UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO – FMRP/USP

MARCELA CRISTINA CORRÊA DE FREITAS

**Caracterização do padrão de integração do retrovírus pMFG-FVIII-
P140K em linhagens celulares humanas produtoras de fator VIII
recombinante**

Ribeirão Preto

2010

MARCELA CRISTINA CORRÊA DE FREITAS

Caracterização do padrão de integração do retrovírus pMFG-FVIII-P140K em linhagens celulares humanas produtoras de fator VIII recombinante

Dissertação de mestrado apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestre em Ciências Médicas – Modalidade Biomédica.

Área de Concentração: Clínica Médica
Opção: Investigação Biomédica
Orientador: Profa. Dra. Aparecida Maria Fontes

Ribeirão Preto
2010

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

FICHA CATALOGRÁFICA

Freitas, Marcela Cristina Corrêa de

Caracterização do padrão de integração do retrovírus pMFG-FVIII-P140K em linhagens celulares humanas produtoras de fator VIII recombinante. Ribeirão Preto, 2010.

57 p. : il. ; 30cm

Dissertação de Mestrado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP – Área de concentração: Clínica Médica – Modalidade Biomédica

Orientador: Fontes, Aparecida Maria

1. Hemofilia A. 2. Fator VIII. 3. LM-PCR. 4. Integração Retroviral.

SUMÁRIO

RESUMO.....	i
1 - Português.....	i
2 - Língua estrangeira.....	ii
LISTA DE FIGURAS.....	iii
LISTA DE TABELAS.....	iv
LISTA DE ABREVIATURAS E SIGLAS.....	v
1 Introdução	1
1.1 Hemofilia A	1
1.2 Fator VIII recombinante	3
1.3 Estratégias de seleção do clone celular recombinante	5
1.4 Mecanismos de integração dos retrovírus	6
1.5 Organização do genoma humano	8
1.6 Sítios de integração preferencial dos retrovírus.....	9
1.7 Objetivos.....	13
2 Materiais e métodos.....	14
2.1 Vetor retroviral pMFG-FVIII-P140K.....	14
2.2 Linhagens celulares.....	15
2.3 Análise da expressão do mRNA do FVIII por RT-PCR.....	15
2.4 Quantificação da atividade biológica do fator FVIII.....	16
2.5 LM-PCR (Ligação Mediada por PCR)	17
2.5.1 Extração de DNA genômico	18
2.5.2 Digestão do DNA genômico com enzimas de restrição.....	19
2.5.3 Precipitação do DNA genômico com acetato de amônio.....	19

2.5.4	Preparo da sequência oligo-adaptadora (Linker) e ligação ao DNA.....	20
2.5.5	Primeira reação de PCR e “ <i>Nested</i> ” PCR	20
2.6	Clonagem no vetor pCR®2.1-TOPO.....	22
2.7	Transformação bacteriana e expansão clonal	22
2.8	Sequenciamento do DNA	22
2.9	Análise das sequências quanto ao padrão de integração	23
2.10	Análises estatísticas	25
3	Resultados	26
3.1	Avaliação da expressão do mRNA do FVIII recombinante nas linhagens celulares HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K	26
3.2	Análise da atividade do FVIII recombinante biologicamente ativo, pelas linhagens celulares HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K	28
3.3	Padrão de integração cromossômica.....	29
3.4	Frequência de integração do vetor pMFG-FVIII-P140K perto de Sítios de Ligação de Fatores de Transcrição (TFBS)	32
3.5	Frequência de integração do vetor pMFG-FVIII-P140K perto de ilhas CpGs	33
3.6	Frequência de integração em regiões gênicas do genoma	34
3.7	Frequência de integração em sítios frágeis do genoma	38
3.8	Frequência de integração do vetor pMFG-FVIII-P140K em regiões repetitivas do genoma	39
4	Discussão.....	42
5	Conclusão	50
6	Referencias Bibliográficas.....	52

RESUMO

A hemofilia A é uma doença caracterizada pela deficiência do fator VIII (FVIII) de coagulação sanguínea e atinge 1 em 5000 homens. Países desenvolvidos, utilizam como terapia o FVIII recombinante (rFVIII), por apresentar maior segurança e consistir uma fonte ilimitada. Estudos realizados em nosso laboratório, utilizando o vetor retroviral pMFG-FVIII-P140K, originaram as linhagens celulares humanas, HepG2FVIIIIdB/P140K (hepática) e Hek293FVIIIIdB/P140K (renal), que foram selecionadas pelo mesmo sistema de seleção, o tratamento *in vitro* com as drogas Benzilguanina e Temozolomide. O presente trabalho teve por objetivo caracterizar e detalhar o padrão de integração do vetor retroviral pMFG-FVIII-P140K nas duas linhagens celulares humanas, e verificar se o perfil de integração está relacionado ao tipo celular específico ou se este sofre influência da estratégia de seleção a qual as células foram submetidas anteriormente. Foi utilizada a técnica de LM-PCR (ligação-mediada por PCR) que permite localizar o sítio de integração do vetor viral, por meio do sequenciamento do produto de PCR obtido após a digestão com enzimas de restrição e ligação de uma sequência adaptadora (LINKER) ao DNA genômico. Após o sequenciamento as sequências foram submetidas a análises em bancos de dados de genomas (Human BLAT, *QuickMap* e DAVID/EASE) para caracterização dos respectivos clones. Também foi realizada uma análise da expressão relativa do mRNA relativo ao rFVIII por RT-PCR e da atividade biológica pelo teste de TTPA. Dessa forma foi possível observar que ambas as linhagens modificadas expressam o mRNA relativo ao rFVIII e que as células HepG2 e Hek293 apresentam níveis de secreção da proteína recombinante biologicamente ativa da ordem de 7,9 UI/mL e 2,1 UI/mL, respectivamente. Foram sequenciados um total de 201 clones da HepG2, sendo 123 similares ao genoma humano e dentre estes 73 considerados verdadeiros e 50 ambíguos. Da Hek293 foram sequenciados 221 clones, sendo 179 similares ao genoma humano e dentre estes 64 verdadeiros e 115 ambíguos. Observamos que o vetor retroviral pMFG-FVIII-P140K apresenta um perfil de integração não randômico e diferente entre as células estudadas, uma vez que na linhagem HepG2 houve uma preferência pelos cromossomos 19, 17 e 11, e na Hek293 pelo cromossomo 9. Em relação a regiões genômicas tais como distância de ilhas CpGs e de sítios de ligação de fatores de transcrição não houve diferença no perfil de integração em ambas as linhagens celulares, visto que nas duas células o vetor se inseriu preferencialmente a uma distância de $\pm$ 30-60Kb dessas regiões. Observamos também que houve uma integração dentro de genes codificadores de proteínas da ordem de 52% e 44% nas células HepG2 e Hek293, respectivamente, contudo em ambos os casos mais de 90% dessas inserções ocorreram em regiões intrônicas. Houve 20% de integrações em sítios frágeis do genoma na linhagem HepG2 e 17% na Hek293. Em suma este trabalho mostrou a integração do vetor retroviral pMFG-FVIII-P140K específica para as duas linhagens em estudo. Além disso, pudemos observar que em ambas as linhagens celulares, HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K, existem regiões gênicas dentro dos cromossomos na qual o vetor exibe uma preferência. O perfil de integração caracterizado aqui difere de outros trabalhos da literatura que utilizaram retrovírus derivados de MLV. Isso nos leva a crer que o padrão de inserção descrito pode estar relacionado ao fato de que essas células foram tratadas anteriormente com drogas quimioterapêuticas para seleção de um clone celular com maior nível de produção de rFVIII, levando conseqüentemente a seleção de um perfil de integração semelhante entre as linhagens celulares, mesmo estas tendo origens teciduais diferentes.

ABSTRACT

Hemophilia A is a disease characterized by deficiency of blood clotting factor VIII (FVIII) and affects 1 in 5000 men. Developed countries use as therapy for Hemophilia A recombinant FVIII (rFVIII) because rFVIII concentrates have proven efficacy and safety and this therapy consist of an unlimited supply.. Studies in our laboratory using the retroviral vector pMFG-FVIII-P140K, originate the human recombinant cell lines, HepG2FVIIIIdB/P140K (liver origen) and Hek293FVIIIIdB/P140K (renal origen) which were selected with high doses of Benzyladenine and Temozolomide. This study aimed to characterize the integration pattern of retroviral vector pMFG-FVIII-P140K in these human cell lines, and verify if the profile of integration is related to specific cell type or it was influenced by the selection strategy which cells have been previously submitted. We used the technique of LM-PCR (ligation-mediated PCR) that allows to locate the site of viral vector integration by the sequencing of PCR products. The sequences were obtained after genomic DNA digestion with restriction enzymes and subsequent connection of an adapter (LINKER) to 5'or 3' of the DNA fragment. The sequences were analyzed in genomic databases (Human BLAT, and Quickmap DAVID / EASE) for characterization of the respective clones. It was also performed an analysis of relative expression of rFVIII mRNA by RT-PCR and the biological activity was measured by APTT test. Thus it was observed that both cell lines express the mRNA relative to rFVIII and HepG2 and HEK293 cells have levels biologically active recombinant protein approximately 7.9 IU / mL and 2.1 IU / mL, respectively . We sequenced a total of 201 clones of HepG2FVIIIIdB/P140K, 123 of these sequences match to the human genome and, among those 73 were considered true and 50 were ambiguous. In Hek293FVIIIIdB/P140K 221 clones were sequenced, these total, 179 were similar to the human genome. 64 were true and 115 were ambiguous. We note that the retroviral vector pMFG-FVIII-P140K presents a profile of integration different and non-random between the studied cells lines, in HepG2FVIIIIdB/P140K was observed a preference of integration for chromosomes 19, 17 and 11, and in Hek293FVIIIIdB/P140K for the chromosome 9. In genomic regions such as CpG islands and transcription factors binding sites (TFBS), there was no difference in the profile of integration in both cell lines. In the two cell lines the vector is inserted preferably at a distance of $\pm 30 - 60$ Kb of these regions. We also observed that there was 52% of integration within genes encoding proteins (RefSeq genes) in HepG2FVIIIIdB/P140K and 44% in Hek293FVIIIIdB/P140K cells. In both cases more than 90% of the insertions occurred in intronic regions. There were 20% of integrations at fragile sites in the genome of HepG2 cell line and 17% in Hek293. In summary this study showed the integration profile of retroviral vector pMFG -FVIII-P140K in two different cell lines that produces FVIIIr protein are very similar. Furthermore, we observed that in both cell lines, HepG2FVIIIIdB/P140K and Hek293FVIIIIdB/P140K, there are genetic regions within the chromosomes in which the vector displays a preference. The integration profile featured described here differs in CpGs island and TFBS from other studies in the literature that used retroviruses derived from MLV. This leads us to believe that the integration pattern described here may be related to the fact that these cells were previously treated with chemotherapeutic drugs for selecting a cell clone with the highest level of production of rFVIII, consequently leading to selection of a similar profile integration between cell lines, even those having different tissue origins.

1 Introdução

1.1 Hemofilia A

A hemofilia A é uma doença genética ligada ao cromossomo X, causada pela ausência ou anormalidade do fator VIII (FVIII), um cofator necessário para a geração da fibrina. A deficiência de fator VIII é a mais comum dentre as coagulopatias, com uma incidência de aproximadamente 1 em cada 5.000 homens e atualmente afeta cerca de 400.000 pessoas no mundo (ALEDORT, 2007; MANNUCCI, 2008). Durante a coagulação, o fator VIII é clivado via trombina para produzir o fator ativado (FVIIIa), que serve como um cofator do fator IX, na ativação do fator X em fator Xa. Uma vez formado, o fator Xa converte a protrombina em trombina, que cliva fragmentos peptídicos de fibrinogênio para a forma insolúvel de polímeros, que resultam na formação do coágulo de fibrina. A fibrina age em conjunto com um agregado de plaquetas para reduzir a perda de sangue (SANDBERG *et al.*, 2001).

A gravidade da doença está relacionada à quantidade de FVIII ativo no plasma do paciente, e, de acordo com o fenótipo clínico, esta dividida em grave (< 1% de fator VIII ativo circulante), moderada (1% a 4%) e leve (5% a 25%) (ANTONARAKIS, 1998; POWELL, 2009) (Tabela I). Os sintomas clínicos são caracterizados pelo surgimento de sangramentos espontâneos ou sangramento pós-trauma. Ocorrem sangramentos nas articulações (joelho, tornozelo e cotovelo) e musculatura, que podem causar grandes inflamações e deformações nas articulações. Tanto os sangramentos espontâneos, quanto sangramentos pós-traumáticos oferecem risco à vida de pacientes sem tratamento (HOYER, 1994; ANTONARAKIS, 1998).

A terapia atual para a hemofilia A consiste na infusão de concentrados de FVIII (FVIII derivado de plasma humano ou de FVIII recombinante). O tratamento ideal, principalmente para a forma grave da doença, deve ter dois objetivos principais: 1) fornecer ao paciente

quantidade suficiente de fator VIII para evitar sangramentos espontâneos e, 2) fornecer fator VIII suficiente para que o paciente tenha uma função de coagulação normal após qualquer tipo de trauma (POWELL, 2009).

Este tratamento requer assistência médica e apoio familiar, principalmente devido à meia-vida relativamente curta da proteína, cerca de 10 horas, o que exige infusões constantes de FVIII (POWELL, 2009). Além disso, outros inconvenientes advindos da administração de FVIII são: o alto custo do tratamento, o desenvolvimento de anticorpos contra o FVIII em aproximadamente 30% dos pacientes, e no caso de concentrados derivados do plasma, a disponibilidade limitada e o risco de contaminação com doenças transmitidas pelo sangue (HAY, 2006; PEERLINCK e HERMANS, 2006),

Em países subdesenvolvidos e em desenvolvimento, como o Brasil, devido ao elevado custo dos tratamentos atuais, não é possível a terapia com infusões constantes visando evitar sangramentos espontâneos. Nesses países as infusões de FVIII são administradas somente após episódios de sangramentos pós-trauma, diminuindo a qualidade de vida desses pacientes, e fazendo necessário o desenvolvimento de novas tecnologias que visam atender essa população.

Tabela I - Correlação entre a atividade do FVIII plasmático e o grau de hemofilia A.

Atividade do FVIII no plasma	Classificação	Sintomas
< 1%	Grave	Sangramentos espontâneos, deformação das articulações.
1-4 %	Moderada	Sangramentos após intervenções operatórias, poucos sangramentos espontâneos.
5-25%	Leve	Sangramentos após intervenções operatórias, poucas dificuldades na vida diária

1.2 Fator VIII recombinante

O fator VIII recombinante (FVIIIr) tem muitas vantagens sobre os concentrados de fator VIII derivados do plasma, incluindo um elevado grau de pureza, não oferecer riscos de uma transmissão viral humano-derivada, e, a princípio, ter um suprimento ilimitado (CHEN *et al.*, 1999).

Tanto na Europa quanto nos Estados Unidos, quatro produtos de FVIIIr são licenciados para o uso clínico no tratamento de Hemofilia A. São eles o KOGENATE e o KOGENATE FS produzidos pela Bayer, RECOMBINATE pela Baxter e REFACTO pela Wyeth/Genetics Institut. Mais recentemente uma terceira geração de produtos, livres de contaminantes de origem animal, está disponível no mercado, sendo eles ADVATE (Baxter) (SHAPIRO, 2007) e XYNTHA (Wyeth).

Estudos pioneiros da estrutura primária da proteína determinada pela clonagem do cDNA (DNA complementar) do fator VIII, mostraram a presença de uma estrutura de domínios: A1-a1-A2-a2-B-a3-A3-C1-C2 (CHURCH *et al.*, 1984, VEHAR *et al.*, 1984; JENNY *et al.*, 1987). Atualmente, a proteína recombinante comercial é produzida em células CHO (Chinese Hamster Ovary) ou em BHK (Baby Hamster Kidney), que são transfectadas estavelmente com o FVIII completo ou sem o domínio B, o qual não é necessário para a atividade da proteína madura (LEE *et al.*, 1999; PIPE *et al.*, 2009). Todos esses recombinantes possuem atividade biológica semelhante ao FVIII derivado do plasma (MANNUCCI & GIANGRANDE *et al.*, 2000). Contudo, essas formas recombinantes de fator VIII que se encontram no mercado mostraram-se inviáveis para o uso dos pacientes nos países subdesenvolvidos por ser um medicamento extremamente caro, não sendo acessível para a maioria da população.

Estudos mostram que o uso de linhagens celulares humanas para a produção do FVIII oferece vantagens devido ao fato de que a maquinaria dessas células pode realizar modificações pós-traducionais necessárias à proteína, de forma mais semelhante ao que ocorre no corpo humano, do que as células murinas (MEI *et al.*, 2006). Além disso, quando

utilizados sistemas de transferência gênica viral, a expressão do transgene é mais duradoura e estável, uma vez que estes integram no genoma celular (PICANÇO *et al.*, 2007; RODRIGUES ES, 2006).

O sistema de transferência gênica retroviral apresenta diversas vantagens, como por exemplo: a) apresentam uma elevada eficiência de transdução em vários tipos celulares; b) permitem a integração do gene de interesse no genoma da célula alvo proporcionando uma expressão estável e duradoura e c) apresentam deleção de genes virais que estão ligados à patogênese e replicação, evitando assim a expressão de antígenos imunogênicos ou proteínas virais (SHERIDAN *et al.*, 2000; VANDAMME *et al.*, 2003).

Em vista disso, estudos realizados no Laboratório de Transferência Gênica da Fundação Hemocentro de Ribeirão Preto, permitiram a geração de linhagens celulares humanas para a produção de FVIII recombinante. Essas linhagens foram transduzidas com o gene do FVIII com domínio B deletado, utilizando o sistema de transferência gênica retroviral e lentiviral (PICANÇO *et al.*, 2007; RODRIGUES ES, 2006).

É de suma importância a caracterização da integração do vetor retroviral em cada uma das linhagens celulares modificadas, para uma posterior seleção de um clone celular que expresse elevados níveis de FVIII recombinante de maneira mais estável e duradoura. Linhagens celulares humanas transduzidas com vetor lentiviral e com gene do FVIII já foram caracterizadas quanto ao sítio de inserção e mostraram um perfil de integração semelhante entre elas, mesmo apresentando origens teciduais diferentes (rim e fígado) (DE SOUZA RUSSO-CARBOLANTE *et al.*, 2010). Uma vez que o padrão de integração do vetor lentiviral utilizado já foi descrito, torna-se necessária a caracterização do perfil de integração do vetor retroviral.

1.3 Estratégias de seleção do clone celular recombinante

Historicamente, a biotecnologia e a geração de vetores bicistrônicos portadores de genes marcadores de seleção em células de mamíferos caminharam juntos. Como consequência inúmeros sistemas vêm sendo explorados objetivando a super-expressão da proteína recombinante, entre eles, Timidina Quinase, Timidilato Sintetase, Diidrofolato Redutase e Glutathione S-transferase (LANDIS *et al.*, 2001; BLAKLEY e SORRENTINO, 1998; HANSSON *et al.*, 1999).

O fator VIII é uma proteína que exhibe fatores limitantes para a obtenção de um elevado nível de sua expressão. Entre os diferentes fatores, se destacam: 1) o tamanho do gene; 2) seu baixo nível de expressão, quando comparado com outras proteínas e 3) a secreção ineficiente do produto desse gene, que foi explicada pelo fato da maior parte do fator VIII sintetizado permanecer retida no retículo endoplasmático em associação com uma proteína residente denominada BiP de 78 kDa (“immunoglobulin-binding protein”) ou GRP78 (“glucose regulated protein”) (KAUFMAN, 1991). Duas maneiras de contornar esses problemas consistiriam em: a) utilização do sistema de transferência gênica viral e b) emprego de diferentes reagentes químicos para a seleção de um clone celular com altos níveis do fator VIII recombinante.

Com relação às estratégias de seleção do clone celular recombinante com níveis elevados do fator VIII, os três principais sistemas empregados em células de mamífero são: a) emprego do gene neomicina que confere resistência a geneticina (OH *et al.*, 1999); b) emprego da proteína GFP, que permite a seleção por citometria de fluxo (PICANÇO *et al.*, 2007; RODRIGUES, 2006) e c) emprego do gene DHFR que permite a seleção do clone celular após o tratamento com metotrexato (GRILLARI *et al.*, 2001).

Um quarto sistema de seleção, ainda não explorado para a produção do fator VIII recombinante é o emprego do gene mutante P140K que codifica uma enzima de reparo de DNA conhecida como AGT (DNA alkyltransferase) e que permite a seleção da população celular portadora do gene de interesse após o tratamento com as drogas quimioterapêuticas

O⁶Benzilguanina (BG) e Temozolomide (TMZ) (FONTES *et al.*, 2006). Este sistema de seleção age da seguinte forma: após a transfecção, a população celular mista é submetida a dois tratamentos com as drogas O⁶ Benzilguanina e o agente metilante Temozolomide. A O⁶ Benzilguanina é uma potente inativadora da enzima DNA alkyltransferase (AGT) selvagem, enzima codificada pelo gene *mgmt* (*O⁶-methylguanine-DNA-methyltransferase*), cuja função é reparar lesões do DNA provocadas por agentes metilantes (Temozolomide). Uma vez que a enzima AGT selvagem encontra-se inativada, não é possível para a célula reparar os danos causados em seu DNA, pelo Temozolomide, levando-a a morte. Porém, a célula portadora do gene P140K expressa uma forma mutante dessa enzima, resistente a O⁶Benzilguanina e, portanto capaz de realizar o reparo do DNA ao dano causado pela droga TMZ permitindo com que estas sobrevivam (KRAMER *et al.*, 2006; DAVIS *et al.*, 1999).

A utilização deste sistema de seleção é interessante visto que é um gene endógeno e não tem como produto a síntese de proteínas advindas de outros organismos, diferentemente do GFP, que produz uma proteína exógena e dos genes que conferem resistência a antibióticos.

1.4 Mecanismos de integração dos retrovírus

Os retrovírus são classificados pela sua capacidade de copiar seu genoma de RNA em DNA, utilizando uma enzima viral denominada transcriptase reversa (RT) (ANDERSON e HOPE, 2005). Os retrovírus podem infectar células em divisão (ROE *et al.*, 1993; LEWIS e EMERMAN, 1994) e, no caso dos lentivírus, células que não estejam em divisão (LEWIS *et al.*, 1992), e de forma eficiente, são capazes de atravessar o citoplasma e chegar ao núcleo da célula, para a integração e expressão dos genes presentes em seu material genético (ANDERSON e HOPE, 2005).

Após entrar no citoplasma da célula infectada, o RNA genômico do retrovírus passa por uma transcrição reversa. O DNA viral, a integrase e certas proteínas virais e celulares

constituem um complexo de pré-integração que é levado para o interior do núcleo (FLINT *et al.*, 2004) (Figura 1). Já dentro do núcleo da célula hospedeira a enzima integrase catalisa a integração do DNA viral ao DNA da célula. Contudo, esse processo se aplica apenas a alguns retrovírus como HIV-1 e ASV (FLINT *et al.*, 2004). O vírus da leucemia murina (MLV) e seus vetores derivados não conseguem passar pelos poros do núcleo celular e só conseguem acessar o DNA da célula hospedeira quando o envelope nuclear encontra-se desintegrado durante a mitose (ROE *et al.*, 1993).

Figura 01: Esquema geral da integração dos retrovírus no genoma das células hospedeiras. Após entrar no citoplasma da célula infectada, o RNA genômico do retrovírus passa por uma transcrição reversa. O DNA viral, a integrase (IN) e certas proteínas virais e celulares constituem um complexo de pré-integração que é levado para o interior do núcleo, local no qual a enzima integrase catalisa a integração do DNA viral ao DNA da célula.

1.5 Organização do genoma humano

Sabe-se hoje que existem aproximadamente 25000 genes no genoma humano (*International Human Genome Sequencing Consortium*, 2004). Os primeiros estudos sugeriam que os retrovírus poderiam se integrar próximo a essas unidades transcricionais (SCHERIDIN *et al.*, 1990). Contudo não estava clara qual a porcentagem do genoma que continha esses sítios de integração favoráveis. Só houve uma mudança nessa situação após o seqüenciamento completo do genoma humano, o que permitiu análises estatísticas confiáveis a respeito dos sítios de integração (DANIEL e SMITH, 2008).

Após vários estudos observou-se que a integração do retrovírus não é aleatória (SCHRODER *et al.*, 2002; WU *et al.*, 2003; MITCHELL *et al.*, 2004) e, dessa maneira, foi possível elucidar os principais sítios de integração preferencial dos retrovírus. Dentre esses sítios de preferência de inserção podemos destacar as regiões repetitivas do genoma (SINE, LINE, LTR, etc.), ilhas CpGs e sítios frágeis, além das regiões gênicas ativas e sítios de início de transcrição (WU *et al.*, 2003; MITCHELL *et al.*, 2004; LEWINSKI *et al.*, 2006; AMBROSI *et al.*, 2008).

As regiões repetitivas estão amplamente distribuídas ao longo do genoma humano constituindo cerca de 50% deste. Podem ser classificadas em dois grandes grupos: as repetições de DNA “*in tandem*” e as repetições de DNA dispersas. Dentre essas ultimas encontram-se os elementos transponíveis, os quais estão subdivididos em famílias: os LINEs que possuem 850.000 cópias (21% do genoma humano), os SINEs que aparecem em número de 1.500.000 cópias (13% do genoma humano), as sequencias de LTR retrotransposons (8% do genoma humano) e os DNA transposons (3% do genoma humano) que apresentam 443.000 e 300.000 cópias respectivamente (RICHARD *et al.*, 2008).

Apesar de não se saber ao certo a função dos DNAs repetitivos, diferentes estudos sugerem o envolvimento dessas sequências no processo de replicação do DNA, recombinação, rearranjos cromossômicos, tais como deleções, duplicações, inversões e variações cariotípicas observadas em muitos grupos. Outro papel significativo das

sequências repetidas pode ser representado por segmentos presentes nos centrômeros e telômeros dos cromossomos, desempenhando funções cruciais na manutenção e propagação do material genético nuclear (LANDER *et al.*, 2001).

As ilhas CpGs são regiões ricas em citosina e guanina (geralmente acima de 50% dos nucleotídeos da região), diferenciando-se do resto do genoma (ILLINGWORTH e BIRD, 2009). Existem cerca de 25.000 ilhas CpGs no genoma humano (ILLINGWORTH *et al.*, 2008). Essas sequências únicas possuem aproximadamente 1Kb de comprimento e sobrepõem regiões promotoras de 60-70% dos genes humanos (LARSEN *et al.*, 1992; LANDER *et al.*, 2001). As ilhas CpGs apresentam colocalização com os promotores de todos os genes constitutivamente expressos e com cerca de 40% dos genes que apresentam um perfil de expressão restrito ao tecido específico (LARSEN *et al.*, 1992; ILLINGWORTH e BIRD, 2009).

Além das sequências repetitivas de DNA e ilhas CpGs, existem regiões do genoma, conhecidas como sítios frágeis, nas quais podem ocorrer a integração do vetor retroviral. Essas regiões apresentam instabilidade cromossômica e não estão distribuídas ao acaso no genoma podendo induzir a formação de “*gaps*”, quebras ou fissuras em cromossomos metafásicos (DURKIN e GLOVER, 2007). O genoma humano apresenta cerca de 110 sítios frágeis localizados em regiões de bandas G e, portanto, estão associados a genes ativos. Podem também estar associados com o locus de muitos oncogenes e correspondem as quebras cromossômicas em muitos tipos de câncer.

1.6 Sítios de integração preferencial dos retrovírus

Devido a sua propriedade de integração no genoma da celular os vetores retrovirais começaram a ser utilizados em estudos de terapia gênica e de introdução de genes exógenos nas células hospedeiras. Contudo, o risco de ocorrer mutagênese e/ou inserção

dentro ou próximo a proto-oncogenes sempre foi eminente, e estudos para mapear os sítios de integração dos vetores começaram a ser realizados.

Um estudo pioneiro, realizado por Rohdewohld *et al.* em 1987, objetivou mapear a integração do retrovírus M-MuLV. Neste trabalho os autores infectaram a linhagem celular NIH3T3 e camundongos Mov. Utilizando técnicas de Southern Blot e enzimas de restrição puderam observar que tanto *in vitro* como *in vivo* houve uma preferência de integração do retrovírus por sítios hipersensíveis a DNase (ROHDEWOHL D *et al.*, 1987).

Historicamente, acreditava-se que os eventos de integração dos retrovírus eram randômicos e considerava-se remota a chance de acidentalmente quebrar ou ativar um gene. Com o sequenciamento do genoma humano completo, um grande número de estudos sobre sítios de integração viral se tornou possível. Schroder *et al.* investigaram a inserção do vírus da imunodeficiência humana (HIV) e os vetores derivados do HIV, em linhagens celulares humanas linfóides (SupT1). Esses autores encontraram que as regiões gênicas são fortes sítios receptores de integração lentiviral. Uma análise global dos transcritos das células SupT1 indicaram que genes ativos são alvos preferenciais para a integração do HIV (SCHRODER *et al.*, 2002).

Em 2003, Wu *et al.* analisaram a integração do retrovírus da leucemia murina (MLV), em células HeLa. Nesses estudos, encontram que o MLV tem o sítio de integração preferencial próximo $\pm 5\text{Kb}$ a regiões de inícios de transcrição (tanto a montante quanto a jusante) e próximos $\pm 1\text{Kb}$ de ilhas CpGs. Em contrapartida, observaram que a integração do HIV-1 pode ocorrer em qualquer lugar da unidade transcricional (WU *et al.*, 2003). Mitchell *et al.* comparou o perfil de integração dos vírus HIV com a do MLV e observaram que nas duas linhagens celulares primárias (células mononucleares de sangue periférico e fibroblasto de pulmão IMR90) a integração do MLV é favorecida por regiões próxima $\pm 1\text{Kb}$ de ilhas CpGs e no caso do HIV ocorre preferencialmente em genes ativos (MITCHELL *et al.*, 2004), sugerindo que os dois vetores possuem um perfil distinto de integração.

No mesmo ano de 2003 um estudo de terapia gênica utilizou vetores derivados do retrovírus MLV para o tratamento da síndrome da imunodeficiência severa ligada ao

cromossomo X (SCID-X1) (HACEIN-BEY-ABINA *et al.*, 2003). Após cinco anos o mesmo grupo de pesquisadores mostraram que 4 dos 11 pacientes tratados desenvolveram leucemia de células T devido a integração do vetor perto do proto-oncogene LMO2 (HACEIN-BEY-ABINA *et al.*, 2008), colocando em dúvida a segurança da terapia gênica.

Estudos de terapia gênica em animais e mesmo em seres humanos comumente empregam vetores derivados de MLV ou HIV-1. Uma análise dos pacientes utilizados nesses estudos mostraram uma distribuição dos sítios de integração não aleatória (DEICHMANN *et al.*, 2007). As integrações ocorreram preferencialmente perto da região 5' de genes e associada a ilhas CpGs (BUSHMAN, 2007; DEICHMANN *et al.*, 2007), o que corrobora com os resultados obtidos com MLV em culturas celulares.

Todos esses estudos citados anteriormente mostraram que além dessas preferências, a sequência de DNA específica do retrovírus parece ter um papel importante na seleção dos sítios de integração, contudo não somente elas que governam a interação viral sendo que este mecanismo pode incluir cofatores e outras estruturas celulares (DANIEL e SMITH, 2008).

Com o intuito de descrever e entender como agem os fatores celulares envolvidos no mecanismo de integração dos retrovírus, muitos estudos começaram a ser desenvolvidos. Em 2003, Cherepanov e colaboradores, utilizando experimentos que demonstram a interação proteína-proteína (*yeast two-hybrid assay*) identificaram uma proteína ligante da integrase do HIV-1 denominada LEDGF/p75. Essa proteína se liga a sítios específicos da integrase viral mediando o contato entre esta e a cromatina da célula hospedeira. Esta associação faz com que o complexo de pré-integração do HIV-1 se ligue perto ou dentro de genes (CHEREpanov *et al.*, 2003, CHEREpanov *et al.*, 2004).

Uma vez descrita a função da LEDGF/p75 na integração do HIV-1, estudos começaram a buscar novas proteínas celulares que fariam parte do processo de integração de outros retrovírus. Em 2008, um estudo realizou um rastreamento das proteínas que estariam participando da interação com integrase dos retrovírus MoMLV e HIV-1. Utilizando ensaios de interação proteína-proteína (*yeast two-hybrid assay*) e ensaios de ligação *in vitro*

foi possível a identificação de 27 proteínas que atuam na integração de ambos os vírus. Dentre elas encontram-se fatores transcricionais, proteínas de reparo de DNA e remodelamento de cromatina, helicases, proteínas ligantes de RNA, entre outras (STUDAMIRE e GOFF, 2008).

Recentemente, um estudo identificou o fator de transcrição ZASC1 como uma das proteínas celulares que interagem exclusivamente com o retrovírus MLV, ligando-se com a região LTR e atuando tanto na infecção quanto na transcrição dos genes virais (BRUCE *et al.*, 2010).

É de suma importância que os fatores celulares que contribuem para a inserção já estejam sendo elucidados e que os sítios de integração preferencial para cada tipo de retrovírus já estejam descritos na literatura. Contudo ainda existem muitos estudos que precisam ser realizados para entender de forma clara e completa os mecanismos de integração dos vetores retrovirais em cada tipo celular específico.

Em 2006, estudos realizados na Fundação Hemocentro de Ribeirão Preto, gerou duas linhagens celulares produtoras de FVIII recombinante, utilizando o vetor retroviral derivado de MLV, denominado pMFG-FVIII-P140K (RODRIGUES, 2006). Estas células foram selecionadas por meio do tratamento com alta estrigência (400 μ M) com as drogas quimioterapêuticas O⁶Benzylguanina (BG) e Temozolomide (TMZ), para a obtenção de uma população celular produtora de elevados níveis de FVIII recombinante. Uma vez que essas células foram tratadas com a mesma concentração de drogas e possuem origens teciduais diferentes (Hek293 é renal e HepG2 é hepática) faz-se necessária a descrição e o entendimento do perfil de integração do vetor pMFG-FVIII-P140K nas linhagens celulares HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K, para uma melhor caracterização das células produtoras de FVIII e posterior seleção de um clone celular capaz de produzir elevados níveis da proteína recombinante de maneira mais estável e duradoura. Além disso, de posse dos conhecimentos a respeito do padrão de integração dos retrovírus e que esta pode estar relacionada com fatores celulares tecido específico, também nos cabe investigar

se o tratamento das células por meio de agentes químicos pode influenciar na seleção de um perfil de integração específico.

1.7 Objetivos

O presente trabalho visa caracterizar e detalhar o padrão de integração do vetor retroviral pMFG-FVIII-P140K em duas linhagens celulares humanas, HepG2 e Hek293, e verificar se o perfil de integração está relacionado ao tipo celular específico ou se este sofre influencia da estratégia de seleção a qual as células foram submetidas anteriormente.

2 Materiais e métodos

2.1 Vetor retroviral pMFG-FVIII-P140K

Foi utilizado nesse estudo o vetor plasmidial retroviral bicistrônico pMFG-FVIII-P140K que possui as regiões 5' LTR e 3' LTR derivadas do *Moloney Murine Leukemia Virus* (MoMuLV). Contém o elemento IRES (*Internal Ribosome Entry Site*) entre o sítio múltiplo de clonagem, onde foi clonado o gene do FVIII e o gene P140K que confere resistência à droga sintética O⁶ Benzilguanina e ao agente metilante Temozolomide. Possui o gene Amp que confere resistência a ampicilina e a região gênica designada a emitir o sinal de empacotamento viral, $\Psi+$ (Figura 02).

Figura 02: Vetor de DNA plasmidial retroviral bicistrônico pMFG-FVIII-P140K.

2.2 Linhagens celulares

Hep-G2 - linhagem celular de hepatócito humano, derivada de carcinoma hepatocelular. Foram adquiridas do American Type Culture Collection (ATCC) sob o número de acesso HB-8065.

Hek-293 - linhagem celular de rim humano da fase embrionária. Foram adquiridas do American Type Culture Collection (ATCC) sob o número de acesso CRL-1573.

Ambas as linhagens celulares foram cultivadas em incubadoras com 5% de CO₂ e 80% de umidade, em meio DMEM contendo 10% de soro fetal bovino inativado, L-Glutamina 4mM, penicilina 100U/mL e streptomycin 100µg/mL.

2.3 Análise da expressão do mRNA do FVIII por RT-PCR

A análise da expressão do mRNA relativo ao fator VIII nas linhagens celulares HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K foi realizada por meio da reação de Transcriptase Reversa (RT-PCR).

Realizou-se a extração de RNA total de cerca de 5×10^6 células utilizando-se o Kit Mini-Spin RNA extraction da QIAGEN, seguindo as instruções do fabricante. Após a extração do RNA, a pureza e concentração do RNA total foram avaliadas por análise de absorvância a 260 nm e 280 nm.

Para a síntese de cDNA foi utilizado o kit SuperScriptTM First-Strand (Invitrogen) no termociclador GeneAmp PCR System 9700 (Applied Biosystems), de acordo com o protocolo de amplificação: 65°C por 5 min (adição de RNA, Primers, dNTPs, H₂O); 4°C por 2 min; 25°C por 2 min (adição de 10X Buffer RT, 25 mM MgCl₂, DTT, RNaseOUT); 25°C por 10 min (adição da Enzima Transcriptase Reversa); 42°C por 50 min (síntese de cDNA) e 70°C por 15 min (inativação da enzima transcriptase reversa).

Após a síntese de cDNA, o primeiro passo consiste em checar se a reação da Transcriptase Reversa funcionou por meio da amplificação de um gene endógeno. Neste

caso o gene de escolha foi a beta-actina. Para esta reação foi utilizado: 3 µl de cDNA, 2,5 µl de 10X PCR Buffer, 0,75 µl de 50 mM MgCl₂, 20 pmoles de “*primer*” 5’ (5’-GGCATCGTGATGGACTCCG-3’) e 20 pmoles de “*primer*” 3’ (3’-GCTGGAAGGTGGACAGCGA-5’), 1 µl de dNTP a 10 mM, 1 µl da enzima Taq DNA Polymerase (1 U/mL) em um volume final de 25 µl. A reação foi submetida à seguinte termociclagem: 94°C por 1 min; 25 ciclos de 94°C por 1min, 55°C por 1min e 72°C por 1 min; 72°C por 1 min.

Os produtos das reações foram analisados após a eletroforese em gel de agarose 1%. Após a confirmação da viabilidade do cDNA, o mesmo foi utilizado para a reação de PCR utilizando agora, um par de primers para a amplificação de uma sequência da cadeia pesada do FVIII. A sequência de DNA do par de “*primer*” utilizada foi P5A1seq-II 5’-CTTGGACGATTTCTACTG-3’ e P3A2seq-I 3’-GACGGACATCAGTGATTC-5’, o que resultou em um produto de 546 pb.

As condições da reação foram similares às descritas para a amplificação do mRNA relativo ao beta-actina, com diferença apenas no ciclo de amplificação, sendo este: 94°C por 1 min; 30 ciclos 94°C por 1min, 52°C por 1min e 72°C por 1 min; 72°C por 1 min. Os produtos das reações foram analisados após a eletroforese em gel de agarose 1%.

2.4 Quantificação da atividade biológica do fator FVIII

Para a detecção do FVIII ativo no sobrenadante das culturas celulares utilizamos o teste do tempo de tromboplastina parcial ativada (TTPA). A dosagem da atividade biológica iniciou-se com a obtenção de uma curva de calibração padrão construída utilizando como referência o FVIII humano derivado do plasma conhecido como *Verify Reference Plasma* (Organon Teknika, Durham). Este plasma referência foi reconstituído em 1mL de água destilada, concentração final de 1 IU/mL de FVIII, e uma série de diluições seriadas na razão de dois foram realizadas utilizando a solução tampão fisiológica recomendada (Owren’s Veronal - Biomeriëux). A partir do tempo de coagulação médio de cada diluição foi traçada

uma curva de calibração, em que o eixo da ordenada indica o tempo de coagulação sangüínea e o eixo da abscissa as respectivas diluições do plasma referência. Ao final a atividade biológica do FVIII foi obtida com a média da inclinação da curva, e convertida automaticamente em porcentagem, considerando que 100% atividade representa 1 IU/mL (CHNG et al., 2005).

Após 24 horas de cultivo das linhagens celulares, as amostras foram diluídas 1:10 em tampão (*Owren's Veronal - Biomérieux*), 100µl dessa diluição foi misturada com 100µl de plasma deficiente em FVIII humano (*Factor Deficient Plasma - Biomérieux*), e a reação incubada por 20 seg a 37°C. Em seguida adicionou-se 100µL do ativador de contato (*Platelin® LS - Biomérieux*) e a reação incubada novamente por 240 seg a 37°C. Ao término dessa reação adicionou-se 100µL do cloreto de cálcio (CaCl_2 0,025M) e imediatamente iniciou-se a contagem do tempo de coagulação sangüínea utilizando o coagulômetro automático (*COAG-A-MATE® XM - Organon Teknika*) conforme as instruções do fabricante.

2.5 LM-PCR (Ligação Mediada por PCR)

Para a caracterização do padrão de integração do vetor pMFG-FVIII-P140K nas linhagens celulares HepG2 e Hek293 foi utilizada a técnica de LM-PCR (Figura 04).

Figura 03: LM-PCR. Técnica que permite localizar o gene de interesse pelo sequenciamento do produto de PCR obtido após a ligação de uma sequência adaptadora (LINKER) ao DNA genômico. Este foi previamente digerido com enzima cujo sítio de restrição está no vetor retroviral a montante do sítio de inserção do gene de interesse. Assim, construiu-se uma biblioteca de DNA, no plasmídeo PCR 2.1 topo, contendo produtos de LM-PCR e realizou-se o sequenciamento do DNA. As sequências foram submetidas à análises em bancos de dados de genomas (Human BLAT, GTSG/Vector integration site mapping e DAVID/EASE) para caracterização dos respectivos clones.

2.5.1 Extração de DNA genômico

As células das linhagens celulares HepG2 e Hek293 transduzidas com o vetor pMFG-FVIII-P140K foram cultivadas em meio DMEM contendo 10% de soro fetal bovino até atingirem a confluência de 80%. Em seguida, 3×10^6 células de cada uma das linhagens, foram submetidas a extração de DNA genômico, utilizando o kit DNAesy (Invitrogen), seguindo as instruções do fabricante. Após a extração, a qualidade do DNA foi verificada por

meio de uma eletroforese em gel de agarose 1% (Figura 04), e quantificado em espectrofotômetro a 260 nm.

Figura 04: DNA genômico das linhagens celulares HepG2FVIIIdB/P140K (1) e Hek293FVIIIdB/P140K (2). Eletroforese em gel de agarose 1% corado com brometo de etídio na qual observa-se bandas de alto peso molecular, sugerindo boa qualidade do DNA genômico.

2.5.2 Digestão do DNA genômico com enzimas de restrição

O DNA genômico das células foi digerido com as enzimas de restrição *Mse I* e *Sac I*. Para a reação de digestão foram utilizados 5 µg de DNA, 5 µl de tampão 10X da enzima, 0,5 µl de BSA 0,1%, 2 µl da enzima *Mse I* (10.000 U/mL), 1,5 µl da enzima *Sac I* (20.000 U/mL) e água mili-Q até completar o volume para 50 µl. A reação foi então incubada por 8 horas a 37°C. Em seguida foi realizada a inativação das enzimas de restrição, colocando-se a reação a 60°C durante 20 min.

2.5.3 Precipitação do DNA genômico com acetato de amônio

O DNA genômico digerido foi precipitado com acetato de amônio. Inicialmente adicionou-se ao DNA digerido TE em pH 8,0 até completar o volume para 100 µl. Após foi adicionado 100 µl de fenol buffer saturado (Invitrogen) e 100 µl de clorofórmio 100%, com o intuito de eliminar as enzimas. A reação foi então agitada em vortex, seguida de centrifugação a 13000 rpm durante 10 minutos. Depois de centrifugada, a fase aquosa

que contem o DNA digerido foi coletada e colocada em tubo eppendorf estéril. Em seguida foi adicionado ao DNA 66,5 ul de acetato de amônio 7,5 M, 1,5 ul de glicogênio (20mg/mL) e 500 ul de etanol 100%. A reação foi incubada a -20°C “*overnight*”. A precipitação do DNA é obtida por centrifugação a 13000 rpm, 4°C durante 30 min. O sobrenadante foi então descartado e o precipitado foi lavado 2x com etanol 70%. Após as lavagens, deixou-se o DNA secar por 30 min e esse foi ressuscitado em 20 ul de água mili-Q. Para avaliar a qualidade do DNA foi realizada um eletroforese em gel de agarose 1%, utilizando-se 5 ul da amostra.

2.5.4 Preparo da sequência oligo-adaptadora (Linker) e ligação ao DNA

O “*linker*” é uma sequência de oligonucleotídeo que se liga na extremidade do DNA previamente digerido com a enzima de restrição *Mse I*. Para sua preparação foram utilizados 240 pmols do “*linker forward*” (5'-GTAATACGACTCACTATAGGGCTCCGCTTAAGGGAC-3') e 240 pmols do “*linker reverse*” (5'-TAGTCCCTTAAGCGGAG-NH₂-3'), em um volume final de 59,29 ul. A reação foi então colocada em termociclador a 95°C por 3 min e após esse período foi diminuído 1°C por minuto até 4°C. Alíquotas de 5,5 ul foram feitas e armazenadas a -20°C.

Após o preparo do “*linker*”, a reação de ligação foi realizada de forma que 7 ul do DNA digerido foi colocado juntamente com 5,5 ul do linker, 1 ul da enzima T4 DNA Ligase (2.000.000 U) (Invitrogen), 2 ul de tampão 10x da enzima e água mili-Q para completar o volume de 20 ul. Essa reação foi incubada a 16°C “*overnight*”.

2.5.5 Primeira reação de PCR e “*Nested*” PCR

A primeira reação de PCR foi realizada em um volume final de 50µL utilizando-se 10 pmoles do DNA ligado ao “*linker*”, 10 pmoles do “*primer*” AP-1 (5'-

GTAATACGACTCACTATAGGGC -3'), 10 pmoles do "*primer*" MseI 5' LTR Retro (5'-TAGCTTGCCAAACCTACAGGT -3'), 2 µl de dNTP a uma concentração de 10 mM, 5 µl de tampão 10x, 1,5 µl de MgCl₂ e 1,5 µl de Taq DNA polimerase (Invitrogen). Essa reação foi levada ao termociclador GeneAmp® PCR System 9700 (Applied Biosystems) com o programa de amplificação: 95°C por 2 min; 35 ciclos de 95°C por 25 s, 55°C por 30 s e 72°C por 1 min; 72°C por 4 min.

O produto da primeira reação de PCR foi diluído na proporção 1:50 e 1 µl dessa solução foi utilizada na reação de "*Nested*" PCR. Essa segunda reação foi realizada utilizando-se os mesmos parâmetros da primeira, contudo os "*primers*" utilizados foram AP-2 (5'-AGGGCTCCGCTTAAGGGAC -3') e "*primer*" MseI 5' LTR Nested (5'-ACCTACAGGTGGGGTCTTTCA -3'). Ao final da reação foi realizada uma eletroforese em gel de agarose 3% para verificar a integridade do produto da reação, no qual foi possível observar um rastro referente a um conjunto de bandas (Figura 05).

Figura 05: Produto de "*nested*" PCR das linhagens HepG2FVIIIdB/P140K e Hek293FVIIIdB/P140K. Eletroforese em gel de agarose 3% do produto da reação de "*nested*" PCR das linhagens celulares HepG2FVIIIdB/P140K (1) e Hek293FVIIIdB/P140K (2). É possível observar um rastro referente a um conjunto de bandas. Foi utilizado os "*primers*" referentes ao LTR viral e a sequência adaptadora ("*linker*"), para a amplificação da região de inserção do retrovírus no genoma humano. (3) controle negativo da reação.

2.6 Clonagem no vetor pCR®2.1-TOPO

A clonagem no vetor de seqüenciamento prosseguiu incubando-se 4 ul do produto da reação de Nested PCR com 1 ul de Salt solution e 1 ul de TOPO vector, totalizando um volume de 6 ul. Essa reação foi incubada a temperatura ambiente por 30 minutos e depois realizada a transformação bacteriana.

2.7 Transformação bacteriana e expansão clonal

Foram adicionados 6 µL da reação de ligação em 150 µL bactérias competentes. Por meio de incubação a 30 min no gelo, choque térmico 42°C por 50 s e 2 min no gelo, procedeu-se à transformação. A seguir foi adicionado 250µL de meio SOC (20mM glicose e meio SOB contendo 1mM MgCl₂, 1mM MgSO₄) e incubado por 1 h à 37°C sob agitação. Posteriormente, as bactérias transformadas foram colocadas em placas de Petri com meio de cultura LB-ágar contendo o antibiótico ampicilina (MERCK) a uma concentração final de 0,1 mg/mL, com a finalidade de selecionar as colônias contendo plasmídeos recombinantes. Cada uma das colônias de bactérias transformadas foram colocadas em 150 ul de meio LB com o antibiótico ampicilina a 0,1 mg/mL (MERCK) em placa de 96 poços e incubadas a 37°C, “overnight”, em um processo denominado expansão clonal. Essas bactérias transformadas foram estocadas utilizando glicerol 70% e armazenadas em freezer -80°C.

2.8 Sequenciamento do DNA

Para o sequenciamento do vetor pCR®2.1-TOPO contendo o DNA de interesse, inicialmente foi realizada uma reação de PCR utilizando 1 ul da expansão clonal bacteriana juntamente com 3,2 pmoles de “*primer M13 forward*” (5’ – GTAAAACGACGGCCAG – 3’), 3,2 pmoles de “*primer M13 reverse*” (5’ – CAGGAAACAGCTATGAC – 3’), 1,5 ul de dNTP a uma concentração de 1,25 mM, 1,5 ul de tampão 10x, 0,45 ul de MgCl₂ e 0,2 ul de Taq DNA

polimerase (Invitrogen) em um volume final de 15 μ l. Essa reação foi submetida a uma termociclagem de 95°C por 10 s, 35 ciclos de 95°C por 40 s, 60°C por 40 s e 72°C por 55 s; 72°C por 5 min.

O produto dessa primeira reação de PCR para o seqüenciamento foi submetido a uma segunda reação na qual 1 μ l deste foi adicionado a 4 μ l de DYEnamic ET Terminator Kit (GE Healthcare) e 3,2 pmoles do “*primer*” M13, em um volume final de 10 μ l. Essa reação foi submetida a uma termociclagem de 95°C por 2 min, 35 ciclos de 95°C por 10 s, 51°C por 15 s e 60°C por 1 min. A seguir o produto de DNA foi precipitado seguindo o protocolo ABI/Isopropanol (MERCK) e submetida à eletroforese no aparelho MegaBACE 1000 (GE Healthcare). Na forma de eletroferograma as seqüências de DNA geradas foram diretamente enviadas, para um computador conectado ao aparelho sequenciador. Estes eletroferogramas foram interpretados pelo “*software*” DNA Sequencing Analyser 4.0 e convertidos em seqüências de DNA, as quais foram posteriormente analisadas.

2.9 Análise das seqüências quanto ao padrão de integração

As seqüências geradas foram submetidas a uma primeira análise utilizando o programa ChromasPro. Nessa primeira etapa foram selecionadas apenas as seqüências que continham em uma das extremidades a seqüência adaptadora (“*linker*”) e na outra a seqüência referente ao LTR retroviral. Em seguida, as seqüências foram convertidas para o formato Fasta e enviadas ao banco de dados do genoma “Human BLAT” (<http://genome.ucsc.edu/cgi-bin/hgBlat>) para análise quanto a homologia ao genoma humano. Apenas as seqüências que apresentaram mais de 95% de homologia com as seqüências depositadas no banco de dados foram consideradas.

Em seguida, as seqüências foram submetidas à ferramenta publica *QuickMap*, acessível no site www.gtsg.org, que permite um mapeamento detalhado do local de integração do vetor retroviral (APPELT *et al.*, 2009). Essa ferramenta analisa uma grande

variedade de dados referentes às características genômicas (genes, sítios de início de transcrição, ilhas CpGs, sítios de ligação de fatores de transcrição, dentre outros) em uma distância de ± 250 Kb em torno do sítio de integração do vetor. Além disso, as fitas *plus* e *minus* foram analisadas separadamente e a distância individual de cada uma das características até a base da integração é calculada em pares de bases (Figura 06).

Figura 06: Perfil dos sítios de integração. Os retrovírus integram seu genoma (5'LTR, transgene, 3'LTR) no genoma das células hospedeiras. Como indicado, o DNA hospedeiro é fita dupla (fitas + e -), e, cada fita pode conter vários elementos genômicos tais como genes, TSS (sítio de início de transcrição), TFBS (sítio de ligação de fatores de transcrição), ilhas CpGs, LINE, SINE, elementos LTR ou elementos simples. Estas características genômicas podem ser rompidas pelo vetor (Gene A), ou se inserir adjacente (Gene B). As distâncias (D) de cada uma das características até o sítio de integração são calculadas. Para cada integração uma variedade de informações, são geradas e acumuladas (retirado e modificado de APPELT *et al*, 2009).

Para a classificação das regiões gênicas nas quais o vetor se inseriu foi utilizada a ferramenta publica DAVID/EASE (<http://david.abcc.ncifcrf.gov/>), a qual permite identificar os genes, suas funções biológicas segundo critérios do Gene Ontology, descobrir grupos de genes relacionados funcionalmente, fazer associações dos genes com doenças, visualizar vias metabólicas relacionadas, entre outros (HUANG *et al.*, 2009).

2.10 Análises estatísticas

O teste do Qui-quadrado com a correção de Yates foi utilizada para analisar se o número de integrações observadas (i_o) segue uma distribuição multinomial como as integrações esperadas (i_e) para os 24 cromossomos (22 autossomos, X e Y). A integração esperada foi calculada assumindo uma distribuição uniforme discreta, corrigida pela distribuição do tamanho do cromossomo [$i_e = (\text{número de bases do cromossomo} / \text{número de bases do genoma}) \times \text{número total de integrações}$]. O nível α de significância utilizado foi de 0,05. Um cut-off de $(i_o - i_e)^2 / i_e \geq 3$ foi utilizado para a determinação de sítios preferenciais de integração.

O teste exato de Fisher foi utilizado para determinar se havia uma associação entre as regiões de ilhas CpGs e TFBS nas duas linhagens celulares.

3 Resultados

3.1 Avaliação da expressão do mRNA do FVIII recombinante nas linhagens celulares HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K

Com o intuito de melhor caracterizar as linhagens celulares produtoras da proteína recombinante foi realizada a avaliação da expressão do mRNA do FVIII por meio da técnica de RT-PCR.

A análise em gel de agarose 1% dos produtos da reação mostrou a presença de uma banda de 546 pb referente ao final do domínio A1 e início do domínio A2 da cadeia pesada do fator VIII em ambas populações transgênicas, HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K, e a ausência da mesma nas linhagens celulares não transduzidas.

A figura 07 mostra a presença do cDNA relativo ao fator VIII na linhagem celular HepG2FVIIIIdB/P140K (raia 3B) e a ausência do mesmo na linhagem celular não transduzida (raia 1B). Como controle, foi avaliada a expressão do mRNA relativo a β -actina em que se observa a presença do mesmo tanto na linhagem celular não transduzida (raia 1A) quanto na linhagem celular recombinante (raia 3A) (Figura 07).

Figura 07: Expressão do fator VIII na linhagem celular HepG2FVIIIdB/P140K. Gel de agarose corado por brometo de etídeo após a eletroforese dos produtos de RT-PCR. Cerca de 2 ug de RNA total HepG2FVIIIdB/P140K e da HepG2 não transduzidas foram submetidos a reação com a transcriptase reversa e, em seguida, o cDNA foi submetido a reação de amplificação por PCR. Em A observamos fragmentos de 612pb representando o mRNAs da β -actina. Em B, fragmentos de 546pb referentes a cadeia leve do FVIII. Raias 1A e 1B: HepG2 não transduzida. Raias 3A e 3B: HepG2FVIIIdB/P140K. Raias 2A e 2B e 4A e 4B: controles negativos da reação, onde não foi adicionado a enzima RT. Raias 5A e 5B: controles negativos das reações de PCR. As raias M mostram o DNA marcador de peso molecular ϕ X 174 (fragmentos 1353; 1078; 872; 603 e 310 pb).

A figura 08 mostra a presença do cDNA relativo ao fator VIII na linhagem celular Hek293FVIIIdB/P140K (3B) e a ausência do mesmo na linhagem celular não transduzida (3A). Da mesma forma que na linhagem HepG2, como controle, foi avaliado a expressão do mRNA relativo a β -actina em que se observa a presença do mesmo tanto na linhagem não transduzida (1A) como na linhagem recombinante (1B).

Figura 08: Expressão do fator VIII na linhagem celular Hek293FVIIIΔB/P140K. Gel de agarose corado por brometo de etídeo após a eletroforese dos produtos de RT-PCR. Cerca de 2 ug de RNA total das linhagens Hek293 não transduzidas (A) e da Hek293FVIIIΔB/P140K (B) foram submetidos a reação com a transcriptase reversa e, em seguida, o cDNA foi submetido a reação de amplificação por PCR. Nas raias 1A e 1B observamos fragmentos de 612pb representando o mRNAs da β-actina. Em 3B observamos fragmentos de 546pb referentes à cadeia leve do FVIII e a ausência do mesmo na raia 3A. Raias 2A e 2B e 4A e 4B: controles negativos da reação, onde não foi adicionada a enzima RT. A raia 5 é o controle positivo da reação de PCR, em que foi utilizado o plasmídeo recombinante pMFG-FVIII-P140K e a raia 6 refere-se ao controle negativo da reação de PCR. As raias M mostram o DNA marcador de peso molecular φX 174 (fragmentos 1353; 1078; 872; 603 e 310 pb).

3.2 Análise da atividade do FVIII recombinante biologicamente ativo, pelas linhagens celulares HepG2FVIIIΔB/P140K e Hek293FVIIIΔB/P140K

Para a caracterização das linhagens celulares recombinantes também foi realizado o teste da atividade biológica seguindo o protocolo do teste do tempo de tromboplastina parcial ativada (TTPA).

Conforme demonstrado na tabela II as linhagens celulares HepG2FVIIIΔB/P140K e Hek293FVIIIΔB/P140K mostraram níveis de secreção de FVIII recombinante biologicamente ativo, da ordem de 7,9 UI/mL e 2,1 UI/mL, respectivamente.

Tabela II – Atividade Biológica do FVIII recombinante em HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K

Amostra	Tempo (seg)	Ativ. (%) 1:10	IU/mL
1- Meio DMEM	101.8	1.5	0,01
2- Sob. HepG2 não transduzida	69.7	4.0	0,04
3- Sob. Hek293 não transduzida	72.7	12.7	0,12
4- Sob. HepG2FVIIIIdB/P140K	36.1	789.7	7,9
5- Sob. Hek293FVIIIIdB/P140K	46.9	209.2	2,1

3.3 Padrão de integração cromossômica

Foram sequenciados 201 clones da linhagem HepG2FVIIIIdB/P140K e 221 da linhagem Hek293FVIIIIdB/P140K. Quando submetidos ao banco de dados do genoma, Human BLAT, das 201 sequências da HepG2, 123 apresentaram similaridade (>95%) ao genoma humano, enquanto que na Hek293, das 221 sequências, 179 foram similares. Dentre os não similares, foi possível observar que estes continham apenas uma pequena sequência referente ao vetor retroviral, e por isso não apresentaram similaridade com o genoma humano.

Foram utilizados nesse estudo apenas clones os quais apresentavam em uma das extremidades a sequência correspondente ao LTR viral e na outra, a sequência adaptadora (linker). Os sítios de integração foram sequenciados até a saturação, ou seja, como temos duas linhagens celulares selecionadas com alta estrigência de drogas quimioterapêuticas, todos os possíveis locais de integração foram mapeados e, subsequente as inserções descritas aqui, os sítios começaram a se repetir. Em vista disso, não foi necessária a análise de um número mais elevado de sequências.

Após essa primeira triagem, as sequências de DNA foram submetidas a uma análise mais detalhada utilizando a ferramenta *QuickMap* (APPELT et al., 2009), que em primeira instância classifica as sequências em “verdadeiras”, quando estas são homologas a apenas um local no genoma humano, ou em “falsas”, quando as sequências apresentam ambiguidade, ou seja, são similares a mais de um local no genoma. Dentre as 123 sequências da linhagem HepG2FVIIIIdB/P140K, 73 foram consideradas verdadeiras enquanto as 50 restantes foram ambíguas. Na linhagem Hek293FVIIIIdB/P140K, das 179 sequências submetidas a análise, 64 foram verdadeiras e 115 ambíguas (Tabela III).

Tabela III – Dados gerais sobre os clones gerados

	HepG2FVIIIIdB/P140K	Hek293FVIIIIdB/P140K
Clones Sequenciados	201	221
Sequências homólogas ao genoma humano		
humano	123	179
Sequências Verdadeiras	73	64
Sequências Ambíguas	50	115

Em seguida, os clones verdadeiros foram caracterizados quanto ao sítio cromossômico de integração. As sequências dos clones gerados da linhagem HepG2/FVIIIIdBP140K mostram que as integrações ocorreram em 18 dos 23 cromossomos, ocorrendo uma preferência por cromossomos específicos. Nessa linhagem observou-se que 19% das integrações ocorreram no cromossomo 19, seguido de 16% no 17 e 9,6% no cromossomo 11. Na Hek293FVIIIIdB/P140K as inserções afetaram 20 dos 23 cromossomos, e as altas frequências de integrações ocorreram nos cromossomos 9 (14%) e X (9%) (Figura 09).

Figura 09: Número de integrações cromossômicas observadas do vetor pMFG-FVIII-P140K nas linhagens celulares HepG2FVIII-B/P140K e Hek293FVIII-B/P140K.

A análise estatística do número de integrações observadas *versus* o número de integrações esperadas é dada pelo RISC. Utilizando um *cut-off* ≥ 3 , observamos que o vetor retroviral pMFG-FVIII-P140K apresenta um perfil de integração não randômico e diferente entre as células estudadas, uma vez que na linhagem HepG2 houve uma preferência pelos cromossomos 19 (Risc = 94,28), 17 (Risc = 49,28) e 11 (Risc = 3,44), e na Hek293 pelo cromossomo 9 (Risc = 10,67) (Figura 10).

Figura 10: RISC Score. Número de integrações cromossômicas observadas *versus* número de integrações esperadas do vetor pMFG-FVIII-P140K nas linhagens celulares HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K (cut-off ≥ 3).

3.4 Frequência de integração do vetor pMFG-FVIII-P140K perto de Sítios de Ligação de Fatores de Transcrição (TFBS)

Os vetores retrovirais derivados de MLV se integram preferencialmente a uma distância de ± 1 Kb de regiões genômicas flanqueadas por conjuntos específicos de TFBSs, independente da sua localização com o respectivo gene ou sítio de início de transcrição (FELICE *et. al.*, 2009).

A frequência de integração do vetor pMFG-FVIII-P140K em regiões próximas ≤ 5 Kb de TFBS foi de 3,6% na linhagem celular HepG2FVIIIIdB/P140K e 1,7% na linhagem Hek293FVIIIIdB/P140K.

Nesse estudo observamos que dentre as integrações ocorridas nas células HepG2FVIIIIdB/P140K, um total de 140 sítios de ligação de fatores de transcrição (TFBS) foram afetados. Dentre essas, 5 integrações ocorreram em locais próximos ± 5 Kb, 9 integrações a ± 10 Kb e 39 a uma distância de ± 30 Kb (Figura 11).

Na linhagem renal Hek293FVIIIIdB/P140K, verificamos que um total de 117 TFBS foram afetados, sendo que dentre esses 2 inserções ocorreram próximas a ± 5 Kb, apenas 1 integração a ± 10 Kb e 9 a uma distância de ± 30 Kb (Figura 11).

O perfil de integração em regiões TFBS não apresentou diferenças estatisticamente significativas entre as duas linhagens quando realizado o teste exato de Fisher ($p=0,606$), sugerindo que ambas possuem um perfil de integração semelhante.

Figura 11: Número de integrações do vetor retroviral pMFG-FVIII-P140K em regiões próximas a TFBS.

3.5 Frequência de integração do vetor pMFG-FVIII-P140K perto de ilhas CpGs

A frequência de integração do vetor pMFG-FVIII-P140K em regiões próximas ≤ 5 Kb de ilhas CpGs foi de 16% na linhagem celular HepG2FVIIIIdB/P140K e 7,4% na linhagem Hek293FVIIIIdB/P140K.

Considerando que a integração do vetor afeta as fitas “*plus*” e “*minus*” do DNA e que cada uma das fitas apresenta uma região *upstream* e outra *downstream*, das 73 integrações do vetor retroviral 106 ilhas CpGs foram afetadas na linhagem HepG2. Dentre essas, 17

ocorreram em locais próximos $\pm 5\text{Kb}$, 8 integrações a $\pm 10\text{Kb}$ e 25 a uma distância de $\pm 30\text{Kb}$ (Tabela IV).

Na linhagem renal Hek293FVIlIdB/P140K, das 64 integrações do vetor retroviral, 81 ilhas CpGs foram afetadas, sendo que 6 inserções ocorreram próximas a $\pm 5\text{Kb}$, apenas 1 integração a $\pm 10\text{Kb}$ e 7 a uma distancia de $\pm 30\text{Kb}$ (Tabela IV).

Nas duas linhagens estudadas não houve diferenças estatísticas no perfil de integração do vetor retroviral pMFG-FVIII-P140K próximo a ilhas CpGs (Teste de Fisher, $p=0,5462$), sugerindo que em relação as ilhas CpGs, as células possuem um padrão de integração semelhante.

Tabela IV – Perfil de integração do vetor pMFG-FVIII-P140K próximos a ilhas CpGs

Distancia das ilhas CpGs	HepG2/FVIlIdBP140K (%)	Hek293/FVIlIdBP140K (%)	p - valor
$\pm 5\text{Kb}$	22,66	22,22	$p=0,5462^{(NS)}$
$\pm 10\text{Kb}$	10,66	3,72	$p=0,5462^{(NS)}$
$\pm 30\text{Kb}$	33,33	25,92	$p=0,5462^{(NS)}$
$\pm 60\text{Kb}$	33,33	48,14	$p=0,5462^{(NS)}$

NS = não significativo

3.6 Frequência de integração em regiões gênicas do genoma

Dentre as 73 sequências analisadas da linhagem HepG2/FVIlIdBP140K que apresentaram homologia ao genoma humano, foi possível observar que 38 (52%) delas estão localizadas em regiões gênicas. Do total das integrações apenas 6% encontram-se em regiões de exons e 94% em regiões intrônicas (Tabela V).

Na linhagem celular Hek293/FVIlIdBP140K, foram observados que dentre os 64 clones, 28 (44%) tiveram seu local de integração em regiões gênicas, sendo que 10% estão em regiões de exons e 90% em regiões de introns (Tabela V).

Tabela V – Frequência de integração em regiões gênicas.

	HepG2/FVIIIIdBP140K	Hek293/FVIIIIdBP140K
Integração em genes	38 (52%)	28 (44%)
Exon	6%	10%
Intron	94%	90%

As frequências de inserção do vetor retroviral em genes das linhagens HepG2 (52%) foi superior aos resultados gerados pelo conjunto de 10.000 integrações randômicas simuladas pelo programa QuickMap, que gira em torno de 40%. Já a célula Hek293 apresentou uma frequência semelhante ao esperado randomicamente (44%).

A tabela VI mostra a relação de genes nos quais houve integração em ambas às linhagens celulares, o número e o local (exon ou intron) das inserções e suas respectivas funções celulares.

Tabela VI – Genes em que houve integração do vetor retroviral pMFG-FVIII-P140K

Genes	HepG2	Hek293	Função
AC007510.6	0	1 (intron)	
C1orf175	1 (intron)	0	Ligante
C20orf58	1 (intron)	0	
C6orf10	1 (intron)	0	
C8orf74	1 (intron)	0	
CCDC138	0	1 (intron)	
CCDC81	1 (intron)	0	
CCDC93	0	1 (intron)	
CHST3	0	1 (exon)	Atividade transferase
COL1A1	12 (8 intron e 4 exon)	0	Componente da matrix extracelular
DSCAML1	5 (intron)	0	Ligante de proteína
ENDOD1	0	1 (intron)	Atividade endonuclease
F5	0	1 (intron)	Atividade de Oxidoreductase
FOXP1	1 (intron)	0	Atividade de fator de transcrição
GPM6B	1 (intron)	0	Função molecular
GRID1	1 (intron)	0	Ligante de proteína
GRIK5	2 (intron)	0	Atividade de receptor
GRM8	1 (intron)	0	Atividade de receptor de glutamato
HERC3	0	1 (intron)	Atividade ligase
IL2RB	0	1 (exon)	Receptor de Interleucina-2
KALRN	0	1 (intron)	Ligante de nucleotideo
KIAA1128	0	1 (intron)	
LIN52	1 (intron)	0	Fator de transcrição

MGMT	1 (exon)	3 (exon)	Atividade de transferase
MSH2	1 (intron)	0	Atividade ATPase
NP_997182.1	0	2 (intron)	
PARD3B	0	1 (intron)	Ligante de protein
PPAPR3	0	8 (intron)	
PRB4	0	1 (intron)	Ligante de protein
PRDM2	1 (intron)	0	Atividade de fator de transcrição
RP1L1	1 (intron)	0	Ligante de protein
SPP2	1 (intron)	0	Atividade de inibidor de endopeptidase
ST3GAL4	0	1 (intron)	Beta-galactoside alpha-2,3-sialyltransferase
TANC2	0	1 (intron)	Ligante
ZNF710	2 (intron)	0	Ligante de ion metal

Quando analisados os genes em que ocorreu a integração do vetor pMFG-FVIII-P140K nas células HepG2, observamos que dentre as 38 integrações em regiões gênicas, o colágeno tipo 1 (COL1A1) foi o gene que teve o maior número de inserções, 32%, seguido de 13% no gene da molécula de adesão celular da síndrome de Down (DSCAML1). Outros genes importantes que tiveram integração do vetor nas células HepG2 foram o FOXP1 que é um fator de transcrição do sistema nervoso central e o MSH2, classificado como gene de reparo do DNA. Esses genes são comumente afetados quando há um desenvolvimento de um tumor e por isso são considerados como genes relacionados ao câncer por algumas bases de dados.

As células Hek293 tiveram um total de 28 integrações em genes sendo que 29% destas ocorreram no gene PPAPR3, que faz parte da família das fosfatases lipídicas (Tabela VI).

Comparando as duas linhagens celulares, podemos observar que houve integração em apenas um gene em comum, o MGMT, sendo que esta ocorreu 1 vez na HepG2 e 3 vezes na Hek293. Este gene codifica uma O⁶-metilguanina-DNA-metiltransferase, cuja função é reparar lesões causadas no DNA por agentes metilantes.

Uma classificação funcional dos genes afetados em ambas as células foi realizada utilizando o *website* DAVID/EASE, que utiliza critérios do *Gene Ontology* (GO), além de avaliar classes e vias metabólicas, bem como doenças que possam estar relacionadas. Após a análise dos dados podemos observar que dentre os genes que tiveram integração

do vetor retroviral, alguns estão relacionados a processos biológicos, como diferenciação celular, desenvolvimento celular, diferenciação neuronal, entre outros, bem como genes relacionados a funções moleculares, tais como receptores de glutamato, como mostra a figura 12.

Figura 12: Classificação do Gene Ontology mostrando categorias afetadas em funções moleculares e processos biológicos.

Além da classificação funcional, foi possível fazer uma análise quanto à participação de alguns genes em vias metabólicas. Assim, dentre os genes afetados na linhagem HepG2, 3 deles (GRID1, GRIK5 e GRM8) estão relacionados com a via de interação receptor-ligante neuroativa, de acordo com KEGG-Pathway. Nesta via esses genes atuam como receptores de glutamato, contudo a integração do vetor retroviral não afetou as funções metabólicas, uma vez que este se inseriu dentro de regiões intônicas dos respectivos genes. Na célula Hek293 não houve integrações em genes que estivessem relacionados a alguma via específica descrita na literatura.

3.7 Frequência de integração em sítios frágeis do genoma

Além das integrações em regiões gênicas, foi possível observar que houve inserções em sítios frágeis do genoma humano, sendo que a linhagem HepG2/FVIIIIdBP140K apresentou 20% das integrações nessas regiões, enquanto que na Hek293/FVIIIIdBP140K foram da ordem de 17%. As regiões de sítio frágil apresentam uma instabilidade cromossômica e consistem em locais nos cromossomos metafásicos suscetíveis a quebras (DURKIN e GLOVER, 2007).

Podem ser classificados em dois grupos, comuns e raros, de acordo com suas características e frequência em que são encontrados nos cromossomos. O vetor retroviral pMFG-FVIII-P140K apresentou integração em ambos os grupos. Conforme mostrado na figura 13, a frequência de integração e os respectivos sítios frágeis afetados diferem entre as duas linhagens sendo que na HepG2 houve uma maior frequência de integração nos sítios FRA11G (5,35%) e FRA11B (5,35%), enquanto que na Hek293 os sítios mais afetados foram FRA2B (3,77%) e FRA1H (3,77%). Quando comparadas com um conjunto de 10.000 integrações randômicas simuladas pelo programa *QuickMap*, houve uma diferença na porcentagem de integração, uma vez que se a integração fosse aleatória as frequências esperadas seriam de 0,18% para esses sítios frágeis na HepG2 e 0,44% para Hek293, sugerindo que há uma preferência de integração do vetor pMFG-FVIII-P140K nesses sítios frágeis.

Figura 13: Frequência de integração do vetor retroviral pMFG-FVIII-P140K em sítios frágeis do genoma das linhagens HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K.

3.8 Frequência de integração do vetor pMFG-FVIII-P140K em regiões repetitivas do genoma

Ao analisarmos as integrações em regiões de DNA repetitivo, podemos observar que os dados sugerem um padrão de integração semelhante entre as duas linhagens (Figura 14), contudo não foi possível a realização de uma análise estatística devido ao baixo número de integrações nessas regiões.

Na linhagem celular HepG2/FVIIIIdBP140K, 18 integrações ocorreram em regiões repetidas do genoma (SINE, LINE, LTRs, etc.). Dentre estas 33,3% se localizaram em regiões LINE/L1, seguido de 22,2% em regiões LTR/ERV1 e 16,6% em SINE/MIR.

Nas células Hek293/FVIIIIdBP140K, 21 integrações ocorreram em regiões repetidas do genoma (SINE, LINE, LTRs, etc.), sendo que dessas, 57% estão localizadas nas regiões

LINE/L1, assim como na HepG2, seguido de 23,8% em LTR/MaLR e 9,5% em regiões LTR/ERV1.

Figura 14: Frequência de integração do vetor retroviral pMFG-FVIII-P140K em regiões repetidas do genoma, nas linhagens celulares HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K.

Com o intuito de uma melhor visualização dos dados gerados, foi realizado um resumo com as principais características analisadas neste trabalho, como mostra a tabela VII.

Tabela VII – Resumo dos dados gerados

Características Genômicas	HepG2FVIIIIdB/P140K (%)	Hek293FVIIIIdB/P140K (%)	Valores Randômicos (%)
Unidades			
Transcricionais	52	44	40
Exons	3,12	4,4	2,22
Introns	48,8	39,6	38,10
SINES			13,7
Alu	0	4,7	
MIR	16,6	0	
LINES	33,3	57	21,35
≤ 5Kb de ilhas CpGs	16	7,4	7,32
≤ 5Kb de TFBS	3,6	1,7	
Sítios Frágeis	20	17	29,45

4 Discussão

O presente estudo teve como principal objetivo a caracterização do padrão de integração do vetor retroviral pMFG-FVIII-P140K em duas linhagens celulares produtoras do FVIII recombinante, derivadas de fígado (HepG2) e de rim (Hek293).

Estudos realizados anteriormente no laboratório de Transferência Gênica do Hemocentro de Ribeirão Preto permitiram a geração das duas linhagens celulares produtoras de FVIII recombinante utilizando sistema de transferência gênica retroviral. O vetor pMFG-FVIII-P140K contém o gene do FVIII de coagulação humana com domínio B deletado e o gene marcador de seleção MGMT-P140K, uma forma mutante do gene que codifica a enzima DNA alquiltransferase e que confere resistência à droga quimioterapêutica O⁶ Benzilguanina (FONTES *et al.*, 2006).

As linhagens utilizadas nesse estudo foram cultivadas em condições artificiais de cultura *in vitro*, e selecionadas por meio do tratamento de alta estrigência com as drogas quimioterapêuticas O⁶ Benzilguanina e Temozolomide. Sendo assim trabalhamos com populações advindas de seleção e o perfil de integração do retrovírus descrito é específico para estas populações.

Os resultados mostraram que ambas as células, HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K, expressam o RNA mensageiro referente ao FVIII recombinante, indicando a estabilidade da integração e transcrição no genoma das células transduzidas. Quando analisamos a atividade biológica da proteína observamos que a linhagem hepática HepG2FVIIIIdB/P140K apresentou níveis de secreção da molécula biologicamente ativa da ordem de 7,9 UI/mL enquanto que a linhagem renal Hek293FVIIIIdB/P140K foi de 2,1 UI/mL. O fator VIII é uma glicoproteína grande e bastante instável. No plasma, sua concentração é baixa (100% de atividade = 150 ng). Deste modo, a linhagem celular HepG2 mostrou níveis

de fator VIII 7 vezes superior aos nível de fator VIII detectado no plasma normal, enquanto que na Hek293 os níveis foram 2 vezes maiores.

Neste estudo, trabalhamos com duas linhagens celulares modificadas geneticamente com o mesmo vetor retroviral, tratadas com a mesma concentração das drogas quimioterapêuticas, e que expressam diferentes níveis de FVIII, o que permitiu pensarmos em duas hipóteses: 1) o perfil de integração do vetor seria diferente entre ambas as células, visto que além dos fatores virais, sabe-se, atualmente, que fatores celulares também estão envolvidos no processo de integração do retrovírus (LEWINSKI *et al.*, 2005; LEWINSKI *et al.*, 2006; STUDAMIRE e GOFF, 2008; BRUCE *et al.*, 2010), e 2) o padrão de inserção do retrovírus seria semelhante em ambas as células, uma vez que as duas linhagens foram selecionadas por meio do mesmo tratamento de alta estrigência, com as drogas O⁶ Benzilguanina e Temozolomide.

Quando analisamos a integração do vetor retroviral pMFG-FVIII-P140K observamos que esta segue um padrão não randômico em ambas as células, visto que houve uma preferência de inserção em cromossomos específicos, e que esta não foi proporcional ao tamanho dos cromossomos. A distribuição dos sítios de integração pode ser influenciada por várias características das populações celulares, dentre eles o fato de que estas células foram selecionadas e possuem um cariótipo de difícil definição, com cromossomos alta e baixamente representados (LEWINSKI *et al.*, 2006).

Observamos que houve uma inserção em cromossomos diferentes entre as linhagens celulares, sendo que na Hek293FVIIIIdB/P140K a maioria das integrações ocorreram no cromossomo 9 e na HepG2FVIIIIdB/P140K ocorreram no 19, 17 e 11. É interessante notar que os genes que tiveram maior número de integrações estão localizados nos cromossomos que apresentaram maior número de inserções, sendo na Hek293, o gene PPAPR3 no cromossomo 9 e na HepG2 os genes COL1A1 e DSCAML1 nos cromossomos 17 e 11, respectivamente.

Em 2003, Laufs e colaboradores caracterizaram o perfil de integração retroviral *in vivo* em camundongos NOD/SCID utilizando células CD34+ selecionadas por *sorting* e

transduzidas com um vetor que contem o gene de resistência a multidrogas 1 (*MDR1*) e observou uma alta frequência de integração nos cromossomos 17 e 19 (LAUFS *et al.*, 2003). Em 2007, um estudo também utilizando células CD34+ de amostras de sangue periférico de pacientes (SCID-X1) tratados com terapia gênica utilizando retrovírus também mostraram uma frequência elevada de integrações nos cromossomos 17 e 19 (DEICHMANN *et al.*, 2007). Ambos os estudos não utilizam estratégias de seleção de um clone celular, diferindo do presente estudo, entretanto os resultados corroboram com as observações descritas aqui para integração na linhagem HepG2, e sugerem que este perfil de integração pode estar relacionado com o elevado número de genes presente nesses cromossomos e não com o tamanho de cada cromossomo, como seria esperado se a integração fosse randômica (LAUFS *et al.*, 2003; DEICHMANN *et al.*, 2007).

As diferenças nas frequências de integração ao longo dos cromossomos estão relacionadas, em parte, em função da densidade gênica (MITCHELL *et al.*, 2004). A posição intracelular dos cromossomos também pode influenciar, uma vez que tem sido proposto que esta é relativamente fixa para cada tipo celular, e difere entre os tipos celulares (BOYLE *et al.* 2001, CHUBB e BICKMORE 2003; MITCHELL *et al.*, 2004).

Com as análises das inserções do vetor retroviral dentro de genes pudemos observar que na linhagem HepG2, 52% das integrações ocorreram em genes codificadores de proteínas, enquanto que na Hek293 tivemos 44%. Esses valores foram acima do esperado randomicamente, 40% (APPLET *et al.*, 2009), e acima dos 34% observados por WU *et al.*, 2003. Todavia, os resultados demonstrados aqui corroboram com os observados por outros autores, que descrevem que para os retrovírus derivados de MLV a integração dentro de genes é de 49,1% em células mononucleares e granulócitos maduros de macacos (HEMATTI *et al.*, 2004), 55% em células CD34+ e 48,1% em células HeLa (CATTOGLIO *et al.*, 2007). Vale à pena ressaltar que apesar da elevada frequência de integração dentro de genes codificadores de proteínas, mais de 90% delas ocorreram em regiões intrônicas em ambas as linhagens celulares, o que sugere que essas inserções não estariam influenciando na codificação das respectivas proteínas.

Quanto a inserção em regiões do genoma como ilhas CpGs e sítios de ligação de fatores de transcrição (TFBS), as linhagens celulares HepG2 e Hek293 mostraram um padrão de integração semelhante entre elas, indicando novamente que o perfil de inserção do vetor pMFG-FVIII-P140K não é aleatório nesses clones celulares.

Vários estudos têm demonstrado que vetores retrovirais derivados de MLV integram preferencialmente ao redor de sítios de início de transcrição (TSS) e ilhas CpGs (WU *et al.*, 2003; MITCHELL *et al.*, 2004; LAUFS *et al.*, 2004; AMBROSI *et al.*, 2008). As ilhas CpGs são regiões cromossômicas enriquecidas em dinucleotídeos CG, que geralmente correspondem a regiões gênicas regulatórias e consequentemente, são mais frequentes em regiões ricas em genes (MITCHELL *et al.*, 2004; LEWINSKI *et al.*, 2006; DERSE *et al.*, 2007). Em 2003, Wu *et al.*, mostraram que 16,8% das integrações de MLV ocorreram a uma distância de $\pm 1\text{Kb}$ de ilhas CpGs, uma integração oito vezes maior que a esperada randomicamente (2,1%). Neste estudo observamos um número elevado de integrações adjacentes a ilhas CpGs em ambas as linhagens celulares, contudo em regiões não tão próximas quanto as descritas na literatura ($\pm 1\text{Kb}$), uma vez que tanto na HepG2 quanto na Hek293 a maior frequência de integração foi a uma distância de $\pm 30\text{Kb}$ e $\pm 60\text{Kb}$ de ilhas CpGs, respectivamente. Isso indica que dentro das características específicas de cultivo e seleção as quais o vetor estudado foi submetido, este apresentou um perfil de integração diferente do descrito na literatura, sugerindo a sua baixa preferência por essas regiões do genoma.

Em relação aos sítios de início de transcrição, não observamos integrações em locais próximos a TSS em nenhuma das duas células, o que difere dos dados encontrados na literatura, que apontam esses sítios como locais preferenciais de inserção de retrovírus MLV e seus vetores derivados (WU *et al.*, 2003; MITCHELL *et al.*, 2004). Contudo, tivemos um grande número de inserções a uma distância de $\pm 30\text{Kb}$ e $\pm 60\text{Kb}$ de sítios de ligação de fatores de transcrição (TFBS). LEWINSKI *et al.*, 2006, e mais recentemente FELICE *et al.*, 2009, utilizando células CD34+, mostraram que as regiões próximas $\pm 1\text{Kb}$ de TFBS também são sítios de integração preferencial de MLV (LEWINSKI *et al.*, 2006; FELICE *et al.*,

2009) e que tanto a integrase quanto a região 3' LTR do retrovírus são os maiores responsáveis virais pela seleção de sítios ricos em TFBS no genoma da célula hospedeira. Além disso, os autores sugerem que a seleção de regiões ricas em TFBS pode, dar suporte a todas as integrações preferenciais conhecidas dos retrovírus, particularmente por TSS, ilhas CpGs e sítios hipersensíveis a DNase-I (WU *et al.*, 2003; MITCHELL *et al.*, 2004; LEWINSKI *et al.*, 2006), nas quais as regiões regulatórias ricas em TFBS estão altamente representadas (FELICE *et al.*, 2009, MITCHELL *et al.*, 2004). Seguindo essa linha de raciocínio, observamos que nossos dados corroboram com essa sugestão, uma vez que as integrações tanto em ilhas CpGs quanto em TFBS ocorreram preferencialmente a uma distância de $\pm 30\text{Kb}$, sugerindo uma associação entre essas regiões regulatórias do genoma.

É interessante observar que quando analisamos as distância de integração do nosso vetor retroviral, em regiões próximas $\pm 1\text{Kb}$ das ilhas CpGs e dos sítios de ligação de fatores de transcrição (TFBS), os resultados aqui descritos não corroboram com os trabalhos existentes na literatura (WU *et al.*, 2003; MITCHELL *et al.*, 2004; LEWINSKI *et al.*, 2006; FELICE *et al.*, 2009) visto que a maioria das nossas integrações ocorreram a uma distância entre $\pm 30\text{-}60\text{Kb}$. Este achado pode estar relacionado ao fato de que na maioria destes estudos, os autores utilizaram células que foram infectadas e analisadas após um período de no máximo 48h e não utilizaram nenhum tipo de estratégia de seleção, diferentemente do presente estudo, no qual as células permaneceram em cultura após a transdução do vetor retroviral e passaram por um tratamento com alta estringência, para a seleção de um clone celular produtor de fator VIII.

Um estudo realizado por Grund *et al.*, 2010, utilizando células CD34+, comparou o perfil de integração do vetor lentiviral SIN, que contem o gene de seleção P140K, antes e após o tratamento das células com as drogas quimioterapêuticas O⁶ benzilguanina e BCNU e, mostrou que não houve diferença no padrão de inserção do vetor entre as células tratadas e não tratadas, sugerindo que a estratégia de seleção não estaria influenciando o perfil de integração viral (GRUND *et al.*, 2010). Contudo, a concentração da droga utilizada

por Grund et al. (12uM), para o tratamento das células foi 33 vezes mais baixa do que a concentração utilizada nas linhagens celulares do presente trabalho (400uM), o que pode explicar o fato de que o autor não observou diferença de integração entre as células tratadas e não tratadas, enquanto que os dados descritos aqui, sugerem que o tratamento com as drogas pode ter uma influencia na seleção de um perfil de integração específico.

Um estudo realizado por Russo-Carbolante *et al.*, 2010, utilizou as linhagens celulares Sk-Hep e 293T, e o GFP como sistema de seleção. Neste trabalho os autores observaram que o vetor lentiviral apresentou um perfil de integração diferente entre os cromossomos nas duas linhagens. Contudo, quando analisada a inserção próxima a regiões específicas do genoma, como ilhas CpGs e TSS, ambas as células mostraram um padrão de integração semelhante, com mais de 50% das integrações ocorrendo a uma distância de $\pm 30\text{Kb}$ (RUSSO-CARBOLANTE *et al.*, 2010), da mesma forma que observado no presente trabalho. Uma vez que foram utilizadas células diferentes e vetores virais diferentes e o padrão de integração mostrou-se semelhante entre as células, sugere-se novamente que é possível que as estratégias de seleção utilizadas possam ter uma influência na seleção de clones celulares com sítios de integração semelhantes.

Observou-se também a integração do vetor pMFG-FVIII-P140K em regiões de sítios frágeis do genoma humano. Os sítios frágeis têm sido associados com instabilidade genômica em células de câncer e com a paralisação da resposta ao dano no DNA durante a replicação. Embora não estejam envolvidos nas translocações mais comuns presentes em alguns cânceres e leucemias, numerosos estudos têm mostrado que os sítios frágeis são sítios recorrentes de deleções e outros rearranjos cromossômicos em células tumorais (DRURKIN e GLOVER, 2007). Este fato pode explicar a frequência de integração do vetor pMFG-FVIII-P140K nessas regiões do genoma (20% de integração na HepG2 e 17% na Hek293), visto que a HepG2 é derivada de um hepatocarcinoma e a Hek293 possui potencial tumorigênico *in vivo* (dados não mostrados). Além disso, Bester *et al.* (2006) em um estudo utilizando células CD3+ e HeLa, relataram que os sítios frágeis são regiões de

integração preferencial de vetores derivados de MLV. A relação custo-benefício de utilizar um retrovírus que possui preferência de integração por sítios frágeis do genoma, precisa ser bem avaliada, principalmente quando existe a intenção de utilizá-lo em ensaios clínicos de terapia gênica, pois existe o risco de ativação de genes relacionados ao desenvolvimento de cânceres nesses locais do genoma. No presente estudo o vetor retroviral foi utilizado para a transferência do gene do FVIII em duas linhagens celulares para uma possível produção da proteína recombinante, e neste caso, acreditamos que a integração em sítios frágeis não irá influenciar no produto final.

Em suma o presente trabalho mostrou o perfil de integração do vetor retroviral pMFG-FVIII-P140K, específico para as linhagens HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K. A inserção em cromossomos específicos mostrou-se diferente nas duas linhagens celulares, podendo ser devido à fatores celulares específicos de cada tipo de célula que possuem origens de diferentes tecidos. Além disso, pudemos observar que em ambas as linhagens celulares, HepG2FVIIIIdB/P140K e Hek293FVIIIIdB/P140K, existem regiões gênicas específicas dentro dos cromossomos na qual o vetor exibe uma preferência. Neste estudo, bem como na literatura, a preferência de integração ocorreu por regiões transcricionais, tais como genes (introns e exons), regiões regulatórias como ilhas CpGs, sítios de ligação de fatores de transcrição e sítios frágeis. O fato da linhagem celular HepG2FVIIIIdB/P140K apresentar um maior número de integrações em regiões ativas do genoma, tais como unidades transcricionais, ilhas CpGs e SINES, pode estar relacionado ao maior nível de expressão de FVIII recombinante nessa linhagem associadas as características do tecido de origem, o fígado, produtor de fator VIII no organismo.

Além disso, vale à pena ressaltar que o tratamento com alta estringência com as drogas quimioterapêuticas pode ter influenciado a seleção de clones específicos, o que poderia explicar o padrão do perfil de integração semelhante entre as duas linhagens, concordando com a nossa hipótese de que o padrão de inserção do retrovírus seria semelhante em ambas as células, uma vez que as duas linhagens foram selecionadas por

meio do mesmo tratamento de alta estringência, com as drogas O⁶ Benzilguanina e Temozolomide.

Em conjunto os resultados proporcionaram uma melhor compreensão do comportamento do retrovírus utilizado juntamente com o tratamento com drogas quimioterapêuticas usadas para a seleção. Estes dados permitirão a seleção posterior de um clone celular capaz de produzir altos níveis FVIII recombinante de maneira mais estável e duradoura.

5 Conclusão

Neste trabalho foi realizada a caracterização do perfil de integração do vetor retroviral pMFG-FVIII-P140K para as linhagens celulares HepG2FVIIIIdB/P140K, de origem hepática e Hek293FVIIIIdB/P140K, de origem renal;

Ambas as linhagens celulares expressaram estavelmente o mRNA do fator VIII e apresentaram uma atividade biológica da proteína da ordem de 7,9 UI/mL para a linhagem HepG2FVIIIIdB/P140K e 2,1 UI/mL para a Hek293FVIIIIdB/P140K;

Foram sequenciados um total de 422 clones sendo 201 das células HepG2 e 221 da Hek293. Dentre as sequências da HepG2, 123 apresentaram similaridade ao genoma humano, sendo 73 consideradas verdadeiras e 50 ambíguas quando analisadas pelo programa *QuickMap*. Na célula Hek293 179 apresentaram similaridade ao genoma humano, sendo 64 consideradas verdadeiras e 115 ambíguas;

A integração em cromossomos mostrou-se não randômica com uma preferência de inserção nos cromossomos 19, 17 e 11 na linhagem HepG2FVIIIIdB/P140K e pelo cromossomo 9 na linhagem Hek293FVIIIIdB/P140K;

Não houve diferença estatística na integração do vetor retroviral pMFG-FVIII-P140K em regiões do genoma como ilhas CpGs e sítios de ligação de fatores de transcrição, mostrando que este vetor tem preferência de integração por sítios distantes a essas regiões;

Houve uma elevada frequência de integração dentro de genes codificadores de proteínas, da ordem de 52% para a HepG2 e 44% para a Hek293, além de 20% e 17% das integrações em sítios frágeis nas células HepG2 e Hek293, respectivamente;

O padrão de inserção descrito pode estar relacionado ao fato de que essas células foram tratadas anteriormente com drogas quimioterapêuticas para seleção de um clone celular com maior nível de produção de rFVIII, levando conseqüentemente a seleção de um perfil de integração semelhante entre as linhagens celulares, mesmo estas sendo origens teciduais diferentes.

Os resultados proporcionaram uma melhor compreensão do comportamento do retrovírus utilizado e permitirão futuramente a seleção de um clone celular capaz de produzir elevados níveis FVIII recombinante de maneira mais estável e duradoura.

6 Referências Bibliográficas

- Aledort LM. 2007. History of haemophilia. *Haemophilia*. 13 Suppl 5:1–2.
- Ambrosi A, Cattoglio C, Di Serio C. 2008. Retroviral integration process in the human genome: is it really non-random? A new statistical approach. *PLoS Comput Biol*. Aug 8;4(8):e1000144
- Anderson JL, Hope TJ. 2005. Intracellular trafficking of retroviral vectors: obstacles and advances. *Gene Ther*. 12(23):1667-78.
- Antonarakis SE. 1998. Molecular genetics of coagulation factor VIII gene and haemophilia A. *Haemophilia*. 4 Suppl 2:1-11. Review.
- Appelt JU, Giordano FA, Ecker M, Roeder I, Grund N, Hotz-Wagenblatt A, Opelz G, Zeller WJ, Allgayer H, Fruehauf S, Laufs S. 2009. QuickMap: a public tool for large-scale gene therapy vector insertion site mapping and analysis. *Gene Ther*. Jul;16(7):885-93.
- Bester AC, Schwartz M, Schmidt M, Garrigue A, Hacein-Bey-Abina S, Cavazzana-Calvo M, Ben-Porat N, Von Kalle C, Fischer A, Kerem B. 2006. Fragile sites are preferential targets for integrations of MLV vectors in gene therapy. *Gene Ther*. 13(13):1057-9.
- Blakley RL, Sorrentino BP. 1998. In vitro mutations in dihydrofolate reductase that confer resistance to methotrexate: potential for clinical application. *Hum Mutat*. 11:259–63.
- Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Mol Genet*. 10(3):211-9.
- Bruce JW, Hierl M, Young JA, Ahlquist P. 2010. Cellular transcription factor ZASC1 regulates murine leukemia virus transcription. *J Virol*. 84(15):7473-83.
- Bushman, F.D. 2007. Retroviral integration and human gene therapy. *Clin. Invest*. 117, 2083–2086.
- Cattoglio C, Facchini G, Sartori D, Antonelli A, Miccio A, Cassani B, Schmidt M, von Kalle C, Howe S, Thrasher AJ, Aiuti A, Ferrari G, Recchia A, Mavilio F. 2007. Hot spots of retroviral integration in human CD34+ hematopoietic cells. *Blood*. 110(6):1770-8.
- Chen C, Fang XD, Zhu J, Wu XF, Zhang ZC, Gu JX, Wang ZY, Chi CW. 1999. The gene expression of coagulation factor VIII in mammalian cell lines. *Thromb Res*. 95(2):105-15.
- Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. 2003. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol. Chem*. 278:372–381.

- Cherepanov, P., Devroe, E., Silver, P.A., and Engelman, A. 2004. Identification of an evolutionarily conserved domain in human lens epithelium-derived growth factor/transcriptional co-activator p75 (LEDGF/p75) that binds HIV-1 integrase. *J. Biol. Chem.* 279: 48883–48892.
- Chng WJ, Sum C, Kuperan P. 2005. Causes of isolated prolonged activated partial thromboplastin time in an acute care general hospital. *Singapore Med J.* 46(9):450-6.
- Chubb JR, Bickmore WA. 2003. Considering nuclear compartmentalization in the light of nuclear dynamics. *Cell.* Feb 21;112(4):403-6.
- Daniel R, Smith JA. 2008. Integration site selection by retroviral vectors: molecular mechanism and clinical consequences. *Hum Gene Ther.* 19(6):557-68. Review.
- de Sousa Russo-Carbolante EM, Picanço-Castro V, Alves DC, Fernandes AC, Almeida-Porada G, Tonn T, Covas DT. 2010. Integration pattern of HIV-1 based lentiviral vector carrying recombinant coagulation factor VIII in Sk-Hep and 293T cells. *Biotechnol Lett.*
- Davis BM, Koc ON, Reese JS, Gerson SL. 1999. O6-benzylguanine-resistant mutant MGMT genes improve hematopoietic cell tolerance to alkylating agents. *Prog Exp Tumor Res.* 36:65-81.
- Deichmann, A., Hacein-Bey-Abina, S., Schmidt, M., Garrigue, A., Brugman, M.H., Hu, J., Glimm, H., Gyapay, G., Prum, B., Fraser, C.C., Fischer, N., Schwarzwaelder, K., Siegler, M.L., De Ridder, D., Pike-Overzet, K., Howe, S.J., Thrasher, A.J., Wagemaker, G., Abel, U., Staal, F.J., Delabesse, E., Villeval, J.L., Aronow, B., Hue, C., Prinz, C., Wissler, M., Klanke, C., Weissenbach, J., Alexander, I., Fischer, A., Von Kalle, C., and Cavazzana-Calvo, M. 2007. Vector integration is nonrandom and clustered and influences the fate of lymphopoiesis in SCID-X1 gene therapy. *J. Clin. Invest.* 117: 2225–2232.
- Durkin SG, Glover TW. 2007. Chromosome fragile sites. *Annu Rev Genet.* 41:169-92. Review.
- Felice B, Cattoglio C, Cittaro D, Testa A, Miccio A, Ferrari G, Luzi L, Recchia A, Mavilio F. 2009. Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. *PLoS One.* 4(2):e4571.
- Flint, S.E., Racaniello, V.R., and Skalka, A.M. 2004. *Principles of Virology.* (ASM Press, Washington, D.C.).
- Fontes AM, Davis B, Encell L, Lingas K, Covas DT, Zago MA, Loeb L, Pegg A, Gerson S. 2006. Differential competitive resistance to methylating versus chloroethylating agents among five O6 –alkylguanine DNA alkyltransferase in human hematopoietic cells. *Mol Cancer Ther.* 5(1) : 1-8.
- Grillari J, Fortschegger K, Grabherr R, Hohenwarter O, Kunert R, Katinger H. 2001. Analysis of alteration in gene expression after amplification of recombinant genes in CHO cells. *J. Biotech* 87: 59-65.
- Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, Sorensen R, Forster A, Fraser P, Cohen JI, de Saint Basile G, Alexander I, Wintergerst U, Frebourg T, Aurias A, Stoppa-Lyonnet D, Romana S, Radford-Weiss I, Gross F, Valensi F, Delabesse E, Macintyre E, Sigaux F, Soulier J, Leiva LE, Wissler M, Prinz C,

- Rabbits TH, Le Deist F, Fischer A, Cavazzana-Calvo M. 2003. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science*. 302(5644):415-9.
- Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, Clappier E, Caccavelli L, Delabesse E, Beldjord K, Asnafi V, MacIntyre E, Dal Cortivo L, Radford I, Brousse N, Sigaux F, Moshous D, Hauer J, Borkhardt A, Belohradsky BH, Wintergerst U, Velez MC, Leiva L, Sorensen R, Wulffraat N, Blanche S, Bushman FD, Fischer A, Cavazzana-Calvo M. 2008. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest*. Sep;118(9):3132-42.
- Hansson LO, Widersten M, Mannervik B. 1999. An approach to optimizing the active site in a glutathione transferase by evolution in vitro. *Biochem J*. 344 Pt 1:93-100.
- Hay CR. 2006. The epidemiology of factor VIII inhibitors. *Haemophilia*. 12 Suppl 6:23-8.
- Hematti P, Hong BK, Ferguson C, Adler R, Hanawa H, Sellers S, Holt IE, Eckfeldt CE, Sharma Y, Schmidt M, von Kalle C, Persons DA, Billings EM, Verfaillie CM, Nienhuis AW, Wolfsberg TG, Dunbar CE, Calmels B. 2004. Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells. *PLoS Biol*. Dec;2(12):e423.
- Hermans C, Claeys D. 2006. Review of the rebound phenomenon in new anticoagulant treatments. *Curr Med Res Opin*. 22(3):471-81.
- Hoyer LW. 1994. Hemophilia A. *N. Engl. J. Med*. 330: 38-47.
- Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 4(1):44-57.
- Illingworth R, Kerr A, Desousa D, Jørgensen H, Ellis P, Stalker J, Jackson D, Clee C, Plumb R, Rogers J, Humphray S, Cox T, Langford C, Bird A. 2008. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol*. 6, e22.
- Illingworth RS, Bird AP. 2009. CpG islands--'a rough guide'. *FEBS Lett*. 583(11):1713-20. Review.
- Jenny RJ, Pittman DD, Toole JJ, Kriz RW, Aldape RA, Hewick RM, Kaufman RJ, Mann KG. 1987. Complete cDNA and derived amino acid sequence of human factor V. *Proc Natl Acad Sci U S A*. 84(14):4846-50.
- Kaufman RJ. 1991. Insight into the structure, function, and biosynthesis of factor VIII through recombinant DNA technology. *Ann Hematol*. 63(3):155-65. Review.
- Kim S, Kim Y, Liang T, Sinsheimer JS, Chow SA. 2006. A high-throughput method for cloning and sequencing human immunodeficiency virus type 1 integration sites. *J Virol*. 80(22):11313-21.
- Kramer BA, Lemckert FA, Alexander IE, Gunning PW, McCowage GB. 2006. Characterisation of a P140K mutant O6-methylguanine-DNA-methyltransferase (MGMT)-expressing transgenic mouse line with drug-selectable bone marrow. *J Gene Med*. 2006 8(9):1071-85.

- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W et al. 2001. Initial sequencing and analysis of the human genome. *Nature*. 409(6822):860-921.
- Landis DM, Heindel CC, Loeb LA. 2001. Creation and characterization of 5-fluorodeoxyuridine-resistant Arg50 loop mutants of human thymidylate synthase. *Cancer Res*. 61:666–72.
- Larsen, F., Gundersen, G., Lopez, R. and Prydz, H. 1992. CpG islands as gene markers in the human genome. *Genomics*. 13:1095–1107.
- Laufs S, Gentner B, Nagy KZ, Jauch A, Benner A, Naundorf S, Kuehlcke K, Schiedlmeier B, Ho AD, Zeller WJ, Fruehauf S. 2003. Retroviral vector integration occurs in preferred genomic targets of human bone marrow-repopulating cells. *Blood*. 101(6):2191-8.
- Laufs S, Nagy KZ, Giordano FA, Hotz-Wagenblatt A, Zeller WJ, Fruehauf S. 2004. Insertion of retroviral vectors in NOD/SCID repopulating human peripheral blood progenitor cells occurs preferentially in the vicinity of transcription start regions and in introns. *Mol Ther*. Nov;10(5):874-81.
- Lee CA. 1999. Recombinant clotting factors in the treatment of hemophilia. *Tromb Haemost* 82: 516-524
- Lewinski MK, Bushman FD. 2005. Retroviral DNA integration--mechanism and consequences. *Adv Genet*. 55:147-81.
- Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, Crawford G, Collins F, Shinn P, Leipzig J, Hannenhalli S, Berry CC, Ecker JR, Bushman FD. 2006. Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog*. 2(6):e60.
- Lewis PF, Emerman M. 1994. Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. *J Virol*. 68: 510–516
- Lewis P, Hensel M, Emerman M. 1992. Human immunodeficiency virus infection of cells arrested in the cell cycle. *EMBO J*. 11: 3053–3058.
- Mannucci PM, Giangrande PL. 2000. Choice of replacement therapy for hemophilia: recombinant products only? *Hematol J*. 1(2):72-6. Review.
- Mannucci PM. 2008. Back to the future: a recent history of haemophilia treatment. *Haemophilia*. 14 Suppl 3:10–18.
- Mei B, Chen Y, Chen J, Pan CQ, Murphy JE. 2006. Expression of human coagulation factor VIII in a human hybrid cell line, HKB11. *Mol Biotechnol*. 34(2):165-78.
- Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR, Bushman FD. 2004. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol*. 2(8):E234.
- Oh SH, Kim SH, Kim HW, Kim YJ. 1999. An efficient retrovirus-mediated transduction of human blood coagulation factor VIII cDNA in regenerating rat liver. *Ann. Hematol* 78:213-218.
- Peerlinck K, Hermans C. 2006. Epidemiology of inhibitor formation with recombinant factor VIII replacement therapy. *Haemophilia*. 12:579–90.

- Picanco V, Heinz S, Bott D, Behrmann M, Covas DT, Seifried E, Tonn T. 2007. Recombinant expression of coagulation factor VIII in hepatic and non-hepatic cell lines stably transduced with third generation lentiviral vectors comprising the minimal factor VIII promoter. *Cytotherapy*. 9:785–794
- Pipe S. 2009. Visions in haemophilia care. *Thromb Res*. 124 Suppl 2:S2-5.
- Powell JS. Recombinant factor VIII in the management of hemophilia A: current use and future promise. *Ther Clin Risk Manag*. 2009 Apr;5(2):391-402. Epub 2009 May 20.
- Richard GF, Kerrest A, Dujon B. 2008. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev*. 72(4):686-727.
- Rodrigues ES. 2006. Clonagem e expressão do fator VIII recombinante em células de mamífero utilizando o sistema retroviral. Dissertação de Mestrado.
- Roe T, Reynolds TC, Yu G, Brown PO. 1993. Integration of murine leukemia virus DNA depends on mitosis. *EMBO J*. 12: 2099–2108.
- Rohdewohld H, Weiher H, Reik W, Jaenisch R, Breindl M. 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol*. 61(2):336-43.
- Sandberg H, Almstedt A, Brandt J, Castro VM, Gray E, Holmquist L, Lewin M, Oswaldsson U, Mikaelsson M, Jankwski MA, Bond M, Scoble HA. 2001. Structure and functional characterization of B-domain-deleted recombinant factor VIII. *Semin Hematol*. 38, 2, 4: 4-12.
- Scherdin, U., Rhodes, K., and Breindl, M. 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol*. 64, 907–912.
- Schröder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*. 23;110(4):521-9.
- Shapiro AD. 2007. Anti-hemophilic factor (recombinant), plasma/albumin-free method (octocog-alpha; ADVATE) in the management of hemophilia A. *Vascular Health and Risk Management*. 3(5) 555-565.
- Sheridan PL, Bodner M, Lynn A, Phuong TK, DePolo NJ, de la Vega DJ Jr, O'Dea J, Nguyen K, McCormack JE, Driver DA, Townsend K, Ibanez CE, Sajjadi NC, Greengard JS, Moore MD, Respess J, Chang SM, Dubensky TW Jr, Jolly DJ, Sauter SL. 2000. Generation of retroviral packaging and producer cell lines for large-scale vector production and clinical application: improved safety and high titer. *Mol Ther* 2(3):262-75.
- Studamire B, Goff SP. 2008. Host proteins interacting with the Moloney murine leukemia virus integrase: multiple transcriptional regulators and chromatin binding factors. *Retrovirology*. 13;5:48
- Vandamme AM. 2003. Viral evolution and molecular epidemiology--evolving viruses and evolving analysis techniques. *FEMS Immunol Med Microbiol*. 28;39(2):95-6.
- Vehar GA, Keyt B, Eaton D et al. 1984. Structure of human factor VIII. *Nature*. 312: 337-342

Wu X, Li Y, Crise B, Burgess SM. 2003. Transcription start regions in the human genome are favored targets for MLV integration. *Science*. 13;300(5626):1749-51.