcritos diferencialmente expressos. O critério para a seleção dos transcritos diferencialmente expressos foi o valor de $p < 0,05$ (AUDIC e CLAVERIE, 1997) e uma diferença de expressão de pelo menos 5 vezes (5 fold) entre as amostras. De acordo com este critério, 42 transcritos foram classificados como diferencialmente expressos, dentre os quais foram encontrados genes conhecidos, ESTs, *no matches* e outros tipos de sequências preditas.

Os transcritos diferencialmente expressos distribuíram-se em duas situações distintas: transcritos que apresentaram maior expressão no paciente talassêmico β intermediário e transcritos que apresentaram maior expressão no paciente talassêmico β maior.

Dentre os genes que apresentaram maior expressão no paciente talassêmico β intermediário destacamos os genes *ABCB10*, *HMBS*, *EYA3* e *EIF5A*.

O gene *ABCB10* (*ATP-binding cassette, sub-family B (MDR/TAP), member 10*) foi primeiramente identificado durante a análise de transcritos na diferenciação de células

G1E-ER2 (WEISS *et al.*, 1997) para avaliar possíveis genes alvo para o gene *GATA-1* (SHIRIHAI *et al.*, 2000). O transcrito foi isolado e caracterizado como um novo membro da família ABC, um grande grupo de proteínas que transporta uma variedade de diferentes compostos através da membrana das células (HIGGINS, 1992; CROOP, 1998). Essa proteína foi chamada de *ABC me* (ABC *mitochondrial erythroid*) expressa, particularmente, em altos níveis nas mitocôndrias dos precursores eritróides. O gene *ABC me* se mostrou predominantemente expresso em células eritróides, especialmente nos estágios finais da maturação, quando o gene *GATA-1* é essencial (WEISS *et al.*, 1994). Em um trabalho recente publicado por nosso grupo de pesquisa, que avaliou a diferença de expressão durante a diferenciação de células eritróides, esse dado foi novamente confirmado, uma vez que a expressão deste gene mostrou-se seis vezes mais expressa na fase final de diferenciação (CUNHA *et al.*, 2010). A superexpressão do gene *ABC me* em células MEL estimuladas à diferenciação por DMSO, elevaram os níveis de hemoglobina nessas células, sugerindo a participação desse gene na síntese do grupo heme.

Em 2009, Chen e colaboradores, demonstraram que *ABCB10* interage fisicamente com a mitoferrina, uma proteína mitocondrial que é a principal importadora de Fe^{+2} para a mitocôndria, formando um complexo, que aumenta a estabilidade protéica e a importação de Fe^{+2} para a mitocôndria. Já em 2010, Chen demonstrou a formação de um complexo oligomérico composto por *ABCB10*, mitoferrina-1 e ferroquelatase, essencial para o biossíntese do grupamento heme. Um outro gene relacionado à síntese do grupo heme e que também encontramos com expressão elevada no paciente intermediário é o gene *HMBS* (*hydroxymethylbilane synthase*) que é responsável pela terceira enzima na cascata da síntese do grupo heme. Possui dois promotores, sendo que um deles é ativo somente em células eritróides e possui grande homologia com o promotor do gene da globina β (CHRETIEN *et al.*, 1988). Mutações nesse gene levam à porfiria aguda intermitente (GRANDCHAMP, 1998).

Devido à grande participação desses genes no processo de formação do grupo heme, podemos sugerir que os níveis aumentados da expressão desses genes no paciente talassêmico β intermediário contribuam para aumentar a eficiência na produção de heme, aumento dos níveis de hemoglobina e consequente diminuição dos níveis de ferro nos tecidos, contribuindo para o quadro intermediário. Ainda, de acordo com Shirihai, a

expressão aumentada do gene *ABCB10* pode ser relacionada ao gene *GATA-1* e ao fato de que as células eritróides do paciente talassêmico β intermediário consigam chegar à um estágio final de maturação avançado em relação ao paciente talassêmico β menor.

Já o gene *EYA* (*Eyes Absent*) foi primeiramente descrito como um componente na maquinaria de reguladores de transcrição no desenvolvimento do olho em *Drosophila*. Encontramos em mamíferos 4 tipos de proteínas EYA (1-4) e sabe-se hoje, que o gene *EYA3* (*Eyes Absent homolog 3*) está envolvido no desenvolvimento de órgãos e tecidos dos mais diversos organismos, atuando como um fator de transcrição (JEMC & REBAY, 2007). O *EYA* é o primeiro exemplo de fator de transcrição com atividade de fosfatase intrínseca (RAYAPUREDI, 2003; TOOTLE, 2003 e LI, 2003). Krishnan e colaboradores (KRISHNAN, 2009) mostraram que o *EYA3* funciona como uma proteína tirosina fosfatase na desfosforilação do resíduo Tyr-142 da Histona H2A.X, atuando no reparo do DNA, apoptose e carcinogênese. Um fato ainda a ser considerado é que os genes *EYA1* e *EYA2* foram identificados em famílias hematopoéticas primordiais (TERSKIKH, 2003), porém nada foi descrito com relação ao *EYA3*. Estudos de camundongos *Knockout* para o gene *EYA3* realizados por Söker e colaboradores, 2008 mostram uma leve anemia nos animais cujo gene foi suprimido (dados não publicados e obtidos através de contato com o grupo responsável pelo trabalho). No trabalho já anteriormente citado e realizado por nosso grupo, Cunha e colaboradores (2010) demonstraram um aumento da expressão desse gene durante a proliferação de linhagens eritróides, indicando uma possível participação deste gene na fase final do processo de diferenciação. Nossos resultados indicam uma possível correlação com a expressão desse gene e o fenótipo dos pacientes e por isso há um trabalho sendo realizado em nosso grupo de pesquisa que visa avaliar a função deste gene através de experimentos de inibição da expressão gênica e superexpressão em linhagem de células K562 e células CD34⁺. Este trabalho também está avaliando a expressão desse gene em reticulócitos, neutrófilos e células mononucleares de pacientes com anemia falciforme e talassemia.

O gene *EIF5A* (*eukaryotic translation initiation factor 5A*), é o único gene que traduz uma proteína que contém um único aminoácido hypusine. Entretanto, seus mecanismos de ação precisam ser melhor estudados. Sabe-se que está envolvido em processos de proliferação celular (PARK, 1998; TOME, 1996), regulação do transporte de

RNAm (ROSORIUS, 1999; LIPOWSKY, 2000), apoptose (Li, 2004; THOMPSON, 2010), tradução protéica (HANAUKE, 1995; JAO, 2006) e inflamação (MOORE, 2008). Em 2009, Flanagan e colaboradores realizaram um *microarray* do fígado de pacientes falciformes e talassêmicos β maior com sobrecarga de ferro, dependentes de transfusão. O gene *EIF5A* se mostrou subexpresso em comparação às amostras de pacientes com baixa concentração de ferro no fígado. Nossos resultados mostraram discordância entre os métodos de avaliação da expressão desse gene, porém, quando avaliamos a expressão pelo método da qRT-PCR, considerado um método “*gold standard*” para avaliação de expressão gênica, observamos que o paciente talassêmico β maior tem uma expressão diminuída em relação ao paciente talassêmico β intermediário, corroborando com os resultados apresentados por Flanagan.

Entre os genes que apresentaram maior expressão no paciente talassêmico β maior encontramos uma relação interessante envolvida em reparo de DNA entre os genes: *HMGB1*, *RAD23B*, *APEX1* e *PCBP2*.

O gene *HMGB1* (*high-mobility group box 1*) codifica uma proteína multifuncional que media uma variedade de processos, dentro e fora da célula, como regulação da transcrição (SINGH, 1990), remodelamento de cromatina (BONALDI, 2002) e inflamação (WANG, 1999). Em 2009, Lange e colaboradores realizaram experimentos com células HeLa CCL4, irradiadas com radiação UVA 365nm para avaliar qual seria o papel do gene *HMGB1* no processo de reparo do DNA. Os resultados mostraram que o gene *HMGB1* aumenta as interações com o gene *RAD23B* (*RAD23 homolog B (S. cerevisiae)*), gene que codifica uma proteína envolvida no reparo por excisão de nucleotídeo (*NER*), ou seja, processos de reparo do DNA, sugerindo a formação de um complexo atuando nas células danificadas pela radiação.

Outro gene também encontrado mais expresso na biblioteca MAIOR foi o gene *APEX1* (*APEX nuclease (multifunctional DNA repair enzyme) 1*), que faz parte do processo de DNA reparo por excisão de base (*BER*). Esse gene codifica uma proteína multifuncional que regula a apoptose e é induzida por espécies reativas de oxigênio (ROS) (ROBERTSON *et al.*, 1997). Sabe-se que, os pacientes portadores de talassemia beta possuem níveis elevados de ROS e que isso contribui para a patogenia da doença (NAGABABU, 2000).

Experimentos realizados com células CD34+ *knockdown* para o gene *APEX1* apontam para uma diminuição na produção de progenitores hematopoéticos primitivo e definitivo e que talvez, esse gene possa exercer um papel regulador durante a eritromielopoese normal (ZOU, 2007).

De posse desses dados, podemos sugerir que a expressão elevada desses genes no paciente portador de talassemia β maior se deve ao grande número de eventos que causam danos às células desses pacientes, como o elevado nível de espécies reativas de oxigênio, a eritropoese ineficaz, a precipitação das cadeias α livres, causando danos à membrana dos eritrócitos e diminuindo a vida média dessas células. Sendo assim, a expressão desses genes estaria ligada ao fato de que há, nesses pacientes, um maior dano celular decorrente do seu estado clínico grave em comparação ao paciente talassêmico β intermediário.

Outro dado interessante encontrado nesse estudo é o aumento da expressão do gene *PCBP2*. Foi sugerido que esta proteína seja parte de um complexo de ligação a uma sequência específica na região 3' do RNA mensageiro para α globina que é rica em citosina e responsável pela ligação a ribossomos para a formação da proteína. Estudos mostraram que deleções nesta região estão associadas com a instabilidade do RNAm da α globina, resultando em sua degradação antes da formação da proteína (WEISS, 1994; WEISS, 1995; WANG *et al.*, 1995). Nossos resultados mostram uma diferença na expressão deste gene entre os pacientes, duas vezes superior no paciente talassêmico β maior. Como os dados de literatura sugerem que este gene esteja associado com maior produção de α globina, o aumento da expressão observado poderia influenciar um aumento na produção desta proteína que irá precipitar nos eritrócitos, lesando as hemácias e provocando um quadro de hemólise mais acentuado no paciente de clínica mais grave. No momento, estamos aprofundando os estudos para verificar a função desse gene para tentar correlacioná-lo com o fenótipo da doença.

Assim, neste trabalho, identificamos a expressão diferencial de novos genes que podem atuar como moduladores genéticos nesta hemoglobinopatia, levando genótipos iguais a expressarem fenótipos diferentes. Esses resultados necessitam de um melhor aprofundamento, com experimentos que inibam e aumentem a expressão de cada um desses genes, em cultura de células β talassêmicas e normais, com o objetivo de avaliar sua função e indicar potenciais alvos terapêuticos.

CONCLUSÃO

O transcriptoma entre os fenótipos de talassemia β maior e intermediária é muito semelhante, uma vez que um pequeno grupo de genes foi encontrado como diferencialmente expresso.

Nesse grupo, demonstramos que existe uma possível relação entre alguns genes: o aumento de expressão dos genes *ABCB10* e *HMBS* no paciente talassêmico β intermediário sugere uma possível relação com o aumento da eficiência da produção de heme e consequente aumento dos níveis de hemoglobina. Já o aumento de expressão de genes relacionados a reparo de DNA no paciente com talassemia β maior (*APEX1* e *HMGB1*) corroboram o achado com os frequentes danos causados às células desses pacientes pelas sucessivas hemólises causadas por diferentes eventos como estresse oxidativo, precipitação de α globina e danos à membrana. Outros dois genes promissores também foram identificados nesse estudo, o gene *EYA3*, que se mostrou mais expresso no paciente talassêmico β intermediário e parece ter uma relação importante com a diferenciação eritróide e o gene *PCBP2*, que tem influência na estabilidade do RNA mensageiro de α globina.

Assim os resultados desse trabalho contribuem para um melhor entendimento dessa hemoglobinopatia e sugerem novos genes que podem estar envolvidos nos diferentes fenótipos observados.

REFERÊNCIAS BIBLIOGRÁFICAS

Ashburner M.; Ball C.A.; Blake J.A.; Botstein D.; Butler H.; Cherry J.M. et al. Geneontology: toll for the unification of biology. The gene ontology consortium. **Nat Genet.** 2000; 25 (25-29).

Audic, S.; Claverie, J.M. The significance of digital gene expression profiles. **Genome Res.** 1997; 7:986-95.

Bank, A. The thalassemia syndrome. **Blood.** 1978; (51):369-384.

Bonaldi, T.; Langst, G.; Strohn, R.; Becker, P.B.; Bianchi, M.E. The DNA chaperone HMGB1 facilitates ACF/CHRAC-dependent nucleosome sliding. **EMBO J.** 2002; 21:6865-6873.

Boon K.; Osorio E.C.; Greenhut S.F.; Schafer C.F.; Shoemaker J.; Polyak K. et al. An anatomy of normal and malignant gene expression. **Proc Natl Acad Sci USA.** 2002; 99(17): 11287-92.

Bunn, H.F.; Forget, B.G. Molecular Genetics and Biosynthesis of Hemoglobin. In: Bunn, H.F.; Forget, B.G. **Hemoglobin: Molecular, Genetic and Clinical Aspects.** 1986; 1st edition, Saunders Company, pp 169-222.

Camaschella, C.; Cappellini, M.D. Thalassemia intermedia. **Haematologica.** 1995; v80, pp 58-78.

Cao, A. Phenotype-Genotype Relationships in Mendelian Disorders: the Example of β -Thalassemias. **Italian Journal of Pediatrics.** 2002; v28, n 6, pp 440-452.

Cao, A.; Galanello, R.; rosatelli, M.C. Genotype-phenotype correlations in β -thalassemias. **Blood Reviews.** 1994; v8, pp 1-12.

Cao, A.; Saba, L.; Galanello, R.; Rosatelli, M.C. Molecular diagnosis and carrier screening for β -thalassemia. **JAMA.** 1997; v278, pp 1273-1277.

Cappellini, M.D.; Comino, A. In: Clinica e Terapia della Talassemia. **Società Editrice Europea di Nicodemo Maggiulli & C.**, Firenze, Itália. 2000; Cap 3, pp 61-94.

Chen, W.; Paradkar, P.N.; Li, L.; Pierce, E.L.; Langer, N.B.; Takahashi-Makise, N.; Hyde, B.B.; Shirihai, O.S.; Ward, D.M.; Kaplan, J.; Paw, B.H. Abcb10 physically interacts with mitoferrin-1 (Slc25a37) to enhance its stability and function in the erythroid mitochondria. **PNAS.** 2009; vol 106, 38:16263-16268.

Chen, W.; Dailey, H.A.; Paw, B.H. Ferrochelatase forms an oligomeric complex with mitoferrin-1 and Abcb10 for erythroid heme biosynthesis. **Blood.** 2010; 116(4):628-30.

Chretien, S.; Dubart, A.; Beaupain, D.; Raich, N.; Grandchamp, B.; Rosa, J.; Goossens, M.; Romeo, P-H. Alternative transcription and splicing of the human porphobilinogen

deaminase gene result either in tissue-specific or in housekeeping expression. **Proc. Natl. Acad. Sci.** 1988; 85:6-10.

Costa, F.F.; Tavella, M.H.; Zago, M.A. β -Thalassemia Intermedia and IVS-1 NT6 Homozygosis in Brazil. **Braz. J. Med. Biol. Res.** 1991; v24, pp 157-161.

Croissille, L.; Chiron, M.; Vainchenker, W.; Romeo, P. H.; Rince, P. et al. Diamond-blackfan anemia with mutations in RPS 19 gene: a different pattern of in vitro erythroid differentiation? **Blood.** 1999; 94(10) suppl 1, 414a.

Croop, J.M. Evolutionary relationships among ABC transporters. **Methods Enzymol.** 1998; 292:101-116.

Cunha, A.F.; Brugnerotto, A.F.; Duarte, A.S.; Lanaro, C.; Costa, G.G.L.; Saad, S.T.O.; Costa, F.F. Global gene expression reveals a set of new genes involved in the modification of cells during the erythroid differentiation. **Cell Proliferation.** 2010; 43:297-230.

Forget, B.G.; Higgs, D.R.; Nagel, R.N. Disorders of Hemoglobin: Genetics, Pathophysiology and Clinical Management. **Cambridge University Press: NY, USA.** 2001; Cap 12, pp 252-276.

Flanagan, J.M.; Steward, S.; Hankins, J.S.; Howard, T.M.; Neale, G.; Ware, R.E. Microarray analysis of liver gene expression in iron overloaded patients with sickle cell disease and beta-thalassemia. **Am J. Hematol.** 2009; 84:328-334.

Galanello, R.; CAO, A. Relationship Between Genotype and Phenotype. **Annals New York of Sciences.** 1998; v850, pp 325-333.

Grandchamp, B. Acute intermittent porphyria. **Semin. Liver Dis.** 1998; 18(1):17-24.

Hanauske-Abel, H.M.; Slowinska, B.; Zagulska, S.; *et al.* Detection of a sub-set of polyssomal mRNAs associated with modulation of hypusine formation at the G1-S boundary. Proposal of a role for eIF-5A in onset of DNA replication. **FEBS Lett.** 1995; 366:92-98.

Higgins, C.F. ABC transporters: from microorganisms to man. **Annu. Rev. Cell Biol.** 1992; 8:67-113.

Higgs, D.R. α -Thalassemia. In: **Baillière's Clinical Haematology.** 1993; v6, n1, pp 117-150.

Ho, P.J.; Hall, G.W.; Luo, L.Y.; Weatherall, D.J.; Thein, S.L. β -thalassemia intermedia: is it possible consistently to predict phenotype from genotype? **British Journal of Haematology.** 1998; v100, pp 70-78.

Hoffman, R.; Benz, E.J.; Shattil, A.J.; Furie, B.; Cohen, H.J. Hematology Basic Principles and Practice. **Churchill Livingstone Inc.** 1991; pp252-300.

Jao, D.L.; Chen, K.Y. Tandem affinity purification revealed the hypusine-dependent binding of eukaryotic initiation factor 5A to the translation 80S ribosomal complex. **J Cell Biochem.** 2006; 97:583-598.

Jemc, J., and Rebay, I. **Annu. Rev. Biochem.** 2007; 76: 513–53.

Krishnan, N., Jeong, D.G., Jung S.K., Ryu, S.E., Xiao, A., Allis, C.D., Kim, S.J. and Tonks, N.K. Dephosphorylation of the C-terminal Tyrosyl Residue of the DNA Damage-related Histone H2A.X is Mediated by the Protein Phosphatase Eyes Absent. **The Journal of Biological Chemistry.** 2009; 24:16066-16070.

Lange, S.S.; Reddy, M.C.; Vasquez, K.M. Human HMGB1 directly facilitates interactions between nucleotide excision repair protein on triplex-directed psolaren interstrand crosslink. **DNA Repair (Amst).** 2009; 8(7):865-872.

Li, A.L.; Li, H.Y.; Jin, B.F.; *et al.* A novel eIF5A complex functions as a regulator of p53 and p53-dependent apoptosis. **J Biol Chem.** 2004; 279:49251-49258.

Lipowsky, G.; Bischoff, F.R.; Schwarzmaier, P.; *et al.* Exportin 4: A mediator of nuclear export pathway in higher eukaryotes. **EMBO J.** 2000; 19:4362-4371.

Margulies, E. H.; Innis, J.W. eSAGE: managing and analyzing data generated with serial analysis of gene expression (SAGE). **Bioinformatics.** 2000; 16(7): 650-1.

Moore, C.C.; Martin, E.N.; Lee, G.; *et al.* Eukaryotic translation initiation factor 5A small interference RNA-liposome complexes reduce inflammation and increase survival in murine models of severe sepsis and acute lung injury. **J. Infect Dis.** 2008; 198:1407-1414.

Nagababu, E.; Rifkind, J.M. Heme degradation during autooxidation of oxyhemoglobin. **Biochem. Biophys. Res. Commun.** 2000; 273:839-845.

Olivieri, N.F. The beta-thalassemias. **New England Journal of Medical.** 1999; v341, pp 99-109.

Owens, M. A. & Loken, M. R. flow cytometry principles for clinical laboratory practice. New York – EUA. **Ed. Wile-Liss.** 1995; 224p.

Park, M.H.; Joe, Y.A.; Kang, K.R. Deoxyhypusine synthase activity is essential for cell viability in the yeast *saccharomyces cerevisiae*. **J Biol Chem** 1998; 273:1677-1683.

Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. **Nucleic Acids Res.** 2001; 29(9): e45.

Pontius, J.U.; Wagner L.; Schuler G.D. UniGene: a unified view of the transcriptome. In: Information NCfB. **The NCBI Handbook.** 2003; Bethesda (MD).

Rayapureddi, J. P., Kattamuri, C., Steinmetz, B. D., Frankfort, B. J., Ostrin, E. J., Mardon, G., and Hegde, R. S. **Nature**. 2003; 426:295–298.

Rosorius, O.; Reichart, B.; Kratzer, F.; *et al.* Nuclear pore localization and nucleocytoplasmic transport of eIF-5A: Evidence for direct interaction with the export receptor CRM1. **J Cell Sci**. 1999; 112:2369-2380.

Shirihai, O.S.; Gregory, T.; Xu, C.; Orkin, S.H.; Weiss, M.J. ABC-me: a novel mitochondrial transporter induced by GATA-1 during erythroid differentiation. **EMBO J**. 2000; v19, 11:2492-2502.

Singh, J.; Dixon, G.H. High mobility group proteins 1 and 2 function as general class II transcription factors. **Biochemistry** 1990; 29:6295–6302.

Söker Torben, and Dalke Claudia, and Puk Oliver, and Floss Thomas, and Becker Lore, and Bolle Ines, and Favor Jack, and Hans Wolfgang, and Hölter Sabine, and Horsch Marion, and Kallnik Magdalena, and Kling Eva, and Moerth Corinna, and Schrewe Anja, and Stigloher Christian, and Topp Stefanie, and Gailus-Durner Valerie, and Naton Beatrix, and Beckers Johannes, and Fuchs Helmut, and Ivandic Boris, and Klopstock Thomas, and Schulz Holger, and Wolf Eckhard, and Wurst Wolfgang, and Bally-Cuif Laure, and de Angelis, Martin and Graw, Jochen. Pleiotropic effects in Eya3 knockout mice. **BMC Developmental Biology**. 2008; 8:118.

Steinberg, M.H.; Forget, B.G.; Higgs, D.R.; Nagel, R.L. Disorders of Hemoglobin. **Genetic, Pathophysiology and Clinical Manegement**, Cambridge University Press, pp 231-356, 2001.

Taher, A.; Isma'eel, H.; Cappellini, M.D. Thalassemia Intermedia: Revisited. **Blood Cells, Molecules and Diseases**. 2006; V37, pp 12-20.

Tamagnini, G.P.; Lopes, M.C.; Castanheira, M.E.; Wainscoat, J.S. β thalassemia – Portuguese type: clinical, haematological and molecular studies of a newly defined form of β thalassemia. **Bristish Journal of Hematology**. 1993; v54, pp 189-200.

Terskikh, A.V., et al. Gene expression analysis of purified hematopoietic stem cells and committed progenitors. **Blood**. 2003; 102: 94-101.

Thein, S.L. β -Thalassemia. In: Rodgers, G.P. Baillière's Clinical Haematology. **Bailliere Tindall**. 1998; v11, pp 91-126.

Thein, S.L. Genetic insigths into clinical diversity of β thalassemia. **Bristish Journal of Haematology**. 2004; v124, pp 264-274.

Thein, S.L. Genetic Modifiers of β -Thalassemia. **Haematologica**. 2005; v90, pp 649-660.

- Tome, M.E.; Gerner, E.W. Hypusine modification in eukaryotic initiation factor 5A in rodent cells selected for resistance to growth inhibition by ornithine decarboxylase-inhibiting drugs. **Biochem J.** 1996; 320(Part1):55-60.
- Tootle, T. L., Silver, S. J., Davies, E. L., Newman, V., Latek, R. R., Mills, I. A., Selengut, J. D., Parlikar, B. E., and Rebay, I. **Nature.** 2003; 426: 299–302.
- Treisman, R.; Orkin, S.H.; Maniatis, T. Specific transcription and RNA splicing defects in five cloned β thalassemia genes. **Nature.** 1983; v302, pp 591-596.
- Velculescu, V.E.; Zhang, L.; Vogelstein, B.; Kinzler, K.W. Serial Analysis of Gene Expression. **Science.** 1995; 270(5235):484-7.
- Wang, X.; Kiledjian, M.; Weiss, I.M.; Liebhaber, S.A. Detection and characterization of a 39 untranslated region ribonucleoprotein complex associated with the α -globin mRNA stability. **Mol Cell Biol.** 1995.
- Wang, H.; Bloom, O.; Zhang, M.; Vishnubhakat, J.M.; Ombrellino, M.; Che, J.; Frazier, A.; Yang, H.; Ivanova, S.; Borovikova, L.; Manogue, K.R.; Faist, E.; Abraham, E.; Andersson, J.; Andersson, U.; Molina, P.E.; Abumrad, N.N.; Sama, A.; Tracey, K.J. HMG-1 as a late mediator of endotoxin lethality in mice. **Science.** 1999; 285:248–251.
- Weatherall, D.J.; Clegg, J.B. The Thalassemia Syndromes. **Ed Oxford: Blackwell Scientific Publications**, 1981; 3rd.
- Weatherall, D.J. The Thalassemias. In: Williams, W.J.; Beutler, E.; Erslev, A.J.; Rundles, R.W. Hematology. **Ed New York: McGraw Hill Book Company**, 1990; 4th.
- Weatherall, D.J.; Clegg, J.B. The thalassemia syndrome. **Ed Oxford: Blackwell Scientific Publications**, 2001; 4th.
- Weiss, M.J.; Keller, G.; Orkin, S.H. Novel insights into erythroid development revealed through *in vitro* differentiation of GATA-1 embryonic stem cells. **Gens Dev.** 1994; 8:1184-1197.
- Weiss, I.M.; Liebhaber, S.A. Erythroid cell-specific determinants of a α -globin mRNA stability. **Mol Cell Biol.** 1994; 8123-8132.
- Weiss, I.M.; Liebhaber, S.A. Erythroid cell-specific mRNA stability elements in the α -globin 39 nontranslated region. **Mol Cell Biol.** 1995; 2457-2465.
- Weiss, M.J.; Yu, C.; Orkin, S.H. Erythroid cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell-line. **Mol Cell Biol.** 1997; 17:1642-1651.
- Wickramasinghe, S.N.; Hughes, M. Ultrastructural studies of erythropoiesis in β -thalassemia trait. **British Journal of Haematology.** 1980; (46):401-407.

Zago, M.A. Talassemias. In: Zago, M.A.; Falcão, R.P.; Pasquini, R. Hematologia Fundamentos e Prática. São Paulo: **Ed Atheneu**. 2004; Cap. 31, pp 309-328.

Zou, G.M.; Luo, M.H.; Reed, A.; Kelly, M.R.; Yoder, M.C. Ape1 regulates hematopoietic differentiation of embryonic stem cells through its redox functional domain. **Blood**. 2007; vol 109, 5:1917-1922.

Global gene expression reveals a set of new genes involved in the modification of cells during erythroid differentiation

A. F. da Cunha*, A. F. Brugnerotto†, A. S. Duarte†, C. Lanaro†, G. G. L. Costa†, S. T. O. Saad† and F. F. Costa†

*Departamento de Genética e Evolução, Centro de Ciências Biológicas e da Saúde, Universidade Federal de São Carlos, São Carlos, São Paulo, Brazil and †Centro de Hematologia e Hemoterapia, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

Received 14 July 2009; revision accepted 2 September 2009

Abstract

Objectives: Erythroid differentiation is a dynamic process in which a pluripotent stem cell undergoes a series of developmental changes that commit it to a specific lineage. These alterations involve changes in gene expression profiles. In this study, gene expression profiles during differentiation of human erythroid cells of a normal blood donor were evaluated using SAGE.

Materials and methods: Global gene expression was evaluated in cells collected immediately before addition of erythropoietin (0 h) and 192 and 336 h after addition of this hormone. Real-time PCR was used to evaluate activation of differentially expressed genes.

Results: The data indicate that global aspects of the transcriptome were similar during differentiation of the majority of the genes and that a relatively small set of genes is probably involved in modification of erythroid cells during differentiation. We have identified 93 differentially expressed genes during erythroid development, and expression of some of these was confirmed by qPCR. Various genes including *EYA3*, *ERH*, *HES6*, *TIMELESS* and *TRIB3* were found to be homologous to those of *Drosophila melanogaster* and here are described for the first time during erythroid development. An important and unique carboxypeptidase inhibitor described in mammals, *LXN*, was also identified.

Conclusions: The results of this study amplify previously published data and may contribute to comprehension of erythroid differentiation and identification

of new target genes involved in some erythroid concerning diseases.

Introduction

Haematopoiesis is maintained by pluripotent, long-term repopulating stem cells that generate progenitors capable of differentiating into all three haematopoietic lineages. Erythroid cell maturation, known as erythropoiesis, is mediated by a combination of regulatory proteins acting in concert. These direct development of progenitor cells into mature erythrocytes, which are one of the most highly specialized cell types in the human body (1,2). This process can be reproduced in an *in vitro* study using a two-phase liquid culture. Using this technique, stem cells differentiate into erythroid cells by addition of the hormone erythropoietin (EPO) in the culture (3).

Extensive studies have led to a considerable understanding of the cellular and molecular control of haemoglobin production during red blood cell differentiation (4–7); however, identification of the genes expressed as part of the erythroid differentiation programme remains an important goal because of the insights that these data will bring to erythrocyte biology and disease (8). One of the first studies evaluating gene expression in human erythroid cells was carried out by Gubin *et al.* (9). These authors made a subtractive library before and after addition of erythropoietin in a two-phase liquid culture, and obtained a transcriptional profile of genes arising only in response to EPO. Following this study, several related ones on similar themes evaluated global gene expression in haematopoietic stem cells (10–13) and in reticulocytes (14,15).

Using a microarray strategy, Komor *et al.* (16) evaluated gene expression during differentiation of erythroid cells, megakaryocytes and platelets. This work identified several genes that were differentially expressed during differentiation of cells. However, during microarray analysis, knowledge of presence and sequence of genes to be

Correspondence: A. F. da Cunha, Departamento de Genética e Evolução, Centro de Ciências Biológicas e da Saúde, Universidade Federal de São Carlos, Rodovia Washington Luis km 235, 13565-905, São Carlos, São Paulo, Brazil. Tel./Fax: +55 16 3351 8377; E-mail: andferf@ufscar.br

analysed is required and, thus, only genes spotted on slides are studied, making it difficult to find new genes not evaluated in the analysis (17).

Although several studies have been performed on haematopoietic cells, global gene expression during erythroid differentiation has been poorly evaluated. As such, identification of all genes expressed as part of the erythroid differentiation programme remains an important goal (8).

Here, we report global gene expression during differentiation of human erythroid cells from a normal blood donor, in a two-phase liquid culture, using Serial Analysis of Gene Expression (SAGE) (18). Global gene expression was evaluated in cells collected immediately before addition of erythropoietin (0 h) and 192 and 336 h after addition of this hormone. We identified 93 differentially expressed genes and development and expression of some of these genes was confirmed by qPCR.

Our data amplify previously published research and will contribute to understanding the pattern of gene expression during erythroid differentiation. In addition, these results contribute to the comprehension of erythroid differentiation and identification of new target genes involved in haematopoietic diseases.

Materials and methods

Erythroid cell cultures

Blood from normal volunteers was cultured using a two-phase liquid culture procedure, as described previously (3). Briefly, mononuclear cells were isolated from peripheral blood samples by centrifugation over a Ficoll-Hypaque gradient and cultured for 7 days (phase I) in IMDM medium (Invitrogen, Rockville, MS, USA) supplemented with 20% foetal calf serum (Invitrogen), 1 µg/ml cyclosporin A (Sandoz, Holzkirchen, Germany) and 10% conditioned medium, collected from culture of the human bladder carcinoma 5637 cell line. Cells were incubated at 37 °C in an atmosphere of 5% CO₂ and 92% extra humidity. After 7 days, non-adherent cells were harvested and re-cultured in phase II medium, IMDM supplemented with 30% foetal calf serum (Invitrogen), 1% deionized bovine serum albumin (BSA; Sigma, St Louis, MO, USA), 10⁻⁵ M 2-mercaptoethanol (Sigma), 1.5 nmol/l glutamine (Invitrogen), 300 µg iron-saturated transferrin (Sigma), 10⁻⁶ M dexamethasone, 5 ng/ml human stem cell factor (SCF; Calbiochem, Darmstadt, Germany), 1 U/ml human recombinant erythropoietin (Cilag, Beerse, Belgium), 2.5 µg/ml funzigone (Invitrogen), 50 µg/ml streptomycin (Invitrogen) and 25 µg/ml glutamicin (Invitrogen). Cell samples were collected from phase II cultures at 0, 192 and 336 h after erythropoietin addition. Cell numbers and viability were determined by

trypan blue exclusion. Samples of 5×10^6 cells were pelleted and resuspended in Trizol (Invitrogen) and stored at -80 °C for total RNA extraction and cDNA synthesis. For morphological analyses of cell differentiation stages, cytopsin slides were prepared and stained with Leishman's stain before examination using an Eclipse E-600 microscope (Nikon, Tokyo, Japan) with Image Pro-Express 4.0 software (Media Cybernetic, Bethesda, MD, USA).

RNA extraction

Total RNA was extracted with TRIzol reagent (Invitrogen, Rockville, MS, USA), according to the manufacturer's protocol. Samples were quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc, Wilmington, DE, USA).

SAGE libraries and data analysis

Libraries were constructed using the I-SAGE kit (Invitrogen) with *Nla* III enzyme, as described by the manufacturer. To produce libraries, 10 µg of total RNA was prepared. Sequencing was carried out in a Dynamic ET Terminator cycle sequencer (GE Healthcare, Uppsala, Sweden) and MEGA-BACE automated DNA sequencer (Amersham Pharmacia, Bucks, UK). Vector sequences were trimmed with *Phred/Phrap* software. Automatic tag detection and differential gene expression analyses were performed using *eSAGE* software v1.2 (19). Only tags presenting $P < 0.01$ and fold ≥ 10 between comparisons were considered to be differently expressed. Data bank 'Best Gene for a tag', from SAGEGenie, CGAP (<http://cgap.nci.nih.gov/SAGE>), downloaded on April 2007, was used for tag-to-gene mapping. According to their identification, tags were further classified as 'no match' (no correspondence found in the data bank), 'known genes' or 'putative genes/proteins', including ESTs (expressed sequence tags), ORFs (open reading frames), cDNA clones and hypothetical proteins. Functional classification of transcripts was performed according to Gene Ontology Consortium criteria (<http://www.geneontology.org>). Hierarchical clustering analysis by Spearman's confidence correlation was used to identify gene clusters. The separation ratio was set at 0.5.

Quantitative real time polymerase chain reaction

RNA samples were subjected to DNase I treatment (Invitrogen) and reverse transcription using SuperScript III (Invitrogen). Primers were designed using PrimerExpress™ programme (Applied Biosystems, Foster City, CA, USA) (Table S1). Ideal concentration for use was

determined for each pair of primers and amplification efficiency was calculated according to the equation $E^{(-1/\text{slope})}$, to confirm accuracy and reproducibility of the reactions (Table S1). Amplification specificity was verified by running a dissociation protocol. Quantitative real time polymerase chain reaction (qRT-PCRs) were performed in duplicate, using 12.5 μ l SYBR Green Master Mix (Applied Biosystems), 25 ng cDNA and ideal quantities of each primer, in a final volume of 25 μ l. Samples were run in MicroAmp Optical 96-well plates (Applied Biosystems) in a 5700 Sequence Detection System (Applied Biosystems). To validate SAGE profiles, *GAPDH* was used as a reference gene. Gene expressions in SAGE samples are presented as mean $\pm$ SEM.

Results

We performed a large-scale gene expression study of erythroid differentiation using SAGE. Samples of cultures were collected at 0 (SAGE-0H), 192 (SAGE-192H) and 336 (SAGE-336H) hours after erythropoietin addition and typical morphology was detectable during cell differentiation (Fig. S1). Cells were collected at these points and their RNA was prepared for SAGE library construction.

After sequencing and tag extraction, 30 512 tags for SAGE-0H, 30 117 tags for SAGE-192H and 30 189 for SAGE-336H profiles were generated, representing 12 026, 11 709 and 11 337 unique tags respectively. Identification of tags in the libraries demonstrated that 28%, 26.2% and 26.7% respectively, had no correspondence in the data bank (no matches) and could represent novel genes. A complete list of tags is available for download at <http://www.lge.ibi.unicamp.br/~anderf>.

To investigate reliability of the profiles designed by SAGE, we arbitrarily selected 18 genes to be studied by qRT-PCR in the same samples used to generate the libraries. Both techniques were consistent in identifying expression of 17 of 18 genes studied (*HBA*, *HBB*, *HBG*, *RNAseI*, *TIMP1*, *TIMP2*, *LYZ*, *B2M*, *MMP9*, *NFE2*, *AHSP*, *SI00A8*, *SI00A9*, *BCR*, *GATA1*, *PFN1* and *CEBPB*). Only expression of the *STAT5A* gene demonstrated discordant results between the techniques (Fig. 1).

For subsequent analysis, only tags present at least five times in one of the libraries were considered (20–22). Using these data, expression profiles of libraries were compared using the Gene Ontology Consortium Database. Most abundant genes expressed at the beginning of differentiation were found to be related to various pathways including immune response, lysozyme activity, iron homeostasis, cell proliferation and apoptosis. At 192 h after erythropoietin addition, the most abundant genes were related to ribosomal activity, reflecting intense and dynamic protein production in this intermediate phase. At

the end of differentiation we observed high expression of genes involved in haemoglobin synthesis, such as *HBA*, *HBB* and *HBG*, and these represent the most expressed proteins in reticulocytes and in red cells. Summaries of the most expressed genes in each library are described in Table 1.

Differential gene expression between the libraries was further analysed using $P < 0.01$ criterion, and fold higher than 10 to select tags that presented differential expression with a statistically significant level. Ninety-three genes were identified and these were hierarchically clustered by Spearman's confidence correlation, with a separation ratio set at 0.5. We identified 32 up-regulated genes in the 0H library, 29 in 196H and 32 in 336H (Fig. 2). Tag number found for each gene is displayed in Table 2. Differentially expressed genes were categorized by Molecular Function and Biological Process using the gene ontology consortium. At the beginning of differentiation (0H), processes such as cell adhesion, cell proliferation, cell development and apoptosis regulation were found to be up-regulated. After 192 h of erythropoietin addition, processes like structural constituents of ribosomes, transcription factor activity and RNA polymerase II activity were up-regulated. At the end of differentiation, these processes were down-regulated and cells demonstrated restriction of expression of pathways, like transport, biosynthetic processes, oxygen binding pathways plus ion, tetrapyrrole, nucleotide, protein and cofactors (Fig. S2).

Of the differentially expressed genes, we found several with homology to *Drosophila melanogaster* genes (Fig. 3). These genes have been identified in humans, however, most of them do not have any function yet described. We also found high expression of *TIMELESS*, *HES6*, *EYA3*, *ERH* and *TRIB3* genes during the intermediate phase and at the end of differentiation.

To understand whether expressions of these genes are related to erythroid lineage expression, we evaluated them in further two-phase liquid cultures (Fig. 4a) and in CD34+ culture (Fig. 4b). CD34+ cell culture was used as contamination with other cell types such as lymphocytes and monocytes/macrophages is lower than that seen in two-phase culture and all cells are committed to the erythroid lineage. Results confirmed SAGE data in both cultures and demonstrated that probably, differences observed are related to erythroid lineage and not to other cell types. We also evaluated expression of *LXN* gene, the only known carboxypeptidase inhibitor in mammals (23), because its expression was observed only after the intermediate stage of differentiation and was lower at the end of differentiation.

To verify whether expressions of these genes were ubiquitous, we also evaluated them in several tissues using a cDNA tissue library (Clontech Laboratories Inc.,

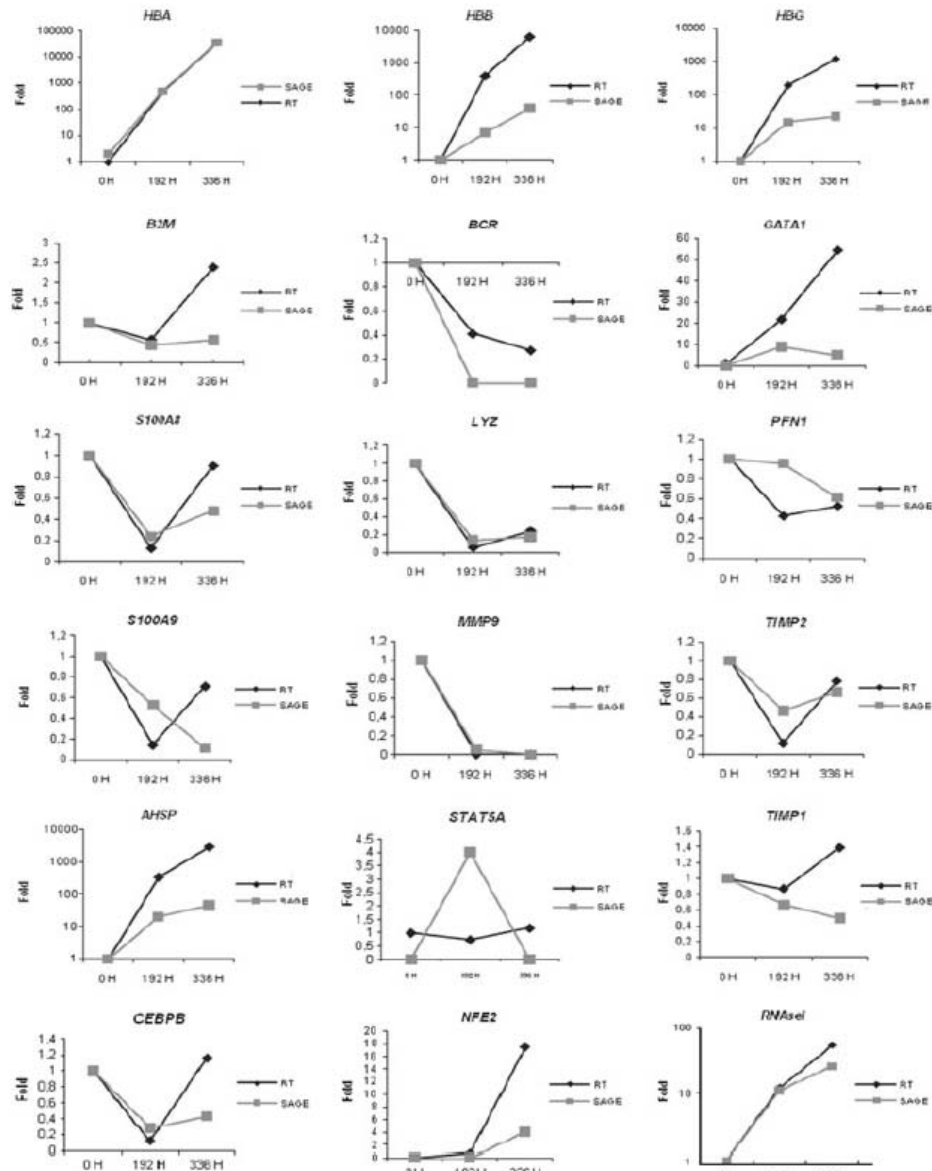


Figure 1. Validation of SAGE technique – eighteen genes arbitrarily selected for study by qRT-PCR in the same samples used to generate libraries. Results showed a 95% concordance (17 of 18).

Mountain View, CA, USA). We observed high expression of *EYA3* and *LXN* in bone marrow, and for genes *ERH*, *TRIB3* and *TIMELESS*, we observed high expression in other haematopoietic islands such as placenta and liver. Exceptionally, expression of *HES6* gene was not observed in these tissues and highest expression was observed in intestine and brain (Fig. 5).

Discussion

Gene expression during erythroid differentiation is poorly understood. Study of the global pattern of gene expression that accompanies erythroid differentiation could help improve understanding of erythroid-specific mechanisms

Table 1. The most expressed genes (more than 100 copies) in 0H, 192 and 336H library respectively

Tag	0H	192H	336H	Hs	Symbol	Description	Ontology
GGGCATCTCT	203	32	39	Hs.520048	HLA-DRA	Major histocompatibility complex, class II, DR alpha	Immune response
ATCAAGAATC	238	26	27	Hs.14623	IFI30	Interferon, gamma-inducible protein 30	Lysozyme activity
ATGTAAAAAA	253	37	45	Hs.524579	LYZ	Lysozyme (renal amyloidosis)	Lysozyme activity
GTTGTGGTTA	304	131	166	Hs.534255	B2M	Beta-2-microglobulin	Immune response
CCCTGGGTTC	361	115	156	Hs.433670	FTL	Ferritin, light polypeptide	Cellular iron ion homeostasis
TTGGGGTTTC	386	274	458	Hs.524910	FTH1	Ferritin, heavy polypeptide 1	Cellular iron ion homeostasis
GTTCACATTA	420	99	116	Hs.436568	CD74	CD74 molecule, major histocompatibility complex, class II invariant chain	Cell proliferation/negative regulation of apoptosis/ signal transduction
GAAATACAGT	648	121	148	Hs.67201	NT5C	5', 3'-nucleotidase, cytosolic	5'-nucleotidase activity
GGATTTGGCC	174	152	101	Hs.437594	TSPAN4	Tetraspanin 4	Membrane fraction
CACAAACGGT	140	66	71	Hs.504517	TSPAN9	Tetraspanin 9	Membrane fraction
TTGGTGAAGG	169	18	45	Hs.522584	TMSB4X	Thymosin, beta 4, X-linked	Cytoskeleton organization and biogenesis
CTGACCTGTG	128	26	50	Hs.77961	HLA-B	Major histocompatibility complex, class I, B	Immune response
CCACTGCACT	139	22	46	Hs.107003	CCNB1IP1	Cyclin B1 interacting protein 1	Apoptosis
ACATTCTTTT	103	21	41	Hs.190495	GPNMB	Glycoprotein (transmembrane) nmb	Negative regulation of cell proliferation
AGGGCTTCCA	105	99	43	Hs.534404	RPL10	Ribosomal protein L10	Ribosomal subunit
GTGAAACCCC	107	62	86	Hs.590913	PAFAH2	Platelet-activating factor acetylhydrolase 2, 40 kDa	Phospholipid binding
AGTTTCTTGT	108	33	45	Hs.647419	CD68	CD68 molecule	Transmembrane glycoprotein
CCTGTAATCC	108	34	50	Hs.591920	NT5C2	5'-nucleotidase, cytosolic II	5'-nucleotidase activity
CCCATCGTCC	192	279	102	Hs.559716		Transcribed locus, weakly similar to XP_220207.3 similar to serine/ arginine repetitive matrix 2 [<i>Rattus norvegicus</i>]	RNA/protein binding
GAGGGAGTTT	151	191	98	Hs.523463	RPL27A	Ribosomal protein L27a	Ribosomal subunit
GAAAAATGGT	80	183	80	Hs.449909	RPSA	Ribosomal protein SA	Ribosomal subunit
GCATAATAGG	139	178	89	Hs.381123	RPL21	Ribosomal protein L21	Ribosomal subunit
CTGGGTAAAT	96	172	135	Hs.438429	RPS19	Ribosomal protein S19	Ribosomal subunit
ATAATTCTTT	153	174	120	Hs.156367	RPS29	Ribosomal protein S29	Ribosomal subunit
GGGCTGGGGT	85	163	72	Hs.425125	RPL29	Ribosomal protein L29	Ribosomal subunit
TTGGTCTCT	116	149	105	Hs.632703	RPL41	Ribosomal protein L41	Ribosomal subunit
TTCAATAAAA	92	145	71	Hs.356502	RPLP1	Ribosomal protein, large, P1	Ribosomal subunit
CAATAAATGT	91	141	71	Hs.558601	RPL37	Ribosomal protein L37	Ribosomal subunit
TGCACGTTTT	92	141	70	Hs.265174	RPL32	Ribosomal protein L32	Ribosomal subunit
TGTGTTGAGA	125	135	80	Hs.644639	EEF1A1	Eukaryotic translation elongation factor 1 alpha 1	Translational elongation/GTPase activity
TAATAAAGGT	72	130	65	Hs.512675	RPS8	Ribosomal protein S8	Ribosomal subunit
TGTACCTGTA	43	111	61	Hs.524390	TUBA3	Tubulin, alpha 3	Microtubule-based movement/GTPase activity
GCAAGAAAGT	36	253	1391	Hs.523443	HBB	Haemoglobin, beta	Haemoglobin synthesis
CTTCTTGCCC	20	147	1264	Hs.449630	HBA1	Haemoglobin, alpha 1	Haemoglobin synthesis
CCCAACGCGC	7	26	473	Hs.449630		Haemoglobin, alpha 1	Haemoglobin synthesis
TAGGTTGTCT	198	191	211	Hs.374596	TPT1	Tumour protein, translationally controlled 1	Anti-apoptosis/cellular calcium ion homeostasis
ATGCAGAGCT	4	120	178	Hs.295459	HBG1	Haemoglobin, gamma A	Haemoglobin synthesis
ATTCAGAGCT	2	105	154	Hs.295459		Haemoglobin, gamma A	Haemoglobin synthesis
TTAACCCCTC	5	56	130	Hs.78224	RNASE1	Ribonuclease, RNase A family, 1 (pancreatic)	RNA binding/ endonuclease activity

that are required for optimal function of erythrocytes and therefore, identify targets for treatment of erythrocyte disorders (8).

To understand this mechanism, global gene expression during erythroid differentiation was evaluated using SAGE. By this strategy, 93 genes were identified that

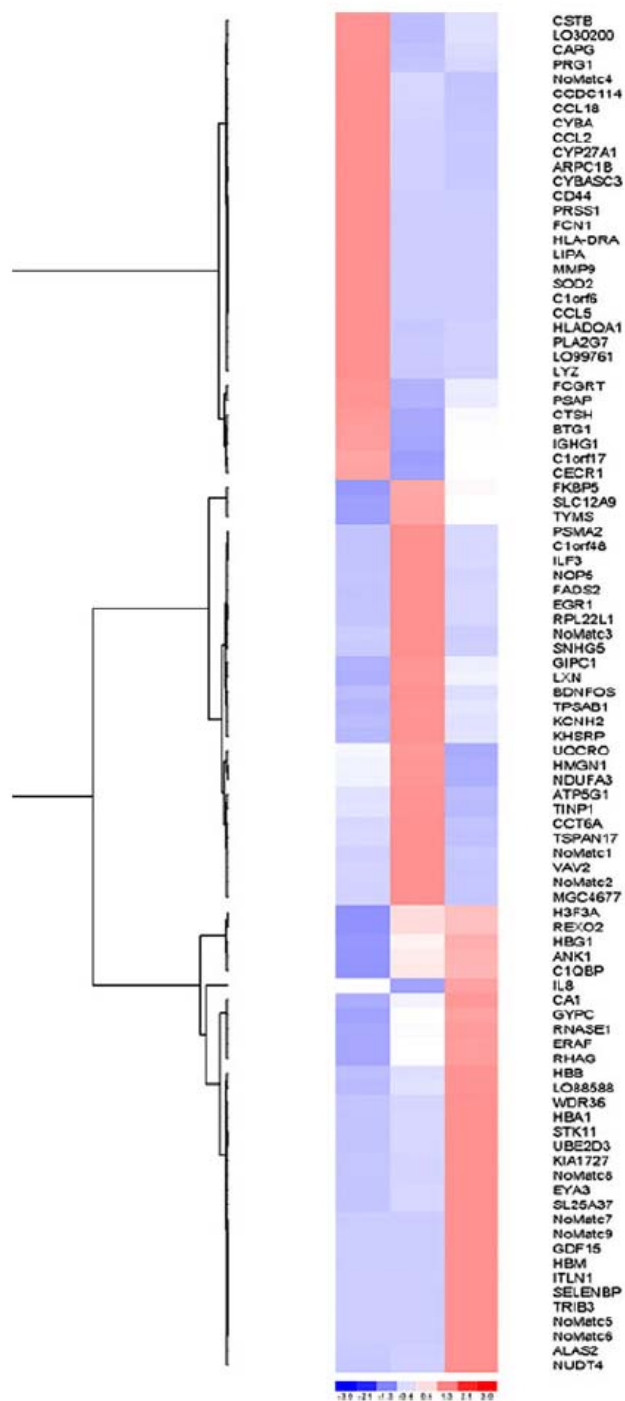


Figure 2. Cluster analysis of differentially expressed genes associated with erythroid differentiation. Three clusters were found according to up-regulation of each stage of development. Colour code: blue, low expression; red, high expression. Intensity of colour reflects reliability of expression data.

Table 2. Differentially expressed genes found during erythroid development

Gene symbol	Hs number	Description	Number of tags		
			0H	192H	336H
No Match 1			5	107	0
No Match 2			1	14	0
No Match 3			0	10	0
No Match 4	Hs.605719	CDNA clone IMAGE:3927515	10	1	0
No Match 5		Unclustered ESTs	0	0	25
No Match 6	Hs.623908	Transcribed locus, strongly similar to XP_001072910.1 similar to Oligodendrocyte transcription factor 3 (Oligo3) - (Oligodendrocyte-specific bHLH transcription factor 3) (Basic helix-loop-helix domain-containing class B protein 7) [<i>Rattus norvegicus</i>]	0	0	15
No Match 7			0	0	13
No Match 8		Unclustered ESTs	0	3	36
No Match 9			0	0	11
ALAS2	Hs.522666	Aminolevulinate, delta-, synthase 2 (sideroblastic/hypochromic anaemia)	0	2	71
ANK1	Hs.491558	Ankyrin 1, erythrocytic	0	18	25
ARPC1B	Hs.489284	Actin-related protein 2/3 complex, subunit 1B, 41 kDa	18	1	0
ATP5G1	Hs.80986	ATP synthase, H ⁺ transporting, mitochondrial F0 complex, subunit C1 (subunit 9)	2	10	0
BDNFOS	Hs.577179	Brain-derived neurotrophic factor opposite strand	2	20	5
BTG1	Hs.255935	B-cell translocation gene 1, anti-proliferative	38	5	19
C11orf17	Hs.131180	Chromosome 11 open reading frame 17	10	1	6
C19orf48	Hs.256301	Chromosome 19 open reading frame 48	0	9	1
C19orf6	Hs.515003	Chromosome 19 open reading frame 6	10	0	0
C1QBP	Hs.555866	Complement component 1, q subcomponent binding protein	3	12	15
CA1	Hs.23118	Carbonic anhydrase I	0	21	61
CAPG	Hs.516155	Capping protein (actin filament), gelsolin-like	41	3	8
CCDC114	Hs.112645	Coiled-coil domain containing 114	11	1	0
CCL18	Hs.143961	Chemokine (C-C motif) ligand 18 (pulmonary and activation-regulated)	12	1	0
CCL2	Hs.303649	Chemokine (C-C motif) ligand 2	17	1	0
CCL5	Hs.514821	Chemokine (C-C motif) ligand 5	11	0	0
CCT6A	Hs.82916	Chaperonin-containing TCP1, subunit 6A (zeta 1)	2	10	1
CD44	Hs.502328	CD44 molecule (Indian blood group)	10	0	0
CECR1	Hs.170310	Cat eye syndrome chromosome region, candidate 1	17	1	9
CSTB	Hs.695	Cystatin B (stefin B)	53	5	13
CTSH	Hs.148641	Cathepsin H	20	0	8
CYBA	Hs.513803	Cytochrome b-245, alpha polypeptide	14	1	0
CYBASC3	Hs.22546	Cytochrome b, ascorbate dependent 3	18	1	0
CYP27A1	Hs.516700	Cytochrome P450, family 27, subfamily A, polypeptide 1	19	1	0
EGR1	Hs.326035	Early growth response 1	0	11	1
ERAF	Hs.274309	Erythroid associated factor	1	20	45
EYA3	Hs.185774	Eyes absent homologue 3 (<i>Drosophila</i>)	1	7	69
FADS2	Hs.502745	Fatty acid desaturase 2	0	10	1
FCGR1	Hs.111903	Fc fragment of IgG, receptor, transporter, alpha	21	4	9
FCN1	Hs.440898	Ficolin (collagen/fibrinogen domain containing) 1	18	0	0
FKBP5	Hs.407190	FK506 binding protein 5	1	19	12
GDF15	Hs.616962	Growth differentiation factor 15	0	0	11
GIPC1	Hs.631639	GIPC PDZ domain containing family, member 1	1	11	4
GYPC	Hs.59138	Glycophorin C (Gerbich blood group)	0	10	20
H3F3A	Hs.533624	H3 histone, family 3A	11	23	25
HBA1	Hs.449630	Haemoglobin, alpha 1	0	176	1780
HBB	Hs.523443	Haemoglobin, beta	36	270	1489
HBG1	Hs.295459	Haemoglobin, gamma A	6	225	332
HBM	Hs.647389	Haemoglobin, mu	0	0	15
HLA-DQA1	Hs.387679	Major histocompatibility complex, class II, DQ alpha 1	39	3	5

Table 2. (Continued)

Gene symbol	Hs number	Description	Number of tags		
			0H	192H	336H
HLA-DRA	Hs.520048	Major histocompatibility complex, class II, DR alpha	10	1	1
HMGNI	Hs.356285	High-mobility group nucleosome-binding domain 1	5	13	1
IGHG1	Hs.510635	Immunoglobulin heavy constant gamma 1 (G1m marker)	11	2	6
IL8	Hs.443948	Interleukin 8	42	11	73
ILF3	Hs.465885	Interleukin enhancer binding factor 3, 90 kDa	1	10	2
ITLN1	Hs.50813	Intelectin 1 (galactofuranose binding)	0	0	10
KCNH2	Hs.647099	Potassium voltage-gated channel, subfamily H (eag-related), member 2	0	20	4
KHSRP	Hs.646750	KH-type splicing regulatory protein (FUSE binding protein 2)	1	10	3
KIAA1727	Hs.132629	KIAA1727 protein	0	1	13
LIPA	Hs.643030	Lipase A, lysosomal acid, cholesterol esterase (Wolman disease)	10	1	1
LOC388588	Hs.22047	Hypothetical gene supported by BC035379; BC042129	1	13	59
LOC399761	Hs.647203	Hypothetical protein LOC399761	23	0	1
LOC730200	Hs.553015	Hypothetical protein LOC730200	17	0	3
LXN	Hs.478067	Latexin	0	15	5
LYZ	Hs.524579	Lysozyme (renal amyloidosis)	280	38	49
MGC4677	Hs.446688	Hypothetical protein MGC4677	1	15	0
MMP9	Hs.297413	Matrix metalloproteinase 9 (gelatinase B, 92 kDa gelatinase, 92 kDa type IV collagenase)	16	0	0
NDUFA3	Hs.198269	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 3, 9 kDa	3	9	0
NOP5/NOP58	Hs.471104	Nucleolar protein NOP5/NOP58	2	15	3
NUDT4	Hs.591008	Nudix (nucleoside diphosphate-linked moiety X)-type motif 4	0	3	57
PLA2G7	Hs.584823	Phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)	30	1	2
PRG1	Hs.1908	Proteoglycan 1, secretory granule	12	1	2
PRSS1	Hs.622865	Protease, serine, 1 (trypsin 1)	23	1	1
PSAP	Hs.523004	Prosaposin (variant Gaucher disease and variant metachromatic leukodystrophy)	111	14	41
PSMA2	Hs.333786	Proteasome (prosome, macropain) subunit, alpha type, 2	1	10	2
REXO2	Hs.7527	REX2, RNA exonuclease 2 homologue (<i>S. cerevisiae</i>)	0	12	14
RHAG	Hs.120950	Rh-associated glycoprotein	0	12	28
RNASE1	Hs.78224	Ribonuclease, RNase A family, 1 (pancreatic)	5	56	130
RPL22L1	Hs.380933	Ribosomal protein L22-like 1	0	11	1
SELENBP1	Hs.632460	Selenium-binding protein 1	0	0	12
SLC12A9	Hs.521087	Solute carrier family 12 (potassium/chloride transporters), member 9	2	24	14
SLC25A37	Hs.122514	Solute carrier family 25, member 37	0	2	23
SNHG5	Hs.292457	Small nucleolar RNA host gene (non-protein coding) 5	3	27	4
SOD2	Hs.487046	Superoxide dismutase 2, mitochondrial	10	1	1
STK11	Hs.515005	Serine/threonine kinase 11	0	1	10
TINP1	Hs.482526	TGF beta-inducible nuclear protein 1	3	10	1
TPSAB1	Hs.405479	Tryptase alpha/beta 1	0	45	11
TRIB3	Hs.516826	Tribbles homologue 3 (<i>Drosophila</i>)	0	0	11
TSPAN17	Hs.532129	Tetraspanin 17	2	9	1
TYMS	Hs.592338	Thymidylate synthetase	0	13	7
UBE2D3	Hs.518773	Ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homologue, yeast)	0	1	10
UQCRCQ	Hs.146602	Ubiquinol-cytochrome c reductase, complex III subunit VII, 9.5 kDa	4	9	1
VAV2	Hs.369921	Vav 2 oncogene	1	12	0
WDR36	Hs.533237	WD repeat domain 36	0	3	32

presented differential expression at statistically significant levels. As such, these genes may easily be involved in several important processes that lead to differentiation of haematopoietic stem cells into erythrocytes and may constitute therapeutic targets for haematopoietic diseases.

Several genes found in this study as differentially expressed are well described in the literature; these include *ALAS2*, *ANK1*, *GDF15*, *NUDT4* and *AHSP* (16,24), and validate the results found in our libraries. In addition, some genes are described for the first time. Among them, an interesting finding was presence of some

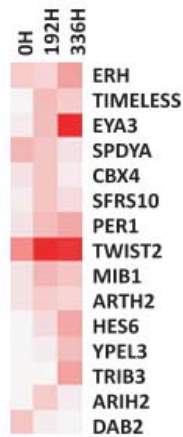


Figure 3. Cluster analysis of 15 differentially expressed genes homologous to *D. melanogaster* found during erythroid differentiation. These genes have been identified in humans; however, most of them do not have any described function. Genes *HES6*, *EYA3*, *ERH* and *TRIB3* were found with high expression at the end of differentiation, while *TIMELESS* showed high expression in the intermediate phase. With the exception of *ERH*, expression of these genes were hardly observed at the beginning of differentiation. Intensity of colour reflects reliability of expression data.

genes homologous to genes of *D. melanogaster* and that were highly and differentially expressed during erythroid differentiation here (Fig. 3). Most differentially expressed genes were *TIMELESS*, *TRIB3*, *EYA3*, *HES6* and *ERH* identified in humans, but some of them do not have any described function in people and none has been reported during erythroid differentiation.

Timeless protein is mainly known for its essential role in circadian rhythm in *Drosophila*; however, a recent study in humans suggests an intimate connection between the circadian cycle and DNA damage checkpoints that is partly mediated by Timeless protein. Timeless protein interacts with Chk1 kinase, which regulates DNA damage-induced G₂/M arrest and is mainly activated by BRCA1 (25,26). The gene was also identified among a common prognostic signature of 29 genes that are associated with patient survival in breast cancer (27) and as a candidate to predict response to tamoxifen, the most common endocrine agent used to treat women at all stages of breast cancer (28). To date, there are no studies demonstrating the relationship of this gene with erythropoiesis, and our data suggest its participation during erythroid maturation, as increase in its expression was observed from the intermediate stage of differentiation onwards, being more evident in CD34⁺ cells (Fig. 4).

Tribbles 3 homologue (*TRIB3*), is a putative protein kinase that, in *Drosophila*, appears to play a role in

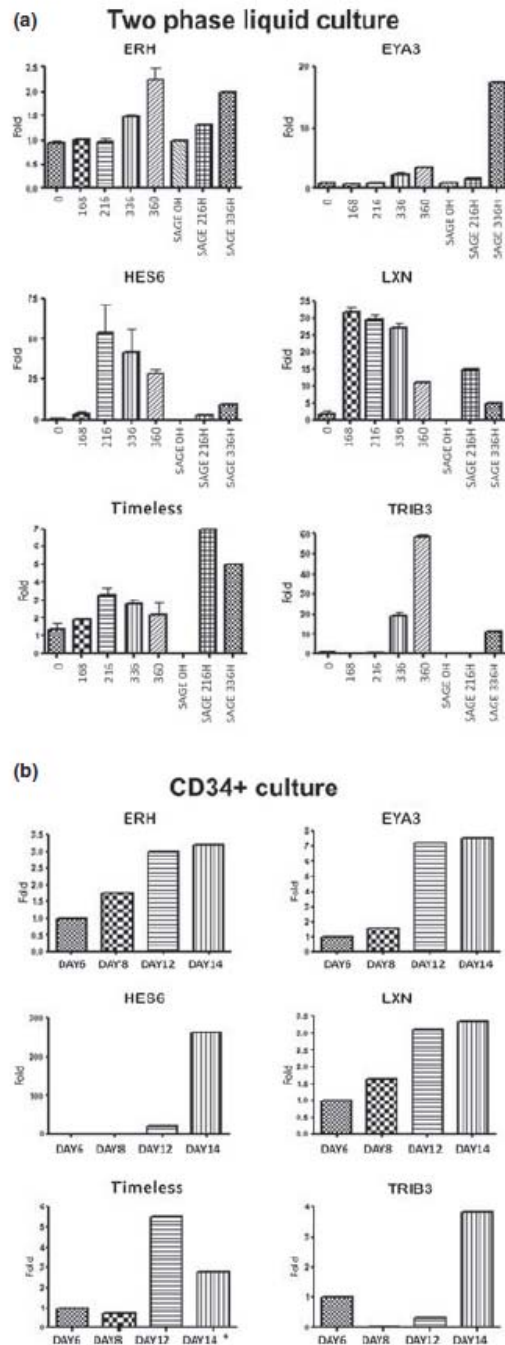


Figure 4. Gene expression of selected genes during erythroid differentiation. Gene expressions of six selected genes were evaluated by qPCR in three different two-phase liquid cultures (a) and in a CD34⁺ culture (b). Expressions observed in both cultures are the same as those identified by SAGE analysis. The pattern observed in SAGE libraries is displayed together with two-phase culture.

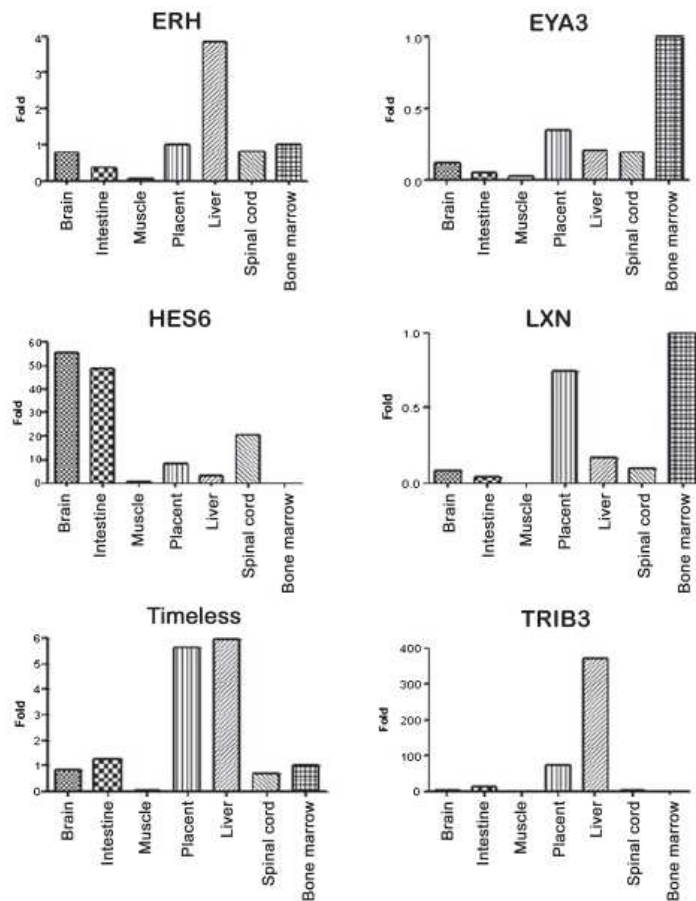


Figure 5. Differential expression of selected genes in several tissues using a cDNA tissue library (Clontech Laboratories Inc).

regulation of the cell cycle and cell migration. In mammals, *TRIB3* was initially cloned as an inducible gene in neuronal PC-3 cells following NGF withdrawal. The protein is emerging as a negative regulator of various signal transducers and has been implicated in several processes, including apoptosis regulation, cell survival, regulation of adipocyte differentiation and insulin resistance (29–32), and also acts as an important participant in tumour cell growth (33). Overexpression of this gene at the end of erythroid differentiation (Fig. 4) demonstrates that this process is finely regulated, as the cells are almost fully differentiated and intense proliferation typically observed in previous stages is controlled. Deregulation of expression of this could be implicated in increase in cell proliferation, in turn inducing a tumour development.

EYA3 (Eyes absent 3) is another gene that demonstrated increase in expression at the end of differentiation,

suggesting a possible role of this transcription factor in maturation of erythroid cells. Li *et al.* (34) demonstrated that the Eya family (*EYA1*, *EYA2* and *EYA3*) has protein phosphatase function, and its enzymatic activity is required for regulating genes that encode growth control and signalling molecules, modulating precursor cell proliferation. Studies with Eya1-deficient mice show that the gene controls critical early inductive signalling events involved in ear and kidney formation and integrate Eya1 into the genetic regulatory cascade controlling kidney formation upstream of Gdnf, which is required to direct ureteric bud outgrowth *via* activation of c-ret Rtk (35). Occasionally, anaemic embryos of these mice are seen, suggesting a haematopoietic defect (12). In a study analysing gene expression of purified haematopoietic stem cells (HSC), the authors identified expression of *EYA1* and *EYA2* and suggested that they could be involved in HSC

self-renewal (12); however, in our study, expression of *EYA3* was not identified. *EYA3* is mapped to chromosome 1 and no studies have been carried out on it in humans. Recently, Soker *et al.* (36) studied pleiotropic effects in *Eya3*-knockout mice and showed that homozygous mutants displayed decreased bone mineral content and shorter body length; furthermore, apparently no haematopoietic effects were observed. Our results suggest that this transcription factor could be important at the end of differentiation as its expression was observed to be high at the end of differentiation and high expression was found in bone marrow.

HES6 (Hairy/Enhancer of Split 6) is another *Drosophila* homologous gene that encodes a member of a subfamily of basic helix-loop-helix transcription repressors (37). The protein encoded by this gene functions as a cofactor, interacting with other transcription factors through a tetrapeptide domain in its C-terminus (38), and may be involved in neurogenesis (39) and cell proliferation in promyelocytic leukaemia (40). However, precise molecular mechanism of *Hes6*-mediated control of differentiation remains to be elucidated (40). This transcription factor was found to be highly expressed at the end of differentiation here, but was not observed at the beginning of CD34+ differentiation (Fig. 4) in bone marrow (Fig. 5), showing that its expression is stage-specific and finely regulated.

ERH (Expression of Enhancer of Rudimentary) gene was found to be continuously regulated during erythropoiesis and its expression increased during differentiation (Fig. 4). The product of this gene is a small, highly conserved, nuclear protein with a unique three-dimensional structure. Involvement of *ERH* in fundamental processes such as regulation of pyrimidine metabolism, cell cycle progression, transcription and cell growth control has been suggested (41–44); however, none of these interactions has been verified experimentally. To date, the mechanism of action of *ERH* remains unclear, and our result needs to be studied in detail to identify its function in erythroid differentiation.

In addition to these *Drosophila* homologous genes, *LXN* (latexin) gene was observed to be continuously expressed from the beginning of differentiation and was highly expressed in bone marrow (Figs 4 and 5). *LXN* is the only known carboxypeptidase inhibitor in mammals and despite several structure–function studies of latexin, there is little knowledge of its biological roles in stem cells and ageing. Recent studies have shown that *LXN* is a negative regulator of stem cell number and acts through at least two mechanisms to modulate stem cell pool size: (i) it decreases HSC cell replication and (ii) it increases HSC apoptosis. Thus, in the haematopoietic system, and perhaps other organs, latexin influences ageing and lifespan

through its action on stem cells (23). Continuous expression of the gene, found in this study, showed that its regulation was directly related to differentiation of the cells; during cell proliferation and consequent maturation, expression of the gene increased then began to decline. Further studies on gene expression using inhibition and superexpression of these genes in CD34+ cultures are being carried out and results will provide new insights to the relationship of its expression to haematopoiesis.

Another important finding in our study was the number of tags that had no correspondence in the data bank and that were denominated 'no matches' (27% approximately); these tags could represent novel genes. Several studies observed the same results and have shown that approximately 35% of total SAGE tags are unmapped or unidentified. Several authors have suggested that this could be explained by several reasons: for instance, tags overlapping two exons, tags extended into the polyA tail and tags that differ from the genome sequence due to polymorphism. These tags could also correspond to antisense transcripts or new variants of known transcripts, suggesting that many transcripts are still to be annotated and that the human transcriptome seems to be more complex than shown in current genome annotations (45). Study on non-identified SAGE tags could help improve the annotation process and identify genes with important functions that could potentially be used as targets for disease therapies.

One of these tags (No Match 1 – Table 2) demonstrated a large increase in expression during the intermediate phase of differentiation and could be very important in metabolic pathways involved in differentiation of erythroid cells. Two other tags (No Match 5 and 8 – Table 2) demonstrated increases at the end of differentiation and could be involved in maturation of haematopoietic cells. Identification of these tags could identify new genes or new isoforms of genes involved in differentiation of erythroid cells.

Results shown in this study amplify previously published data and present new clues concerning gene regulation and dynamic organization of genes in chromosomes of cells contributing to comprehension of erythroid differentiation, and to identification of new target genes involved in some erythroid diseases.

Acknowledgements

This study was supported by Grants from FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo, São Paulo, Brazil, 02/13801-7). A.F.C. was also supported by FAPESP (05/51222-7). The authors thank Dr. Nicola Conran, HEMOCENTRO-UNICAMP, for help with English revision.

References

- Baron MH (1997) Transcriptional control of globin gene switching during vertebrate development. *Biochim. Biophys. Acta* **1351**, 51–72.
- Shivdasani RA, Orkin SH (1996) The transcriptional control of hematopoiesis. *Blood* **87**, 4025–4039.
- Pope SH, Fibach E, Sun J, Chin K, Rodgers GP (2000) Two-phase liquid culture system models normal human adult erythropoiesis at the molecular level. *Eur. J. Haematol.* **64**, 292–303.
- Hu M, Krause D, Greaves M, Sharkis S, Dexter M, Heyworth C *et al.* (1997) Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11**, 774–785.
- Jimenez G, Griffiths SD, Ford AM, Greaves MF, Enver T (1992) Activation of the beta-globin locus control region precedes commitment to the erythroid lineage. *Proc. Natl. Acad. Sci. USA* **89**, 10618–10622.
- Ford AM, Bennett CA, Healy LE, Navarro E, Spooncer E, Greaves MF (1992) Immunoglobulin heavy-chain and CD3 delta-chain gene enhancers are DNase I-hypersensitive in hemopoietic progenitor cells. *Proc. Natl. Acad. Sci. USA* **89**, 3424–3428.
- Heyworth C, Pearson S, May G, Enver T (2002) Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *EMBO J.* **21**, 3770–3781.
- Ney PA (2006) Gene expression during terminal erythroid differentiation. *Curr. Opin. Hematol.* **13**, 203–208.
- Gubin AN, Njoroge JM, Bouffard GG, Miller JL (1999) Gene expression in proliferating human erythroid cells. *Genomics* **59**, 168–177.
- Georgantas RW III, Tanadve V, Malehorn M, Heimfeld S, Chen C, Carr L *et al.* (2004) Microarray and serial analysis of gene expression analyses identify known and novel transcripts overexpressed in hematopoietic stem cells. *Cancer Res.* **64**, 4434–4441.
- Zhou G, Chen J, Lee S, Clark T, Rowley JD, Wang SM (2001) The pattern of gene expression in human CD34(+) stem/progenitor cells. *Proc. Natl. Acad. Sci. USA* **98**, 13966–13971.
- Terskikh AV, Miyamoto T, Chang C, Diatchenko L, Weissman IL (2003) Gene expression analysis of purified hematopoietic stem cells and committed progenitors. *Blood* **102**, 94–101.
- Steidl U, Kronenwett R, Rohr UP, Fenk R, Kliszewski S, Maercker C *et al.* (2002) Gene expression profiling identifies significant differences between the molecular phenotypes of bone marrow-derived and circulating human CD34+ hematopoietic stem cells. *Blood* **99**, 2037–2044.
- Bonafoux B, Combes T (2009) Serial analysis of gene expression adapted for downsized extracts (SAGE/SADE) analysis in reticulocytes. *Methods Mol. Biol.* **496**, 299–311.
- Bonafoux B, Lejeune M, Piquemal D, Quere R, Baudet A, Assaf L *et al.* (2004) Analysis of remnant reticulocyte mRNA reveals new genes and antisense transcripts expressed in the human erythroid lineage. *Haematologica* **89**, 1434–1438.
- Komor M, Guller S, Baldus CD, de Vos S, Hoelzer D, Ottmann OG *et al.* (2005) Transcriptional profiling of human hematopoiesis during *in vitro* lineage-specific differentiation. *Stem Cells* **23**, 1154–1169.
- Wang SM (2007) Understanding SAGE data. *Trends Genet.* **23**, 42–50.
- Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. *Science* **270**, 484–487.
- Audic S, Claverie JM (1997) The significance of digital gene expression profiles. *Genome Res.* **7**, 986–995.
- Hashimoto S, Nagai S, Sese J, Suzuki T, Obata A, Sato T *et al.* (2003) Gene expression profile in human leukocytes. *Blood* **101**, 3509–3513.
- Chabardes-Garonne D, Mejean A, Aude JC, Cheval L, Di Stefano A, Gaillard MC *et al.* (2003) A panoramic view of gene expression in the human kidney. *Proc. Natl. Acad. Sci. USA* **100**, 13710–13715.
- de Chaldee M, Gaillard MC, Bizat N, Buhler JM, Manzoni O, Bockaert J *et al.* (2003) Quantitative assessment of transcriptome differences between brain territories. *Genome Res.* **13**, 1646–1653.
- Liang Y, Van Zant G (2008) Aging stem cells, latexin, and longevity. *Exp. Cell Res.* **314**, 1962–1972.
- Pina C, Enver T (2007) Differential contributions of haematopoietic stem cells to foetal and adult haematopoiesis: insights from functional analysis of transcriptional regulators. *Oncogene* **26**, 6750–6765.
- Yarden RI, Pardo-Reoyo S, Sgagias M, Cowan KH, Brody LC (2002) BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage. *Nat. Genet.* **30**, 285–289.
- Unsal-Kacmaz K, Mullen TE, Kaufmann WK, Sancar A (2005) Coupling of human circadian and cell cycles by the timeless protein. *Mol. Cell. Biol.* **25**, 3109–3116.
- Naderi A, Teschendorff AE, Barbosa-Morais NL, Pinder SE, Green AR, Powe DG *et al.* (2007) A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene* **26**, 1507–1516.
- Tozlu-Kara S, Roux V, Andrieu C, Vendrell J, Vacher S, Lazar V *et al.* (2007) Oligonucleotide microarray analysis of estrogen receptor alpha-positive postmenopausal breast carcinomas: identification of HRPAP20 and TIMELESS as outstanding candidate markers to predict the response to tamoxifen. *J. Mol. Endocrinol.* **39**, 305–318.
- Prudente S, Scarpelli D, Chandalia M, Zhang YY, Morini E, Del Guerna S *et al.* (2009) The TRIB3 Q84R polymorphism and risk of early-onset type 2 diabetes. *J. Clin. Endocrinol. Metab.* **94**, 190–196.
- Takahashi Y, Ohoka N, Hayashi H, Sato R (2008) TRIB3 suppresses adipocyte differentiation by negatively regulating PPARgamma transcriptional activity. *J. Lipid Res.* **49**, 880–892.
- Zhou Y, Li L, Liu Q, Xing G, Kuai X, Sun J *et al.* (2008) E3 ubiquitin ligase SLAH1 mediates ubiquitination and degradation of TRIB3. *Cell. Signal.* **20**, 942–948.
- Rzymiski T, Paantjens A, Bod J, Harris AL (2008) Multiple pathways are involved in the anoxia response of SKIP3 including HuR-regulated RNA stability, NF-kappaB and ATF4. *Oncogene* **27**, 4532–4543.
- Bowers AJ, Scully S, Boylan JF (2003) SKIP3, a novel Drosophila tribbles ortholog, is overexpressed in human tumors and is regulated by hypoxia. *Oncogene* **22**, 2823–2835.
- Li X, Oghi KA, Zhang J, Krones A, Bush KT, Glass CK *et al.* (2003) Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis. *Nature* **426**, 247–254.
- Xu PX, Adams J, Peters H, Brown MC, Heaney S, Maas R (1999) Eyal-deficient mice lack ears and kidneys and show abnormal apoptosis of organ primordia. *Nat. Genet.* **23**, 113–117.
- Soker T, Dalke C, Puk O, Floss T, Becker L, Bolle I *et al.* (2008) Pleiotropic effects in Eya3 knockout mice. *BMC Dev. Biol.* **8**, 118.
- Qian D, Radde-Gallwitz K, Kelly M, Tyrberg B, Kim J, Gao WQ *et al.* (2006) Basic helix-loop-helix gene Hes6 delineates the sensory hair cell lineage in the inner ear. *Dev. Dyn.* **235**, 1689–1700.
- Bae S, Bessho Y, Hojo M, Kageyama R (2000) The bHLH gene Hes6, an inhibitor of Hes1, promotes neuronal differentiation. *Development* **127**, 2933–2943.
- Jacobsen KX, Vanderluit JL, Slack RS, Albert PR (2008) HES1 regulates 5-HT1A receptor gene transcription at a functional polymorphism: essential role in developmental expression. *Mol. Cell. Neurosci.* **38**, 349–358.
- Eun B, Lee Y, Hong S, Kim J, Lee HW, Kim K *et al.* (2008) Hes6 controls cell proliferation via interaction with cAMP-response element-binding protein-binding protein in the promyelocytic leukemia nuclear body. *J. Biol. Chem.* **283**, 5939–5949.
- Isomura M, Okui K, Fujiwara T, Shin S, Nakamura Y (1996) Cloning and mapping of a novel human cDNA homologous to DROER,

- the enhancer of the *Drosophila melanogaster* rudimentary gene. *Genomics* **32**, 125–127.
- 42 Pogge von Strandmann E, Senkel S, Ryffel GU (2001) ERH (enhancer of rudimentary homologue), a conserved factor identical between frog and human, is a transcriptional repressor. *Biol. Chem.* **382**, 1379–1385.
 - 43 Wojcik E, Murphy AM, Fares H, Dang-Vu K, Tsubota SI (1994) Enhancer of rudimentaryp1, e(r)p1, a highly conserved enhancer of the rudimentary gene. *Genetics* **138**, 1163–1170.
 - 44 Gelsthorpe M, Pulumati M, McCallum C, Dang-Vu K, Tsubota SI (1997) The putative cell cycle gene, enhancer of rudimentary, encodes a highly conserved protein found in plants and animals. *Gene* **186**, 189–195.
 - 45 Keime C, Semon M, Mouchiroud D, Duret L, Gandrillon O (2007) Unexpected observations after mapping LongSAGE tags to the human genome. *BMC Bioinformatics* **8**, 154.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Morphology of cells during erythroid differentiation using two-phase liquid culture after erythropoietin addition. Typical morphology was detectable during differentiation (0 h, proerythroblast; 192 h, basophilic erythroblasts and 336 h, orthochromatic erythroblasts).

Fig. S2 Gene ontology categorization. Differentially expressed genes were categorized by Molecular Function and Biological Process using the gene ontology classification.

Table S1 Sequence and ideal concentration for the primers used in qPCR.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

ORIGINAL ARTICLE

HIGH LEVELS OF HUMAN γ -GLOBIN ARE EXPRESSED IN ADULT MICE CARRYING A TRANSGENE OF THE BRAZILIAN TYPE OF HEREDITARY PERSISTENCE OF FETAL HEMOGLOBIN (A_{γ} -195)

Anderson F. da Cunha,^{1,2} Ana F. Brugnerotto,¹ Marcus A. Finzi Corat,³ Emily E. Devlin,⁴ Ana P. Gimenes,³ Mônica Barbosa de Melo,¹ Luiz A. Corrêa Passos,³ David Bodine,⁴ Sara T.O. Saad,¹ and Fernando F. Costa¹

¹Hematology and Hemotherapy Center, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

²Genetic and Evolution Department, Universidade Federal de São Carlos, São Carlos, São Paulo, Brazil

³Centro Multidisciplinas para Investigação Biológica, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil

⁴Hematopoiesis Section, National Human Genome Research Institute, Bethesda, Maryland, USA

□ Hereditary persistence of fetal hemoglobin (HPFH) is characterized by increased levels of Hb F during adult life. Nondeletional forms of HPFH are characterized by single base mutations in the A_{γ} and G_{γ} promoters, resulting in an increase of Hb F ranging from 3 to 20% in heterozygotes. Many point mutations in this region have been described, including the A_{γ} -195 (C>G) mutation that causes the Brazilian type of HPFH (HPFH-B). To better understand this mechanism, we have developed HPFH-B transgenic mice. mRNA levels of human γ -globin of -195 transgenic mice were clearly higher when compared with control transgenic mice bearing a wild type sequence of the γ promoter. Thus, our data indicate that the -195 mutation is the unique cause of elevation of Hb F in Brazilian HPFH. These results could provide us with an opportunity to study the modifying effects of the Hb F in the phenotype of sickle cell disease and β -thalassemia (β -thal).

Keywords Hb F, Hereditary persistence of fetal hemoglobin (HPFH), Transgenic mice, -195 Mutation, γ -Globin

Received 3 June 2009; Accepted 14 July 2009.

Address correspondence to Anderson F. da Cunha, Rodovia Washington Luís, km 235, Departamento de Genética e Evolução, CCBS, UFSCar, São Carlos, SP 13565-905, Brazil; Tel: +551633518377; Fax: +551633518377; E-mail: anderf@ufscar.br

INTRODUCTION

The human β -globin gene cluster consists of five functional globin genes (ϵ , $\text{A}\gamma$, $\text{G}\gamma$, δ and β) arranged in the locus according to the order of their expression during development (1,2). The switch from fetal (HbG) to adult (HbD and HbB) globin gene expression occurs at birth, leading to the gradual replacement of HbF ($\alpha_2\gamma_2$) with HbA ($\alpha_2\beta_2$) and a small amount of HbA₂ ($\alpha_2\delta_2$). The exact mechanisms that regulate globin gene switching are incompletely understood. However, developmental regulation of this multigene locus appears to be mediated by complex developmental stage- and tissue-specific interactions between *cis*- and *trans*-acting regulatory elements within the gene cluster. Understanding the molecular basis of globin gene switching is of great interest as the hereditary persistence of fetal hemoglobin (HPFH), which is characterized by an inappropriate, persistent expression of the fetal γ -globin genes in adult life, ameliorates the effects of hemoglobinopathies (3–5).

There are two types of inherited HPFH mutations: point mutations in the proximal promoter of either of the HbG genes, and deletions that remove the HBB gene plus extensive sequences either side of it (6). Structural studies have shown that nondeletional HPFH point mutations are clustered in three regions of the γ gene promoters centered around positions –200 (+C), –175 (T>C) and –115 (C>A), relative to the transcriptional start site. These regions are highly GC-rich, known to be the target for different, but closely spaced, point mutations affecting the nucleotides in $\text{G}\gamma$ promoters at positions –202 (C>G), –196 (C>T), –175 (T>C), –161 (A>G) and –158 (C>T), and the $\text{A}\gamma$ promoters at positions –202 (C>T), –201 (C>T), –198 (T>C), –196 (C>T), –195 (C>G), –175 (T>C) and –117 (G>A), respectively (1,2,5,7).

Two hypotheses have been proposed to explain the increased γ -globin gene expression in adults carrying nondeletional HPFH mutations. The first proposes that these point mutations decrease the binding of a transcriptional repressor or a complex that is involved in the silencing of γ -globin expression in the adult. Alternatively, these mutations may create binding sites that enhance the binding for a transcriptional activator or complex, thus increasing γ gene expression in the adult. Early *in vitro* studies focused on characterizing the effects of these mutations on the binding of different DNA-binding proteins (1).

A previous study showed that the Brazilian HPFH mutation [$\text{A}\gamma$ –195 (C>G)] (8) alone was not able to increase gene expression *in vitro*, and a modest increase was observed when a locus control region (LCR) element fragment (HS2) was introduced in the construction (3,9). To better understand this mechanism, *in vivo*, we have developed transgenic mice carrying the Brazilian HPFH mutation ($\text{A}\gamma$ –195). Our data showed an

increase in γ -globin expression in all development stages of these mice, indicating that the -195 mutation is the unique cause of elevation of Hb F. These results could provide an opportunity to study the modifying effects of the Hb F in the phenotype of sickle cell disease and β -thalassemia (β -thal).

MATERIALS AND METHODS

An engineered C to G substitution at position -195 of the promoter region of γ gene was generated by site direct mutagenesis and introduced into the *Xho*I/*Not*I site of pet32a (+) plasmid. The mutation was confirmed by DNA sequencing and then the 5.2 kb fragment was inserted in the same region of the μ LCRA $\gamma\psi\beta\delta\beta$ cosmid (10), generating the cosmid μ LCRA $\gamma\psi\beta\delta\beta$ -B-HPFH (Figure 1).

The μ LCRA $\gamma\psi\beta\delta\beta$ -B-HPFH, with the μ LCR and the γ , $\psi\beta$, δ and β genes with the -195 mutation at the γ promoter, was injected into mouse egg cells, which were then introduced into pseudo pregnant mouse females

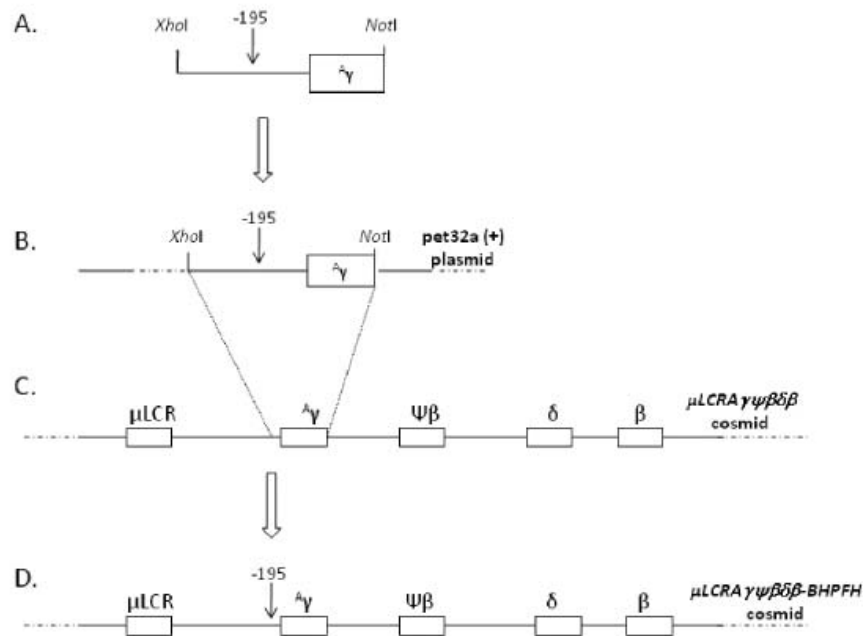


FIGURE 1 Scheme of μ LCRA $\gamma\psi\beta\delta\beta$ -B-HPFH cosmid construction: a site directed mutagenesis introduced a C>G substitution at position -195 of the promoter region of γ . (A). This insert was cloned into the *Xho*I/*Not*I site of the pet32a (+) plasmid (B). This plasmid was digested with *Xho*I and *Not*I enzymes and the resulting 5.2 kb fragment was inserted into the same region of the μ LCRA $\gamma\psi\beta\delta\beta$ cosmid (C), generating the μ LCRA $\gamma\psi\beta\delta\beta$ -B-HPFH cosmid (D).

[transgenic mice were generated by Umea Transgene Core Facility (Umea, Sweden) using F1 hybrid (CBA X B6/CBA) females obtained from Charles River, Kisslegg, Germany]. The transgenic selection was performed using specific primers for human γ -globin gene. Three lineages of transgenic lines were generated. These lineages were crossed with B6 wild type to generate F1 and those that were positive for the transgene were crossed to generate F2. Wild type transgenic mice (WTTg) obtained from Hematopoiesis Section, National Human Genome Research Institute, Bethesda, MD, USA were analyzed as controls. These mice were described in previous studies by Sabatino et al. (11) and Zoueva et al. (12). All experimental procedures in this study in Brazil were approved by the Brazilian College for Animal Experimentation (COBEA) of the State University of Campinas and in the United States. The protocol number was G-97-4.

Total cellular RNA was extracted from the day 10.5 yolk sac, and day 13.5 fetal liver using TRIZOL reagent (Invitrogen Inc., Carlsbad, CA, USA). RNA was extracted from adult and newborn reticulocytes, obtained by collecting 200 μ L of blood from phlebotomized animals or by decapitation of newborn mice, respectively, also using TRIZOL reagent (Invitrogen Inc.). Samples were quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies Inc., Wilmington, DE, USA). A riboprobe was generated containing sequences for both exons 2 of the human γ - and of the murine α -globin genes, and a second riboprobe was generated containing both exons 2 of the human β - and of the murine α -globin genes. Both probes were linearized with *Bgl*II. 32 P-Labeled RNA probes were transcribed using the MAXIscriptTM in vitro transcription kit (Ambion Inc., Austin, TX, USA). RNA (0.1–0.25 μ g) was hybridized to the probe, overnight, according to standard procedures (RPA II; Ambion Inc.). RNase digestion was performed using an RNase A/RNase T1 mixture and the protected fragments were separated on an 8% non denaturing polyacrylamide gel (13). To quantitate levels of mRNA, the gel was exposed to a PhosphorImager screen and scanned on a Molecular Dynamics PhosphorImager (General Electric Healthcare, Uppsala, Sweden). The relative amounts of the human globin expression were estimated by the ratio with the mouse α -globin.

RESULTS

Construction and Selection of B-HPFH Transgenic Strain

Transgenic mice were generated by the introduction of a fragment with the μ LCR and the γ , $\psi\beta$, δ and β genes with the –195 mutation at the γ promoter (Figure 1) in mouse egg cells, as described in Materials and Methods. To evaluate the presence of the transgene, we performed an amplification of a specific region of the human γ -globin gene. The mice that presented

an amplification of this fragment were used in RNase protection assays (RPA) (data not shown).

RNase Protection Assay in Transgenic Mice Carrying the -195 Promoter Mutation of the Human γ -Globin Gene

We evaluated the expression of human $\Lambda\gamma$ -globin in transgenic mice (-195HPFHTg) that carried a mutation in the $\Lambda\gamma$ -globin [$\Lambda\gamma$ -195 (C>G)] gene of the β -globin gene cluster. As a control, a wild type transgenic mouse (WTTg) that had the intact β -globin gene cluster was used. In the WTTg mouse, the γ -globin expression occurred strongly in the yolk sac and was observed at a low level in the liver 13 days after the embryonic phase. Murine α -globin gene expression, which is continually expressed, and human β -globin gene, which is expressed in wild transgenic mice only from the 10th day in the fetal liver, were used as internal controls. These results are similar to those observed by other authors (10–14). Our results showed a clear increase in the expression of human $\Lambda\gamma$ -globin in all samples of transgenic mice carrying the mutation, compared to the wild type transgenic control (Figure 1A). Comparing the -195HPFHTg mice samples, the yolk sac and newborn blood showed the highest expression of human $\Lambda\gamma$ -globin (Figure 1B).

Flow Cytometry for Human Fetal Hemoglobin (Hb F) in -195 Transgenic Mice

To evaluate if the γ -globin protein was produced as a tetramer with mouse α -globin to form a hybrid HbF, we performed a flow cytometry procedure using an antibody to fetal hemoglobin (HbF) in the mice with the -195 mutation. The results showed a strong presence of this hemoglobin (Hb), corroborating our results that the -195 alone is able to develop HPFH in these mice (Figure 3).

DISCUSSION

Many point mutations in the promoter region of the γ -globin gene have been reported as being responsible for causing the persistent expression of this gene during adult life in humans. Of these mutations, $\Lambda\gamma$ -195 (C>G), was first identified in a Brazilian population and showed an increase in Hb F of between 4–7% in heterozygotes (8).

To better understand this mechanism, we investigated whether the -195 mutation in the γ -globin promoter gene, is sufficient to cause the persistence of γ -globin expression during mouse development. Our results indicate that the γ -globin gene expression remained active at the adult

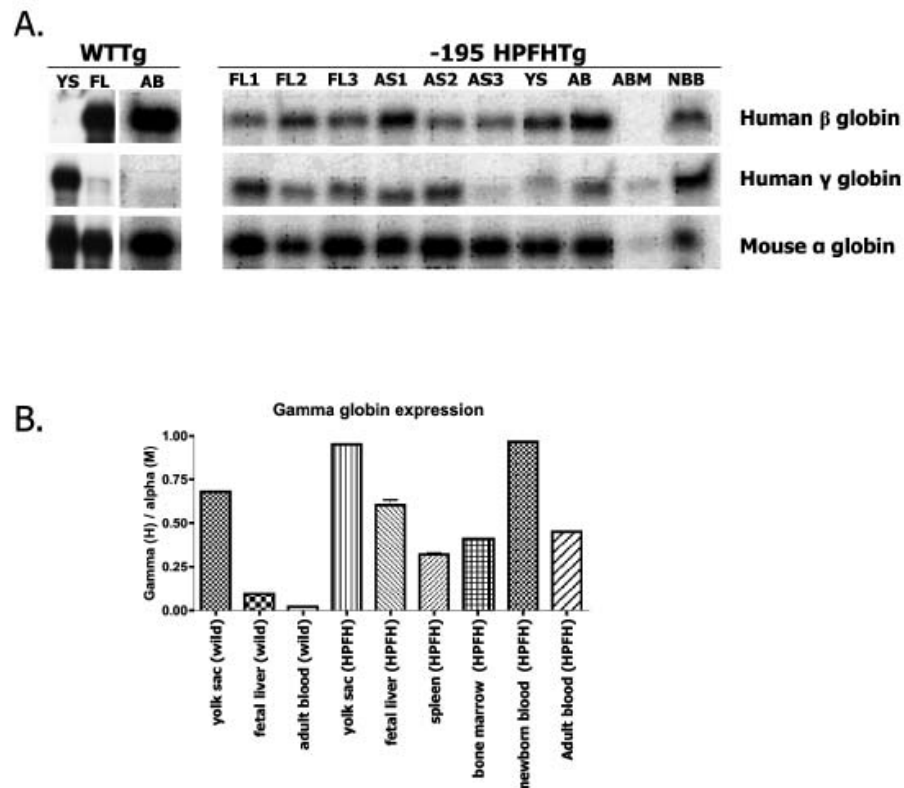


FIGURE 2 A) RNAse protection assay in transgenic mice carrying the mutation in the -195 promoter of the human γ -globin gene. RNA of yolk sac (YS), fetal liver (FL), and adult blood (AB) were extracted from wild type transgenic mice that carried the cluster of human β -globin (wild) and from the -195 transgenic mice (HPFH). Adult spleen (AS), adult bone marrow (ABM) and newborn bone marrow (NBM) were extracted only from -195 transgenic mice (HPFH). The probes used in the experiment were for γ and β human globins and for murine α -globin. The numbers 1 to 3 represent RNA samples of tissues of three different transgenic lines that were used. The expression of γ -globin increased in all samples of mice carrying the -195 mutation (HPFHTg). B) The human γ - and α -globin ratios showed that the expression of γ -globin could be observed in all samples of the -195 transgenic mice (HPFH), and that the higher expression could be observed in the yolk sac and newborn bone marrow. In the wild type, γ expression was high only in the yolk sac, decreasing in fetal liver and almost undetectable in adult blood.

stage and that the point mutation was responsible for this HPFH phenotype. Several studies have used the same strategy to evaluate mutations in the γ promoter, providing a powerful indication that a single point mutation was able to cause HPFH (14–17). In the -117 (G>A) mutation, Berry et al. (16) speculated that the change in γ -globin expression during development could be correlated with the loss of GATA-1 binding to the γ promoter region, and deduced that GATA-1 might act as a negative regulator

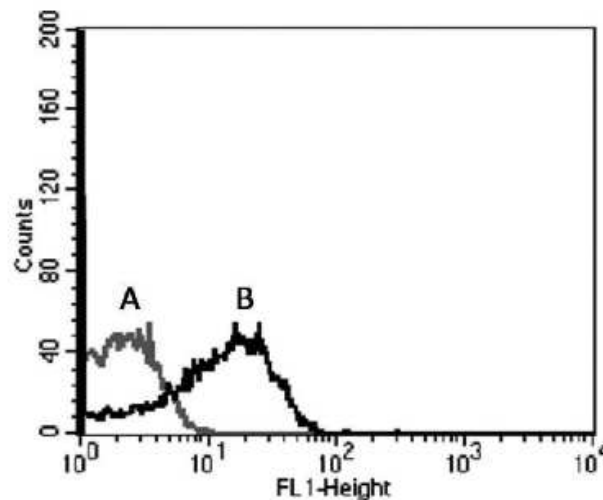


FIGURE 3 Evaluation of the presence of a hybrid fetal Hb in a B-HPFH transgenic mouse by flow cytometry. Peripheral blood of this mouse was incubated with IgG (A) and human Hb F (B) antibodies. A significant increase was observed in the sample incubated with Hb F antibodies, showing the presence of this Hb in the mouse with -195 mutation.

of the γ -globin gene in normal adults. However, earlier in vitro experiments on a Black form of the HPFH point mutation, a T>C substitution at position -175 of the γ -globin promoter, showed no significant difference in the GATA-1 affinity for the promoter. Nevertheless, this mutation dramatically decreased the affinity of the Oct-1 transcription factor, which binds to an area partially overlapping the GATA-1 binding (14).

Another study showed that the HPFH -198 (T>C) mutation increased the γ -globin gene expression in adult transgenic mice by the creation of a new CACCC box and that this increase was mediated by the Sp-1 transcription factor (18). In a recent study, the same group identified a set of other proteins (DNMT1, CDC5-like protein, RAP74, SNEV, the coactivator p52, and a protein of unknown function) that bound to the same region in an in vitro study. Using the same strategy, in our laboratory, studies showed that the -195 mutation alone was not able to increase gene expression in vitro and a modest increase was observed when an LCR element fragment (HS2) was introduced in the construction (3,9). Studies also showed that the mechanism of Hb F elevation by the -195 mutation was neither mediated by the Sp-1 transcription factor nor by the creation of a CACCC box, as described for the -198 mutation (3). These studies showed that several proteins may be involved in the over expression of the γ -globin chain and/or may depend on DNA structure (1).

In this report we show that the -195 mutation alone was probably the unique cause of elevation of Hb F in the transgenic mice carrying this mutation. The results in Figure 2 showed that γ -globin expression remained high during the development of the mice, since its expression was observed from the yolk sac to adult blood, and the results in Figure 3 show that the γ -globin protein could tetramerize with mouse α -globin forming a hybrid fetal Hb. Many proteins could be involved in the persistence of the γ -globin expression in an adult affected by the -195 nondeletional HPFH, and this mouse model may provide the means to study these interactions and understand the changes that occur during the reactivation of the γ -globin genes. The identification of proteins that could be involved in this process will be helpful in better understanding the complex mechanisms that regulate globin gene expression, as well as providing an opportunity to study the modifying effects of the Hb F in the phenotype of sickle cell disease and β -thal.

ACKNOWLEDGMENTS

We thank Qiliang Li (Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA) for providing the μ LCRA $\gamma\psi\beta\delta\beta$ cosmid, Vinicius Motta and Umea Transgene Core Facility (Umea, Sweden) for help in the generation of the -195PHHF transgenic mice and Dr. Nicola Conran (Hematology and Hemotherapy Center, Universidade Estadual de Campinas, São Paulo, Brazil) for English revision. This study was funded by grants from Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP grants 02/13801-7 and 05/51222-7).

Declaration of Interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

REFERENCES

1. Olave IA, Doneanu C, Fang X, Stamatoyannopoulos G, Li Q. Purification and identification of proteins that bind to the hereditary persistence of fetal hemoglobin -198 mutation in the γ -globin gene promoter. *J Biol Chem*. 2007;282(2):853-862.
2. Stamatoyannopoulos G, Grosfeld FG. *Hemoglobin Switching, 3rd ed. Vol. II*. Philadelphia: W.B. Saunders Company, 2001.
3. Takahashi T, Schreiber R, Krieger JE, Saad ST, Costa FF. Analysis of the mechanism of action of the Brazilian type (γ -195 C>G) of hereditary persistence of fetal hemoglobin. *Eur J Haematol*. 2003;71(6):418-424.
4. Katsantoni EZ, Langeveld A, Wai AW, et al. Persistent γ -globin expression in adult transgenic mice is mediated by HPFH-2, HPFH-3, and HPFH-6 breakpoint sequences. *Blood*. 2003;102(9):3412-3419.
5. Forget BG. Molecular basis of hereditary persistence of fetal hemoglobin. *Ann NY Acad Sci*. 1998;850:38-44.
6. Keys JR, Tallack MR, Zhan Y, et al. A mechanism for Ikaros regulation of human globin gene switching. *Br J Haematol*. 2008;141(3):398-406.

7. Tasiopoulou M, Boussiou M, Sinopoulou K, Moraitis G, Louradi-Anagnostou A, and Karababa P. γ -196 C>T, γ -201 C>T: two novel mutations in the promoter region of the γ -globin genes associated with nondeletional hereditary persistence of fetal hemoglobin in Greece. *Blood Cells Mol Dis.* 2008;40(3):320–322.
8. Costa FF, Zago MA, Cheng G, Nechuman JF, Stoming TA, Huisman THJ, The Brazilian type of non-deletional γ -HPFH has a C→G substitution at nucleotide –195 of the γ -globin gene. *Blood.* 1990;76(9):1896–1990.
9. Schreiber R, Goncalves MS, Junqueira ML, Saad ST, Krieger JE, Costa FF. The γ -195 (C>G) mutation in hereditary persistence of fetal hemoglobin is not associated with activation of a reporter gene in vitro. *Braz J Med Biol Res.* 2001;34(4):489–492.
10. Enver T, Raich N, Ebens AJ, Papayannopoulou T, Costantini F, Stamatoyannopoulos G. Developmental regulation of human fetal-to-adult globin gene switching in transgenic mice. *Nature.* 1990;344(6264):309–313.
11. Sabatino D, Cline A, Gallagher P, et al. Substitution of the human β -spectrin promoter for the human γ -globin promoter prevents silencing of a linked human β -globin gene in transgenic mice. *Mol Cell Biol.* 1998;18(11):6634–6640.
12. Zoueva O, Garrett L, Bodine D, Rodgers G. BP1 motif in the human β -globin promoter affects β -globin expression during embryonic/fetal erythropoiesis in transgenic mice bearing the human β -globin gene. *Blood Cells Mol Dis.* 2008;41(3):244–251.
13. Gallagher P, Sabatino D, Basseres D, et al. Erythrocyte ankyrin promoter mutations associated with recessive hereditary spherocytosis cause significant abnormalities in ankyrin expression. *J Biol Chem.* 2001;276(45):41683–41689.
14. Liu LR, Du ZW, Zhao HL, et al. T to C substitution at –175 or –173 of the γ -globin promoter affects GATA-1 and Oct-1 binding in vitro differently but can independently reproduce the hereditary persistence of fetal hemoglobin phenotype in transgenic mice. *J Biol Chem.* 2005;280(9):7452–7459.
15. Gumucio DL, Rood KL, Blanchard-McQuate KL, Gray TA, Saulino A, Collins FS. Interaction of Sp1 with the human γ globin promoter: binding and transactivation of normal and mutant promoters. *Blood.* 1991;78(7):1853–1863.
16. Berry M, Grosveld F, Dillon N. A single point mutation is the cause of the Greek form of hereditary persistence of fetal haemoglobin. *Nature.* 1992;358(6386):499–502.
17. Peterson KR, Li QL, Clegg CH, et al. Use of yeast artificial chromosomes (YACs) in studies of mammalian development: production of β -globin locus YAC mice carrying human globin developmental mutants. *Proc Natl Acad Sci USA.* 1995;92(12):5655–5659.
18. Li Q, Duan ZJ, Stamatoyannopoulos G. Analysis of the mechanism of action of non-deletion hereditary persistence of fetal hemoglobin mutants in transgenic mice. *EMBO J.* 2001;20(1–2):157–164.