Naja Vergani

**Triagem funcional de genes envolvidos no processo
de manutenção da inativação do cromossomo X em
humanos**

***Functional screening of genes involved in the
maintenance of X chromosome inactivation
in humans***

São Paulo

2014

Naja Vergani

**Triagem funcional de genes envolvidos no processo
de manutenção da inativação do cromossomo X em
humanos**

***Functional screening of genes involved in the
maintenance of X chromosome inactivation in
humans***

Tese apresentada ao Instituto de
Biociências da Universidade de
São Paulo, para a obtenção de
Título de Doutor em Biologia, na
Área de Genética.

Orientadora: Profa. Dra. Lygia da
Veiga Pereira Carramaschi

São Paulo

2014

Vergani, Naja

Triagem funcional de genes envolvidos no processo de manutenção da inativação do cromossomo X em humanos.

Número de páginas 241p.

Tese (Doutorado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Inativação do cromossomo X, 2. Triagem genômica funcional, 3. Epigenética, I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Profa. Dra. Lygia da Veiga Pereira Carramaschi
Orientadora

Este trabalho foi realizado com auxílio financeiro das agências de fomento Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – PROEX – Projeto 33002010021P7), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Dedicatória

Para você meu vizinho querido Syllas Ermel,
por todo imenso carinho, pela maravilhosa ajuda sempre
e pelos lindos exemplos de vida. Vou carregá-los sempre comigo.

E também para minha mãe Vivian , meu irmão Guido,
minha vó Glacy e meu querido Thomas.
Vocês são parte de mim.

Amo muito vocês.

“If you can keep your head when all about you
Are losing theirs and blaming it on you,
If you can trust yourself when all men doubt you,
But make allowance for their doubting too;
If you can wait and not be tired by waiting,
Or being lied about, don't deal in lies,
Or being hated, don't give way to hating,
And yet don't look too good, nor talk too wise;

If you can dream - and not make dreams your master;
If you can think - and not make thoughts your aim;
If you can meet with Triumph and Disaster
And treat those two impostors just the same;
If you can bear to hear the truth you've spoken
Twisted by knaves to make a trap for fools,
Or watch the things you gave your life to, broken,
And stoop and build 'em up with worn-out tools;

If you can make one heap of all your winnings
And risk it on one turn of pitch-and-toss,
And lose, and start again at your beginnings
And never breathe a word about your loss;
If you can force your heart and nerve and sinew
To serve your turn long after they are gone,
And so hold on when there is nothing in you
Except the Will which says to them: 'Hold on!';

If you can talk with crowds and keep your virtue,
Or walk with Kings - nor lose the common touch,
If neither foes nor loving friends can hurt you,
If all men count with you, but none too much;
If you can fill the unforgiving minute
With sixty seconds' worth of distance run,
Yours is the Earth and everything that's in it,
And - which is more - you'll be a Man, my son!”

Rudyard Kipling, If.

Agradecimentos

À Profa. Dra. Lygia da Veiga Pereira, pela grande oportunidade, por todo o carinho, compreensão e apoio durante todos esses anos de trabalho.

À Profa. Dra. Tatiana Teixeira Torres pela imensa ajuda, boa vontade e toda a paciência em me ensinar.

À Profa. Dra. Jeanne Loring, pela acolhida e colaboração.

À Mara, Profa. Dra. Ângela Maria Vianna Morgante, aos colegas Rob, Rathí e à Profa. Dra. Louise Laurent pela importante ajuda recebida nesse trabalho.

Aos amigos de laboratório e do coração Simone, Ana Maria, Fabiano e Gustavo pela grande e constante ajuda, pela amizade de todo dia, pelos momentos de diversão e pelo carinho de vocês. E também à Érica e Joana pela super ajuda e à Fernanda, Luis, Max, Cynthia, Tiago, Bruno, Juliana, Giovana, Karla por todos os bons momentos compartilhados.

À todos os amigos do departamento, entre eles Maraísa, Deisy, Helenice, Susy, Paulo, Lilian, Marianne, Marcos e Estela que muito me ajudaram. Especialmente à Fátima, pelas deliciosas conversas, pelo carinho e por todo o apoio.

Aos Professores Regina Mingroni, Luciana Haddad e Luís Netto.

Ao grande amigo Edmar Zanoteli por todas as conversas, apoio e incentivo.

Um agradecimento especial ao meu querido “bomdrasto” Edmundo Roberto Doura (Lê) por todo imenso apoio, pelas risadas e por todo carinho.

A Deus por colocar em meu caminho pessoas tão maravilhosas.

Dedicatória.....	iv
Epígrafe.....	v
Agradecimentos.....	vi
Lista de abreviaturas.....	xi
Lista de figuras.....	xiii
Lista de tabelas.....	xvi
1. Introdução.....	1
1.1. Estrutura da cromatina e mecanismos de regulação epigenética.....	2
1.1.1. Variantes de histonas.....	4
1.1.2. Modificações pós-traducionais em histonas.....	7
1.1.3. Modificações epigenéticas no DNA.....	15
1.1.4. Papel dos RNAs não-codificadores.....	18
1.2. Inativação do cromossomo X em mamíferos.....	23
1.3. Tipos de XCI.....	25
1.4. Genes que escapam à XCI.....	25
1.5. O centro de inativação do cromossomo X.....	27
1.6. Etapas da XCI.....	30
1.6.1. Contagem do número de cromossomos X.....	30
1.6.2. Escolha dos futuros Xa e Xi.....	32
1.6.3. Silenciamento do cromossomo X.....	33
1.6.4. Manutenção do estado de inativação.....	36
1.7. Uso de células HPRT ^{+/-} para o estudo da manutenção da XCI.....	39
1.8. Uso de Bibliotecas de shRNA para Triagem Genômica Funcional.....	42
2. Objetivos.....	46
3. Material e Métodos.....	49
3.1. Laboratórios e financiamentos.....	50
3.2. Linhagens celulares.....	50
3.3. Cultivo de fibroblastos.....	52
3.4. Passagem de fibroblastos.....	53

3.5.	Congelamento de fibroblastos.....	53
3.6.	Descongelo de fibroblastos.....	54
3.7.	Análise da eficiência de transdução viral em linhagens primárias de fibroblastos dérmicos humanos heterozigotas para o gene <i>HPRT</i>	55
3.8.	Extração de DNA de fibroblastos humanos primários.....	57
3.9.	Extração de RNA de fibroblastos humanos primários.....	57
3.10.	Produção de cDNA a partir de RNA.....	58
3.11.	Quantificação de ácidos nucleicos para sequenciamento direto.....	58
3.12.	Genotipagem de SNPs por sequenciamento direto.....	58
3.13.	Teste de sensibilidade de fibroblastos humanos primários à droga 6-tioguanina.....	65
3.14.	Seleção de fibroblastos humanos primários <i>HPRT</i> ⁺ e <i>HPRT</i> ⁻	65
3.15.	Cariotipagem de células da linhagem GM1661 selecionadas em meio 6TG.....	67
3.16.	Geração e análise de curvas de crescimento celular.....	69
3.17.	Escolha da biblioteca de shRNAs utilizada.....	72
3.18.	Otimização da transdução de fibroblastos primários humanos da linhagem GM1661 com partículas lentivirais copGFP.....	74
3.19.	Transdução de fibroblastos primários humanos com a biblioteca de shRNAs.....	77
3.20.	Extração e quantificação de DNA dos fibroblastos transduzidos com a biblioteca de shRNA.....	80
3.21.	<i>Multiplex</i> PCR para amplificação dos insertos de shRNAs.....	80
3.22.	Purificação dos amplicons contendo shRNAs.....	85
3.23.	Sequenciamento de próxima geração (NGS).....	86
3.24.	Análise dos dados de NGS.....	87
3.24.1.	Corrida teste.....	87
3.24.2.	Corrida final.....	88
3.25.	Estudo das características e funções dos genes candidatos para manutenção da XCI identificados na triagem funcional.....	92
4.	Resultados.....	93
4.1.	Escolha das linhagens celulares utilizadas.....	94
4.2.	Análise da eficiência de transdução viral em linhagens femininas de fibroblastos dérmicos humanos primários <i>HPRT</i> ^{+/-}	95

4.3.	Genotipagem de SNPs por sequenciamento direto.....	96
4.4.	Teste de sensibilidade de fibroblastos humanos primários à droga 6-tioguanina.....	100
4.5.	Geração e análise de curvas de crescimento celular.....	103
4.6.	Cariotipagem de células da linhagem GM1661 selecionadas em meio 6TG.....	106
4.7.	Escolha da biblioteca de shRNA utilizada.....	107
4.8.	Otimização da transdução de fibroblastos primários humanos da linhagem GM1661 com partículas lentivirais copGFP.....	109
4.9.	Transdução de fibroblastos primários humanos com a biblioteca de shRNAs.....	112
4.10.	<i>Multiplex</i> PCR para amplificação dos insertos de shRNAs.....	114
4.11.	Purificação dos amplicons contendo shRNAs.....	116
4.12.	Sequenciamento de próxima geração (NGS).....	118
4.13.	Análise dos dados de NGS.....	121
4.13.1.	Corrida teste.....	121
4.13.2.	Corrida final.....	129
5.	Discussão.....	145
5.1.	Metodologia utilizada.....	146
5.2.	Seleção em 6TG e HAT.....	147
5.3.	Representatividade da biblioteca de shRNAs.....	148
5.4.	Escolha do método de análise dos dados gerados por NGS.....	153
5.5.	Comparação dos dados obtidos a partir de alinhamentos contra diferentes bancos de dados.....	155
5.6.	Genes candidatos.....	158
5.6.1.	H3.3B e ASF1A.....	168
5.6.2.	Candidatos envolvidos em mecanismos de <i>splicing</i>	171
5.6.3.	Outras possíveis correlações entre diferentes candidatos.....	172
5.7.	Genes envolvidos na manutenção da XCI.....	174
5.8.	Direções futuras.....	177
6.	Conclusões.....	179
7.	Resumo.....	182
8.	<i>Abstract</i>	184
9.	Referências Bibliográficas.....	186

10. Anexos.....	212
Anexo 1.....	215
Anexo 2.....	217
Anexo 3.....	221

Lista de Abreviaturas

A	Adenina
C	Citosina
°C	Graus Celsius
Cat#	Número de catálogo
cDNA	DNA complementar
copGFP	Proteína verde fluorescente 2 (originária de <i>Pontellina plumata</i>)
DMSO	Dimetilsulfóxido
DNA	Ácido dexoxirribonucléico
dNTP	Desoxirribonucleotídeo trifosfato
EDTA	Ácido etilenodinitrotetracético
EGFP	<i>Enhanced green fluorescent protein</i> (proteína verde fluorescente)
ESC	<i>Embryonic stem cell</i> (células tronco-embrionárias)
FBS	<i>Fetal bovine serum</i> (soro fetal bovino)
G	Nucleotídeo contendo base guanina
g	Grama
HAT	Hipoxantina, Aminopterina e timidina. (Pode se referir também a histona acetiltransferase).
hESC	<i>Human embryonic stem cell</i> (células tronco-embrionária humanas)
HPRT	Hipoxantina guanina fosforibosiltransferase 1
IFU	<i>Infection units</i> (unidades de infecção)
kb	Quilobase
M	Molar
MgCl ₂	Cloreto de magnésio
mg	Miligrama
µg	Micrograma
mL	Mililitro
µL	Microlitro
mM	Milimolar
µM	Micromolar
MOI	<i>Multiplicity of infection</i> (multiplicidade de infecção)

n	Número amostral
NaCl	Cloreto de sódio
ncRNA	non-coding RNA (RNA não-codificador)
ng	Nanograma
NGS	<i>Next generation sequencing</i> (sequenciamento de próxima geração)
nM	Nanomolar
pb	Pares de base
PCR	Reação em cadeia da polimerase
pH	Potencial hidrogeniônico
pmol	Picomol
RNA	Ácido ribonucleico
rpm	Rotações por minuto
SDS	Dodecil sulfato de sódio
shRNA	<i>Short hairpin RNA</i>
SNP	<i>Single nucleotide polymorphism</i> (polimorfismo de base única)
T	Timina
TBE	Tampão Tris-Borato-EDTA
Tris	Tris (hidroximetil) aminometano
U	Unidade
U.V.	Ultravioleta
V	Volts
Xa	Cromossomo X ativo
Xi	Cromossomo X inativo
XCI	<i>X chromosome inactivation</i> (Inativação do cromossomo X)
6TG	6-tioguanina
8AZAG	8-azaguanina

Lista de Figuras

Figura 1. Localização de modificações em histonas ao longo dos genes e sua relação com a regulação da transcrição.....	9
Figura 2. Interações entre as modificações de histonas.....	14
Figura 3. Distribuição dos genes que escapam à XCI e genes submetidos a inativação em humanos e camundongos.....	27
Figura 4. Localização cromossômica e organização genômica do centro de inativação do cromossomo X em humanos e camundongos.....	28
Figura 5. Fatores que regulam a expressão de Tsix e Xist no futuro cromossomo X inativo e no cromossomo X que permanecerá ativo.....	33
Figura 6. Modificações epigenéticas no cromossomo X inativo ao longo da diferenciação celular em camundongos.....	35
Figura 7. Síntese <i>de novo</i> de purinas e papel de HPRT na via de recuperação de purinas.....	40
Figura 8. Monitoramento da atividade do Xi através do uso de meio de cultura seletivo.....	41
Figura 9. Esquematização das etapas experimentais desenvolvidas neste trabalho para a triagem funcional de genes envolvidos na manutenção da XCI em humanos.....	45
Figura 10. Sequência de nucleotídeos dos plasmídeos contendo shRNAs que compõem a biblioteca de shRNAs e localização dos primers utilizados nas duas rodadas de PCR para amplificação dos insertos de shRNAs.....	81
Figura 11. Cromatogramas evidenciando os genótipos heterozigotos encontrados na linhagem GM1661 para diferentes SNPs presentes em genes submetidos à XCI e genes que escapam à XCI em humanos.....	98

Figura 12. Cromatogramas evidenciando os genótipos heterozigotos encontrados na linhagem GM2226 para diferentes SNPs presentes em genes submetidos à XCI em humanos.....	98
Figura 13. Cromatogramas evidenciando os genótipos encontrados para dois diferentes SNPs presentes em genes submetidos à XCI em DNA e cDNA de células da linhagem GM1661.....	99
Figura 14. Visualização por microscopia de luz de fibroblastos dérmicos humanos primários da linhagem FibN1 não selecionados e selecionados em HAT ou 6TG após dois, quatro e doze dias do início de tratamento com as drogas 6TG e HAT.....	101
Figura 15. Visualização por microscopia de luz de fibroblastos humanos primários de prepúcio da linhagem Hs27 não selecionados e selecionados em HAT ou 6TG após dois, quatro e doze dias do início de tratamento com as drogas 6TG e HAT.....	102
Figura 16. Curvas de crescimento das células da linhagem GM1661 quando expostas aos tratamentos com 6TG ou HAT.....	104
Figura 17. Curva de crescimento das células da linhagem GM1661/HPRT ⁻ expostas ao tratamento com HAT.....	105
Figura 18. Cariotipagem de células da linhagem GM1661 selecionadas em meio contendo 6TG.....	106
Figura 19. Visualização por microscopia de fluorescência de fibroblastos primários humanos da linhagem GM1661 copGFP-positivos cerca de 48h após a transdução com partículas lentivirais pSIH1-H1-siLuc-copGFP em diferentes confluências e MOI.....	111
Figura 20. Visualização por microscopia de luz de fibroblastos primários humanos da linhagem GM1661 transduzidos com partículas lentivirais da biblioteca de shRNA e submetidos a 10 dias de seleção em meio contendo puromicina.....	113
Figura 21. Fotografias de géis de agarose mostrando os amplicons esperados após amplificação dos insertos de shRNA da biblioteca de shRNAs.....	115

Figura 22. Fotografia do gel de agarose após primeiro corte para excisão das regiões contendo os amplicons desejados nas amostras controle (6TG) e experimental (HAT).....	117
Figura 23. Fotografia do gel de agarose mostrando os amplicons presentes nas amostras controle (6TG) e experimental (HAT) antes e depois da purificação dos amplicons de interesse contidos nos fragmentos de agarose excisados.....	117
Figura 24. Resultado obtido após corrida das bibliotecas de amplicons em aparelho LabChip® GX-LST.....	120
Figura 25. Representação gráfica dos parâmetros avaliados pelo programa FastQC para as sequencias geradas na corrida teste de NGS da amostra controle 6TG...	126
Figura 26. Representação gráfica dos parâmetros avaliados pelo programa FastQC para as sequencias geradas na corrida teste de NGS da amostra experimental HAT.....	127
Figura 27. Representação gráfica dos parâmetros avaliados pelo programa FastQC para as sequencias geradas na corrida final de NGS da amostra controle 6TG....	132
Figura 28. Representação gráfica dos parâmetros avaliados pelo programa FastQC para as sequencias geradas na corrida final de NGS da amostra experimental HAT.....	133

Lista de Tabelas

Tabela 1. Principais Variantes de Histonas do Núcleo em Humanos.....	6
Tabela 2. Principais PTMs Observadas em Histonas em Humanos.....	11
Tabela 3. Principais Tipos de <i>Housekeeping</i> ncRNAs Encontrados em Mamíferos.....	19
Tabela 4. Principais Tipos de sncRNAs Encontrados em Mamíferos.....	21
Tabela 5. Principais Tipos de lncRNAs Encontrados em Mamíferos.....	22
Tabela 6. Linhagens de Fibroblastos Utilizadas.....	51
Tabela 7. Lista dos SNPs analisados.....	64
Tabela 8. Taxas de Eficiência de Transdução por Retrovírions Contendo o Gene Repórter EGFP nas Linhagens Celulares Estudadas.....	96
Tabela 9. Genótipos Referentes à SNPs em Genes Ligados ao Cromossomo X nas Linhagens Estudadas (DNA).....	97
Tabela 10. Genótipos dos Alelos Expressos (cDNA) na Linhagem GM1661 Referentes a SNPs em Genes Submetidos à XCI.....	100
Tabela 11. Eficiência de Transdução Lentiviral em Células das Linhagens HDF48, HDF49 e HDF51 em Diferentes MOI e Confluências Celulares.....	108
Tabela 12. Correspondência Entre o Número de Células copGFP Positivas e o Número de Cópias do Gene Réporter por Célula.....	109
Tabela 13. Eficiência de Transdução por Lentivíriões Baseados em HIV e Número de Cópias do Gene Repórter em Células da Linhagem GM1661/HPRT ⁻ em Diferentes MOI e Confluências Celulares.....	110
Tabela 14. Resumo dos Dados Obtidos nos Alinhamentos pelo Programa Bowtie Para a Corrida Teste de NGS.....	129

Tabela 15. Resumo dos Dados Obtidos nos Alinhamentos pelo Programa Bowtie Para a Corrida Final de NGS.....	134
Tabela 16. Lista de Candidatos Obtidos por Análise em Programa DESeq para o Banco de Dados de Sequências-Alvo.....	136
Tabela 17. Lista de Candidatos Obtidos por Análise em Programa DESeq para o Banco de Dados de Sequências de shRNAs.....	137
Tabela 18. Lista de Genes Candidatos Obtidos por Análise em Programa DESeq para o Banco de Dados RefMrna.....	138
Tabela 19. Identificação dos Transcritos e Regiões Genômicas Associados às Sequências Candidatas Obtidas nos Alinhamentos Contra os Bancos de Dados de Sequências-Alvo e Sequências de shRNAs.....	140
Tabela 20. Principais Características e Funções dos Genes Candidatos de Representação mais Elevada na Amostra Experimental Tratada em HAT.....	160

1. Introdução

Em mamíferos, uma grande diferença entre o número de genes presentes em células masculinas e femininas é observada. Isso se dá devido à presença de dois cromossomos X em fêmeas e apenas um cromossomo X em machos. Como o cromossomo X é visivelmente maior que o cromossomo Y e, portanto, contém maior quantidade de genes, grande parte dos genes presentes em um dos cromossomos X femininos sofre inativação a fim de que haja compensação da dosagem gênica entre os dois sexos. Por envolver silenciamento gênico em um cromossomo quase inteiro, a inativação do cromossomo X (XCI, do inglês, *X chromosome inactivation*) constitui o exemplo mais drástico de controle epigenético em mamíferos (Payer & Lee, 2008).

Discutiremos a seguir os principais mecanismos de regulação epigenética observados em mamíferos. Tal discussão ilustra a complexidade desses mecanismos e a necessidade em se compreender melhor esses fenômenos. Em seguida discutiremos os principais aspectos da XCI e descreveremos os principais eventos epigenéticos relacionados a cada uma das etapas desse processo. Finalmente, mencionaremos os modelos de estudo para XCI em humanos e descreveremos como o método de triagem funcional poderia ser utilizado para identificar novos candidatos envolvidos no processo de manutenção da XCI.

1.1. Estrutura da Cromatina e Mecanismos de Regulação Epigenética

O estudo da epigenética abrange o controle da expressão gênica através de diversos fatores que agem sob a cromatina. Em células eucariontes a cromatina é formada através da associação física do DNA com proteínas histonas e não-histonas e é organizada através de múltiplos níveis de compactação (Junqueira & Carneiro, 2000; Alberts *et al.*, 2002).

No nível mais baixo de compactação, 145pb a 147pb de DNA são enovelados ao redor de um octâmero de histonas constituído por duas histonas H2A, duas histonas H2B, duas histonas H3 e duas histonas H4, formando assim o

núcleo da estrutura básica da cromatina, o nucleossomo (Junqueira & Carneiro, 2000; Alberts *et al.*, 2002). Durante a formação do nucleossomo, dois heterodímeros de H3-H4 são primeiramente ligados formando um tetrâmero que reconhece e se liga a sequências específicas de DNA denominadas sequências posicionadoras de nucleossomos (NPS). Além das NPS, outras proteínas de ligação ao DNA também podem influenciar o posicionamento dos nucleossomos. Seguindo-se a ligação do tetrâmero de histonas H3-H4 ao DNA, dois dímeros de histonas H2A-H2B são incorporados e o DNA é então enovelado ao redor do octâmero de histonas (Alberts, 2002; Kumari *et al.*, 2012). Após o enovelamento do DNA, uma unidade de histona H1 se liga externamente a cada octâmero de histonas. Nucleossomos adjacentes são conectados por porções de DNA livre de histonas. Quando histonas H1 ligadas a nucleossomos vizinhos interagem entre si, a cromatina é arranjada em fibras de 10nm. Essa estrutura pode se enovelar ainda mais formando a estrutura helicoidal denominada solenóide ou fibra de 30nm. A fibra de 30nm se compacta ainda mais – provavelmente com o auxílio de proteínas não-histonas – formando alças de 300nm. A ligação dessas alças a um esqueleto protéico formado por proteínas não-histonas durante a divisão celular culmina na geração da mais condensada forma de cromatina, as cromátides de cromossomos metafásicos, que possuem cerca de 700nm de largura (Junqueira & Carneiro, 2000; Alberts *et al.*, 2002; Pierce, 2004).

Entretanto, o arranjo dos nucleossomos no genoma não serve apenas o papel de empacotamento do DNA no núcleo celular, mas possui também significado funcional. Os nucleossomos possuem propriedades repressivas inerentes, o que não pode ser mantido durante todas as fases do ciclo celular. O genoma deve ser capaz de afrouxar ou compactar a cromatina em cada uma de suas regiões específicas para que as diversas funções celulares possam ser desempenhadas e orquestradas no organismo. A presença de variantes de histonas, histona-chaperones (responsáveis pela adição e retirada de histonas do nucleossomo), modificações pós-transducionais (PTMs) em histonas, modificações no DNA, enzimas modificadoras de histonas, remodeladores de cromatina (proteínas que reconhecem as PTMs em histonas), proteínas efetoras, enzimas modificadoras de DNA, e RNAs não-codificadores (ncRNA) afetam a mobilidade dos nucleossomos e o dinamismo da cromatina, e uma colaboração muito precisa e complexa entre esses diversos efetores, ainda não completamente entendida, existe para orquestrar

os diversos eventos que ocorrem nos diferentes tipos celulares e fases do desenvolvimento do organismo (Alberts *et al.*, 2002; Kumari *et al.*, 2012).

1.1.1. Variantes de Histonas

As histonas H2A, H2B, H3 e H4 são classificadas como histonas canônicas ou histonas replicativas e estão entre as mais conservadas formas de proteínas presentes em todos os eucariontes. São codificadas por genes de múltiplas cópias organizados em grupos que não possuem íntrons e são transcritos em mRNAs que não sofre poliadenilação. Ao invés da cauda poli-A, os transcritos de histonas canônicas contém uma haste em alça na extremidade 3' que é responsável por restringir a expressão dessas moléculas à fase S da divisão celular, quando novas moléculas de histonas são produzidas para se associar ao DNA recém-sintetizado (Khare *et al.*, 2012; revisto por Law & Cheung, 2012).

Além das histonas canônicas, diversas formas variantes, estão presentes em humanos, sendo que todas as histonas canônicas possuem formas variantes correspondentes. Em humanos, existem no total, 29 formas de histona H2A, 19 formas de histonas H2B, seis formas de histonas H3 e duas formas de histona H4 (Khare *et al.*, 2012; revisto por Law & Cheung, 2012).

As variantes de histonas são isoformas não-alélicas das histonas canônicas e diferem das mesmas em pequenas porções da sequência de aminoácidos. São codificadas por genes de cópias únicas que podem conter íntrons, seus transcritos sofrem poliadenilação, são constitutivamente transcritos durante todas as fases do ciclo celular e alguns possuem expressão tecido-específica, o que sugere que essas variantes possuem funções biológicas adicionais além de compactação do DNA (Khare *et al.*, 2012; revisto por Law & Cheung, 2012). A presença de variantes de histonas no nucleossomo é suficiente para causar alteração na estrutura da cromatina, tornando-a mais compacta ou mais relaxada devido à alteração do posicionamento do nucleossomo e de sua afinidade pelo DNA e por outras proteínas, como proteínas remodeladoras da cromatina e proteínas envolvidas no reparo de DNA, divisão celular e transcrição gênica (revisto por Kumari *et al.*, 2012; Law & Cheung, 2012). A maioria das variantes de histona H2B possui expressão tecido-específica enquanto variantes de histona H3 e H2A são mais uniformemente

encontradas e estão envolvidas em diversas funções celulares. As diferentes características e funções das principais variantes de histonas do núcleo encontradas em mamíferos estão resumidas na Tabela 1 (Ausió, 2006; Loyola & Almouzni, 2007; Banaszynski *et al.*, 2010; Gonzáles-Romero *et al.*, 2010; Jufvas *et al.*, 2011; Kumari *et al.*, 2012; Law & Cheung, 2012).

Embora não faça parte do núcleo do nucleossomo, a histona H1 também desempenha um importante papel na estabilidade do nucleossomo e estrutura da cromatina. No total, 10 formas de histona H1 podem ser encontradas em humanos sendo que todas são codificadas por genes de cópia única. Cinco formas são transcritas de maneira dependente de replicação e não sofrem poliadenilação dos pré-mRNAs (H1.1, H1.2, H1.3, H1.4 e H1.5) e outras cinco formas são transcritas de maneira independente de replicação e os pré-mRNAs são poliadenilados (H1.0, H1oo, H1t, H1x e H1 testículo-específica). De acordo com suas propriedades e funções, as histonas H1 são classificadas como condensadores de cromatina fortes (H1.0, H1.4, H1.5, H1x), intermediários (H1.3) e fracos (H1.1, H1.2). Muitas variantes de H1 apresentam padrão de expressão tecido-específico. Além de sua função na estrutura da cromatina, as histonas H1 estão envolvidas em processos como espermatogênese, fertilização, apoptose e regulação da expressão gênica (Khare *et al.*, 2012).

Tabela 1. Principais Variantes de Histonas do Núcleo em Humanos

Variante de Histona	Principais Características e Funções
Variantes de H2A	
H2AZ	Principalmente relacionada à ativação da transcrição gênica. Confere maior mobilidade ao nucleossomo – cromatina mais relaxada – e recruta fisicamente elementos envolvidos no processo de transcrição. No entanto está também envolvida na formação de heterocromatina e silenciamento gênico (provavelmente quando ubiquitinada), segregação cromossômica e regulação do ciclo celular.
H2AX	Reparo do DNA e estabilidade cromossômica. Marca a região de dano no DNA e recruta fatores envolvidos no reparo. Apoptose, progressão no ciclo celular.
macroH2A	Três isoformas (macroH2A1.1, macroH2A1.2 e macroH2A2). Relacionada principalmente à repressão da transcrição. Enriquecida na heterocromatina facultativa incluindo Xi e alelos inativos de genes imprintados. Confere cromatina mais fechada. Inibe ligação de fatores de transcrição e complexos protéicos de remodelamento da cromatina associados ao estabelecimento de estado mais transcricionalmente ativo. No entanto pode também ser encontrada associada a genes transcricionalmente ativos, onde a função dessa variante ainda precisa ser elucidada.
H2A.Bdb tipos 1 e 2/3	Facilita a ativação da transcrição. Excluída do corpúsculo de Barr ou Xi. Cromatina mais relaxada. Envolvida na espermiogênese.
Variantes de H2B	
TH2A	Específica do testículo.
TH2B	Específica do testículo. Diminui a estabilidade do nucleossomo. Enriquecida nos telômeros.
H2BFWT	Específica do testículo. Relacionada com a reorganização da cromatina nos estágios finais da espermatogênese.
subH2BV	Específica do testículo. Relacionada a formação do acrossomo presente na cabeça de espermatozoides.
H2BL1	Específica do testículo. Relacionada com a reorganização da cromatina pericêntrica durante a espermatogênese.
H2BL2	Específica do testículo. Relacionada com a reorganização da cromatina pericêntrica durante a espermatogênese.
Variantes de H3	
CENP-A	Substitui H3 canônica no centrômero. Importante para formação do cinetócoro e segregação cromossômica. Também envolvida no reparo do DNA.
H3.1	Uma das isoformas canônicas. Diferença da H3.2 em apenas um aminoácido. Deposição dependente de replicação. Associada tanto à ativação como repressão transcricional. Preferencialmente localizada em genes que podem alterar do estado ativo ao reprimido dependendo do estado metabólico da célula e estágio de desenvolvimento do organismo. Envolvida no reparo do DNA.
H3.2	Uma das isoformas canônicas. Diferença da H3.1 em apenas um aminoácido. Deposição dependente de replicação. Enriquecida em heterocromatina. Geralmente associadas à repressão da transcrição.
H3.3	Enriquecida principalmente em genes ativos. Confere maior mobilidade ao nucleossomo (fibra mais relaxada). No entanto, encontrada também em regiões intergênicas, telômeros e centrômeros onde sua função ainda precisa ser elucidada.
H3.3C	Mitose.
H3.1t	Específica do testículo. Envolvida no empacotamento da cromatina em espermatozoides primários.
Variantes de H4	
HIST2H4B	Identificada recentemente em células adiposas humanas.

1.1.2. Modificações Pós-Traducionais em Histonas

As histonas são proteínas básicas, ricas em lisinas e argininas, formadas por uma porção globular e uma cauda N-terminal altamente flexível que se projeta para fora do DNA enovelado ao octâmero de histonas podendo fazer contato com nucleossomos adjacentes. Geralmente as modificações pós-traducionais (PTMs) ocorrem na cauda N-terminal das histonas, embora possam ocorrer também na região globular. A adição de PTMs pode afetar direta ou indiretamente – através do recrutamento ou repelência de proteínas que se associam à cromatina – o estado de condensação da cromatina, de forma a organizar o genoma em regiões transcricionalmente ativas e inativas, denominadas respectivamente, eucromatina e heterocromatina (revisto por Bannister & Kouzarides, 2011). As PTMs de histona podem recrutar proteínas que escrevem novas PTMs, apagam outras PTMs ou lêem essas modificações (como fatores de transcrição e fatores remodeladores de cromatina), e o repertório desses modificadores de cromatina é bastante grande (cerca de 150 diferentes proteínas já descritas em humanos) (revisto por Bannister & Kouzarides, 2011; Khare *et al.*, 2012).

As modificações pós-traducionais (PTMs) mais comuns em histonas são acetilação, metilação (mono-, di- e tri- metilação), fosforilação e ubiquitinação, embora outros tipos de modificações como SUMOilação, ADP-ribosilação, citrulinação, biotinilação, adição de β -N-acetilglucosamina e isomerização de prolinas tenham sido também descritas (Kothapalli *et al.*, 2005; Kouzarides, 2007; Smith & Denu, 2009; Sakabe *et al.*, 2010; Banerjee & Chakravarti, 2011; Bannister & Kouzarides, 2011; Inbar-Feigenberg *et al.*, 2013). Embora algumas PTMs possam ser adicionadas a histonas livres, a grande maioria dos padrões de PTMs de histonas é estabelecida durante e após a deposição das histonas no nucleossomo e depende das condições do microambiente onde essas histonas são incorporadas. Enquanto algumas marcas estão mais comumente relacionadas à ativação da transcrição gênica outras estão mais comumente relacionadas à repressão gênica (Figura 1, Tabela 2). No entanto, as histonas podem ser submetidas simultaneamente a mais de um tipo de PTM em diferentes resíduos de aminoácidos, o que gera uma grande gama de modificações únicas ou combinadas possíveis, cada uma com suas respectivas consequências funcionais (revisto Banerjee & Chakravarti, 2011; Miller & Grant, 2012).

As principais PTMs de histonas estão descritas a seguir. Uma lista abrangente de modificações em histonas observadas em humanos e suas principais funções é fornecida na Tabela 2 (Shiio & Eisenman, 2003; Khotapalli *et al.*, 2005; Kouzarides, 2007; Zemleni *et al.*, 2009; Baneerjee & Chakravarti, 2011; Messner & Hottiger, 2011; Zhang *et al.*, 2011; Fong *et al.*, 2012; Khare *et al.*, 2012; Miller & Grant, 2012; Rossetto *et al.*, 2012; Wright *et al.*, 2012).

A acetilação de resíduos de lisina (K) em histonas é a PTM de histona mais bem estudada. Essa reação é catalisada pelas enzimas histona-acetiltransferases (HATs) que transferem um grupo acetil do acetilCoA ao grupo ϵ -amino de lisinas das cadeias laterais das histonas, o que resulta em neutralização da carga da lisina e redução da afinidade entre o DNA, negativamente carregado, e a histona. Sendo assim, a acetilação geralmente leva a um afrouxamento da cromatina e predispõe a ativação da transcrição gênica através da exposição de sítios de ligação ao DNA (Figura 1, Tabela 2) (revisto por Smith & Denu, 2009; Khare *et al.*, 2012; Miller & Grant, 2012). As histona-desacetilases (HDACs), presentes em diversos complexos proteicos, reverterem a reação química desempenhada pelas HATs, removendo as marcas de acetilação das lisinas para restaurar suas cargas positivas e, portanto, estão geralmente associadas à repressão da transcrição gênica (revisto por Miller & Grant, 2012).

A metilação de histonas ocorre nos resíduos lisina (K) e arginina (R) e é catalisada pelas histona-metiltransferases (HMTs). As lisinas podem ser mono-, di-, ou trimetiladas enquanto que as argininas podem ser mono- ou dimetiladas. As HMTs que metilam lisinas (HKMTs) catalisam a transferência de um grupo metil da S-adenosilmetionina (SAM) para o grupo ϵ -amino de lisinas. As HKMTs possuem alta especificidade e geralmente modificam apenas uma única lisina em um único tipo de histona. A metilação de lisinas pode levar tanto a ativação como repressão da transcrição gênica (Figura 1, Tabela 2) e pode ser revertida pela ação de histona-demetilases (HDMTs). A metilação de argininas também pode levar tanto a ativação como repressão da transcrição gênica (Tabela 2). Esse tipo de metilação é realizada por enzimas denominadas arginina-metiltransferases de proteínas (PRMTs) que transferem um grupo metil de SAM para o grupo ω -guanidino de argininas (revisto por Kouzarides, 2007; Smith & Denu, 2009; Bannister & Kouzarides, 2011; Miller & Grant, 2012). Embora a existência de um tipo de histona-arginina demetilase tenha sido relatada (Chang *et al.*, 2007),

estudos posteriores não confirmaram o papel de histona-arginina demetilase dessa enzima (Webby *et al.*, 2009; Han *et al.*, 2012). Por ser uma modificação pequena e neutra, a metilação provavelmente não afeta diretamente a estrutura da cromatina e age indiretamente através do recrutamento de proteínas que regulam a transcrição gênica ou de proteínas modeladoras da cromatina (revisto por Bannister & Kouzarides, 2011).

Figura 1. Localização de modificações em histonas ao longo dos genes e sua relação com a regulação da transcrição. O padrão de enriquecimento de histonas é dado em um modelo gênico e sua região reguladora. A região reguladora corresponde a pequena região que precede a região livre de nucleossomos (região vazia) que, por sua vez, precede os sítios de início da transcrição representados por uma seta. Os efeitos na expressão gênica relacionados a cada tipo de modificação estão indicados por cores diferentes: azul para ativação da transcrição, vermelho para repressão da transcrição e violeta para enriquecimento de marcas presentes tanto em genes ativos como em genes reprimidos. Extraído de Miller & Grant, 2012.

A fosforilação em histonas pode ocorrer em resíduos de serina (S), treonina (T), tirosina (Y) ou histidina (H) sendo que todos os tipos de histona do núcleo e também a histona H1 estão sujeitas a esse tipo de modificação. É mediada por diferentes quinases através da transferência de um grupo fosfato do ATP para o grupo hidroxil do aminoácido alvo – o que adiciona carga negativa às histonas – e pode ser revertida pela ação de diferentes fosfatases. A fosforilação de histonas foi associada a uma variedade de processos celulares, incluindo ativação e repressão da transcrição gênica, apoptose, progressão no ciclo celular, reparo de DNA, condensação cromossômica na mitose e meiose e regulação gênica durante o desenvolvimento (Tabela 2) (revisto por Banerjee & Chakravarti, 2011; Bannister & Kouzarides, 2011; Khare *et al.*, 2012; Rossetto *et al.*, 2012).

A ubiquitinação de histonas envolve a ligação de um (mono-ubiquitinação) ou mais (poli-ubiquitinação) polipeptídeo de 76 aminoácidos denominado ubiquitina aos resíduos lisina (K) através da ação seqüencial de três enzimas – a *E1-ubiquitin activating enzyme*, *E2-ubiquitin conjugating enzyme* e *E3-ubiquitin ligase*. Em mamíferos, embora possa ocorrer em todos os tipos de histonas do núcleo, ocorre mais frequentemente na forma de mono-ubiquitinação nas lisinas K119 da histona H2A e K120 da histona H2B, sendo menos prevalente em outras histonas. A mono-ubiquitinação de H2AK119 em humanos está associada a repressão da transcrição gênica, enquanto que a mono-ubiquitinação de H2BK120 está ligada a ativação da transcrição gênica. Além disso, a ubiquitinação de histonas está também associada ao reparo do DNA e ao processamento de certos pré-mRNAs (Tabela 2). O mecanismo de ação da ubiquitinação de histonas envolve tanto o recrutamento de fatores modificadores de histonas adicionais à cromatina como age fisicamente mantendo a cromatina mais aberta devido ao grande tamanho desse tipo de modificação. A ubiquitinação de histonas pode ser revertida pela ação de ubiquitina-proteases (revisto por Kouzarides, 2007; Bannister & Kouzarides, 2011; Khare *et al.*, 2012; Wright *et al.*, 2012).

Tabela 2. Principais PTMs Observadas em Histonas em Humanos

Tipos de PTMs e Resíduos de Histonas	Papel na Transcrição ou Atividade Celular
Acetilação de Lisina (Kac)	
H1.4 (K25)	Ativação da transcrição gênica
H2A (K5)	Ativação da transcrição gênica, reparo do DNA
H2A (K9)	Ativação da transcrição gênica
H2B (K5, 12, 15, 16, 20, 46, 120)	Ativação da transcrição gênica
H3 (K4, 27, 36)	Ativação da transcrição gênica
H3 (K9, 14, 23)	Ativação da transcrição gênica, deposição de histona no nucleossomo, reparo do DNA
H3 (K18)	Ativação da transcrição gênica, replicação do DNA, reparo do DNA
H3 (K56)	Ativação da transcrição gênica, reparo do DNA
H4 (K5)	Ativação da transcrição gênica, deposição de histona no nucleossomo
H4 (K8, 16)	Ativação da transcrição gênica, reparo do DNA
H4 (K12)	Ativação da transcrição gênica, deposição de histona no nucleossomo, reparo do DNA, silenciamento telomérico
H4 (K91)	Ativação da transcrição gênica
Fosforilação de Serina (Sph)	
H1 (S17, 171, 172, 188)	
H1.4 (S26)	Ativação da transcrição gênica, descondensação da cromatina
H1.4 (S186)	Ativação da transcrição gênica
H2A (S1)	Repressão da transcrição gênica, mitose
H2A (S16)	Sinalização por EGF
macroH2A.1 (S137)	Promove exclusão de macroH2A.1 da heterocromatina. Enriquecida na mitose.
H2AX (S139)	Reparo do DNA, apoptose
H2B (S14)	Meiose, apoptose, reparo do DNA
H2B (S32)	Sinalização por EGF, possível papel na apoptose
H2B (S36)	Ativação da transcrição gênica e regulação da cromatina em resposta celular ao stress
H3 (S10, 28)	Mitose e meiose, ativação da transcrição gênica
H3.3 e H3.3C (S31)	Mitose
CENP-A (S36)	Citocinese
H4 (S1)	Reparo do DNA, mitose
H4 (S47)	Deposição de dímeros H3.3-H4 no nucleossomo

**Tabela 2. Principais PTMs Observadas em Histonas em Humanos
(Continuação)**

Tipos de PTMs e Resíduos de Histonas	Papel na Transcrição ou Atividade Celular
Fosforilação de Treonina (Tph)	
H1 (T10, T17, T30, T137, T145, T153, T154)	Mitose
H2A (T120)	Mitose, meiose, estabilidade cromossômica, reparo do DNA
H3 (T3)	Localização de CPC (<i>chromosome passenger complex</i>) no centrômero durante a mitose.
H3 (T6)	Ativação da transcrição gênica mediada por andrógenos, proliferação celular
H3 (T11)	Ativação da transcrição gênica mediada por andrógenos, mitose, reparo do DNA
H3 (T45)	Apoptose
Fosforilação de Tirosina (Yph)	
H2AX (Y142)	Decisão entre reparo do DNA/sobrevivência celular e apoptose
H2B (Y37)	Repressão da transcrição gênica
H3 (Y41)	Repressão da transcrição gênica e diferenciação celular durante hematopoiese
Fosforilação de Histidina (Hph)	
H4 (H18, H75)	Pode facilitar a replicação do DNA
Metilação de Lisina (Kme)	
H1.4 (K25me1)	Repressão da transcrição gênica
H1.2 (K126me1)	
H2B (K5me1)	Associada a promotores de genes ativos
H3 (K4me1,2,3, K23, K27me1, K79me1,2,3)	Ativação da transcrição gênica
H3 (K9me1,2,3, 27me2,3)	Repressão da transcrição gênica, mas pode também estar presente em genes ativos.
H3 (K36me1)	Ativação da transcrição gênica, mas pode também estar presente em genes inativos.
H4 (K20me1,2,3)	Repressão da transcrição gênica (me1,3), mitose (me1,2)
Metilação de Arginina (Rme)	
H2A (R3me2)	
H3 (R2me1)	Ativação da transcrição gênica
H3 (R2me2)	Repressão da transcrição gênica
H3 (17me1,2, 26me1)	Ativação da transcrição gênica
H3 (R8me2)	Repressão da transcrição gênica
H4 (R3me1,2)	Ativação da transcrição gênica
Ubiquitinação de Lisina (Kub)	
H2A (K119)	Repressão da transcrição gênica mediada por polycomb (monoub), reparo do DNA (poliub)
H2A.Z (K121ub1)	Enriquecida no Xi, associada a genes silenciados, se associa a H3K4me2 e H3K4me3 no nucleossomo
H2B (K120ub1)	Ativação da transcrição gênica, reparo do DNA
H4 (K91ub1)	Proteção contra danos ao DNA

**Tabela 2. Principais PTMs Observadas em Histonas em Humanos
(Continuação)**

Tipos de PTMs e Resíduos de Histonas	Papel na Transcrição ou Atividade Celular
Sumoilação de Lisina (Ksu)	
H4 (?)	Repressão da transcrição gênica
Citrulinização de Arginina (Rcit)	
H2A (R3)	
H3 (R8)	Repressão da transcrição gênica
H3 (R2, 17, 26)	Repressão da transcrição gênica
H4 (R3)	Repressão da transcrição gênica
Biotinilação de Lisinas (Kbio)	
H2A (K9, 13, 125, 127, 129)	
H3 (K4, 9, 18, 23)	
H4 (K8, 12, 16)	Repressão da transcrição gênica, meiose, reparo do DNA
ADP-Ribosilação de Glutamatos (Ear) e Lisinas (Kar)	
H1 (E2, 14, 116 e K213)	
H2A (K13)	
H2B (E2 e K30)	
H3 (K27, K37)	
H4 (K16)	Replicação do DNA, reparo do DNA
O-GlcNAcilação de Serinas (Sglc) e Treoninas (Tglc)	
H2A (T101)	
H2B (S36)	Possível papel na progressão do ciclo celular
H2B (S112)	Ativação da transcrição gênica
H3 (S10, T32)	Mitose e progressão no ciclo celular
H4 (S47)	

De maneira geral, a eucromatina é caracterizada por altos níveis de acetilação e altos níveis de H3K4me1/2/3, H3K36me3 e H3K79me1/2/3, ao passo que a heterocromatina é caracterizada por baixos níveis de acetilação e altos níveis de H3K9me2/3, H3K27me2/3 e H4K20me3. Entretanto, o padrão de marcação de histonas é muito mais complexo do que primeiramente se imaginava e estudos recentes desafiam essa distinção simplificada entre modificações associadas a estados gênicos transcricionalmente ativos ou reprimidos (revisto por Bannister & Kouzarides, 2011; Miller & Grant, 2012).

De fato, algumas marcas podem estar associadas tanto a ativação como repressão gênica (Figura 1, Tabela 2). Outras podem também estar associadas a um estado transcricional bivalente, no qual o padrão de ativação ou repressão

gênica pode ser rapidamente alterado, como por exemplo, promotores de genes que regulam o desenvolvimento em células pluripotentes. Essas regiões promotoras apresentam as marcas de metilação H3K4me3 (frequentemente associada a genes ativos) e H3K27me3 (frequentemente associada a genes inativos) em alta proximidade. Além disso, parece haver uma interação entre as diversas PTMs de histonas presentes no nucleossomo ou em nucleossomos próximos (Figura 2), de maneira que as diferentes proteínas associadas à cromatina que se ligam a PTMs de histonas podem interagir sinergisticamente ou antagonisticamente resultando em vários gradientes de ativação ou repressão transcricional através do genoma. Esses exemplos ilustram a complexidade dos padrões de marcação de histonas que ainda precisam ser melhor compreendidos (revisto por Bannister & Kouzarides, 2011; Miller & Grant, 2012).

Figura 2. Interações entre as modificações de histonas. Várias PTMs de histonas afetam a ocorrência de outras PTMs de histonas. As setas indicam um efeito positivo e as barras um efeito negativo. Extraído de Bannister & Kouzarides, 2011.

1.1.3. Modificações Epigenéticas no DNA

A metilação do DNA é o tipo mais frequente de modificação epigenética que afeta diretamente a molécula de DNA em eucariontes. É essencial para o desenvolvimento normal do organismo e possui importante função na regulação da transcrição gênica e manutenção da estabilidade genômica (Shen & Zhang, 2013).

O estabelecimento do perfil genômico de metilação de DNA ocorre cedo na embriogênese, no período de implantação do embrião, após as marcas de metilação contidas nos DNAs dos gametas terem sido apagadas, um processo que se inicia ainda no zigoto. Esse fenômeno é um componente-chave do sistema de marcação epigenética e simboliza o apagamento da programação contida nas células da linhagem germinativa como um prelúdio para o estabelecimento da totipotência. Uma vez estabelecido, o padrão de metilação é mantido através de todas as divisões celulares subsequentes e funciona principalmente como um mecanismo de repressão da expressão gênica. Ajustes são realizados durante toda a vida do organismo através de eventos de metilação *de novo* e demetilação altamente direcionados (revisto por Auclair & Weber, 2012; Bergman & Cedar, 2013).

A adição de grupamentos metil ao DNA é catalisada pelas DNA metiltransferases (DNMTs) que transferem um grupamento metil ($-\text{CH}_3$) da SAM ao carbono 5' do anel de pirimidina da citosina, gerando 5-metilcitosina (5-mC). Quatro DNMTs foram caracterizadas em humanos – DNMT1, DNMT2, DNMT3a, DNMT3b e DNMT3L. A DNMT1 é responsável por manter as marcas de metilação durante a replicação do DNA. O processo de replicação gera sítios de DNA hemimetilado, o tipo de substrato preferencial de DNMT1, que copia as marcas de metilação de CpG presentes na cadeia molde de DNA na nova fita recém-sintetizada. A DNMT1 é recrutada aos sítios de replicação do DNA pelo antígeno nuclear de proliferação celular (PCNA) e pela proteína nuclear NP95, que se liga as CpGs hemimetiladas (revisto por Auclair & Weber, 2012; Miller & Grant, 2012).

A metilação *de novo* do DNA é estabelecida durante o desenvolvimento pelas DNMT3A e DNMT3B. Essas duas enzimas são altamente homólogas, mas possuem alvos específicos diferentes e padrões de expressão distintos. A DNMT3B é mais prevalente durante os estágios iniciais da embriogênese e é a principal enzima envolvida na metilação do DNA durante o período de implantação, enquanto que a DNMT3A é expressa em estágios mais tardios do desenvolvimento embrionário e

em células diferenciadas. Além disso, as DNMT3A e DNMT3B parecem participar também da manutenção das marcas de metilação, metilando sítios que escapam da metilação por DNMT1 (revisto por Auclair & Weber, 2012; Miller & Grant, 2012).

Outra proteína, a DNMT3L, é relacionada em sequência a DNMT3A e DNMT3B, porém não apresenta domínio catalítico. A DNMT3L é abundante durante o início da embriogênese e em células germinativas, e sua ausência leva a falha em se estabelecer a metilação do DNA em muitos alvos genômicos em células germinativas, incluindo regiões de controle de imprinting (ICRs) e muitos elementos de DNA repetitivos. O estudo bioquímico mais aprofundado mostrou que a DNMT3L é de fato um cofator que interage fisicamente com DNMT3A e DNMT3B estimulando a atividade dessas enzimas e as recrutando para a cromatina (revisto por Auclair & Weber, 2012).

Geralmente a metilação do DNA está associada à repressão da transcrição gênica e age através de dois mecanismos – bloqueando o acesso de proteínas de ligação ao DNA que agem como, ou recrutam, ativadores transcricionais, ou recrutando proteínas de ligação a metil (do inglês, *methyl-binding protein*, MBPs) que, por sua vez, recrutam complexos co-repressores transcricionais. Os ativadores e repressores da transcrição, por sua vez, recrutam enzimas modificadoras de histonas e remodeladores de cromatina, levando a mudanças no estado transcricional dos genes (revisto por Miller & Grant, 2012).

Em humanos, cerca de 3% das citosinas presentes no genoma são metiladas e a metilação ocorre quase que exclusivamente no contexto de dinucleotídeos CpG, embora possa ocorrer também em baixa abundância em dinucleotídeos CpA e trinucleotídeos CpNpG (revisto por Auclair & Weber, 2012; Miller & Grant, 2012). Os dinucleotídeos CpG são mais raros no genoma do que o esperado (1%) e a grande maioria desses dinucleotídeos (70 a 80%) encontra-se metilada (revisto por Miller & Grant, 2012).

No entanto, existe uma correlação inversa entre o nível de metilação e a densidade de CpGs – regiões pobres em CpGs, o que compreende a maioria do genoma humano, são altamente metiladas, enquanto que porções de DNA ricas em CpG permanecem praticamente demetiladas (revisto por Auclair & Weber, 2012). De fato, os CpGs não-metilados tendem a se agrupar em ilhas (CGI) – regiões de pelo menos 200pb com um conteúdo de GC maior que 55% e uma proporção entre o conteúdo de GC observado e conteúdo de GC esperado maior que 0.65 (Takai &

Jones, 2002; revisto por Miller & Grant, 2012). Desta forma, o padrão de metilação do genoma é basicamente bimodal em todas as células do organismo, com a grande maioria dos sítios CpG altamente metilados (>85%) e com CGIs altamente demetiladas (<10%). Entretanto, existem raras CGI metiladas em células somáticas. Algumas estão associadas com silenciamento gênico alelo-específico nos contextos de *imprinting* genômico parental e compensação de dosagem gênica através da XCI em células somáticas feminas (revisto por Auclair & Weber, 2012). Além disso, a limitação da expressão gênica em diferentes tipos de tecido parece ser causada pela metilação diferencial em *shores* de ilhas CpG, regiões de até 2kb ao redor das CGIs (revisto por Miller & Grant, 2012).

Aproximadamente 60% das regiões promotoras e reguladoras de genes humanos e primeiros éxons gênicos estão associados à CGIs. Essas CGIs frequentemente são hipometiladas e estão relacionadas a uma estrutura de cromatina mais permissiva que predispõe os genes à ativação da transcrição (revisto por Miller & Grant, 2012). Cerca de metade das CGIs não estão associadas a regiões promotoras e reguladoras de genes humanos e primeiros éxons gênicos. Essas CGIs restantes – denominadas CGI órfãs – estão localizadas de forma intra ou intergênica, onde a metilação do DNA parece ter também um papel importante. De fato, os perfis de metilação de DNA associados a tipos celulares diferentes variam mais frequentemente em seqüências intergênicas do que em regiões promotoras. É provável que esses sítios diferenciais de metilação regulem a atividade de regiões reguladoras distantes (*enhancers*) ou a transcrição de ncRNAs. Além disso, CGIs intragênicas localizadas em regiões codificadoras de genes são metiladas e frequentemente encontradas em genes de expressão alta e constitutiva. É possível que esses sítios de metilação intragênicos modulem sítios alternativos de início de transcrição e/ou sítios de *splicing*. Além disso, como o nível de metilação do DNA é geralmente maior em éxons do que em íntrons, pode ser que a metilação do DNA também tenha um papel no processo de definição das regiões exônicas (revisto por Auclair & Weber, 2012; Miller & Grant, 2012).

Os mecanismos envolvidos na remoção das marcas de metilação do DNA são diversos e controversos e podem depender (remoção passiva) ou não (remoção ativa) da síntese de DNA. A conversão de 5mC em 5-hidroxi-metil-citosina (5hmC) mediada por enzimas da família *ten-eleven translocation* (TET) de dioxigenases foi recentemente relacionada como um mecanismo de iniciação dos processos de

remoção passiva e ativa das marcas de metilação. Uma vez formada a 5hmC, as TET podem catalisar a oxidação das mesmas em 5-formil-citosina (5fC) e 5-carboxil-citosina (5caC), que também poderiam agir como intermediárias adicionais no processo de demetilação. É importante notar que a presença de 5hmC por si só tem um impacto funcional pois bloqueia a ligação de proteínas associadas ao DNA metilado. Por outro lado, as 5hmCs também recrutam seus próprios ligantes e demonstram um perfil de distribuição particular no genoma. Sendo assim, é provável que, além de exercer um importante papel como intermediário na demetilação do DNA, a 5hmC também sirva como marca epigenética associada a funções regulatórias próprias (revisto por Auclair & Weber, 2012; Shen & Zhang, 2013).

1.1.4. Papel dos RNA não-codificadores

Ao contrário do que por muito tempo se acreditou, estudos realizados na última década, incluindo os projetos ENCODE (*Encyclopedia of DNA Elements*) e FANTOM (*Functional Annotation of the Mammalian Genome*), demonstram que uma grande porção do genoma de eucariontes superiores é de fato transcrita (FANTOM Consortium *et al.*, 2005; ENCODE Project Consortium, 2007; ENCODE Project Consortium, 2012; Djebali *et al.*, 2012; Hangauer *et al.*, 2013). Enquanto apenas 1.2% do genoma humano codifica proteínas – 1.9% se considerarmos as regiões não-traduzidas de genes codificadores de mRNAs (International Human Genome Sequencing Consortium, 2004) – 74.7% a 85% do genoma humano produz transcritos primários (The ENCODE Project Consortium, 2012; Djebali *et al.*, 2012; Hangauer *et al.*, 2013) sendo que 61,2% do genoma produz transcritos que sofrem processamento pós-transcricional (Djebali *et al.*, 2012).

Os RNAs que não possuem nenhuma função de codificação de proteínas evidente são denominados RNAs não-codificadores (ncRNAs) e podem ser divididos em dois grandes grupos – *housekeeping* ncRNAs e ncRNAs regulatórios (Morey & Avner, 2004; revisto por Dhanasekaran *et al.*, 2012). Os *housekeeping* ncRNAs são expressos constitutivamente e estão envolvidos em processos como processamento de RNA, modificações de RNA, tradução, tráfego de proteínas e estabilidade genômica. Os principais tipos de *housekeeping* ncRNAs e suas

principais funções estão descritos na Tabela 3 (revisto por Dhanasekaran *et al.*, 2012).

Tabela 3. Principais Tipos de *Housekeeping* ncRNAs Encontrados em Mamíferos

Tipo de ncRNA	Principais Funções
RNAs ribossômicos (rRNAs)	Associam-se a proteínas para constituir as diferentes subunidades ribossômicas. Constituem os sítios ativos dos ribossomos e auxiliam a manter a fidelidade da tradução.
RNAs transportadores (tRNAs)	Envolvidos no processo de tradução, transportando aminoácidos até os ribossomos e participando na incorporação específica de aminoácidos de acordo com os códons presentes na molécula de mRNA.
Pequenos RNAs específicos de telômero (tel-sRNAs)	Associados às extremidades dos telômeros onde auxiliam no estabelecimento e manutenção da heterocromatina telomérica.
RNA de partícula reconhecadora de sinal (SRP RNA)	Fazem parte do complexo ribonucleoprotéico SRP que atua direcionando as proteínas celulares recém-sintetizadas ao retículo endoplasmático durante ou após sua tradução
Pequenos RNAs nucleares (snRNAs) <i>housekeeping</i>	Associam-se a proteínas e formam o centro catalítico de complexos ribonucleoprotéicos conhecidos como spliceossomo que atuam no processo de <i>splicing</i> .
Pequenos RNAs nucleolares (snoRNAs) <i>housekeeping</i>	Moléculas de RNA retidas no nucléolo que auxiliam na incorporação de modificações químicas como metilação e pseudouridilação em outras moléculas de RNA, como rRNAs, tRNAs, snRNAs, entre outras.

Os ncRNAs regulatórios geralmente possuem expressão tecido-específica durante diferentes estágios do desenvolvimento ou são expressos em resposta a estímulos externos e estão envolvidos em uma variedade de importante processos celulares, como regulação da expressão e função gênica, processamento e modificação de RNAs, estabilidade de mRNAs, tráfego de proteínas, degradação de proteínas, estabilidade genômica, memória epigenética e organização da cromatina no ambiente nuclear. Esses ncRNAs formam uma complexa rede de interações regulatórias dentro do ambiente celular, interagindo entre eles, com o genoma e com uma ampla quantidade de proteínas. Como as moléculas de ncRNA são fita simples, podem facilmente se dobrar em estruturas secundárias que podem ser reconhecidas por proteínas. Além disso, as moléculas de ncRNAs podem fazer

pareamento de bases com o DNA ou outras moléculas de RNA para reconhecimento de seus alvos. Essas características fazem dos RNAs excelentes candidatos como uma ponte entre o DNA e as proteínas que remodelam a cromatina, alterando assim o padrão de acessibilidade à cromatina e coordenando a expressão gênica (revisto por Kugel & Goodrich, 2012).

Os ncRNAs regulatórios podem ser classificados de acordo com o tamanho do transcrito – os ncRNAs regulatórios curtos (sncRNAs) são definidos como sendo transcritos de menos de 200pb enquanto que os ncRNAs regulatórios longos (lncRNAs) são definidos como sendo transcritos de mais de 200bp (Batista & Chang, 2012; Danasekaran *et al.*, 2013). Uma grande variedade de sncRNAs regulatórios de expressão espaço-temporal única pode ser encontrada em mamíferos. Os principais tipos de sncRNAs e suas respectivas funções estão descritos na Tabela 4 (Esteller, 2011; Wright & Bruford, 2011; Zhang *et al.*, 2012; Dhanasekaran *et al.*, 2012; Huang *et al.*, 2013).

Tabela 4. Principais Tipos de sncRNAs Encontrados em Mamíferos

Tipo de sncRNA	Principais Funções
Pequenos RNAs nucleares (snRNAs) regulatórios	Ligam-se a fatores de transcrição regulando a transcrição gênica.
Pequenos RNAs nucleolares (snoRNAs) regulatórios	Associam-se a regiões genômicas submetidas a <i>imprinting</i> , o que sugere um papel nesse processo epigenético. Além disso, podem regular <i>splicing</i> alternativo.
micro-RNAs (miRNAs)	Depois de processados, se associam a proteínas do complexo denominado RISC (<i>RNA-induced silencing complex</i>) que regulam a expressão gênica através de supressão translacional (bloqueando a tradução do mRNA pós-transcrição) e/ou degradação do mRNA.
Pequenos RNAs de interferência endógenos (esiRNAs)	Podem estar implicados na supressão de expressão de elementos genéticos móveis (transposons).
Y RNAs	Parte integrante de complexos denominados Ro RNP. Atuam durante a replicação do DNA, onde foram implicados no estabelecimento e alongamento das forquilhas de replicação, assegurando assim que a replicação aconteça de forma semi-conservativa em todo o genoma.
RNAs associados a sítios de início da transcrição (TSSa-RNAs)	Moléculas de RNA transcritas de maneira senso ou antisenso a partir de regiões que flanqueiam promotores de outros genes. Podem estar associados à manutenção da transcrição gênica.
Pequenos RNAs associados a promotores (PASRs)	Função ainda desconhecida.
Pequenos RNAs associados a terminadores (TASRs)	Função ainda desconhecida.
RNAs de iniciação da transcrição (tiRNAs)	Parecem estar envolvidos na regulação da transcrição gênica.
Vault RNAs (vRNAs)	Parte integrante dos complexos ribonucleoprotéicos <i>vault</i> implicados na resistência a drogas e transporte intracelular.
RNAs que interagem com Piwi (piRNAs)	Embora os piRNAs formem a maior classe de sncRNAs em vertebrados (aproximadamente 23000 piRNAs foram mapeados no genoma humano), a sua função ainda não está bem estabelecida. Sabe-se que os piRNAs formam complexos com proteínas da família Piwi, um subtipo de proteína Argonauta (componente do complexo RISC) e que estão envolvidos no silenciamento de retrotransposons.

Os lncRNAs regulatórios estão frequentemente localizados no núcleo e sua expressão parece ser mais específica em relação ao tipo celular do que a de genes que codificam proteínas (Ravasi *et al.*, 2006; Cabili *et al.*, 2011; Djebali *et al.*, 2012). Muitos tipos de lncRNAs podem ser encontrados em mamíferos e os principais

deles estão descritos na Tabela 5, juntamente com suas principais funções (revisto por Esteller, 2011; Zhang *et al.*, 2012; Dhanasekaran *et al.*, 2012).

Tabela 5. Principais Tipos de lncRNAs Encontrados em Mamíferos

Tipo de lncRNA	Principais Funções
Longos ncRNAs intergênicos (lincRNAs)	Um dos maiores tipos de lncRNA, são altamente conservados entre as espécies. Existem mais de 8000 lincRNAs que estão envolvidos em diversos processos biológicos, como regulação da transcrição gênica, reparo do DNA, transmissão de sinal, regulação do ciclo celular, manutenção da pluripotência em células-tronco e apoptose.
ncRNAs totalmente intrônicos (TINs)	Possuem expressão tecido-específica e estão envolvidos na regulação da transcrição dos genes a partir dos quais são transcritos. Agem através de interferência na atividade do promotor ou modulação da estrutura da cromatina.
ncRNAs transcritos a partir de regiões ultra-conservadas (T-UCRs)	Estão envolvidos na regulação da transcrição gênica e podem estar envolvidos também na regulação dos níveis de miRNA e mRNA.
Transcritos de regiões <i>upstream</i> a promotores (PROMPTs)	Produzidos a partir de genes ativos. Expressão desses ncRNAs está relacionada com modificações na estrutura da cromatina.
RNAs longos associados a promotores (PALRs)	Produzidos a partir de genes ativos. Expressão desses ncRNAs está relacionada com modificações na estrutura da cromatina.
lncRNAs ricos em repetições de trincas de polipurina (GRC-RNAs)	Componentes da matriz nuclear onde interagem com diversas proteínas associadas à matriz. Sua expressão parece diminuir com a diferenciação celular.
<i>Enhancer</i> RNAs (eRNAs)	Transcritos a partir de regiões reguladoras gênicas e sua expressão parece ser necessária para a ativação da transcrição de genes vizinhos, o que pode ocorrer através da estimulação da formação de uma estrutura cromatínica mais aberta ou através da promoção de comunicação entre as regiões reguladoras e promotoras dos genes.
mRNA- <i>like</i> ncRNAs (mlncRNAs)	Função ainda não é conhecida. Parecem ter a expressão finamente regulada em compartimentos subcelulares específicos em tecidos também específicos.

Outra importante função dos lncRNAs está relacionada à formação de domínios nucleares, regiões dentro do núcleo onde funções específicas são desempenhadas. Acredita-se que esses domínios sejam formados e dinamicamente regulados através das interações moleculares exercidas entre seus componentes, incluindo ncRNAs. De fato, a função de proteínas moduladoras da expressão gênica, como por exemplo Pc2, pode ser alterada de acordo com o tipo de ncRNA

que interage com ela. Pc2 metilada pode se ligar a ncRNAs presentes em domínios onde a expressão gênica é reprimida, fazendo com que Pc2 reconheça sinais de repressão da cromatina, enquanto que Pc2 demetilada se associa a ncRNAs presentes em domínios nucleares mais permissivos à expressão gênica que podem fazer com que Pc2 reconheça sinais de cromatina ativa. Dessa forma, as regiões promotoras de genes nos quais a proteína Pc2 encontra-se ligada pode mudar de localização dentro do núcleo, passando de uma região nuclear de maior atividade gênica para outra de menor atividade e vice-versa (revisto por Batista & Chang, 2013).

1.2. Inativação do Cromossomo X em Mamíferos

O primeiro indício da ocorrência de XCI em mamíferos registrado na literatura veio de um trabalho publicado por Barr & Bertram em 1949. Esses autores observaram a presença de uma estrutura nuclear esférica fortemente marcada por cristal violeta em células nervosas de gatas. Através de outras técnicas de coloração, os autores puderam notar ainda que essa estrutura continha grande quantidade de DNA (Barr & Bertram, 1951). Essas estruturas são hoje conhecidas como “Corpúsculos de Barr”, ou cromatina sexual.

Foi apenas com os trabalhos de Ohno e colaboradores (1959) e Ohno & Hauschka (1960), que a natureza dessas estruturas foi elucidada. Os autores interpretaram a presença desse corpúsculo em células diplóides femininas de ratos (Ohno *et al.*, 1959) e camundongos (Ohno & Hauschka, 1960) como sendo um dos cromossomos X. Esse cromossomo X apresentava-se condensado e heteropicnótico em contraste com o outro cromossomo X presente que apresentava-se descondensado e era mais similar aos autossomos. Ohno & Hauschka (1960) observaram ainda que em células tetraplóides, um desses corpúsculos poderia estar presente em células masculinas enquanto dois desses corpúsculos poderiam ser encontrados nas células femininas. Essa informação, juntamente com outros trabalhos que demonstraram haver múltiplos corpúsculos de Barr em humanos que apresentavam distúrbios intersexuais, forneceu evidências de que o número de corpúsculos de Barr presentes em uma célula está relacionado não só com o sexo da mesma, mas também com a sua ploidia (revisto por Harper, 2011).

A hipótese de inativação do cromossomo X foi apresentada primeiramente por Mary Frances Lyon em um breve artigo publicado em 1961 como forma de explicar a presença de um cromossomo X condensado em células femininas de camundongo. As observações sobre a XCI foram extendidas para outros mamíferos, incluindo humanos, em um segundo trabalho publicado pela autora (Lyon, 1962).

Para formular sua teoria, Mary Lyon se baseou em diversas observações além daquelas relacionadas ao corpúsculo de Barr, descritas acima. Outras evidências da ocorrência de XCI partiram de observações sobre fenótipos relacionados a anomalias no número de cromossomos sexuais em camundongos e humanos. Havia sido demonstrado, por exemplo, que camundongos X0 apresentavam fenótipo normal e eram férteis (Welshons & Russell, 1959). Em humanos, a mesma frequência de daltonismo observada em indivíduos do sexo masculino havia também sido observada em pacientes com Síndrome de Turner, que carregavam apenas um cromossomo X e não apresentavam cromatina sexual (Polani *et al.*, 1954; Polani *et al.*, 1956; Ford *et al.*, 1959). Indivíduos portadores da Síndrome de Klinefelter (47,XXY) apresentavam fenótipo masculino mesmo possuindo dois cromossomos X em suas células (Jacobs & Strong, 1959). Além disso, mulheres heterozigotas para mutações no cromossomo X eram apenas parcialmente afetadas e apresentavam um fenótipo em mosaico, como observado para características associadas a pele e aos olhos (Lyon, 1962). Outras evidências vieram de estudos descrevendo o efeito de mutações no cromossomo X de camundongos que demonstraram haver um fenótipo em mosaico relacionado a cor da pelagem em fêmeas heterozigotas, o que não era observado em machos ou fêmeas X0 (Dickie, 1954; Lyon, 1960; Lyon, 1961; Lyon, 1962).

Com base nas evidências descritas, Mary Lyon sugeriu que cada célula somática feminina de mamífero possui apenas um cromossomo X ativo (Xa) e que o corpúsculo de Barr corresponde a um cromossomo X inativo (Xi). Em mamíferos, a XCI ocorreria cedo na embriogênese e o Xi poderia ter tanto origem materna como paterna, o que seria determinado de forma aleatória. Entretanto, uma vez definido qual o Xi em uma determinada célula, todas as células descendentes apresentariam o mesmo Xi. Dessa forma as fêmeas seriam mosaicos em relação ao cromossomo X. Mary Lyon sugeriu ainda que apenas um cromossomo X permaneceria ativo por conjunto diplóide (2n), sendo que os demais seriam inativados, e que a XCI

funcionaria como um mecanismo de compensação da dosagem gênica entre os sexos feminino e masculino em mamíferos (Lyon, 1961; Lyon 1962).

1.3. Tipos de XCI

Embora a hipótese de Lyon tenha sugerido que a XCI seja aleatória, estudos posteriores demonstraram que a seleção do cromossomo X a ser inativado pode ser tanto aleatória como não-aleatória. Essa escolha é tecido-específica, com alguns tecidos apresentando seleção aleatória e outros dependente da origem parental do cromossomo X (imprintada). Em camundongos, a XCI é não-aleatória no trofoectoderma embrionário e endoderma primitivo, o que ocorre durante o desenvolvimento pré-implantacional. Nesses tecidos o cromossomo X de origem materna é ativo enquanto que o de origem paterna é silenciado. Em contraste, as células do epiblasto sofrem inativação aleatória (Takagi & Sasaki, 1975). A forma imprintada de XCI é observada também em tecidos extraembrionários e somáticos de marsupiais e no trofoectoderma embrionário e endoderma primitivo de bovinos, na qual o cromossomo X inativado é sempre de origem paterna (Sharman, 1971; Xue *et al.*, 2002). Em humanos, ao contrário do observado em camundongos, a inativação do X ocorre de forma aleatória tanto no trofoectoderma como em derivados do epiblasto (Ropers *et al.*, 1978; Harrison, 1989; Heard, 1997; Plath *et al.*, 2002; Moreira de Mello *et al.*, 2010).

1.4. Genes que Escapam à XCI

Em seu artigo publicado em 1962, Mary Lyon nota ainda que a presença de anormalidades fenotípicas em indivíduos com Síndrome de Turner, em contraste do observado para camundongos XO, sugere que alguns loci ligados ao cromossomo X devem escapar a inativação completa ou que essas anormalidades sejam determinadas após a XCI ocorrer. De fato, foi posteriormente demonstrado que algumas regiões do cromossomo X escapam ao processo inativação (Russell, 1963). Embora esse fenômeno possa ser observado tanto em camundongos como em humanos, o número de genes que escapam a inativação nesses dois grupos de

mamíferos é diferente. Estima-se que cerca de 3% dos genes ligados ao cromossomo X em camundongos escapam à XCI (Disteche *et al.*, 2002). Em humanos, esse número é bem maior. Cerca de 15% dos genes escapam à inativação e outros 10% parecem exibir um padrão variável de inativação, sendo bialélicamente expressos somente em um grupo de mulheres ou determinados tecidos (Carrel & Willard, 2005; Berletch *et al.*, 2010). Isso explicaria a diferença observada entre camundongos e humanos X0 (Disteche *et al.*, 2002).

Interessantemente, o número de genes ligados ao X que escapa ao processo de inativação e permanece ativa nos dois cromossomos X homólogos é equivalente ao número de genes encontrados no cromossomo Y. Esses genes compreendem os genes denominados “pseudo-autossômicos” (genes que possuem correspondentes homólogos tanto no cromossomo X como no cromossomo Y), genes localizados fora da região pseudoautossômica que possuem correspondentes funcionais no cromossomo Y e pequena parte dos genes exclusivos ao cromossomo X sem correspondentes funcionais no cromossomo Y. O escape à XCI desses genes contribui para garantir o balanço gênico entre os sexos (Gartler & Goldman, 2001; revisto por Payer & Lee, 2008).

Em humanos, genes exclusivos ao cromossomo X que escapam à XCI podem ser encontrados ao longo do cromossomo embora a maioria esteja agrupada no braço curto (Figura 3) (revisto por Berletch *et al.*, 2010). Duas regiões pseudoautossômicas podem ser encontradas nas extremidades dos braços curtos (PAR1) e longos (PAR2) dos cromossomos X e Y. Essas duas regiões pequenas (juntas compreendem apenas 4,6% do cromossomo Y) se pareiam e se recombinam durante a meiose de maneira similar aos cromossomos autossomos. PAR1 possui cerca de 2,7Mb e contém 24 genes e PAR 2 possui cerca de 0,33Mb e contém apenas cinco genes (revisto por Flaquer *et al.*, 2008; Otto *et al.*, 2011).

Em camundongos, os genes exclusivos ao cromossomo X que escapam à XCI estão distribuídos ao longo do cromossomo e não tendem a se agrupar (Figura 3) (revisto por Berletch *et al.*, 2010). Assim como nos demais mamíferos, uma única região pseudoautossômica (PAR, do inglês, *pseudoautosomal region*) é encontrada na região distal dos cromossomos X e Y de camundongos. Essa região é muito pequena, correspondendo a menos de 1% do cromossomo Y. Possui aproximadamente 700kb e parece conter apenas três genes (Perry *et al.*, 2001; Berletch *et al.*, 2010; Otto *et al.*, 2011; White *et al.*, 2012).

Figura 3. Distribuição dos genes que escapam à XCI (em laranja) e genes submetidos a inativação (azul) em humanos e camundongos. As posições das regiões pseudoautossômicas (PAR), centrômeros (cen, em violeta) e do gene Xist/XIST (em preto) estão indicadas. Alguns genes não estão representados. Como o centrômero do cromossomo X de camundongos é localizado em uma das extremidade, não há divisão em braço curto e braço longo. Modificado de Berletch et al., 2010.

1.5. O Centro de Inativação do Cromossomo X

O centro de inativação do cromossomo X (XIC, do inglês *X inactivation center*) é uma importante região envolvida na XCI, definida como sendo a região mínima necessária para desencadear esse processo (revisto por Augui *et al.*, 2011). Sua existência foi evidenciada através de estudos de translocações ocorridas entre o cromossomo X e autossomos em camundongos e humanos nos quais foi observado que certas porções do cromossomo X, quando translocadas para autossomos, levavam à inativação do autossomo que continha a translocação (Jacobs *et al.*, 1959; Russell, 1963; Russell, 1964; Lyon, 1971; Rastan, 1983; Rastan, 1985). A caracterização dos pontos de quebra do cromossomo X envolvidas

nessas translocações permitiu mapear o Xic/XIC na banda XD em camundongos (Rastan, 1983) e no braço longo do cromossomo X de humanos (Xq13) (Brown *et al.*, 1991). A organização genômica do centro de inativação do X em humanos e camundongos está representada na Figura 4 (Romito & Rougeulle, 2011).

Figura 4. Localização cromossômica e organização genômica do centro de inativação do cromossomo X (definido por um retângulo azul) em (A) humanos e (B) camundongos. Essa região contém tanto genes que codificam proteínas (em azul) como genes que codificam ncRNAs (em preto). Em camundongos, *Nap1/2* está representado em vermelho por estar contido dentro de *Ppnx*. As setas indicam a direção da transcrição gênica. Extraído de Romito e Rougeulle, 2011.

Estudos posteriores demonstraram que essa região contém diversos elementos importantes para o processo de XCI. Um dos genes presentes no Xic é o gene *Xist* (*X inactivation specific transcript*), que codifica um grande ncRNA de 15kb (17kb em humanos) expresso exclusivamente a partir do X escolhido para ser inativado (Borsani *et al.*, 1991; Brown *et al.*, 1991; Brockdorff *et al.*, 1991). Esse lncRNA age em *cis* e sua função principal, tanto em camundongos quanto em

humanos, é associar-se à cromatina do futuro Xi e à matriz nuclear que o rodeia a fim de recobrir o X a ser inativado. O acúmulo de *Xist* no futuro Xi desencadeia o estabelecimento de outras modificações epigenéticas que resultam na formação de heterocromatina característica do Xi e silenciamento gênico. Além disso, a expressão de *Xist* também induz a reorganização espacial do futuro Xi e cria um compartimento dentro do núcleo livre de fatores de transcrição (revisto por Erwin & Lee, 2008; Payer & Lee, 2008; Augui *et al.*, 2011).

Entretanto, o silenciamento do cromossomo X mediado pela expressão de *Xist* não pode ocorrer a qualquer momento da vida do organismo, sendo restrito a um período específico do desenvolvimento embrionário. Isso indica que a expressão de *Xist* por si só não é suficiente para promover o silenciamento do futuro Xi, sendo necessário um contexto celular apropriado no quais fatores epigenéticos regulatórios críticos estão presentes na cromatina do cromossomo X a ser inativado (revisto por Payer & Lee, 2008).

Em camundongos, um importante repressor de *Xist* é o seu gene antisense *Tsix* (*X inactivation specific transcript antisense*) (Debrand *et al.*, 1999; Lee *et al.*, 1999; Mise *et al.*, 1999). Esse gene é transcrito em um lncRNA antisense de aproximadamente 40Kb de sequência complementar ao *Xist* que ocupa todo o *locus* de *Xist*. A transcrição de *Tsix* através do *Xist* modula a estrutura da cromatina, o que determina se a expressão de *Xist* vai ser ativada ou reprimida. Assim, no futuro Xi, a inibição da expressão de *Tsix* permite o aumento da expressão de *Xist* e silenciamento do cromossomo em *cis*, enquanto que, no Xa, a persistência da expressão de *Tsix* previne o aumento de expressão do *Xist*, evitando que o cromossomo seja silenciado (revisto por Erwin & Lee, 2008; Payer & Lee, 2008; Augui *et al.*, 2011).

Embora desempenhe um importante papel na XCI em murinos, *Tsix* não parece ser conservado em humanos, pelo menos no que diz respeito à sua sequência. Um transcrito antisense ao *XIST* foi identificado em humanos mas esse transcrito não cobre todo o *locus* de *XIST*. Além disso, seu padrão de expressão difere do padrão observado para *Tsix* em camundongos, sendo que *TSIX* parece ser expresso a partir do Xi (Migeon *et al.*, 2001; Migeon *et al.*, 2002). Sendo assim, não se sabe ao certo se *TSIX* equivale ao *Tsix* murino e se desempenha alguma função na XCI em humanos (revisto por Payer & Lee, 2008; Romito & Rougeulle, 2011).

Outros fatores importantes para o processo de XCI também estão localizados no Xic, dentre eles o elemento de DNA *DXPas34*, a região Xpr (*X pairing region*, que sobrepõe, em parte o gene *Slc16A2*) e outros genes produtores de ncRNAs como *Jpx* (*Just proximal to Xist*) e *Xite* (*X-inactivation intergenic transcription element sequence*) (revisto por Payer & Lee, 2008; Romito & Rougeulle, 2011). As funções desses outros elementos serão descritas a seguir no contexto das etapas relacionadas à XCI.

1.6. Etapas da XCI

O processo de inativação aleatória do cromossomo X pode ser dividido em quatro fases distintas: (1) contagem do número de cromossomos X presentes na célula, (2) escolha do X que permanecerá ativo e do X que será inativado, (3) silenciamento do futuro X inativo e (3) manutenção dos estados de ativação e inativação (Pereira & Vasques, 2000; Payer & Lee, 2008).

1.6.1. Contagem do Número de Cromossomos X

A XCI se inicia cedo na embriogênese concomitantemente a diferenciação celular. Na primeira fase do processo de XCI, a proporção do número de cromossomos X em relação ao número de autossomos presentes na célula (X:A) é avaliada, de modo que a quantidade apropriada de cromossomos X (um por conjunto diplóide) possa ser inativada. Desta maneira, em fêmeas normais um dos cromossomos é silenciado e em machos normais nenhum cromossomo é destinado à sofrer inativação. Em células aneuplóides, todos os cromossomos X presentes, exceto um por conjunto diplóide, sofrem inativação. Já em células poliplóides, apenas um cromossomo X por conjunto diplóide permanece ativo (revisto por Payer & Lee, 2008; Barakat & Gribnau, 2010; Augui *et al.*, 2011; Yang *et al.*, 2011).

Diferentes modelos foram propostos para explicar o processo de contagem do número de cromossomos X, embora seja também possível que esses modelos sejam interligados. No modelo do “fator bloqueador”, um complexo formado por componentes provenientes de autossomos e possivelmente também do

cromossomo X se liga preferencialmente ao Xic do futuro Xa, inibindo a expressão de *Xist* (Nicodemi & Prisco, 2007; revisto por Erwin & Lee, 2008; Payer & Lee, 2008). Em um genoma diploide haveria apenas quantidade suficiente de fator bloqueador para prevenir a XCI em um dos cromossomos X presentes. Foi sugerido que essa quebra de simetria seria responsável pela determinação dos futuros Xa e Xi. Fatores de pluripotência, como Oct3/4, Sox2 e Nanog, agem como fatores bloqueadores e se ligando ao Xic do cromossomo X que permanecerá ativo inibindo a expressão de *Xist* (Navarro *et al.*, 2008).

Além dos fatores bloqueadores, acreditava-se que pudesse existir um fator de competência ligado ao X que, quando produzido por mais de um cromossomo X por conjunto diplóide, causaria ativação da expressão do gene *Xist* pelo futuro Xi (Monkhorst *et al.*, 2008; Payer & Lee, 2008). De fato, Rnf12, uma ubiquitina E3-ligase cujo gene está presente no cromossomo X, foi identificada como sendo um importante fator de competência para a inativação do cromossomo X, induzindo a expressão de *Xist* pelo futuro Xi de maneira dose dependente (Jonkers *et al.*, 2009; Barakat *et al.*, 2011). Rnf12 tem como alvo o fator de pluripotência Rex1 que, por sua vez, se liga as regiões reguladoras tanto de *Xist* como de *Tsix* e inibe a XCI. Foi demonstrado que Rnf12 causa a ubiquitinação e degradação de Rex1 via proteossomo de maneira dose dependente (Gontan *et al.*, 2012).

O “modelo estocástico” sugere que cada cromossomo X teria uma probabilidade independente de iniciar a XCI. A concentração de fatores autossômicos inibidores da XCI durante a diferenciação celular determinaria um limite a ser ultrapassado pela concentração de fatores ativadores da XCI ligados ao cromossomo X como *Xist*) em cada cromossomo X. Dependendo da distribuição aleatória dos fatores inibidores pelos diferentes cromossomos X, células femininas diploides, por exemplo, poderiam ter nenhum, um ou seus dois cromossomos X inativados. Por seleção, somente as células que apresentam número adequado de cromossomos X inativos sobreviveriam (revisto por Barakat & Gribnau, 2010).

Outro modelo de contagem proposto baseia-se no pareamento dos cromossomos X presentes na célula. Essa hipótese foi levantada a partir da constatação de que, antes do início do processo de XCI, pode-se observar co-localização dos Xic presentes em ambos os cromossomos X no núcleo celular de células femininas, o que provavelmente facilita a troca de informação entre esses

cromossomos, importante para determinar seus futuros estados ativo ou inativo (Xu *et al.*, 2006; Augui *et al.*, 2007; revisto por Barakat & Gribnau, 2010).

A análise de camundongos transgênicos e *knockouts* demonstrou que o elemento de DNA *DXPas34* e regiões dos genes *Tsix* e *Xite* são componentes do cromossomo X importantes para a determinação do número de cromossomos X na célula já que quando essas regiões são deletadas, números anormais de Xi são observados. De fato, foi demonstrado que essas regiões e também a região *Xpr* contém elementos de DNA importantes para o pareamento e comunicação entre os cromossomos X. Além disso, foi também demonstrado que o fator de transcrição e insulador de cromatina *Ctcf* (*CCCTC-binding factor*) é requerido para o pareamento dos cromossomos X em regiões próximas a *DXPas34* e aos genes *Tsix* e *Xite* (Xu *et al.*, 2007). De fato, o *knockdown* de *Ctcf* abole o processo de pareamento (revisto por Payer & Lee, 2008). Notou-se que também que a depleção de Oct3/4 bloqueia o pareamento. Oct3/4 se liga diretamente aos genes *Tsix* e *Xite* onde interage também com *Ctcf* e seu co-fator e fator de transcrição *Yy1* (*Yin Yang 1*), embora *Yy1* pareça ser dispensável para o pareamento (Xu *et al.*, 2007; Donohoe *et al.*, 2009).

1.6.2. Escolha dos Futuros Xa e Xi

Após a contagem do número de cromossomos X presentes na célula, um cromossomo X é selecionado para permanecer ativo enquanto o outro (caso mais de um cromossomo X por conjunto diplóide esteja presente) é escolhido para ser inativado. A escolha do futuro Xi está intimamente relacionada ao processo de contagem e depende principalmente dos fatores que regulam a expressão alelo-específica do gene *Xist* (revisto por Payer & Lee, 2008).

No futuro X inativo, *Tsix* não é expresso permitindo o aumento da expressão de *Xist* e silenciamento do cromossomo. Já no X ativo *Tsix* é expresso e reprime a expressão de *Xist*, evitando que o cromossomo seja silenciado. *Tsix*, por sua vez, é regulado através de regiões não-codificadoras do gene *Xite* e pelo elemento de DNA *DXPas34* que funcionam como *enhancers* de sua expressão (Figura 5) (revisto por Erwin & Lee, 2008).

Além do efeito potencializador exercido por DXPas34, esse elemento participa da escolha do Xa através de sua associação ao fator de transcrição e insulador da cromatina Ctfc e seu co-fator Yy1. O complexo DXPas34, Ctfc e Yy1 no futuro Xa age como ativador transcricional de *Tsix* e também bloqueia o acesso do gene *Xist* aos seus *enhancers* (revisto por Payer & Lee, 2008). De maneira contrária, Ctfc parece agir no futuro Xi reprimindo a expressão de *Xist* (Figura 5) (Sun *et al.*, 2013).

Outra interação observada no Xi ocorre entre o gene *Xist* e seu vizinho não-codificador *Jpx* (*Just proximal to Xist*) em momentos nos quais o gene *Xist* é induzido à transcrição ou sua expressão é aumentada, de modo que *Jpx* atua como um regulador positivo ativando a expressão de *Xist* (Tian *et al.*, 2010; Sun *et al.*, 2013). Foi demonstrado que o ncRNA *Jpx* age no futuro Xi onde se liga a proteína Ctfc retirando-a do promotor de *Xist* e estimulando sua transcrição (Figura 5) (Sun *et al.*, 2013).

Figura 5. Fatores que regulam a expressão de *Tsix* e *Xist* no futuro cromossomo X inativo (Xi) e no cromossomo X que permanecerá ativo (Xa). No cromossomo Xa, regiões não codificadoras do gene *Xite* e o elemento de DNA DXPas34 estimulam a expressão de *Tsix*. Ctfc e seu co-fator Yy1 ligados a DXPas34 ativam a expressão de *Tsix* e inibem a expressão de *Xist*. Por sua vez, a expressão de *Tsix* reprime a expressão de *Xist*. No Xi, *Tsix* não é expresso. *Jpx* age retirando Ctfc da região promotora de *Xist* estimulando assim sua expressão.

1.6.3. Silenciamento do Cromossomo X

Uma vez selecionado o cromossomo X a ser inativado, a expressão do *Xist* no futuro X inativo (Xi) é aumentada enquanto que o outro alelo do gene, presente no cromossomo X escolhido para permanecer ativo, sofre metilação de seu

promotor e é silenciado. Dois dias após o início da diferenciação celular em embriões murinos, uma grande quantidade de RNA *Xist* já pode ser observada rodeando o futuro Xi. O recobrimento do futuro Xi pelo RNA de *Xist* leva a uma rápida deprivação de RNA polimerase II e fatores de transcrição, o que resulta em uma redução imediata da transcrição gênica nesse cromossomo. Além disso, a presença do RNA de *Xist* leva ao desencadeamento de uma cascata de modificações em histonas que, por sua vez, culminam no silenciamento cromossômico, formação do estado de heterocromatina facultativa no Xi e criação de um compartimento nuclear silenciado (observado no núcleo de células interfásicas como corpúsculo de Barr) (revisto por Hall *et al.*, 2008; Barakat & Gribnau, 2010) (Figura 6).

Pouco antes da ativação da transcrição, a região promotora do gene *Xist* assume um estado heterocromático, evidenciado através de análise alelo-específica pela presença de histonas H3K27 trimetiladas na região promotora do *Xist* presente no futuro Xi. Por outro lado, no cromossomo X que permanecerá ativo, a transcrição persistente de *Tsix* através da região promotora de *Xist* gera uma configuração eucromática que está relacionada à repressão do gene *Xist* (revisto por Payer & Lee, 2008).

Os complexos *polycomb* (*polycomb repressive complex 1*) e PRC2 (*polycomb repressive complex 2*) são elementos chave no processo de silenciamento do cromossomo X. O complexo PRC2 estabelece a marcação característica do Xi por H3K27me3 e o complexo PRC1 promove monoubiquitinação de H2AK119. Outras modificações epigenéticas podem ser observadas no Xi durante o processo de silenciamento e incluem presença de histona H3K9me2 e H4K20me1 e exclusão de marcadores de cromatina ativa como H3K4me1, H3K4me2, H3K9ac e H4ac. Mais tarde, ocorre também incorporação de macroH2A1 e metilação de promotores de genes ligados ao X (revisto por Payer & Lee, 2008; Barakat & Gribnau, 2010). A DNA metiltransferase DNMT3b parece desempenhar um papel importante na metilação durante o estabelecimento da XCI (Gendrel *et al.*, 2012) e complexo ubiquitina E3-ligase formado pelas proteínas SPOP e Cullin3 estão envolvidos na deposição de macroH2A1 no futuro Xi. Já o cromossomo Xi humano parece estar organizado em dois tipos diferentes de heterocromatina facultativa, um caracterizado pela presença de RNA de *XIST* e incorporação de H3K27me3 e

macroH2A1, e outro caracterizado pela associação de HP1, H3K9me3 e H4K20me3 (revisto por Payer & Lee, 2008; Barakat & Gribnau, 2010) (Figura 6).

Figura 6. Modificações epigenéticas no cromossomo X inativo ao longo da diferenciação celular em camundongos. A XCI se inicia com a diferenciação de células-tronco embrionárias femininas. A primeira mudança no futuro Xi é o espalhamento de RNA Xist (vermelho) que é seguida de exclusão de RNA polimerase II (cinza). Logo após ocorre perda de marcas de cromatina ativa (azul) e ganho de marcas de cromatina inativa (violeta). Essas mudanças na cromatina são acompanhadas por acúmulo de proteínas (verde) e são seguidas por outras modificações (amarelo). Modificado de Barakat & Gribnau, 2010.

Os genes que escapam à inativação estão localizados fora do compartimento silenciador gerado pelo recobrimento da cromatina pelo RNA Xist. Tais genes estão frequentemente separados dos genes inativados adjacentes por domínios de cromatina. O fator de transcrição e insulador de cromatina Ctfc provavelmente desempenha um papel importante em estabelecer esses diferentes domínios cromatínicos, prevenindo a metilação de ilhas CpG presentes em genes que escapam a inativação (revisto por Berletch *et al.*, 2010). Além disso, foi sugerido a

presença de sequências LINE-1 facilita o ancoramento e espalhamento do RNA de *XIST* pelo cromossomo (Lyon, 1998). Em humanos (mas não em camundongos) foi observado que o número de sequências LINE-1 no cromossomo X é maior do que nos autossomos. Genes destinados a inativação contém grande número de repetições LINE-1 enquanto que a densidade dessas repetições em genes que escapam a inativação é bem mais baixa (Lyon, 1998; Bailey *et al.*, 2000; Ross *et al.*, 2005; Carrel *et al.*, 2006). Em camundongos, um outro tipo de repetição, as LTRs (*long terminal repeats*) parece ocorrer com menos frequência em genes que escapam à XCI (Tsuchiya *et al.*, 2004).

Embora muito já se conheça sobre o mecanismo de estabelecimento da XCI em murinos, pouco se sabe sobre esse mecanismo em células humanas. Isso se deve ao fato de que muitas das observações relativas à esse processo em camundongos vieram de estudos com células tronco embrionárias (ESCs), um modelo celular de fácil obtenção que permitia manipulação *in vitro* no laboratório. Já em humanos o estudo da inativação do X só era possível em células masculinas de carcinoma embrionário com resultados de cariótipo anormais, híbridos de células somáticas camundongo/humano induzidos a expressar *XIST*, ou em células somáticas transgênicas de linhagem tumoral. Embora o isolamento das primeiras linhagens de células tronco embrionárias humanas (hESCs) em laboratório (Thomson *et al.*, 1998) tenha criado grande expectativa de que os diversos aspectos do mecanismo de estabelecimento da XCI em células humanas pudessem finalmente ser elucidados (Hall *et al.*, 2008), a grande maioria das linhagens de hESC analisadas até o momento demonstrou já apresentar um certo grau de inativação de um dos cromossomos X mesmo em células femininas nunca induzidas à diferenciação (Dvash & Tan, 2009; Shen *et al.*, 2008; Silva *et al.*, 2008; Hall *et al.*, 2008; Hoffman *et al.*, 2005). Sendo assim, as linhagens de hESCs hoje disponíveis não constituem um modelo adequado para o estudo do processo de estabelecimento da XCI em humanos.

1.6.4. Manutenção do Estado de Inativação

Uma vez que o silenciamento de um cromossomo X é estabelecido em uma determinada célula, o mesmo cromossomo permanece estavelmente silenciado em

todas as células descendentes e por toda a vida do organismo (com exceção da linhagem germinativa, onde ocorre reativação do Xi) (Heard, 1997; Plath *et al.*, 2002). Esse fenômeno exige uma maquinaria celular capaz de manter as modificações epigenéticas adquiridas durante o estabelecimento da XCI. Durante a fase de manutenção da XCI, uma variedade de fatores modificadores de histonas age em conjunto para manter o Xi em seu estado reprimido. De fato, o Xi continua a apresentar modificações de histonas que são características de repressão da transcrição gênica, como hipoacetilação de histonas H3 e H4, ausência de H3K4me2 e H3K4me3, e presença de macroH2A1, H3K9me2, H3K9me3, H3K27me3, H4K20me3, e H2AK119ub1, entre outras (Plath *et al.*, 2003; Chadwick & Willard, 2004; Heard, 2004; Heard, 2005; Brinkman *et al.*, 2006; Chadwick, 2007; Payer & Lee, 2008; Chow & Heard, 2010).

Em camungos, o gene *Xist* continua a ser expresso no Xi o que deve contribuir de alguma forma para a manutenção do seu estado inativo, uma vez que na ausência de *Xist* alguns dos genes inativados sofrem reativação (Csankovszki *et al.*, 2001; Zhang *et al.*, 2007). Entretanto, a expressão de *Xist* pelo Xi aparentemente não é um evento crucial para a manutenção do silenciamento cromossômico já que a deleção do gene *Xist* após ocorrência da XCI não resulta em reativação global dos genes antes silenciados (Csankovszki *et al.*, 1999; Csankovszki *et al.*, 2001). Enquanto a expressão *Xist* é mantida no Xi durante a manutenção do silenciamento, a expressão de *Tsix* cessa no cromossomo ativo, não sendo portanto necessária para manter o gene *Xist* presente no Xa em estado reprimido (revisto por Erwin & Lee, 2008).

Em humanos, o gene *XIST* também continua a ser expressado no Xi durante a fase de manutenção da XCI. Entretanto, essa expressão parece ser dispensável para manter o Xi em seu estado silenciado. De fato, a deleção de *XIC* em híbridos de células somáticas não resulta em reativação dos genes silenciados presentes no Xi (Brown & Willard, 1999).

A metilação do DNA é um fator chave para estabilização da inativação do X em camundongos. A deleção do gene *Dnmt1* (*DNA methyltransferase 1*), envolvido na manutenção da metilação do DNA, resulta em reativação do cromossomo X antes silenciado (Sado *et al.*, 2000). No cromossomo Xa, a manutenção do silenciamento do gene *Xist* também é dependente de metilação e é parcialmente mediada pela proteína de ligação ao DNA metilado Mbd2, que age através do

recrutamento de histona-desacetilases (Barr *et al.*, 2007; revisto por Payer & Lee, 2008). A ausência de metilação na região promotora do gene *Xist* presente no Xa durante a fase de manutenção leva à reversão do estado reprimido do gene e inativação inapropriada do Xa (Beard *et al.*, 1995; Panning & Jaenisch, 1996).

Em humanos, o papel da metilação do DNA na manutenção da XCI ainda não está muito claro. O tratamento de células de uma linhagem de fibroblastos dérmicos primários humanos com 5-aza-2'-deoxicitidina – um potente inibidor de DNMT1 – não é suficiente para promover a reativação dos genes submetidos à XCI (Wolf & Migeon, 1982), embora em uma linhagem masculina de células transformadas esse tratamento resulte na expressão parcial de *XIST* a partir do único cromossomo X presente nessas células (Vasques *et al.*, 2005). Além disso, a redução de 97% dos níveis de metilação global observada em uma linhagem de células de carcinoma embrionário duplo-*knockout* para DNMT1 e DNMT3 (*DNA methyltransferase 3*) (Rhee *et al.*, 2002) não é suficiente para promover a expressão de *XIST* a partir do único cromossomo X presente. Entretanto, uma vez que esta linhagem também é masculina, o efeito da falta de metilação do DNA em genes submetidos a XCI não pode ser avaliado nessas células (Vasquez *et al.*, 2005).

A proteína SmcHD1 (*SMC hinge domain containing 1*) também desempenha papel importante na manutenção da XCI, já que em embriões murinos mutantes para essa proteína as ilhas CpG de genes submetidos a XCI apresentam-se hipometiladas e o estado reprimido desses genes é perturbado (Blewitt *et al.*, 2008). De fato, foi demonstrado que essa proteína desempenha um papel importante na metilação do DNA durante o processo de manutenção da XCI (Gendrel *et al.*, 2012). Em humanos, foi demonstrado que o *knockdown* de SMCHD1 resulta em uma estrutura cromatínica menos condensada no Xi e na reorganização da cromatina enriquecida em H3K9me3 durante a intérfase (Noozawa *et al.*, 2013).

Em camundongos, a proteína Atrx (*α -thalassemia mental retardation X-linked protein*) também desempenha um papel na manutenção da XCI. Atrx está associada ao longo de todo Xi onde sua deposição aumenta durante estágios mais avançados de diferenciação celular, posteriores ao estabelecimento da XCI. Além disso a deposição de Atrx é enriquecida em regiões contendo PTMs de histonas características do Xi como, por exemplo, macroH2A (Baumann & De La Fuente, 2009; De La Fuente *et al.*, 2011).

Em conclusão, parece claro que as múltiplas marcas de repressão presentes no Xi trabalham em sinergia para garantir a manutenção apropriada da XCI. De fato, foi demonstrado em camundongos que a expressão de *Xist*, o padrão de metilação em ilhas CpG e a hipoacetilação de histonas são fatores que interagem sinergisticamente para manter o Xi em seu estado silenciado (Csankovszki *et al.*, 2001). Além dos fatores descritos anteriormente, é possível que outros componentes ainda não identificados também apresentem um papel importante nesse processo. Triagens genômicas funcionais em células somáticas humanas femininas poderiam contribuir na identificação de tais componentes.

1.7. Uso de Células *HPRT*^{+/-} para o Estudo da Manutenção da XCI

O gene *HPRT* (MIM *308000) localizado no locus Xq26-q27.2, codifica a proteína hipoxantina guanina fosforibosiltransferase 1 (HPRT), uma enzima que cataliza a conversão de hipoxantina para inosina monofosfato e de guanina para guanosina monofosfato através da transferência do grupo 5-fosforibosil presente na molécula de 5-fosforibosil-1-pirofosfato para um de seus substratos. Essa enzima apresenta papel importante na síntese de purinas por recuperação das mesmas a partir de moléculas de ácidos nucleicos submetidas à degradação celular (Figura 7) (Stout & Caskey, 1985). A deficiência de HPRT leva à perda da capacidade de reaproveitamento de purinas pelas células e em humanos é responsável pela Síndrome de Lesch-Nyhan (MIM #300322).

Quando células completamente deficientes em HPRT são impedidas de realizar a síntese *de novo* de nucleotídeos de purina (Figura 7) – através da inibição de uma outra via metabólica que envolve a participação de folato e da enzima diidrofolato redutase - essas células senescem e morrem. Já as células *HPRT*⁺ são capazes de sobreviver em cultura devido à capacidade de reaproveitamento de purinas presentes no ambiente intracelular (Figura 7), embora sua taxa de divisão celular nessas condições esteja levemente diminuída. Desta forma, células *HPRT*⁺ podem ser facilmente selecionadas em cultura quando expostas ao meio HAT, que contém em sua formulação aminopterina (potente inibidor da enzima diidrofolato redutase), hipoxantina e timidina (intermediários na síntese de DNA). Por outro lado, células *HPRT*⁺ são incapazes de sobreviver em cultura quando 6-tioguanina (6TG)

ou 8-azaguanina (8AZA) são adicionadas ao meio de cultura devido à incorporação de metabólitos tóxicos - produzidos somente na presença de HPRT - ao DNA ou RNA, respectivamente, enquanto que a falta de HPRT permite a sobrevivência da célula nessas condições (Figura 7) (Fujimoto *et al.*, 1971; Watson *et al.*, 1988).

Figura 7. Síntese *de novo* de purinas e papel de HPRT na via de recuperação de purinas. O uso de aminopterina no meio de cultura afeta os passos indicados no esquema pois a aminopterina bloqueia a enzima diidrofolato redutase envolvida na síntese de folatos. As análogos de purina 6-tioguanina (6TG) e 8-azaguanina (8AZAG) são substratos alternativos de HPRT cuja incorporação em ácidos nucleicos possui efeito tóxico para a célula. PRPP=5-fosforibosil-1-pirofosfato, IMP= inosina monofosfato, GMP=guanósina monofosfato, AMP=adenosina monofosfato. Modificado de <http://hemedicalbiochemistrypage.org/nucleotide-metabolism.php#>.

Por esses motivos, a heterozigose para mutação no gene *HPRT* em células femininas, tem sido utilizada na literatura como ferramenta para a detecção da ocorrência de reativação do cromossomo X inativo. Normalmente, quando o gene *HPRT* mutado está localizado no cromossomo X ativo e o gene selvagem no X inativo, as células são capazes de sobreviver em meio de cultura contendo 6TG ou 8AZAG, mas são incapazes de sobreviver em meio HAT. Se ocorrer indução da reativação do cromossomo X, tais células tornam-se capazes de sobreviver em meio HAT (e incapazes de sobreviver em meio contendo 6TG ou 8AZAG), permitindo a seleção da população de células na qual esse *locus* do cromossomo X foi reativado com sucesso (Figura 8) (Migeon, 1970; Vasques & Pereira, 2001; Stabellini *et al.*, 2009; Mekhoubad *et al.*, 2012).

Figura 8. Monitoramento da atividade do Xi através do uso de meio de cultura seletivo. Uma população mista de células 46,XX *HPRT*^{+/-}, *HPRT*⁺ ou *HPRT*⁻ submetida ao tratamento em meio 6-tioguanina ou 8-azaguanina passa a apresentar apenas células *HPRT*⁻. Quando essas células perdem a capacidade de manutenção da XCI, o alelo *HPRT*⁺ é reativado possibilitando a sobrevivências dessas células em meio de cultura contendo HAT. Os cromossomos X ativos (Xa) estão representados em vermelho e os inativos (Xi) em cinza. O alelo *HPRT* selvagem está representado em letras maiúsculas e o alelo *HPRT* mutado em letras minúsculas.

1.8. Uso de Bibliotecas de shRNA para Triagem Genômica Funcional

Um dos meios de se estudar a função de um determinado gene em uma célula, tecido ou organismo é através da comparação entre sistemas em que produto gênico é expresso com sistemas onde o produto gênico foi depletado. A depleção do produto de um gene pode ser adquirida de diversas maneiras como, por exemplo, geração de *knockouts*, *knockins* ou ainda através de um processo denominado *knockdown*. Esse último consiste na diminuição da expressão de um determinado gene e pode ser obtido através do tratamento de células ou tecidos em cultura, ou até mesmo *in vivo*, de interferência de RNA (RNAi) através da utilização, por exemplo, de pequenos RNAs de interferência (siRNA) e *short-hairpin* RNAs (shRNA), específicos para o produto gênico em questão. Uma das grandes vantagens do método *knockdown* por RNAi é a possibilidade de se alterar a expressão de um determinado gene com alta especificidade sem que seja necessário utilizar técnicas de engenharia genética para alterá-lo fisicamente (como ocorre em métodos de *knockout* ou *knockins* que envolvem a construção de vetores para inserção da sequência gênica alterada nas células-alvo), o que torna a técnica mais fácil. Isso é possível porque esses tipos de RNA funcionam como RNAs antisense que reconhecem, se ligam e desencadeiam uma reação que leva a degradação ou inibem a tradução do mRNA alvo (revisto por González-González *et al.*, 2008; Rao *et al.*, 2009).

Os siRNAs são moléculas de RNA dupla-fita (dsRNA) de 21pb a 23pb gerados por síntese química, clivagem enzimática ou sistemas de expressão. Esses oligonucleotídeos são geralmente transfectados nas células-alvo onde demonstram um efeito transiente. Os shRNAs são dsRNAs que geralmente mimetizam pré-micro RNAs endógenos (pré-miRNA). Os shRNAs consistem em duas sequências palindrômicas de 19 a 29 nucleotídeos conectadas por uma pequena haste (*hairpin*) de RNA fita simples em uma das extremidades. Esses shRNAs são incorporados em vetores que podem ser transfectados ou transduzidos nas células-alvo e, portando, possuem um efeito mais duradouro. A enzima RNase III denominada Dicer se liga e cliva tanto os pré-miRNAs como os shRNAs em sua forma madura (miRNAs) de 21pb a 25bp. Uma das fitas desses miRNAs ou de siRNAs é incorporada no complexo e RISC (*RNA-interfering silencing*) que então identifica, se liga e cliva o RNAm correspondente ou reprime sua tradução. A clivagem do mRNA

alvo ocorre em um único sítio localizado a aproximadamente dez nucleotídeos a partir da região 5' da sequência antisense do siRNA. Estudos demonstram que os shRNAs são indutores mais potentes de interferência mediada por RNA do que os siRNA (revisto por González-González *et al.*, 2008; Rao *et al.*, 2009).

Hoje em dia, estão disponíveis no mercado bibliotecas de siRNA ou shRNA que podem ser utilizadas para triagens genômicas funcionais. Nessa estratégia, genes são silenciados aleatoriamente em busca de um determinado fenótipo. Uma vez obtido/selecionado o fenótipo de interesse, o gene silenciado, envolvido no fenótipo, é identificado (Sims *et al.*, 2011; Hu & Luo, 2012). O método permite a identificação de genes envolvidos em diferentes fenômenos biológicos sem o conhecimento prévio de seus mecanismos.

O método de triagem funcional por bibliotecas de RNAi mostrou-se efetivo na identificação de genes envolvidos em diversos mecanismos celulares e vias bioquímicas. A título de exemplificação, o método teve sucesso em estudos epigenéticos relacionados ao Ras (Gazin *et al.*, 2007), identificação de alvos moleculares para câncer (Ngo *et al.*, 2006), identificação de supressores tumorais (Westbrook *et al.*, 2005), identificação de genes envolvidos na proliferação de células normais do epitélio mamário humano, células cancerosas de linhagens de câncer de cólon e câncer de mama (Silva *et al.*, 2008; Schlabach *et al.*, 2008), entre outros.

Experimentos com bibliotecas de shRNAs em células somáticas femininas poderiam levar também a identificação de genes necessários para a manutenção do estado de inativação, uma vez que, com o *knockdown* desses genes, esperar-se-ia observar reativação do cromossomo X que antes se encontrava inativo, gerando assim, células nas quais os dois cromossomos X presentes apresentariam atividade transcricional.

Sendo assim, o trabalho descrito nessa tese buscou empregar uma biblioteca lentiviral de shRNAs de ampla cobertura genômica como ferramenta para a triagem funcional de genes que pudessem estar envolvidos na regulação da manutenção da XCI em células humanas primárias. Como necessitávamos de um modelo celular que nos permitisse selecionar as células femininas em cultura de acordo com o cromossomo X ativo presente e que também nos permitisse selecionar as células que sofreram reativação do cromossomo X após transdução com a biblioteca de shRNAs, optamos por trabalhar com fibroblastos dérmicos humanos 46,XX

heterozigotos para uma mutação de perda de função no gene *HPRT*, já que essas células podem ser facilmente selecionadas, de acordo com o alelo *HPRT* presente no X ativo, em meio contendo HAT (células *HPRT*⁺) ou 6TG (células *HPRT*⁻).

Uma população de células *HPRT*^{+/-}, *HPRT*⁻ foi transduzida pela biblioteca de shRNAs e subsequentemente dividida em duas amostras: uma amostra controle que foi mantida em meio contendo 6TG para a seleção positiva de células *HPRT*⁻ e e uma amostra experimental que foi mantida em meio HAT para a seleção positiva de células *HPRT*⁺. Dessa forma, era esperado que, células que recebessem qualquer tipo de shRNA que tivesse como alvo um gene de expressão importante para a manutenção do estado de inativação do cromossomo X (cujo *knockdown* acarretaria em reativação do cromossomo X contendo o alelo *HPRT*⁺ antes inativado), seriam depletadas de nossa amostra controle mantida em meio 6TG, e enriquecidas em nossa amostra experimental tratada em meio HAT. Seguindo o mesmo raciocínio, esperava-se que as células que não receberam nenhum shRNAs assim como células cujos shRNAs recebidos não eram capazes de resultar na reativação do cromossomo X (células *HPRT*⁻), seriam depletadas da amostra selecionada em HAT e enriquecidas na amostra mantida em 6TG. Após tratamento com as drogas seletivas, foi aplicado sequenciamento de próxima geração (NGS) para determinar a identidade e nível de representatividade de cada shRNA presente em cada uma de nossas amostras (Figura 9).

Caso o envolvimento dos candidatos identificados nesse trabalho no processo de manutenção da XCI seja confirmado em futuros experimentos de validação, novos componentes envolvidos na manutenção da XCI em humanos serão identificados.

Figura 9. Esquematização das etapas experimentais desenvolvidas neste trabalho para a triagem funcional de genes envolvidos na manutenção da XCI em humanos. Fibroblastos primários 46,XX HPRT^{+/-}, HPRT⁻ foram transduzidos com uma biblioteca lentiviral de shRNAs. Após 72h, a população de células transduzidas foi dividida em duas populações: a amostra controle mantida em meio 6TG para seleção de células HPRT⁻ e a amostra experimental mantida em HAT para seleção de células HPRT⁺ (células com reativação do Xi). O DNA genômico foi coletado e submetido ao sequenciamento de próxima geração (NGS) e análise computacional para identificação e contagem de cada shRNA presente em cada amostra. Os dados obtidos foram comparados entre as amostras para determinação dos candidatos.

Objetivos

2. Objetivos

O presente trabalho teve como principal objetivo identificar genes envolvidos na manutenção do estado de inativação do cromossomo X em humanos. Para isso, foi desenvolvida uma estratégia para a triagem genômica funcional desses genes que envolveu a uso de uma biblioteca lentiviral de shRNAs em fibroblastos primários humanos 46,XX heterozigotos para uma mutação no gene *HPRT*.

Objetivos específicos:

- a) Identificar, dentre as diferentes linhagens de fibroblastos dérmicos humanos primários *HPRT*^{+/-} disponíveis em nosso laboratório, linhagens informativas que permitam acessar a presença de reativação do cromossomo X;
- b) Isolar, através de seleção em meio contendo 6TG, uma população de fibroblastos dérmicos humanos primários cujo alelo *HPRT*⁻ encontre-se localizado no cromossomo X ativo e cujo alelo *HPRT*⁺ encontre-se localizado no cromossomo X inativo, a fim de possibilitar a seleção das células nas quais o cromossomo X foi reativado tornando-se *HPRT*⁺;
- c) Determinar, através de geração de curvas de crescimento celular, o tempo necessário para que ocorra depleção de células *HPRT*⁻ (e consequente enriquecimento de células *HPRT*⁺) mediante ao tratamento em meio de cultura contendo HAT;
- d) Realizar a triagem funcional de genes envolvidos no processo de manutenção da inativação do cromossomo X nas células da população de fibroblastos *HPRT*⁻ por meio de uso de uma biblioteca lentiviral de shRNAs seguida por seleção positiva das células que sofreram reativação do cromossomo X (*HPRT*⁺) em meio de cultura contendo HAT e seqüenciamento de próxima geração;
- e) Analisar os dados gerados pelo seqüenciamento de próxima geração, a fim de identificar genes candidatos para o processo de manutenção do estado de inativação do cromossomo X em células somáticas humanas femininas;

- f) Estudar e discutir isoladamente os genes candidatos selecionados pela triagem funcional com base na função relatada para cada um desses genes na literatura e/ou em bancos de dados disponíveis na internet, a fim de construir uma lista dos melhores genes candidatos com possível papel no processo de manutenção do estado de inativação do cromossomo X em células somáticas humanas femininas, a serem validados futuramente, em trabalho de continuidade ao abordado nessa tese.

Material e Métodos

3. Material e Métodos

3.1. Laboratórios e Financiamentos

O presente trabalho foi desenvolvido no Laboratório de Genética Molecular do Departamento de Genética e Biologia Evolutiva do Instituto de Biociências da Universidade de São Paulo (IB-USP) sob orientação da Profa. Dra. Lygia da Veiga Pereira Carramaschi e utilizou verbas provenientes das agências de fomento Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – Programas Demanda Social (DS – Bolsista) e Programa de Excelência Acadêmica (PROEX – Projeto 33002010021P7 – USP Ciências Biológicas, Biologia, Genética)), CNPq e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Parte do trabalho foi também desenvolvida durante estágio de quatro meses no *The Scripps Research Institute (TSRI)*, *Department of Chemical Physiology*, sob coordenação da Profa. Dra. Jeanne F. Loring, em La Jolla, California, Estados Unidos da América. O trabalho também contou com a colaboração da Profa. Dra. Tatiana Teixeira Torres (Laboratório de Genômica e Evolução de Artrópodes, Departamento de Genética e Biologia Evolutiva, IB-USP), Maria R. Lisboa Santana Pinheiro, do grupo da Profa. Dra. Ângela Maria Vianna Morgante (Laboratório de Genética Humana, Departamento de Genética e Biologia Evolutiva, IB-USP) e dos pesquisadores Robert E. Morey e Dra. Rathi D. Thiagarajan ambos integrantes do Laboratório da Profa. Dra. Louise Laurent, MD, *Department of Reproductive Medicine, University of California San Diego (UCSD)*.

3.2. Linhagens Celulares

Linhagens de fibroblastos dérmicos humanos primários 46,XX *HPRT*^{+/-} denominadas GM13A, GM135, GM1660, GM1661, e GM2226 foram adquiridas do Centro de Pesquisa Coriell Institute for Medical Research®. As linhagens HDF48, HDF49 e HDF51 foram estabelecidas no laboratório da Dra. Jeanne F. Loring a partir de biópsia de pele de indivíduos normais. A linhagem FibN1 é proveniente do

Centro de Estudos do Genoma Humano da USP e também foi estabelecida a partir de biópsia de pele de mulher normal. As linhagens Hs27 e 3T3 foram adquiridas comercialmente da organização de pesquisa American Type Culture Collection (ATCC®). As informações referentes a cada uma das linhagens celulares utilizadas estão sumarizadas na Tabela 6.

Tabela 6. Linhagens de Fibroblastos Utilizadas

Linhagem Celular	Fornecedor	Origem	Informações Adicionais
GM13A (Cat# GM00013)	Coriell Institute - NIGMS	<i>Homo sapiens</i>	Linhagem primária, feminina (46,XX) de paciente de 34 anos e caucasóide, portadora clinicamente não-afetada da Síndrome de Lesch-Nyhan, mãe de afetado. Testes bioquímicos indicaram atividade enzimática reduzida (55% de atividade de HPRT em fibroblastos). Fornecida na passagem 10.
GM135 (Cat# GM00135)	Coriell Institute - NIGMS	<i>Homo sapiens</i>	Linhagem primária, feminina (46,XX) de paciente adulta de raça negra, portadora clinicamente não-afetada da Síndrome de Lesch-Nyhan, mãe e tia de afetados. HPRT+, G6PDB / HPRT-, G6PDA. Fornecida na passagem 8.
GM1660 (Cat# GM01660)	Coriell Institute - NIGMS	<i>Homo sapiens</i>	Linhagem primária, feminina de paciente de 11 anos e caucasóide, portadora clinicamente não-afetada da Síndrome de Lesch-Nyhan, irmã de afetado. Nível de HPRT sanguíneo normal, portadora por ensaio de raiz de cabelo. Fornecida na passagem 3.
GM1661 (Cat# GM01661)	Coriell Institute - NIGMS	<i>Homo sapiens</i>	Linhagem primária, feminina de paciente de 14 anos e caucasóide, portadora clinicamente não-afetada da Síndrome de Lesch-Nyhan, irmã de afetado. Nível de HPRT sanguíneo normal, portadora por ensaio de raiz de cabelo. Fornecida na passagem 3.
GM2226 (Cat# GM02226)	Coriell Institute - NIGMS	<i>Homo sapiens</i>	Linhagem primária, feminina de paciente de 40 anos, portadora clinicamente não-afetada da Síndrome de Lesch-Nyhan, mãe de afetado. Duas populações de células identificadas: uma mostrando absorção normal e outra mostrando nenhuma absorção significativa de hipoxantina tritiada por autorradiografia (HPRT+/HPRT-). Fornecida na passagem 6.
HDF48	TSRI	<i>Homo sapiens</i>	Fibroblasto dérmico adulto primário de indivíduo do sexo feminino de raça negra.
HDF49	TSRI	<i>Homo sapiens</i>	Fibroblasto dérmico adulto primário de indivíduo do sexo feminino caucasóide.
HDF51	TSRI	<i>Homo sapiens</i>	Fibroblasto dérmico neonatal primário.
FibN1	Centro de Estudos do Genoma Humano - USP	<i>Homo sapiens</i>	Fibroblasto dérmico adulto primário de indivíduo do sexo feminino clinicamente normal.
Hs27 (Cat# CRL-1634)	ATCC	<i>Homo sapiens</i>	Fibroblasto de prepúcio neonatal.
3T3	ATCC	<i>Mus musculus</i>	Fibroblastos embrionários espontaneamente imortalizados.

NIGMS: National Institute of General Medical Sciences, TSRI: The Scripps Research Institute, ATCC: American Type Culture Collection, USP: Universidade de São Paulo.

3.3. Cultivo de Fibroblastos

As linhagens de fibroblastos foram cultivadas a 37°C, 5%CO₂ em placas aderentes pré-tratadas ou não com solução de 0,1% gelatina em água (Stem Cell Technologies™, Cat# 07903). A formulação dos meios de cultura utilizados foi baseada nos protocolos previamente padronizados nos laboratórios envolvidos e/ou nas recomendações dadas pela empresa Coriell Institute for Medical Research®. Sendo assim, e dependendo da linhagem celular em questão, foi usado meio de fibroblasto contendo 10% (linhagens GM2226, HDF48, HDF49, HDF51, FibN1, Hs27 e 3T3) ou 15% de soro fetal bovino (FBS) (linhagens GM13A, GM135, GM1660 e GM1661), de acordo com as formulações descritas a seguir.

Meio de Fibroblasto 10% FBS

1XDMEM *High Glucose* (Life Technologies™ - Gibco®, Cat# 11965-092)
10% FBS (Life Technologies™ - Gibco®, Cat# 12657-029)
0.1mM MEM NEAA (Life Technologies™ - Gibco®, Cat# 11140-050)
1mM *Sodium Piruvate* (Life Technologies™ - Gibco®, Cat# 11360-070)
2mM GlutaMAX™ Supplement (Life Technologies™ - Gibco®, Cat# 35050-061)
100U/mL Penicillin/ 100ug/mL Streptomycin (Life Technologies™ - Gibco®, Cat# 15140-122)

Meio de Fibroblasto 15% FBS

1XDMEM *High Glucose* (Life Technologies™ - Gibco®, Cat# 11965-092)
15% FBS (Life Technologies™ - Gibco®, Cat# 12657-029)
0.1mM MEM NEAA (Life Technologies™ - Gibco®, Cat# 11140-050)
1mM *Sodium Piruvate* (Life Technologies™ - Gibco®, Cat# 11360-070)
2mM GlutaMAX™ Supplement (Life Technologies™ - Gibco®, Cat# 35050-061)
100U/mL Penicillin/ 100ug/mL Streptomycin (Life Technologies™ - Gibco®, Cat# 15140-122)

3.4. Passagem de Fibroblastos

A metodologia de passagem de fibroblastos primários humanos empregada durante o desenvolvimento desse trabalho seguiu os padrões adotados pelo nosso laboratório que consiste nas seguintes etapas:

- 1) O meio de cultura dos frascos é removido e as células lavadas duas vezes em 1XDPBS (Life Technologies™ - Gibco®, Cat# 14190-144),
- 2) A solução de 0,25% tripsina-EDTA (Life Technologies™ - Gibco®, Cat# 25200-072) é adicionada ao frasco de cultura e as células incubadas a 37°C, 5%CO₂ por 3 a 5 minutos,
- 3) A reação de tripsinização é parada através da adição de meio de cultura de fibroblasto fresco pré-aquecido em banho-maria a 37°C,
- 4) As células em suspensão são transferidas para um tubo Falcon de 15 ou 50mL de capacidade e centrifugadas a 250g, temperatura ambiente, por 5min,
- 5) O sobrenadante é descartado e o pellet ressuspensionado em volume apropriado de meio de cultura de fibroblasto fresco pré-aquecido em banho-maria a 37°C,
- 6) Quando necessário, as células são contadas em hemocítmetro e plaqueadas em novos frascos ou placas de cultura na densidade desejada,
- 7) Os novos frascos ou placas de cultura são retornados a incubadora a 37°C, 5%CO₂.

3.5. Congelamento de Fibroblastos

Sempre que adquirido um número excedente de células em cultura e/ou com o intuito de gerar um estoque de células com diferentes números de passagem para uso futuro, as diferentes linhagens de fibroblastos utilizadas nesse trabalho foram congeladas utilizando-se o meio de congelamento de fibroblastos descrito no final desse item e seguindo-se o protocolo padrão adotado em nosso laboratório que consiste nas seguintes etapas:

- 1) O meio de cultura é removido e as células são lavadas 2X em 1XDPBS (Life Technologies™ - Gibco®, Cat# 14190-144),
- 2) A solução de 0,25% Tripsina-EDTA (Life Technologies™ - Gibco®, Cat# 25200-072) é adicionada ao frasco de cultura (0.5mL por frasco T75) e as células incubadas a 37°C, 5%CO₂ por 3 a 5 min,
- 3) A reação de tripsinização é parada através da adição de meio de cultura de fibroblasto fresco pré-aquecido em banho-maria a 37°C,
- 4) As células em suspensão são transferidas para tubos Falcon de 15 ou 50mL de capacidade e centrifugadas a 250g, temperatura ambiente, por 5min,
- 5) O sobrenadante é descartado e o pellet ressuspensionado em meio de congelamento de fibroblastos,
- 6) As células em meio de congelamento de Fibroblastos são rapidamente transferidas para um criotubo de 2mL de capacidade devidamente identificado (1mL por criotubo), colocadas em recipiente criogênico (*Mr. Frosty Freezing Container*, Thermo-Scientific™ - Nalgene®, Cat# 5100-001) contendo isopropanol a temperatura ambiente e imediatamente transferidas para o freezer -80°C,
- 7) Após três dias a uma semana a -80°C, os criotubos são transferidos rapidamente para o galão de nitrogênio líquido.

Meio de Congelamento de Fibroblastos

70% meio de fibroblasto 10% ou 15% FBS (dependendo da linhagem celular)

20% FBS (Life Technologies™ - Gibco®, Cat# 12657-029)

10% DMSO (Sigma-Aldrich®, Cat# D2650)

3.6. Descongelamento de Fibroblastos

Sempre que necessário, as diferentes linhagens de fibroblastos utilizadas nesse trabalho foram descongeladas segundo o protocolo padrão utilizado em nosso laboratório que consiste nas seguintes etapas:

- 1) As amostras estocadas em nitrogênio líquido (volume=1mL) são transferidas imediatamente para o banho-maria a 37°C até que o descongelamento ocorra ou, no máximo, por 3min,
- 2) Os tubos criogênicos são rapidamente transferidos para o fluxo laminar e, 8mL de meio de cultura de fibroblasto fresco e pré-aquecido em banho-maria à 37°C são adicionados gentilmente ao tubo,
- 3) A amostra é passada para um tubo Falcon de 15mL de capacidade e centrifugada a 250g, temperatura ambiente, por 5min,
- 4) O pellet é ressuspensionado em volume apropriado de meio de cultura de fibroblasto fresco e pré-aquecido em banho-maria a 37°C, as células ressuspensionadas são transferidas para o frasco ou placa de cultura apropriado e, então, transferidas para a incubadora a 37°C, 5%CO₂.
- 5) No dia seguinte, o meio de cultura é trocado por meio de cultura de fibroblasto fresco e pré-aquecido em banho-maria a 37°C e o frasco ou placa de cultura é retornado novamente à incubadora a 37°C, 5%CO₂.

3.7. Análise da Eficiência de Transdução Viral em Linhagens Primárias de Fibroblastos Dérmicos Humanos Heterozigotas para o gene *HPRT*

As linhagens de fibroblastos primários humanos 46,XX *HPRT*^{+/-} disponíveis em nosso laboratório foram testadas para determinar a taxa de eficiência de transdução viral com retrovírions contendo o gene repórter EGFP cedidos pela Dra. Patrícia Cristina Baleeiro Beltrão Braga do Instituto Butantan. A linhagem de fibroblastos embrionários murinos imortalizados 3T3 foi utilizada como controle positivo de transdução já que esta linhagem já havia anteriormente demonstrado ser uma linhagem altamente susceptível à transdução pelas partículas retrovirais utilizadas.

Dia 0:

Para o experimento de transdução com retrovírions EGFP, 2,5x10⁵ células de cada linhagem foram plaqueadas em meio de fibroblasto 15% FBS da seguinte maneira:

- Frasco 1:** Linhagem GM13A (amostra experimental)
Frasco 2: Linhagem GM135 (amostra experimental)
Frasco 3: Linhagem GM1660 (amostra experimental)
Frasco 4: Linhagem GM1661 (amostra experimental)
Frasco 5: Linhagem GM2226 (amostra experimental)
Frasco 6: Linhagem 3T3 (controle positivo)
Frasco 7: Linhagem GM13A (controle negativo – não transduzidas)
Frasco 8: Linhagem GM135 (controle negativo – não transduzidas)
Frasco 9: Linhagem GM1660 (controle negativo – não transduzidas)
Frasco 10: Linhagem GM1661 (controle negativo – não transduzidas)
Frasco 11: Linhagem GM2226 (controle negativo – não transduzidas)

Dia 1:

Para cada frasco T25 experimental e controle positivo, o meio de cultura foi removido e 3mL de meio de fibroblasto 15% FBS fresco, 1mL de sobrenadante contendo retrovírions EGFP (enhanced green fluorescent protein) e 8ug/mL de polibreno (*Hexadimethrine bromide*, Sigma-Aldrich®, Cat# H9268) foram adicionados. O meio de cultura dos frascos T25 contendo células controle negativo foi substituído por 4mL de meio de fibroblasto 15% FBS fresco contendo 8ug/mL de polibreno. Os frascos foram devolvidos a incubadora a 37°C, 5%CO₂ onde os vírions foram deixados em contato com as células por 24h horas.

Dia 2 (24h após transdução viral):

O meio de cultura de todos os frascos T25 foi removido e 5mL de meio de cultura de fibroblastos 15% FBS fresco foi adicionado a cada frasco. Os frascos foram retornados a incubadora a 37°C, 5%CO₂.

Dia (72h após transdução viral):

As células foram visualizadas em microscópio de fluorescência Axioplan 2® (Carl Zeiss®) para verificação da viabilidade celular (todos os frascos) e presença de EGFP (células experimentais e controle positivo). Depois disso, as células foram coletadas por tripsinização em tubos eppendorf de 1,5mL de capacidade e os *pellets*

foram ressuspensos em 500uL de 1XDPBS contendo 10% FBS. Os *pellets* celulares foram colocados em gelo e encaminhados para a citometria de fluxo em aparelho Guava easyCyte™ (EMD Millipore®) e a porcentagem de células EGFP positivas em cada amostra foi então determinada. As células contidas nos frascos controle negativo foram utilizadas para calibração do citômetro.

3.8. Extração de DNA de Fibroblastos Humanos Primários

A extração de DNA genômico total dos fibroblastos dérmicos humanos primários foi realizada a partir de células das linhagens GM1660, GM1661 e GM2226 não-selecionadas. Para tanto, o kit “*Flexigene DNA Kit*” (Qiagen®, Cat# 51206) foi utilizado. As células foram cultivadas em frascos T25, coletadas por tripsinização, lavadas duas vezes em 1XDPBS, peletadas e o DNA genômico total foi extraído de acordo com as instruções fornecidas pelo fabricante do kit de extração. As amostras de DNA genômico total foram armazenadas em freezer a -20°C até o momento de utilização.

3.9. Extração de RNA de Fibroblastos Humanos Primários

A extração de RNA total dos fibroblastos dérmicos humanos primários foi realizada tanto a partir de células da linhagem GM1661 não-selecionadas (passagem 11) como selecionadas por 25 dias em meio contendo 6TG (passagem 15). Para isso, o kit “*RNeasy Mini Kit*” (Qiagen®, Cat# 74104) foi utilizado. As células foram cultivadas em frascos T25, coletadas por tripsinização, lavadas duas vezes em 1XDPBS e o RNA total foi extraído de acordo com as instruções fornecidas pelo fabricante do kit de extração. Após a extração do RNA total, as amostras foram colocadas em gelo, quantificadas e utilizadas imediatamente nos procedimentos de remoção de DNA total contaminante e produção de cDNA. O restante das amostras de RNA total foi armazenado em freezer a -80°C.

3.10. Produção de cDNA a partir de RNA

As amostras de RNA total extraídas dos fibroblastos dérmicos humanos primários foram primeiramente submetidas ao tratamento com DNase para remoção de possível DNA genômico total contaminante. Para esse tratamento foi utilizado o kit “*Turbo DNA-Free™ Kit*” (Life Technologies™ - Invitrogen™ - Ambion®, Cat# AM1907) de acordo com as instruções fornecidas pelo fabricante.

O RNA previamente tratado com DNase foi utilizado para a produção de cDNA com o emprego do kit “*SuperScript® III Reverse Transcriptase*” (Life Technologies™ - Invitrogen™, Cat# 18080-093) e utilização do inibidor de ribonucleases “*RNaseOUT™ Recombinant Ribonuclease Inhibitor*” (Life Technologies™ - Invitrogen™, Cat# 10777-019), de acordo com o protocolo fornecido pelo fabricante. A fim de verificar, por PCR, se as amostras de RNA estavam contaminadas com DNA genômico, um tubo controle RT⁻, no qual a transcriptase reversa não foi adicionada à reação de produção de cDNA, foi gerado para cada uma das amostras de RNA obtidas.

3.11. Quantificação de Ácidos Nucléicos para Seqüenciamento Direto

As amostras de DNA, RNA e cDNA utilizadas nos experimentos envolvendo seqüenciamento convencional de Sanger foram quantificadas em espectrofotômetro GeneQuant Pro version 2.1 (Amershan Biosciences®, Cat# 80-2110-98) após diluídas de 10 a 50X em água MiliQ autoclavada. Durante cada quantificação foi também utilizada uma amostra “branco” para calibração da máquina que consistiu em água MiliQ autoclavada sem adição de ácido nucléico.

3.12. Genotipagem de SNPs por Seqüenciamento Direto

As amostras de DNA ou cDNA obtidas foram primeiramente submetidas a reações de PCR com *primers* e condições previamente estabelecidas em nosso laboratório para amplificação dos fragmentos gênicos contendo os SNPs de interesse (Tabela 7). O kit da enzima “*Taq DNA Polymerase, Recombinant*” (Life

Technologies™ - Invitrogen™, Cat# 10342-053) foi utilizado nas reações de amplificação pré-seqüenciamento. No caso de reações realizadas a partir de cDNA, as reações controle RT⁻ também foram submetidas a amplificação por PCR a fim de se verificar a presença ou ausência de contaminação por DNA em cada uma das amostras de cDNA obtidas.

O sucesso da reação de amplificação bem como o tamanho dos amplicons obtidos foi verificado correndo-se parte das amostras em gel de agarose 1% (Life Technologies™ - Invitrogen™, Cat# 16500-100) contendo brometo de etídeo (*UltraPure™ 10 mg/ml Ethidium Bromide*, Life Technologies™ - Invitrogen™, Cat# 15585-011). No caso de amplicons gerados a partir de cDNA, uma amostra dos produtos de PCR foi corrida, juntamente com os controles RT⁻, em gel de acrilamida 10% que foi subsequentemente corado por solução de nitrato de prata.

Os amplicons foram então submetidos novamente à amplificação por PCR (reação de seqüenciamento) contendo nucleotídeos modificados e marcados com fluoróforos utilizando-se o kit de reagentes “*BigDye® Terminator v3.1 Cycle Sequencing Kit*” (Life Technologies™ - Applied Biosystems®, Cat# 4337457), purificados em colunas de *Sephadex™* e encaminhados para corrida e leitura em aparelho de seqüenciamento automático *ABI PRISM® 3700 Genetic Analyzer* (Life Technologies™ - Applied Biosystems®). Os cromatogramas obtidos foram então analisados e os genótipos determinados. A lista de SNPs analisados bem como a localização cromossômica de cada gene, e os primers e programas de PCR utilizados para amplificação de cada SNP estão relacionados na Tabela 7.

Reação de PCR:

DNA (cerca de 50ng a 100ng) ou 2uL cDNA
1.5uL 10X *Taq polymerase, Recombinant Buffer*
0.6uL 50mM MgCl₂
0.15uL mistura de dNTPs (contendo 10mM de cada dNTP)
0.45uL 5uM *Primer Forward*
0.45uL 5uM *Primer Reverse*
0.15uL *Taq DNA polymerase, Recombinant*
Água MiliQ autoclavada qsp 15uL

As reações foram preparadas em gelo e as amostras transferidas imediatamente para o termociclador.

Programas Utilizados no Termociclador:

Touch55

Termociclador: *PTC-200 Thermal Cycler* (MJ Research®)

Control Method: BLOCK

94°C 3min

94°C 30seg, 60°C 30seg (-0.5°C per cycle), 72°C 1min (15 ciclos)

94°C 30seg, 55°C 30seg, 72°C 1min (22 ciclos)

72°C 5min

10°C *hold*

Touch57

Termociclador: *PTC-200 Thermal Cycler* (MJ Research®)

Control Method: CALCULATED

94°C 3min

94°C 20seg, 62°C 20seg (-0.5°C per cycle), 72°C 1min (15 ciclos)

94°C 20seg, 57°C 20seg, 72°C 1min (22 ciclos)

72°C 5min

10°C *hold*

Touch60

Termociclador: *PTC-200 Thermal Cycler* (MJ Research®)

Control Method: BLOCK

94°C 3min

94°C 30seg, 65°C 30seg (-0.5°C per cycle), 72°C 1min (15 ciclos)

94°C 30seg, 60°C 30seg, 72°C 1min (22 ciclos)

72°C 7min

10°C *hold*

Verificação da qualidade dos cDNAs amplificados por PCR

Os produtos de amplificação por PCR gerados a partir de amostras de cDNA e de seus respectivos controles RT⁻, foram submetidos a eletroforese em gel de poliacrilamida 10% em tampão 1XTBE à 100V por 1h. Após a corrida, o gel de poliacrilamida foi fixado em solução contendo 10% de etanol e 0,5% de ácido acético por 10min, corado em solução de nitrato de prata (2g/mL) por 10min, brevemente lavado em água destilada e mergulhado em solução reveladora (hidróxido de sódio 30 g/L, 0,002% formaldeído a 37%) até que as bandas pudessem ser visualizadas, quando o gel foi transferido novamente para a solução fixadora por 10 min a fim de parar a reação de revelação e, então, transferido e mantido em água destilada durante a análise visual do gel.

Reação de Seqüenciamento:

O produto de PCR foi diluído de 2 a 4X (de acordo com a intensidade da banda em gel de agarose) em água MiliQ autoclavada. O kit “*BigDye® Terminator v3.1 Cycle Sequencing Kit*” (Life Technologies™ - Applied Biosystems®, Cat# 4337457) foi utilizado nas reações de seqüenciamento. Esse kit fornece os componentes necessários para a reação de seqüenciamento já misturados em um *Ready Reaction Mix* e em tampão 5X concentrado, sendo apenas necessário acrescentar o DNA ou cDNA molde e o *primer* desejado à reação. Por conter dNTPs marcados com fluoróforos, o *BigDye® Ready Reaction Mix* é armazenado e manipulado protegido da luz.

1uL DNA ou cDNA diluído

0.6uL 5uM *Primer* de seqüenciamento

3uL 5X *BigDye® Terminator v1.1/3.1 Sequencing Buffer*

1uL *BigDye® Ready Reaction Mix*

13.4uL Água MiliQ autoclavada

As reações foram preparadas em gelo e as amostras transferidas imediatamente para o termociclador. O programa utilizado para as reações de amplificação destinadas ao seqüenciamento está especificado a seguir:

Seqüenciamento

Termociclador: *PTC-200 Thermal Cycler* (MJ Research®)

Control Method: BLOCK

96°C 10seg, 52°C 20seg, 60°C 4min (40 ciclos)

4°C *hold*

Purificação da Reação de Seqüenciamento por Coluna Sephadex™:

- 1) Os poços da placa de metal (*MultiScreen Column Loader 45uL*, EMD Millipore®, Cat# MACL09645) própria para Sephadex™ foram preenchidos com o pó *Sephadex™* (Ilustra Sephadex G-50, DNA grade, Fine, GE Healthcare Life Sciences®, Cat# 17-0573-02) e o excesso do produto foi retirado,
- 2) A placa plástica de 96-well (*MultiScreen_{HTS}-HV, 0.45um*, EMD Millipore®, Cat# MSHVN4550) própria para purificação com colunas de *Sephadex™* foi colocada sobre a placa de metal preenchida e as placas foram rapidamente para que o *Sephadex™* fosse transferido da placa de metal para a placa plástica,
- 3) O *Sephadex™* presente em cada pocinho da placa plástica 96-well foi tratado com 300uL de água MiliQ autoclavada de 3 horas a overnight,
- 4) Uma placa comum 96-well (placa coletora) foi colocada em baixo da placa plástica contendo o *Sephadex™* e as placas foram centrifugadas a 900g, temperatura ambiente, por 8min,
- 5) A água da placa coletora foi descartada e mais 150uL de água MiliQ autoclavada foram adicionados a cada pocinho da placa contendo *Sephadex™*,

- 6) A placa contendo *Sephadex*[™] foi novamente colocada sobre a placa coletora e as placas foram novamente centrifugadas a 900g, temperatura ambiente, por 10min,
- 7) A água da placa coletora foi descartada e a placa contendo *Sephadex*[™] foi então encaixada em uma nova placa 96-well, limpa e estéril, própria para seqüenciamento (*MicroAmp optical 96-well reaction Plate*, Life Technologies[™] - Applied Biosystems®, Cat# N8010560),
- 8) Cada amostra de produto de PCR para seqüenciamento foi pipetada em um dos pocinhos contendo *Sephadex*[™] e a placa foi centrifugada a 900g, temperatura ambiente, por 5min,
- 9) A placa contendo *Sephadex*[™] foi descartada e a placa contendo as amostras coletadas foi colocada em termociclador com tampa semi-aberta e coberta por papel alumínio (para proteger as amostras da luz) por 30min a 94°C ou até que o líquido contido nos pocinhos fosse evaporado,
- 10) A placa contendo as amostras para seqüenciamento foi armazenada protegida da luz em freezer -20°C até a hora da corrida no aparelho de seqüenciamento,
- 11) Momentos antes do envio do material para sequenciamento, 10uL de formamida (*Hi-Di*[™] *Formamide*, Life Technologies[™] - Applied Biosystems®, Cat# 4311320) foram adicionados por pocinho, a placa foi coberta com o adesivo próprio para placa de seqüenciamento (*MicroAmp® Optical Adhesive Film*, Life Technologies[™] - Applied Biosystems®, Cat# 4311971), e passada em vórtex por 1min e 30seg,
- 12) A placa contendo as amostras ressuspensas em formamida foram enviadas protegidas da luz para o seqüenciamento.

Tabela 7. Lista dos SNPs analisados

Gene*	Localização Cromossômica e Coordenada Genômica	Identificação dos SNPs	Amplicon (pb)/ Tm	Primer Forward (5' – 3')	Primer Reverse (5' – 3')	Primer para Sequenciamento
GYG2 MIM*300198	Xp22.33 X:2,746,862 - 2,800,860	rs211659	206pb Touch57	CGCTGTTGAACCTTGTGCCT	ACCACCATCGAGAGTCTACT	Forward
PIGA MIM*311770	Xp22.2 X:15,337,572 - 15,353,675	rs1804165 rs11448948 rs3087965	255pb Touch55	CGACCAGTGCAAAGTACAG	AGGTTTTAGAGGCATGCCAC	Reverse
DYNLT3 MIM*300302	Xp11.4 X:37,698,088 - 37,706,888	rs12849	200pb Touch55	GCCAAAAACAAAGATCACTGC	GAGAGCGGAACCATAACTCA	Forward
TIMP1 MIM*305370	Xp11.23 X:47,441,689 - 47,446,189	rs6520277	169pb Touch55	GCACATCACTACCTGCAGATT	AGCCCACGATTTAGGGAATG	Reverse
EBP MIM*300205	Xp11.23 X:48,380,163 - 48,387,103	rs1049367 rs3048	161pb Touch57	CCATCAGCCCACAAAGACAT	AGCCATGTGGTCACGACTAA	Reverse
SUV39H1 MIM*300254	Xp11.23 X:48,555,130 - 48,567,406	rs3373	292pb Touch57	GCATAGGGTTGAGGGGTGTA	TTGTGCTCACCTGGTTC	Forward
ATRX MIM*300032	Xq21.1 X:76,760,355 - 77,041,718	rs3088074	186pb Touch57	TGGTTTTGAGATGCTTGCTC	CAGCAGAAGGCACAGTTGAT	Forward
TCEAL4	Xq22.3 X: 102,831,159-102,842,657	rs11010 rs221933	186pb Touch55	TGGCTCATTACCTCAAGGAG	CTTGGACCTCTAGGGTAGAA	Reverse
ZNF75D MIM*314997	Xq26.3 X:134,419,722 - 134,429,964	rs1129093	200pb Touch60	GGAGACACTAGACATGCTTC	CACACAGGAGAGAAGCCCTT	Reverse
VBP1 MIM*300133	Xq28 X:154,444,700 - 154,468,097	rs11887	201pb Touch55	AGTTTACTTTTCCCTTTCCAAT	GTAGTTAGTTATTCGTTAAGC	Reverse
HDHD1 MIM*306480	Xp22.31 X:6,966,960 - 7,066,230	rs10458	117pb Touch55	TCTCAGAAGTCAAGCTGATGGA	GCATGTAATGCCTACCTGCAT	Reverse
GPR143 MIM*300808	Xp22.2 X:9,693,452 - 9,734,004	rs3044	138bp Touch55	TCCATTCTCACACGTGTGTGT	CTGTGTTCTCACCTTCCCAA	Reverse
ZFX MIM*314980	Xp22.11 X:24,167,761 - 24,234,371	rs5949242	158pb Touch60	ACATTATGGGGGCCAAGAAG	TGGCAAACCTCTGAAGTGCT	Forward
SYTL4 MIM*300723	Xq22.1 X:99,929,487 - 99,987,134	rs11557084 rs8780	136pb Touch55	ATTAAAGCACACATACATGTCAG	TGAACCGAATGGAGGAATG	Reverse

* Genes submetidos a XCI em humanos estão realçados em letras verdes. Genes que escapam a XCI em humanos estão realçados em letras vermelhas.

3.13. Teste de Sensibilidade de Fibroblastos Humanos Primários à Droga 6-Tioguanina

Fibroblastos humanos primários normais livres de mutações causadoras de perda de função no gene *HPRT* das linhagens FibN1 e Hs27 foram cultivados em meio contendo 6TG a fim de verificar a eficácia do tratamento com essa droga na depleção de células $HPRT^+$ em cultura. O experimento foi realizado conforme descrito abaixo:

- 1) Células das linhagens FibN1 e Hs27 foram plaqueadas em poços de placa de 12-well (6 poços para cada linhagem, 5×10^5 células por poço) em meio de fibroblasto 15% FBS e incubadas a 37°C, 5%CO₂;
- 2) Após 48h, o meio de cultura foi trocado. Para cada linhagem, dois poços receberam meio 6TG, dois poços receberam meio HAT e dois poços receberam meio de fibroblasto 15% FBS sem adição de nenhuma dessas duas drogas. Os poços HAT e não-tratado foram utilizados como controle durante o monitoramento do efeito de 6TG sobre nas linhagens analisadas. A placa de cultura foi retornada a incubadora a 37°C, 5%CO₂;
- 3) A partir daí, os pocinhos contendo células em cada uma das condições estudadas foram visualizados diariamente em microscópio de luz invertido e fotografados no segundo, quinto e décimo segundo dias de tratamento com as drogas.

3.14. Seleção de Fibroblastos Humanos Primários $HPRT^+$ e $HPRT^-$

A seleção de células $HPRT^+$ e $HPRT^-$ a partir de linhagens *HPRT*[±]-femininas de fibroblastos primários humanos foi realizada para células das linhagem GM1661. As células foram primeiramente colocadas em cultura para expansão em meio GM15%FBS. Uma vez expandidas, as células foram passadas por tripsinização na proporção 1:3 em frascos T75 com o seguinte propósito:

- **Frasco 1:** Controle não-tratado (também para geração de estoque congelado)
- **Frasco 2:** Tratamento com meio HAT (seleção positiva das células HPRT⁺ e depleção das células HPRT⁻)
- **Frasco 3:** Tratamento com meio contendo 6-Tioguanina (seleção positiva das células HPRT⁻ e eliminação das células HPRT⁺)

Meio HAT

1XDMEM High Glucose (Life Technologies™ - Gibco®, Cat# 11965-092)
 15% FBS (Life Technologies™ - Gibco®, Cat# 12657-029)
 0.1mM MEM NEAA (Life Technologies™ - Gibco®, Cat# 11140-050)
 1mM Sodium Piruvate (Life Technologies™ - Gibco®, Cat# 11360-070)
 2mM GlutaMAX™ Supplement (Life Technologies™ - Gibco®, Cat# 35050-061)
 100U/mL Penicillin/ 100ug/mL Streptomycin (Life Technologies™ - Gibco®, Cat# 15140-122)
 1X HAT Supplement (Life Technologies™ - Gibco®, Cat# 21060-017)

Meio 6-Tioguanina

1XDMEM High Glucose (Life Technologies™ - Gibco®, Cat# 11965-092)
 15% FBS (Life Technologies™ - Gibco®, Cat# 12657-029)
 0.1mM MEM NEAA (Life Technologies™ - Gibco®, Cat# 11140-050)
 1mM Sodium Piruvate (Cat# 11360-070, Life Technologies™ - Gibco®)
 2mM GlutaMAX™ Supplement (Life Technologies™ - Gibco®, Cat# 35050-061)
 100U/mL Penicillin/ 100ug/mL Streptomycin (Life Technologies™ - Gibco®, Cat# 15140-122)
 60uM Hybri-Max™ 6-Thioguanine (Sigma-Aldrich®, Cat# A-4660)

As soluções estoque (250X) de 6-Tioguanina foram preparadas diluindo-se o produto em pó (2.5mg) em 100uL de solução 0,1M NaOH fresca e estéril em fluxo laminar. Logo após a dissolução do pó, 900uL de água estéril livre de nucleases (*Water for Embryo Transfer*, Sigma-Aldrich®, Cat# W1503) foram adicionados ao frasco e a solução foi homogeneizada. As soluções estoque foram aliquotadas em

tubos eppendorfs escuros de 1.5mL de capacidade (100uL por eppendorf) e armazenadas protegidas da luz em freezer -20°C até o momento de sua utilização.

Uma vez preparados, os meios de cultura contendo HAT ou 6TG foram armazenados a 4°C por não mais de uma semana. O tratamento seletivo em meio HAT e meio 6-Tioguanina foi realizado trocando-se o meio dos frascos de cultura a cada dois dias e durou, pelo menos, 30 dias.

3.15. Cariotipagem de Células da Linhagem GM1661 Selecionadas em Meio 6TG

As células da linhagem GM1661 selecionadas em meio 6TG (passagem 14) foram submetidas à análise de cariótipo por bandamento G para se verificar a integridade do número e estrutura cromossômica na população de células selecionadas. As etapas relacionadas ao cultivo e coleta das células foram realizadas sob supervisão da técnica de laboratório em citogenética Maria R. Lisboa Santana Pinheiro, do grupo da Profa. Dra. Ângela Maria Vianna Morgante (Laboratório de Genética Humana, Departamento de Genética e Biologia Evolutiva, IB-USP), que realizou também o exame de cariótipo.

Para essa finalidade, cerca de 6×10^5 células presentes em um frasco T25 foram utilizadas. As células foram cariotipadas conforme protocolo descrito a seguir:

- 1) 30uL de colchicina ($C=4.10^{-5}M$) foram adicionados diretamente a 4mL de meio de cultura celular e o frasco T25 foi retornado a incubadora a 37°C, 5%CO₂ por 1h30min,
- 2) O meio de cultura contendo a colchicina foi removido e as células lavadas em 5mL de 1XDPBS (Life Technologies™ - Gibco®, Cat# 14190-144) por duas vezes,
- 3) 0.5mL de 0,25% tripsina-EDTA (Life Technologies™ - Gibco®, Cat# 25200-072) foi adicionado ao frasco e as células foram incubadas a 37°C, 5%CO₂ por 3min,

- 4) A reação de tripsinização foi neutralizada pela adição de 5mL meio de fibroblasto 10% FBS, as células em suspensão foram transferidas para um tubo Falcon de 15mL de capacidade e centrifugadas a 200g por 5min,
- 5) O sobrenadante foi descartado e o *pellet* ressuspensionado em cerca de 5.5mL de solução hipotônica (0,075M KCl) a 37°C e incubado por 15min a 37°C,
- 6) 0.5mL de solução fixadora (3 metanol:1 ácido acético) foram adicionados ao tubo falcon contendo solução hipotônica e as células foram ressuspensionadas e em seguidas centrifugadas a 1200rpm por 8min,
- 7) O sobrenadante foi descartado, o pellet ressuspensionado em aproximadamente 4mL de solução fixadora e centrifugado a 1200rpm por 8min,
- 8) O passo 7 foi repetido mais duas vezes,
- 9) As células foram ressuspensionadas em 0.5mL de solução fixadora e duas gotas de 14uL cada foram pingadas em cada lâmina de microscopia limpa, molhada e gelada,
- 10) As lâminas foram incubadas a 37°C por uma semana,
- 11) As lâminas foram imersas e levemente agitadas por 7seg em uma cuba contendo 1 parte de solução de tripsina-EDTA 0,5% em água destilada (*Trypsin (1:250), powder, Life Technologies™ - Gibco®, Cat# 27250-018*) para 9 partes de tampão PBS de Dulbecco pH7.0 pré-aquecidos em banho-maria a 37°C por 30min,

Tampão PBS de Dulbecco pH7.0

0,2g Cloreto de Potássio (KCl)

0,2g Fosfato de Potássio Monobásico (KH₂PO₄)

8,0g Cloreto de Sódio (NaCl)

1,6g Fosfato de Sódio Dibásico diidratado (Na₂HPO₄.2H₂O)

Qsp 1L

- 12) As lâminas foram mergulhadas por 3seg em outra cuba contendo 1 parte de FBS para 9 partes de tampão PBS de Dulbecco pH7.0 também pré-aquecidos em banho-maria a 37°C por 30min,
- 13) As lâminas foram transferidas para uma cuba contendo 1 parte de corante (Giemsa-Wright) para 4 partes de tampão fosfato 0.01M pH 6,8-6,9. A cuba foi coberta e as lâminas foram coradas por 2min,

Corante Giemsa-Wright

0,45g Azure B, Eosina, Azul de Metileno segundo Giemsa

0,75g Eosina, Azul de Metileno segundo Wright

Dissolver em 300mL de glicerina e incubar a 60°C por 12 horas com agitação

Trazer a temperatura ambiente e adicionar 300mL de Metanol

Filtrar e estocar protegido da luz

- 14) As lâminas foram lavadas em água destilada para remoção do excesso de corante, secadas ao ar em posição vertical e analisadas sob microscópio de luz.

Dezoito metáfases foram cuidadosamente analisadas para investigação de possíveis alterações cromossômicas numéricas e/ou estruturais, sendo que imagens referentes a cinco dessas metáfases foram geradas com o auxílio do programa “Ikaros version 3” (MetaSystems Germany®). Os cromossomos presentes em outras 82 metáfases foram contados para verificar a presença de mosaicismo estrutural ou numérico.

3.16. Geração e Análise de Curvas de Crescimento Celular

Curvas de crescimento celular foram traçadas para células da linhagem GM1661 não selecionadas quando expostas ao meio de cultura contendo 6TG ou HAT, assim como para células da linhagem GM1661 pré-selecionadas em meio 6TG quando expostas ao meio de cultura contendo HAT, conforme descrito a seguir:

Dia 0:

As células da linhagem GM1661 foram plaqueadas em placas de 24 poços (7000 células por pocinho) de acordo com o esquema a seguir:

Placa 1: GM1661 passagem 11, não selecionadas. Plaqueadas em meio de fibroblasto 15% FBS (placa controle para curva de crescimento de células não selecionadas quando expostas a 6TG)

Placa 2: GM1661 passagem 11, não selecionadas. Plaqueadas em meio de fibroblasto 15%FBS (placa experimental para curva de crescimento de células não selecionadas quando expostas a 6TG)

Placa 3: GM1661 passagem 12, não selecionadas. Plaqueadas em meio de fibroblasto 15% FBS (placa controle para curva de crescimento de células não selecionadas quando expostas a HAT)

Placa 4: GM1661 passagem 12, não selecionadas. Plaqueadas em meio de fibroblasto 15% FBS (placa experimental para curva de crescimento de células não selecionadas quando expostas a HAT)

Placa 5: GM1661 passagem 15, previamente selecionadas em meio 6TG por 33 dias. Plaqueadas em meio 6TG (placa controle para curva de crescimento de células pré-selecionadas em 6TG quando expostas a HAT)

Placa 6: GM1661 passagem 15, previamente selecionadas em 6TG por 33 dias. Plaqueadas em meio 6TG (placa experimental para curva de crescimento de células pré-selecionadas em 6TG quando expostas a HAT)

Após o plaqueamento as células incubadas a 37°C, 5%CO₂.

Dia 2:

Após 48 horas em cultura, as células contidas em três poços (triplicatas) de cada uma das placas foram coletadas por tripsinização, contadas em câmara de Neubauer na presença de *trypan blue* (apenas células viáveis foram computadas) e o número de células presentes em cada poço foi anotado (Dia 0).

Nesse mesmo dia, o meio das placas experimentais foi trocado para o meio de seleção correspondente. No caso das placas controle, o meio foi trocado por meio fresco do mesmo tipo daquele utilizado durante o plaqueamento das células, ou seja, meio de fibroblasto 15% FBS. Após a troca do meio de cultura as células foram retornadas a incubadora a 37°C, 5%CO₂. A troca de meio está especificada a seguir:

Placa 1: GM1661 passagem 11, não selecionadas. Mantidas em meio de fibroblasto 15% FBS (placa controle para curva de crescimento de células não selecionadas quando expostas a 6TG)

Placa 2: GM1661 passagem 11, não selecionadas. Meio substituído para meio 6TG (placa experimental para curva de crescimento de células não selecionadas quando expostas a 6TG)

Placa 3: GM1661 passagem 12, não selecionadas. Mantidas em meio de fibroblasto 15%FBS (placa controle para curva de crescimento de células não selecionadas quando expostas a HAT)

Placa 4: GM1661 passagem 12, não selecionadas. Meio substituído por meio HAT (placa experimental para curva de crescimento de células não selecionadas quando expostas a HAT)

Placa 5: GM1661 passagem 15, previamente selecionadas em 6TG por 33 dias. Mantidas em meio 6TG (placa controle para curva de crescimento de células pré-selecionadas em 6TG quando expostas a HAT)

Placa 6: GM1661 passagem 15, previamente selecionadas em 6TG por 33 dias. Meio substituído por meio HAT (placa experimental para curva de crescimento de células pré-selecionadas em 6TG quando expostas a HAT)

Dia 4 a Dia 14 (Dias 2 a 4 de tratamento em meio seletivo):

A partir daí, a cada dois ou três dias, as células contidas em três poços (triplicatas) de cada placa foram coletadas por tripsinização, contadas em câmara de Neubauer na presença de *trypan blue* (apenas células viáveis foram computadas) e o número de células presentes em cada poço foi também anotado (Dias 2, 4, 6, 8, 10 e 12 de tratamento para placas das curvas de crescimento de células não selecionadas quando expostas ao meio 6TG ou meio HAT, e dias 2, 4, 6, 9 e 12 de tratamento para placas das curvas de crescimento de células pré-selecionadas em 6TG quando expostas ao meio HAT). O meio de cultura foi trocado a cada dois dias por meio fresco correspondente. Sempre que as células das placas controle tornavam-se altamente confluentes, as células tanto da placa controle como da placa experimental correspondente foram passadas para poços de maior área (primeiro para placas de 12 poços e depois para placas de 6 poços). Ao final do período de tratamento, a média do número de células presentes em cada triplicata

foi computada e os dados obtidos foram plotados em Excel para geração das curvas de crescimento correspondentes.

3.17. Escolha da Biblioteca de shRNAs utilizada

A biblioteca de shRNA que elegemos para desenvolver o presente trabalho (*GeneNet™ Lentiviral shRNA Library Human Genome-Wide Screen*, System Biosciences®) é oferecida em duas diferentes versões, *FIV-based* e *HIV-based*, e a escolha da versão a ser utilizada, conforme recomendação da própria empresa, deve ser baseada na taxa de eficiência de transdução de cada um desses tipos de partículas lentivirais no tipo celular alvo. Sendo assim, a fim de se eleger o tipo de partícula lentiviral que possui melhor capacidade de transdução em fibroblastos dérmicos humanos primários, realizamos um experimento piloto com partículas lentivirais contendo o gene repórter copGFP de ambas as versões, FIV e HIV [*HIV & FIV Based Packaged Positive Transduction Control Virus Sampler Kit (pSIF1- and pSIH1-H1-siLuc-copGFP)*, System Biosciences®, Cat# LV401A-1], em diferentes linhagens de fibroblastos dérmicos humanos primários, sendo uma linhagem de fibroblastos humanos fetais (HDF51 de passagem 12) e duas linhagens de fibroblastos humanos adultos (HDF48 de passagem 4 e HDF49 de passagem 12), essas duas últimas plaqueadas em diferentes confluências celulares. O experimento foi realizado conforme descrito abaixo.

Dia 0:

5×10^4 células foram plaqueadas por poço, tanto em placas de 6-well (linhagens HDF48 e HDF49) como em placas de 24-well (linhagens HDF48, HDF49 e HDF51) em meio de fibroblasto 10% FBS (1mL de meio por poço de placa 24-well e 2mL de meio por poço de placa 6-well).

Dia 1:

Vinte e quatro horas após o plaqueamento das células, o meio de cultura foi removido e substituído por 1mL de meio de transdução de fibroblastos contendo as partículas virais em diferentes multiplicidades de infecção (MOI). A determinação da

quantidade de partículas virais necessárias para cada MOI foi previamente calculada de acordo com as concentrações (IFU/mL) dos estoques de partículas lentivirais (fornecido pela empresa System Biosciences®) e o número de células presentes por poço no momento da transdução viral (MOI 1 = 1 partícula viral por célula plaqueada). Para se estimar o número de células presentes por poço no momento da transdução viral, células das linhagens HDF48 e HDF49 contidas em um pocinho foram coletas por tripsinização seguida por adição de *trypan blue* ao pellet celular ressuspendido e contagem em câmara de Neubauer.

Meio de Transdução de Fibroblasto

1XDMEM High Glucose GlutaMAX-I™ (Life Technologies™ - Gibco®, Cat# 10566-016)

10% FBS Heat-Inactivated (Life Technologies™ - Gibco®, Cat# 10438-026)

0.1mM NEAA (Life Technologies™ - Gibco®, 11140-050)

10ng/mL polibreno (*Hexadimethrine bromide*, Sigma-Aldrich®, Cat# H9268)

As partículas virais foram adicionadas nas seguintes quantidades sendo que o mesmo foi realizado para cada uma das linhagens de fibroblastos primários humanos estudadas (exceto poços controle sem polibreno, os quais foram gerados apenas células da linhagem HDF48 e HDF51) e para as diferentes porcentagens de confluência celular analisadas:

Poço 1: FIV-copGFP MOI=1.

Poço 2: FIV-copGFP MOI=5.

Poço 3: FIV-copGFP MOI=10.

Poço 4: FIV-copGFP MOI=20.

Poço 1: HIV-copGFP MOI=1.

Poço 2: HIV-copGFP MOI=5.

Poço 3: HIV-copGFP MOI=10.

Poço 4: HIV-copGFP MOI=20.

Poço 9: Controle (sem polibreno).

Poço 10: Controle *Mock* (com polibreno).

Após a adição das partículas lentivirais, as placas foram gentilmente homogeneizadas e retornadas a incubadora a 37°C, 5%CO₂.

Dia 2 (24h após transdução):

O meio contendo as partículas lentivirais foi removido, cada poço foi lavado uma vez com 1XDPBS (Life Technologies™ - Gibco®, Cat# 14190-144, 1mL para cada poço de placa 24-well e 2mL para cada poço de placa 6-well), o meio de fibroblasto a temperatura ambiente foi adicionado (1mL para cada poço de placa 24-well e 2mL para cada poço de placa 6-well) e as placas retornadas à incubadora a 37°C, 5%CO₂.

Dia 3 (48h após transdução):

As células foram visualizadas e fotografadas em microscópio de fluorescência em diversas magnitudes de aumento, tanto sob campo de luz comum como sob campo de luz ultravioleta.

Dia 4 (72h após transdução):

As células foram coletadas por tripsinização em tubos eppendorf de 1.5mL de capacidade e os pellets foram ressuspensos em 500uL de 1XDPBS (Life Technologies™ - Gibco®, Cat# 14190-144) contendo 10% FBS. As amostras controle com ou sem polibreno foram utilizadas separadamente sempre que possível ou então combinadas a fim de se obter um maior número de células para calibração do citômetro. Os *pellets* celulares foram colocados em gelo até o momento da análise do número de células copGFP positivas, a qual foi realizada por citometria de fluxo em aparelho *FACSCalibur*™ (BD Biosciences™).

3.18. Otimização da Transdução de Fibroblastos Primários Humanos da linhagem GM1661 com Partículas Lentivirais copGFP

Lentivírus baseados em HIV contendo o gene repórter copGFP (*pSIH1-H1-siLuc-copGFP Packaged Positive Transduction Control*, System Biosciences®, Cat # LV601B-1) fornecidos junto com a biblioteca lentiviral de shRNAs foram utilizados para transduzir fibroblastos primários humanos da linhagem GM1661 de passagem

17 previamente selecionados em 6TG em diferentes porcentagens de confluência celular. Foram testadas diferentes proporções de número de vírions (IFU) por número de células a fim de se determinar a MOI adequada para que cada célula transduzida recebesse, de maneira geral, não mais do que uma partícula lentiviral. Para esse experimento, o reagente *TransDux*[™] (Cat# LV850A-1), comercializado pela System Biosciences® e recomendado pela empresa para a transdução das células-alvo com a biblioteca de shRNAs, foi utilizado ao invés do polibreno como agente facilitador da transdução viral. O experimento de transdução foi realizado conforme descrito a seguir:

Dia 0:

4x10⁵ ou 5x10⁵ células da linhagem GM1661 de passagem 17 previamente selecionadas em meio de cultura contendo 6TG foram plaqueadas por poço em placas 24-well. As células foram plaqueadas em meio de fibroblasto 10% FBS (1mL de meio por poço).

Dia 1:

Vinte e quatro horas após o plaqueamento das células, o meio de cultura foi removido e substituído por 0.3mL de meio *TransDux*[™] contendo as partículas virais em diferentes multiplicidades de infecção (MOI). A determinação da quantidade de partículas virais necessárias para se obter cada MOI foi previamente calculada de acordo com as concentrações (IFU/mL) dos estoques de partículas lentivirais (fornecido pela empresa System Biosciences®) e o número de células presentes por poço no momento da transdução viral (MOI 1 = 1 partícula viral por célula plaqueada). Para se estimar o número de células presentes por poço no momento da transdução viral, células contidas em um pocinho foram coletas por tripsinização seguida por adição de *trypan blue* ao *pellet* celular ressuspensionado e contagem em câmara de Neubauer.

Meio *TransDux*TM

1XDMEM High Glucose (Life TechnologiesTM - Gibco®, Cat# 11965-092)

10% FBS (Life TechnologiesTM - Gibco®, Cat# 12657-029)

0.1mM MEM NEAA (Life TechnologiesTM - Gibco®, Cat# 11140-050)

1mM Sodium Piruvate (Life TechnologiesTM - Gibco®, Cat# 11360-070)

2mM GlutaMAXTM Supplement (Life TechnologiesTM - Gibco®, Cat# 35050-061)

50U/mL Penicillin/ 50ug/mL Streptomycin (Life TechnologiesTM - Gibco®, Cat# 15140-122)

1X TransDuxTM Virus Transduction Reagent (System Biosciences®, Cat# LV850A-1)

As partículas virais foram adicionadas nas seguintes quantidades:

Poço 1: 5×10^4 células/ poço HIV-copGFP MOI=2.

Poço 2: 5×10^4 células/ poço HIV-copGFP MOI=3.

Poço 3: 5×10^4 células/ poço HIV-copGFP MOI=4.

Poço 4: 5×10^4 células/ poço HIV-copGFP MOI=5.

Poço 5: 5×10^4 células/ poço HIV-copGFP MOI=10.

Poço 6: 5×10^4 células/ poço Controle *Mock* (com *TransDux*TM).

Poço 7: 5×10^4 células/ poço Controle (sem *TransDux*TM).

Poço 8: 4×10^4 células/ poço HIV-copGFP MOI=4.

Poço 9: 4×10^4 células/ poço Controle *Mock* (com *TransDux*TM).

Após a adição das partículas lentivirais, as placas foram gentilmente homogeneizadas e retornadas à incubadora a 37°C, 5%CO₂.

Dia 2 (24h após transdução):

O meio contendo as partículas lentivirais foi removido, cada poço foi lavado uma vez com 1XDPBS (Life TechnologiesTM - Gibco®, Cat# 14190-144, 1mL para cada poço de placa 24-well), o meio de fibroblasto 15% FBS a temperatura ambiente foi adicionado (1mL para cada poço de placa 24-well) e as placas retornadas à incubadora a 37°C, 5%CO₂.

Dia 3 (48h após transdução):

As células foram visualizadas e fotografadas em microscópio de fluorescência em diversas magnitudes de aumento, tanto sob campo de luz comum como sob campo de luz ultravioleta.

Dia 4 (72h após transdução):

As células foram coletadas por tripsinização em tubos eppendorf de 1.5mL de capacidade e os pellets foram ressuspensos em 100uL de 1XDPBS (Life Technologies™ - Gibco®, Cat# 14190-144) contendo 10% FBS. As amostras controle com e sem *TransDux*™ plaqueadas foram combinadas a fim de se obter um maior número de células para calibração do citômetro. Os *pellets* celulares foram colocados em gelo até o momento da análise do número de células copGFP positivas, a qual foi realizada por citometria de fluxo em aparelho Guava easyCyte™ (EMD Millipore®).

3.19. Transdução de Fibroblastos Primários Humanos com a Biblioteca de shRNAs

As células GM1661 utilizadas no experimento de transdução com a biblioteca de shRNA foram previamente selecionadas em meio 6TG em frascos aderentes pré-tratados com solução de gelatina a 0,1% (Stem Cell Technologies™, Cat# 07903) por 58 dias. A biblioteca de shRNA escolhida foi a biblioteca “*GeneNet™ Lentiviral shRNA Library Human Genome-Wide Screen*” (System Biosciences®). O experimento de transdução foi realizado com base nas instruções fornecidas pelo fabricante da biblioteca de shRNA e nos experimentos de otimização da transdução de fibroblastos primários descritos acima. O protocolo utilizado está especificado a seguir:

Dia 0:

Dez milhões de células da linhagem GM1661 (passagem 21) previamente selecionadas em meio 6TG foram distribuídas entre nove frascos T25 (sem

gelatina), cada um contendo 4mL de meio de fibroblasto 10% FBS e mantidas em incubadora a 37°C, 5%CO₂ por cerca de 24 horas.

Dia 1:

O criotubo contendo a biblioteca de shRNA foi rapidamente descongelado em banho-maria a 37°C e transferido imediatamente para o fluxo laminar onde foi mantido em gelo. A quantidade apropriada de partículas virais (determinada no experimento de titulação descrito acima como sendo correspondente a MOI=1,8) foi diluída em meio *TransDux*TM (2mL por frasco T25) em tubos Falcon de 50mL de capacidade e a mistura foi homogeneizada gentilmente. O meio de cultura dos frascos contendo as células foi removido e o meio *TransDux*TM contendo as partículas virais foi adicionado. Os frascos foram retornados a incubadora a 37°C, 5%CO₂ por 24 horas.

Dia 2 (24h após transdução):

O meio de cultura contendo as partículas virais foi removido e 4mL de meio de fibroblasto 15% FBS (contendo 50U/mL penicilina e 50ug/mL estreptomicina) foram adicionados a cada frasco. As células foram retornadas a incubadora a 37°C, 5%CO₂ por 24 horas.

Dia 3 (48 após transdução):

As células foram tripsinazadas, peletadas por centrifugação, contadas em câmara de Neubauer e plaqueadas em 24 frascos T125 contendo 24mL de meio de fibroblasto 15% FBS (50U/mL penicilina e 50ug/mL estreptomicina) e em 4 poços de uma placa de 12-well (contendo 2mL de meio de fibroblasto 15% FBS (50U/mL penicilina e 50ug/mL estreptomicina)). As células contidas nos frascos T125 foram destinadas a obtenção das amostras submetidas para NGS, enquanto que as células contidas nos poços da placa de 24-well foram destinadas a seleção por puromicina a fim de que se pudesse estimar a taxa de transdução viral. O número de células presente em cada um dos poços da placa de 12-well está especificado a seguir:

Poço 1: 10^3 células (para tratamento com puromicina)
Poço 2: 5×10^3 células (para tratamento com puromicina)
Poço 3: 10^4 células (para tratamento com puromicina)
Poço 4: 5×10^3 células (controle não-tratado com puromicina)

Após o plaqueamento as células foram retornadas a incubadora a 37°C , $5\%\text{CO}_2$ por 24 horas.

Dia 4 ao Dia 12 (para células destinadas ao NGS):

O meio de fibroblasto foi substituído por 24mL de meio 6TG (12 frascos) ou meio HAT (12 frascos) contendo 50U/mL penicilina e 50ug/mL estreptomicina. A partir daí, o meio 6TG ou meio HAT foi trocado a cada dois dias. Ao final do período de oito dias de tratamento em 6TG ou HAT as células foram coletadas por tripsinização, o *pellet* foi lavado duas vezes em 1XDBBS e congelado em freezer a -80°C até o momento da extração de DNA genômico total.

Dia 4 ao Dia 14 (para células destinadas ao tratamento com puromicina):

O meio de fibroblasto foi substituído por 2mL de meio fibroblasto 15% FBS contendo 400ng/mL de puromicina (*Puromycin Dihydrochloride*, Life Technologies™ - Invitrogen™, Cat# A1113803), 50U/mL penicilina e 50ug/mL estreptomicina, exceto para o poço controle que recebeu meio de fibroblasto 15% FBS contendo 50U/mL penicilina e 50ug/mL estreptomicina sem puromicina. A partir daí, o meio de cultura foi trocado a cada dois dias. Ao final do período de dez dias de tratamento em puromicina, o número de colônias resistentes ao tratamento com puromicina foi contado sob microscópio de luz para cada um dos pocinhos tratados. Adicionalmente, as células foram coletadas por tripsinização, o *pellet* foi ressuspensionado em meio de fibroblasto 15% FBS e as células foram contadas em câmara de Neubauer na presença de *trypan blue*. O número de células presentes em cada um dos pocinhos tratados com puromicina foi proporcionalmente comparado com o número de células presentes no poço não tratado.

3.20. Extração e Quantificação de DNA dos Fibroblastos Transduzidos com a Biblioteca de shRNA

A extração de DNA genômico total dos fibroblastos dérmicos humanos primários transduzidos com a biblioteca de shRNA foi realizada utilizando-se o kit “*Flexigene DNA Kit*” (Qiagen®, Cat# 51206), com base nas instruções fornecidas pelo fabricante com as únicas modificações de que os *pellets* celulares ressuspensos em Buffer FG1 e Buffer FG2 contendo protease foram incubados em banho-maria a 65°C por 45min e em seguida cada lisado foi passado por algumas vezes em seringas conectadas a agulhas de insulina para garantir uma melhor fragmentação das células.

As amostras de DNA extraídas após o experimento com a biblioteca de shRNA foram mantidas em gelo e quantificadas em fluorômetro (*Qubit® Fluorometer*, Life Technologies™ - Invitrogen™), utilizando-se reagentes (*Qubit® dsDNA HS BR assay kit*, Cat# Q32851, Life Technologies™ - Invitrogen™) e tubos plásticos apropriados (*Qubit® Assay Tubes*, Cat# Q32856, Life Technologies - Invitrogen) e seguindo-se as instruções do fabricante. As amostras foram armazenadas em freezer a -20°C até o momento da amplificação por PCR dos insertos de shRNAs.

3.21. Multiplex PCR para amplificação dos insertos de shRNAs

A reação de *multiplex* PCR para amplificação dos insertos de shRNAs presentes nas amostras de fibroblastos primários humanos transduzidos com a biblioteca lentiviral e selecionados em meio contendo 6TG ou HAT foi otimizada e realizada de acordo com as instruções fornecidas no manual do fabricante da biblioteca de shRNA (System Biosciences®). Foram realizadas duas rodadas de amplificação, sendo que a segunda rodada consistiu em uma *nested* PCR na qual foram utilizados primers internos ao produto de PCR proveniente da primeira rodada de amplificação (Figura 10).

plásticos apropriados (*Qubit® Assay Tubes*, Cat# Q32856, Life Technologies - Invitrogen) e seguindo-se as instruções do fabricante.

Primeira rodada de amplificação (*Nested PCR*):

Para a primeira rodada de amplificação, o DNA genômico total extraído das células transduzidas com a biblioteca de shRNAs (selecionadas em 6TG e HAT) e de células da linhagem GM1661 não-transduzidas (controle negativo) ou o DNA plasmidial (controle positivo) foram aliquotados em tubos plásticos de 0.2mL de capacidade para cada amostra obtida. Um tubo “branco”, sem qualquer DNA, também foi preparado. Em seguida, foi preparada uma mistura contendo os demais componentes da PCR. Os tubos foram mantidos em gelo durante todo o processo de preparação da PCR. As quantidades finais utilizadas de cada componente das reações da primeira rodada de PCR estão especificadas a seguir, em ordem de adição ao tubo:

DNA genômico (concentração final: 50ng por uL de reação de PCR) ou DNA plasmidial (concentração final: 100pg por uL de reação de PCR)

Água MiliQ duplamente autoclavada (qsvf)

1X *Titanium® Taq DNA Polymerase Buffer* (Clontech®, Cat # 639208)

200 uM de cada dNTP (dATP, dTTP, dGTP, dCTP)

200uM *Primer Forward* (*Fwd GNH PCR Primer*, fornecido com a biblioteca de shRNA)

200uM *Primer Reverse* (*Rev GNH PCR Primer*, fornecido com a biblioteca de shRNA)

1X *Titanium® Taq DNA Polymerase* (Clontech®, Cat # 639208)

Foram preparados dois tubos contendo 100uL de reação de PCR para cada uma das amostras experimentais (6TG e HAT), um tubo contendo 20uL de reação de PCR para a amostra “controle negativo”, um tubo contendo 20uL de reação de PCR para a amostra “controle positivo” e um tubo contendo 100uL de reação de PCR para a amostra “branco”.

Após adição da mistura contendo os demais componentes da reação de PCR aos tubos contendo DNA, as amostras foram homogeneizadas pipetando-se suavemente o conteúdo do tubo para cima e para baixo e, então, transferidas imediatamente para o termociclador (*Veriti® 96-well Fast Thermal Cycler*, Life Technologies™ – Applied Biosystems®). O programa utilizado para a primeira rodada de amplificação está especificado a seguir:

94°C por 4min (desnaturação inicial)
94°C por 30seg, 68°C por 1min (18 ciclos)
68°C por 3min (extensão final)
15°C hold

Uma vez completado o programa de termociclagem, as amostras foram imediatamente armazenadas a -20°C.

Segunda rodada de amplificação (*Nested* PCR):

Os produtos da primeira rodada de amplificação foram descongelados em um recipiente contendo gelo e então utilizados na segunda reação de amplificação dos insertos de shRNA (*Nested* PCR).

Para a segunda rodada de amplificação, os produtos de PCR provenientes da primeira rodada de amplificação (6TG, HAT, controle positivo, controle negativo e “branco”) foram aliquotados em tubos plásticos de 0,2mL de capacidade para cada amostra obtida. Em seguida, foi preparada uma mistura contendo os demais componentes da PCR. Os tubos foram mantidos em gelo durante todo o processo de preparação da PCR. As quantidades finais utilizadas de cada componente das reações da segunda rodada de PCR estão especificadas a seguir, em ordem de adição ao tubo:

Produto da primeira PCR (0,05uL de produto da primeira PCR por uL de reação da segunda PCR)

Água MiliQ duplamente autoclavada (qsvf)

1X *Titanium® Taq DNA Polymerase Buffer* (Clontech®, Cat # 639208)

200 uM de cada dNTP (dATP, dTTP, dGTP, dCTP)

1mM *Primer Nested Reverse (NRev GNH Universal PCR Primer*, fornecido com a biblioteca de shRNA)

1mM *Primer Nested Forward Biotinilado (NFwd Bio Universal PCR Primer*, fornecido com a biblioteca de shRNA)

1X *Titanium® Taq DNA Polymerase* (Clontech®, Cat # 639208)

Foram preparados seis tubos contendo 100uL de reação de PCR para cada uma das amostras experimentais (6TG e HAT), um tubo contendo 100uL de reação de PCR para a amostra “controle negativo”, um tubo contendo 100uL de reação de PCR para a amostra “controle positivo” e um tubo contendo 100uL de reação de PCR para a amostra “branco”.

Após adição da mistura contendo os demais componentes da reação de PCR aos tubos contendo DNA, as amostras foram homogeneizadas pipetando-se suavemente o conteúdo do tubo para cima e para baixo e então, transferidas imediatamente para o termociclador (*Veriti® 96-well Fast Thermal Cycler*, Life Technologies™ – Applied Biosystems®). O programa utilizado para a segunda rodada de amplificação (*Nested PCR*) está especificado a seguir:

94°C por 2min, 50°C por 2min, 68°C por 1min (1 ciclo)

94°C por 30seg, 68°C por 30seg (18 ciclos)

68°C por 3min (extensão final)

15°C hold

Uma vez completado o programa de termociclagem, 2mL de cada amostra proveniente da segunda rodada de PCR (*Nested PCR*) foram corridas em gel de agarose 2.5% (Life Technologies™ - Invitrogen™, Cat# 16500-100) contendo brometo de etídeo (UltraPure™ 10 mg/ml Ethidium Bromide, Life Technologies™ - Invitrogen™, Cat# 15585-011) em cuba de eletroforese contendo tampão 0.5X TBE

a 100V. O gel foi visualizado e fotografado sob luz U.V.. O restante das amostras foi imediatamente armazenado a -20°C.

3.22. Purificação dos amplicons contendo shRNAs

As amostras HAT e 6TG provenientes da segunda rodada de PCR (*Nested* PCR) foram corridas em gel de agarose 2.5% – contendo 1.25% de agarose comum (Life Technologies™ - Invitrogen™, Cat# 16500-100) e 1.25% de agarose *low-melting point* (Life Technologies™ - Invitrogen™, Cat# 15517-014) e brometo de etídeo (*UltraPure™ 10 mg/ml Ethidium Bromide*, Life Technologies™ - Invitrogen™, Cat# 15585-011) – em cuba de eletroforese contendo tampão 0.5X TBE a 100V. O gel foi rapidamente visualizado e fotografado sob luz U.V. e as bandas correspondentes ao fragmento de interesse foram rapidamente excisadas do gel, também sob luz U.V., com o auxílio de lâminas, e transferidas para um tubo plástico estéril de 15mL de capacidade.

Os amplicons contidos nos fragmentos de agarose excisados do gel foram extraídos utilizando-se o kit “*MinElute Gel Extraction Kit*” (QIAgen®, Cat# 28604) com as seguintes modificações feitas ao protocolo fornecido pelo fabricante – os fragmentos de gel contendo os amplicons foram incubados em seis volumes de QG Buffer em banho-maria por 20min, 10uL de acetato de sódio, pH 5.0 foram adicionados a cada amostra mesmo que a coloração do tampão QG não tenha sido alterada, as colunas contendo as amostras ligadas foram lavadas em 750uL de PE Buffer por duas vezes, sendo que antes da segunda centrifugação para remoção do PE Buffer as colunas foram incubadas a temperatura ambiente por 3min e as colunas foram eluídas duas vezes em 10uL de Buffer EB cada vez, sendo que as amostras foram incubadas a temperatura ambiente por 2min antes de cada centrifugação associadas a eluição, as quais foram feitas por 2min cada. Os procedimentos de eletroforese, excisão das bandas de interesse e purificação de DNA a partir de agarose foram repetidos por três vezes até que todo o volume das reações de *nested* PCR (6 tubos de 100uL para cada amostra, controle e experimental) tivesse sido processado e os amplicons isolados nas três rodadas de purificação foram combinados para cada amostra.

As amostras de DNA eluídas da agarose foram mantidas em gelo e quantificadas em fluorômetro (*Qubit® Fluorometer*, Life Technologies™ - Invitrogen™), utilizando-se reagentes (*Qubit® dsDNA HS BR assay kit*, Cat# Q32851, Life Technologies™ - Invitrogen™) e tubos plásticos apropriados (*Qubit® Assay Tubes*, Cat# Q32856, Life Technologies - Invitrogen) e seguindo-se as instruções do fabricante.

Para verificação dos produtos purificados, pequenas alíquotas das combinações de amostras eluídas (2,5uL) foi corrida novamente, ao lado de alíquotas dos produtos de *nested* PCR correspondentes não purificados (2uL), em gel de agarose 2.5% (Life Technologies™ - Invitrogen™, Cat# 16500-100) contendo brometo de etídeo (*UltraPure™ 10 mg/ml Ethidium Bromide*, Life Technologies™ - Invitrogen™, Cat# 15585-011) em tampão 0.5X TBE. O gel de agarose foi visualizado sob luz U.V. e fotografado. O restante do material foi armazenado em freezer -20°C até o momento de envio das amostras para realização do seqüenciamento de próxima geração (NGS) pela empresa Fasteris na Suíça.

3.23. Seqüenciamento de Próxima Geração (NGS)

As duas amostras de amplicons (amostra controle de células transduzidas com a biblioteca de shRNAs e mantidas em meio 6TG e amostra experimental de células transduzidas com a biblioteca de shRNAs e selecionadas em HAT) purificadas do gel de agarose e quantificadas foram enviadas para a empresa Fasteris, na Suíça, onde foram realizados os experimentos de seqüenciamento de próxima geração. O resumo do plano experimental adotado pela empresa Fasteris está descrito no Anexo 1. A amostra controle 6TG foi denominada pela empresa Fasteris “GZR-1” e a amostra experimental HAT foi denominada “GZR-2”.

Na empresa Fasteris, as duas amostras de amplicons purificados do gel de agarose foram primeiramente submetidas a um teste de controle de qualidade de DNA. As amostras foram então utilizadas para a construção das duas respectivas bibliotecas de amplicons que foram avaliadas por corrida em aparelho LabChip® GX-LST (Perkin-Elmer®, Caliper Life Sciences) com auxílio do programa “LabChip GX Software Version 3.1.938.0 SP1” (Perkin-Elmer®, Caliper Life Sciences) para

que se pudesse verificar a pureza das amostras e o tamanho médio dos amplicons das bibliotecas (~83pb de tamanho do amplicon + 120pb do adaptador).

Uma pequena alíquota de cada biblioteca de amplicons foi então submetida a uma corrida teste em canal 1x50bp de aparelho de seqüenciamento de próxima geração HiSeq 2000 (Illumina®). Após a corrida teste, a empresa forneceu, para cada amostra, 50 seqüências em formato fasta e cerca de vinte mil seqüências em formato fastq, de cerca de 50 nucleotídeos cada, para cada biblioteca de amplicons, com o objetivo de verificar a qualidade das mesmas e confirmar a presença e pureza dos amplicons esperados. Após a análise dos dados obtidos na corrida teste, as bibliotecas de amplicons foram submetidas à corrida final em aparelho de seqüenciamento de próxima geração HiSeq 2500 (Illumina®) em canal 1x100pb de modo que cerca de 100 nucleotídeos presentes em cada amplicon foram seqüenciados. O procedimento de *base-calling* foi realizado através da utilização dos programas “HiSeq Control Software version 2.0.5”, “RTA version 1.17.20.0” e “Consensus Assessment of Sequence and Variation - CASAVA version 1.8.2” (todos da Illumina®) e as seqüências obtidas por NGS foram anotadas em formato fastq. O formato fastq está descrito no documento “*Quality format*” fornecido pela empresa Fasteris (Anexo 2).

3.24. Análise dos Dados de NGS

3.24.1. Corrida Teste

A análise dos dados obtidos na corrida teste de seqüenciamento de próxima geração foi realizada com o auxílio do pesquisador associado Robert E. Morey e da pós-doutoranda Rathi D. Thiagarajan, PhD, ambos do Laboratório da Professora Assistente Louise Laurent, MD, PhD, *Department of Reproductive Medicine, University of California San Diego* (UCSD), a partir dos arquivos contendo 20 mil seqüências correspondentes a cada uma das amostras fornecidos pela empresa Fasteris. Para essa análise, a qualidade das seqüências obtidas para cada amostra foi primeiramente analisada através do programa “FastQC version 0.10.0” (disponível em <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

A sequência de nucleotídeos referente ao *primer forward* utilizado para a amplificação dos insertos de shRNAs, bem como a sequência de nucleotídeos lentiviral que flanqueia os shRNAs na extremidade 3', foram excisadas com o uso do programa "Cutadapt version 1.0" (disponível em <http://code.google.com/p/cutadapt/> ou <https://github.com/marcelm/cutadapt>) conforme exemplificado a seguir:

5'–CTTCCTGTCAGACTGATAGGATATGTTGGTTGCATTTCCTTTTGAATTC– 3'

Onde:

Primer Forward (sequência de 12pb localizada no hairpin do shRNA, cortada)
Sequência de 27pb considerada (referente à segunda metade do shRNA, mantida)
Sequência plasmidial (cortada)

A preparação dos bancos de dados foi realizada utilizando-se o programa Bowtie-Build (distribuído com o Bowtie). As sequências referentes a cada um dos insertos de shRNA presentes em cada amostra foram mapeadas através do uso do programa "Bowtie" (disponível em <http://bowtie-bio.sourceforge.net/index.shtml>) contra o banco de dados das sequências-alvo dos shRNAs que compõem a biblioteca lentiviral (fornecido pela empresa System Biosciences®) de forma a permitir zero ou um *mismatches*, com o objetivo de verificar se as sequências dos amplicons obtidas por NGS realmente correspondiam às sequências dos shRNAs presentes na biblioteca lentiviral. As sequências-alvo de shRNAs contidas nesse banco de dados referem-se às sequências de sondas presentes no chip de microarray *GeneChip® Human Genome U133 Plus 2.0 Array* (Affymetrix®, Cat# 900470) que serviu de base para o desenho dos shRNAs que compõem a biblioteca lentiviral pela empresa que comercializa a biblioteca.

3.24.2. Corrida Final

A análise dos dados obtidos na corrida final de seqüenciamento de próxima geração foi realizada com o auxílio e sob supervisão da Professora Dr Tatiana

Teixeira Torres (Laboratório de Genômica e Evolução de Artrópodes, Departamento de Genética e Biologia Evolutiva, IB-USP). A qualidade das sequências obtidas para cada amostra foi primeiramente analisada através do programa “FastQC version 0.10.0” (disponível em <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

As sequências obtidas por NGS foram alinhadas contra três bancos de dados diferentes: o banco de dados de sequências-alvo de shRNAs, o banco de dados das sequências de shRNAs propriamente ditas (ambos fornecidos pela empresa System Biosciences®) e o banco de dados RefMrna (disponível em <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/>). Antes do alinhamento contra o banco de dados contendo sequências de shRNAs propriamente ditas, as sequências presentes no banco de dados foram cortadas para remoção das regiões correspondentes a primeira metade (região localizada antes do *hairpin*) de cada shRNA, segundo exemplo abaixo:

>C_1007_s_at_0_f (identificação da sequência)

5' –AGCATAGGATCCGCGCCCAGTTGGTCTTGTGGGTGGGATCCTTCCTGTCAGAGATCCC
ATCCACAGGACCAGCTGGGTGTTTTTGAATTCATGCGT– 3'

Onde:

Sequência plasmidial (cortada)

Sequência referente à primeira metade do shRNA (cortada)

Primer Forward (sequência de 12pb localizada no *hairpin* do shRNA, mantida)

Sequência de 27pb considerada (referente à segunda metade do shRNA, mantida)

Sequência plasmidial (mantida)

Sequência incongruente com a obtida por NGS (cortada)

A preparação dos bancos de dados foi realizada utilizando-se o programa Bowtie-Build (distribuído com o Bowtie). O alinhamento das sequências obtidas por NGS para cada amostra, controle e experimental, contra as sequências presentes em cada um dos bancos de dados foi realizada através do programa “Bowtie version 0.12.7” (disponível em <http://bowtie-bio.sourceforge.net/index.shtml>) em plataforma “R” (disponível em <http://www.r-project.org/>) de forma a não permitir nenhum *mismatch* e com o uso dos seguintes comandos:

-q -n 0 -l 27 -5 12 -3 61 -a

onde,

- q: denota que os arquivos de entrada estão no formato fastq
- n 0: número máximo de mismatches na sequência *seed* igual a zero
- l 27: número de nucleotídeos na sequência *seed*
- 5 12: número de nucleotídeos cortados na extremidade 5' igual a 12
- 3 61: número de nucleotídeos cortados na extremidade 3' igual a 61
- a: denota que todos os alinhamentos por *read* devem ser relatados

Dessa forma, os primeiros 12 nucleotídeos presentes em cada amplicon (referentes ao *primer forward*), assim como os 61 nucleotídeos presentes no final de cada amplicon (referentes às seqüências plasmídias e ao *primer reverse*) não foram considerados no alinhamento. Sendo assim, apenas os 27 nucleotídeos presentes na segunda metade de cada shRNA foram alinhados contra os bancos de dados, conforme exemplificado abaixo:

Amplicon

5' –CTTCCTGTCAGACTGATAGGATATGTTGGTTGCATTTCTTTTTGAATTCATGACCAA
TTCTTCGATTCTGCTTTTGTCTTCTAGATCGGAAGAGCGGTT– 3'

Onde:

Primer Forward (seqüência de 12pb localizada no hairpin do shRNA, cortada durante alinhamento)

Seqüência de 27pb considerada durante alinhamento (referente à segunda metade do shRNA)

Seqüência plasmidial (cortada durante alinhamento)

Primer Reverse (cortada durante alinhamento)

Seqüência referente ao adaptador utilizado na preparação da biblioteca de amplicons para NGS (cortada durante alinhamento)

Após os alinhamentos terem sido completados, o número de amplicons alinhados com cada uma das seqüências presentes no banco de dados foi contado com o uso do programa “script bowtie_parser.pl” em plataforma “R” de forma a gerar uma tabela para cada amostra em relação a cada um dos bancos de dados (resumo dos alinhamentos). As duas tabelas de alinhamento obtidas para as amostras controle e experimental foram combinadas, para cada tipo de banco de dados utilizado, através do uso do programa “script combine_tables.pl” em plataforma “R”. Esses dois últimos programas foram desenvolvidos em linguagem de *script* “Perl version” (disponível em <http://www.perl.org/>) pela professora Dr Tatiana Teixeira Torres.

Os dados obtidos para os alinhamentos contra cada banco de dados contidos em cada tabela combinada foram então analisados por distribuição binomial negativa em plataforma “R” através do programa “DESeq” (disponível em <http://bioconductor.org/packages/release/bioc/html/DESeq.html>) (Anders & Huber, 2010) de acordo com os tutoriais fornecidos para as respectivas versões utilizadas. A versão 1.12.0 foi utilizada para análise dos alinhamentos obtidos contra os bancos de dados de seqüências-alvo e contra o banco de dados de seqüências de shRNAs propriamente ditas e a versão 1.14.0 foi utilizada para análise dos alinhamentos obtidos contra o banco de dados RefMrna. As análises foram realizadas utilizando-se os parâmetros “single-end” (uma vez que as sequencias foram geradas no formato 1x100pb) e, para a estimativa de dispersão, *method="blind"*, *sharingMode="fit-only"* (usado para instruir o programa da ausência de réplicas biológicas durante o cálculo da variância).

A análise pelo programa DESeq resultou na construção de tabelas, referentes a cada um dos bancos de dados utilizados, nas quais a identificação das seqüências (nomes das seqüências-alvo, seqüências de shRNA ou RefSeq dos genes) presentes nos bancos de dados que encontravam-se diferencialmente expressas nas amostras controle e experimental, de maneira estatisticamente significativa (com p ajustado <0.1), foram listadas.

As sequencias candidatas obtidas através dos alinhamentos contra os bancos de dados de sequencias-alvo e sequencias de shRNAs propriamente ditos foram recuperadas do banco de dados a partir de suas identificações e posteriormente alinhadas para seqüências altamente similares (megablast) contra o genoma e transcriptoma humanos através da ferramenta disponível no site do

National Center for Biotechnology Information - NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome). Os nomes, números de acesso, localizações cromossômicas e coordenadas genômicas dos genes que continham regiões correspondentes altamente similares (*query coverage* 100% e *identity* >90%) às sequências candidatas, bem como as regiões genômicas altamente similares (*query coverage* 100% e *identity* 100%) às sequências candidatas, foram anotados.

3.25. Estudo das Características e Funções dos Genes Candidatos para Manutenção da XCI Identificados na Triagem Funcional

As características e funções dos genes/proteínas candidatos que apresentaram representação significativamente mais elevada em nossa amostra experimental HAT do que em nossa amostra controle 6TG foram investigadas na literatura e nos bancos de dados GeneCards (disponível em <http://www.genecards.org/>), NCBI AceView (disponível em <http://www.ncbi.nlm.nih.gov/IEB/Research/AceView/index.html?human>) (Thierry-Mieg & Thierry-Mieg, 2006) e *Online Mendelian Inheritance in Man*® - OMIM® (disponível em <http://omim.org/>). As informações coletadas foram utilizadas para eleger os candidatos com maior potencial de envolvimento no processo de manutenção da XCI identificados em nosso trabalho.

Resultados

4. Resultados

4.1. Escolha das linhagens celulares utilizadas

A primeira etapa do trabalho consistiu em eleger uma linhagem-alvo de fibroblastos primários humanos 46,XX $HPRT^{+/-}$ a ser utilizada nos experimentos de transdução com a biblioteca de shRNAs. A escolha dessa linhagem-alvo foi baseada em diversos fatores, como disponibilidade da mesma em nosso laboratório, presença de SNPs informativos em regiões codificantes de genes ligados ao cromossomo X e submetidos ao processo de inativação cromossômica (para que se fosse possível determinar a ocorrência de reativação da expressão desses genes durante validação dos genes candidatos por *knockdown* a partir de shRNAs individuais) e taxa de eficiência de transdução viral.

Dentre as linhagens $HPRT^{+/-}$ disponíveis em nosso laboratório, cinco eram do sexo feminino (linhagens GM13A, GM135, GM1660, GM1661, e GM2226) (Tabela 1) e foram testadas em experimentos de genotipagem para SNPs (tanto a partir de DNA como a partir de cDNA) presentes em diferentes genes ligados ao cromossomo X que são sabidamente submetidos ao processo de inativação cromossômica e experimentos para acessar a taxa de susceptibilidade de cada uma dessas linhagens à infecção viral.

Além dessas linhagens $HPRT^{+/-}$ outras linhagens de fibroblasto foram utilizadas em experimentos realizados para se testar a taxa de susceptibilidade de fibroblastos primários humanos à infecção por diferentes tipos de retrovírions (HDF48, HDF49, HDF51 e 3T3, essa última, de camundongo, utilizada como controle) e também em experimentos realizados para se determinar o tempo necessário de tratamento em meio contendo 6-tioguanina para que células $HPRT^{+}$ sejam depletadas da cultura (linhagens Hs27 e FibN1). Os resultados desses experimentos estão descritos a seguir.

4.2. Análise da Eficiência de Transdução Viral em Linhagens Femininas de Fibroblastos Dérmicos Humanos Primários *HPRT*^{+/-}

A biblioteca de shRNAs escolhida para esse trabalho (*GeneNet 50K Human Genome-Wide Lentiviral shRNA Library*, System Biosciences®) é constituída por cerca de 200 mil diferentes shRNAs pré-empacotados em partículas lentivirais que têm como alvo cerca de 47400 transcritos de mRNA humanos. Sendo assim, e considerando que diferentes linhagens, inclusive de um mesmo tipo celular, podem apresentar diferentes susceptibilidades à infecção por partículas virais em cultura, o primeiro passo desse trabalho consistiu em determinar as taxas de eficiência de transdução viral para as linhagens femininas de fibroblastos dérmicos humanos primários *HPRT*^{+/-} disponíveis em nosso laboratório a fim de selecionar linhagens celulares que apresentassem alta eficiência de transdução viral.

Para esse primeiro experimento, retrovírions contendo o gene repórter EGFP empacotados e cedidos pela Dra Patrícia Cristina Baleeiro Beltrão Braga do Instituto Butantan, foram utilizados para transduzir células das linhagens de fibroblastos humanos primários 46, XX *HPRT*^{+/-} disponíveis em nosso laboratório (GM13A, GM135, GM1660, GM1661 e GM2226). Células da linhagem 3T3, também cedidas pela Dra Patrícia Cristina Baleeiro Beltrão Braga, foram utilizadas como controle, uma vez que já havia sido anteriormente observado que essas células apresentam alta susceptibilidade à infecção por esses retrovírions.

De fato, as diferentes linhagens celulares testadas demonstraram possuir diferentes susceptibilidades à infecção viral (Tabela 8). As linhagens celulares que apresentaram as maiores eficiências de transdução viral foram as linhagens GM1660 (88,16% de células EGFP positivas), GM1661 (72,80% de células EGFP positivas) e GM2226 (68,52% de células EGFP positivas), enquanto que as linhagens GM13A e GM135 demonstraram ser menos susceptíveis à infecção viral (39,50% e 4,60% de células EGFP positivas, respectivamente).

Tabela 8. Taxas de Eficiência de Transdução por Retrovírions Contendo o Gene Repórter EGFP nas Linhagens Celulares Estudadas

Linhagem Celular	Porcentagem de células EGFP-positivas
GM13A	39,50%
GM135	4,60%
GM1660	88,16%
GM1661	72,80%
GM2226	68,52%
3T3 (controle)	71,46%

Sendo assim, descartamos as linhagens GM13A e GM135 dos experimentos subsequentes e utilizamos somente as linhagens GM1660, GM1661 e GM2226 durante os próximos passos do trabalho.

4.3. Genotipagem de SNPs por Seqüenciamento Direto

A presença de SNPs em heterozigose em genes submetidos a XCI pode ser utilizada como uma ferramenta para verificar se o tratamento em cultura de uma população originalmente mista de células $HPRT^{+/-}$, contendo tanto células $HPRT^{+}$ e $HPRT^{-}$, com as drogas 6TG ou HAT foi bem sucedido. Além disso, pode ser utilizada também como ferramenta para detectar a reativação do cromossomo X a partir de uma população originalmente homogênea $HPRT^{-}$ ou $HPRT^{+}$ de células somáticas femininas $HPRT^{+/-}$.

Por essas razões, determinamos os genótipos referentes a diversos SNPs presentes em regiões codificadoras de genes localizados ao longo de todo o cromossomo X nas células das linhagens GM1660, GM1661 e GM2226, o que foi realizado por amplificação por PCR a partir de DNA genômico total e seqüenciamento direto. Foram analisados 14 SNPs associados a dez genes localizados ao longo do cromossomo X sabidamente submetidos ao processo de XCI e cinco SNPs associados a quatro genes que escapam à XCI. Os genótipos referentes aos SNPs analisados obtidos para as três linhagens estudadas estão descritos na Tabela 9.

Tabela 9. Genótipos Referentes à SNPs em Genes Ligados ao Cromossomo X nas Linhagens Estudadas (DNA)

Genes e SNPs*	Genótipos**		
	GM1660	GM1661	GM2226
GYG2 (rs211659)	A/A	A/A	A/A
PIGA (rs1804165 - rs11448948 - rs3087965)	G/G - A/A - A/A	G/G - A/A - A/A	G/G - A/A - A/T
DYNLT3 (rs12849)	T/T	T/T	T/A
TIMP1 (rs6520277)	T/T	T/C	T/C
EBP (rs1049367 - rs3048)	A/A - G/G	A/A - G/G	A/A - G/T
SUV39H1 (rs3373)	A/A	A/A	A/G
ATRX (rs3088074)	C/C	C/G	C/G
TCEAL4 (rs11010 - rs221933)	T/T - C/C	T/T - C/C	T/C - C/C
ZNF75D (rs1129093)	G/G	G/A	G/A
VBP1 (rs11887)	C/C	C/C	C/T
HDHD1 (rs10458)	ND	T/G	T/T
GPR143 (rs3044)	G/G	ND	ND
ZFX (rs5949242)	T/T	T/T	C/C
SYTL4 (rs11557084, rs8780)	C/C - C/C	C/C - C/G	ND

* Genes submetidos a XCI em células humanas estão evidenciados em letras verdes. Genes que escapam a inativação do cromossomo X em humanos estão evidenciados em letras vermelhas.

** Genótipos heterozigotos estão evidenciados em letras vermelhas.

A linhagem GM1660 não mostrou ser heterozigota para nenhum dos SNPs avaliados. A linhagem GM1661 mostrou ser heterozigota para três SNPs presentes em três dos genes submetidos a XCI avaliados e dois SNPs presentes em dois dos genes que escapam a XCI analisados (Tabela 9, Figura 11). A linhagem GM2226 mostrou ser heterozigota para nove SNPs presentes em nove dos genes submetidos a XCI avaliados (Tabela 9, Figura 12), demonstrando que ambas as linhagens celulares podem potencialmente ser empregadas em experimentos de validação. No entanto, as células da linhagem GM2226 apresentaram alteração de

morfologia durante o curso das passagens e dos tratamentos em meio 6TG e meio HAT em cultura. Elegemos a linhagem GM1661 para continuar o trabalho já que essa linhagem apresentou morfologia adequada durante todo o período em cultura e mostrou ser informativa para diversos SNPs analisados.

Figura 11. Cromatogramas evidenciando os genótipos heterozigotos encontrados na linhagem GM1661 para diferentes SNPs presentes em genes submetidos à XCI (destacados em verde) e genes que escapam à XCI (destacados em vermelho) em humanos.

Figura 12. Cromatogramas evidenciando os genótipos heterozigotos encontrados na linhagem GM2226 para diferentes SNPs presentes em genes submetidos à XCI em humanos.

Com o objetivo de verificar se o tratamento com 6TG era efetivo para a seleção positiva de células HPTR⁻ em cultura, assim como verificar quais dos dois diferentes alelos referentes aos genótipos heterozigotos presentes em genes submetidos a XCI estavam sendo expressos na linhagem GM1661 após seleção em meio contendo 6TG, foi realizada a genotipagem dos SNPs informativos a partir de RNA de células cultivadas em meio contendo 6TG e de células não selecionadas (controle).

Conforme o esperado, detectamos a presença de expressão de apenas um alelo para cada um dos SNPs avaliados em nossas amostras de células selecionadas em 6TG (Tabela 10, Figura 13). Interessantemente, detectamos a presença de expressão gênica de apenas um único alelo também na amostra de células não selecionadas para os dois SNPs avaliados (Tabela 10, Figura 13). Isso revela que nossas amostras de células da linhagem GM1661 já apresentavam desvio de XCI mesmo antes de serem submetidas ao tratamento com 6-tioguanina.

Figura 13. Cromatogramas evidenciando os genótipos encontrados para dois diferentes SNPs presentes em genes submetidos à XCI em DNA e cDNA de células da linhagem GM1661. A detecção de um único alelo no cDNA de células não selecionadas sugere desvio de inativação do cromossomo X em nossa amostra.

Tabela 10. Genótipos dos Alelos Expressos (cDNA) na Linhagem GM1661 Referentes a SNPs em Genes Submetidos à XCI

Genes e SNPs	Genótipos*		
	DNA	cDNA	
		6TG	não selecionada
ATR (rs3088074)	C/G	C	C
ZNF75D (rs1129093)	G/A	G	G

* Genótipos heterozigotos estão evidenciados em letras vermelhas.

4.4. Teste de Sensibilidade de Fibroblastos Humanos Primários à Droga 6-Tioguanina

Com o objetivo de melhor verificar se o tratamento com 6-tioguanina era eficaz em eliminar ou, pelo menos, depletar a população de células HPRT⁺ em cultura, tratamos primeiramente fibroblastos primários normais das linhagens FibN1 e Hs27 por 12 dias com essa droga. Para esse experimento, fibroblastos não selecionados e fibroblastos selecionados em meio contendo HAT foram utilizados como controle.

Em relação à linhagem FibN1, foi possível observar, por microscopia de luz, uma alta diminuição no número de células presentes nos poços contendo 6TG a partir de apenas dois dias do início do tratamento em comparação aos poços contendo células não-selecionadas e aos poços tratados com HAT (Figura 14). Essa diminuição no número de células presentes nos poços tratados com 6TG parece ter ocorrido principalmente devido à morte celular, uma vez que a presença de um grande número de células em suspensão no meio de cultura foi constantemente observada nesses poços a partir do primeiro dia de tratamento, ao contrário do observado para os poços contendo células não selecionadas ou selecionadas em meio HAT (Figura 14).

Figura 14. Visualização por microscopia de luz de fibroblastos dérmicos humanos primários da linhagem FibN1 não seleccionados e seleccionados em HAT ou 6TG após dois, quatro e doze dias do início de tratamento com as drogas 6TG e HAT. Aumentos de 40X (dias 5 e 12) e 100X (dia 2).

A linhagem Hs27 também apresentou diminuição do número células nos poços tratados com 6TG em comparação aos poços contendo células não seleccionadas ou tratadas em meio HAT, porém em menor magnitude do que as células da linhagem FibN1. Além disso, a linhagem Hs27 também pareceu apresentar uma leve diminuição no número de células presentes nos poços tratados em meio HAT, o que demonstra que essas células são ligeiramente mais sensíveis ao tratamento com essa droga do que células da linhagem FibN1 (Figura 15). A diminuição no número de células presentes nos poços tratados com 6TG ou HAT na linhagem Hs27 parece não ter sido consequência apenas de um aumento na taxa de morte celular, mas também de diminuição na taxa de proliferação celular, uma vez que o número de células flutuantes observados nos poços durante todo o

tratamento em meio contendo 6TG ou HAT foi consideravelmente menor do que o observado para as células da linhagem FibN1 tratadas em 6TG.

Figura 15. Visualização por microscopia de luz de fibroblastos humanos primários de prepúcio da linhagem Hs27 não seleccionados e seleccionados em HAT ou 6TG após dois, quatro e doze dias do início de tratamento com as drogas 6TG e HAT. Aumento de 40X.

Devido a essas observações, optamos por investigar mais cuidadosamente as consequências do tratamento com as drogas 6TG e HAT em nossa amostra de células da linhagem GM1661, linhagem essa escolhida para o experimento de transdução com a biblioteca de shRNAs.

4.5. Geração e Análise de Curvas de Crescimento Celular

A fim de melhor caracterizar o efeito do tratamento de células da linhagem GM1661 não selecionadas frente à adição das drogas 6TG ou HAT ao meio de cultura, traçamos curvas de proliferação celular para células submetidas a tratamentos em cada uma dessas condições.

Em relação às células mantidas em meio não seletivo, as células expostas ao meio contendo 6TG não diminuíram significativamente em número durante todo o período de tratamento (Figura 16), enquanto que as células expostas ao meio HAT diminuíram consideravelmente em número a partir do quarto dia de tratamento (Figura 16), o que mais uma vez aponta para o fato de que nossa amostra de células da linhagem GM1661 possui um desvio de XCI a favor de células HPRT⁻. O menor número de células presentes nas amostras expostas ao meio HAT em comparação com o número de células presentes nas amostras mantidas em meio não seletivo observado ao longo do período de tratamento parece ter sido decorrente principalmente de uma diminuição da taxa de proliferação celular nas amostras tratadas com HAT. Porém, a partir do sexto de dia de tratamento, a presença de morte celular também foi evidente, uma vez que uma diminuição progressiva do número de células contidas nos poços tratados com HAT foi observada a partir desse ponto do tratamento (Figura 16).

Figura 16. Curvas de crescimento das células da linhagem GM1661 quando expostas aos tratamentos com 6TG ou HAT. Os valores dos eixos X indicam o dia de tratamento e os valores dos eixos Y indicam a média do número de células viáveis presentes nas triplicatas obtidas para cada amostra. As barras verticais indicam o desvio padrão.

Foi traçada também uma curva de crescimento para células da linhagem GM1661 pré-selecionadas em meio 6TG (GM1661/HPRT⁻) frente à exposição ao meio HAT por 12 dias. Para garantir maior homogeneidade da população celular inicial em relação à expressão do alelo HPRT⁻, o pré-tratamento com 6TG foi realizado por cerca de quatro semanas antes que as células fossem utilizadas na geração da curva de crescimento em meio HAT.

O tratamento de células da linhagem GM1661/HPRT⁻ em meio contendo HAT, também causou um intenso declínio na taxa de proliferação celular quando comparado às amostras controle, mantidas em meio 6TG (Figura 17). Embora não tenha sido evidente, não se pode descartar que tal declínio na taxa de proliferação possa ocorrer de maneira concomitante com um pequeno aumento na taxa de morte celular. Além disso, a discrepância entre o número de células presentes nas amostras controle mantidas em 6TG e nas amostras tratadas com HAT pôde ser facilmente detectada a partir do dia 6 de tratamento e, de maneira similar ao observado para células não selecionadas expostas ao meio HAT, tendeu a se evidenciar cada vez mais ao longo do restante do período de tratamento em cultura (Figura 17).

Figura 17. Curva de crescimento das células da linhagem GM1661/HPRT⁻ expostas ao tratamento com HAT. Os valores do eixo X indicam o dia de tratamento e os valores do eixo Y indicam a média do número de células viáveis presentes nas triplicatas obtidas para cada amostra. As barras verticais indicam o desvio padrão.

Esses experimentos nos permitiram estimar o tempo mínimo de tratamento necessário em meio contendo HAT para que haja seleção negativa significativa de células HPRT⁻ em cultura, o que, por sua vez, deve permitir o enriquecimento e mais fácil detecção da população de células HPRT⁺ presentes na mesma placa de cultivo. Vale salientar que a determinação do tempo mínimo necessário de tratamento em HAT para que se possa detectar uma diferença significativa entre o número de células HPRT⁺ e HPRT⁻ presentes em uma mesma placa de cultura é de fundamental importância, uma vez que não se sabe ao certo se células primárias que sofreram a reativação do cromossomo X são capazes de sobreviver em cultura a médio ou longo prazo.

Com base nos resultados obtidos pudemos concluir que, uma vez que as células pré-selecionadas em meio 6TG transduzidas pela biblioteca lentiviral de shRNA fossem expostas ao meio contendo HAT, poderíamos detectar com segurança a diminuição da taxa de proliferação celular/aumento de taxa de morte de células HPRT⁻ na placa experimental tratada com meio HAT, em comparação com a placa controle, também transduzida mas mantida em meio 6TG, no oitavo dia após o início do tratamento com HAT. Nesse ponto do tratamento podemos esperar encontrar cerca de duas a três vezes menos células HPRT⁻ na amostra tratada em meio HAT.

4.6. Cariotipagem de Células da Linhagem GM1661 Selecionadas em Meio 6TG

A linhagem GM1661, escolhida para dar continuação ao trabalho, foi submetida ao exame de cariótipo por bandamento G com o objetivo de verificar se as células selecionadas em meio 6TG apresentavam integridade cromossômica numérica e/ou estrutural, já que a presença de anormalidades cromossômicas poderia influenciar ou confundir a triagem de genes candidatos envolvidos no processo de manutenção da XCI. Esse procedimento foi realizado em colaboração com a técnica de laboratório em citogenética Maria R. Lisboa Santana Pinheiro. Nenhuma alteração cromossômica, estrutural ou numérica, foi encontrada nas 100 metáfases analisadas (Figura 18). Esse resultado demonstra que essas células são adequadas para uso nos experimentos envolvendo o uso da biblioteca de shRNAs.

Figura 18. Cariotipagem de células da linhagem GM1661 selecionadas em meio contendo 6TG. (A) Imagem dos cromossomos presentes em uma célula metafásica. (B) Cariótipo gerado a partir da imagem da célula em metáfase mostrando conjunto cromossômico normal (46,XX).

4.7. Escolha da Biblioteca de shRNA utilizada

Conforme mencionado anteriormente, a biblioteca de shRNA que elegemos para desenvolver o presente trabalho (*GeneNet™ Lentiviral shRNA Library Human Genome-Wide Screen*, System Biosciences®) é oferecida em duas diferentes versões, *FIV-based* e *HIV-based*, e a escolha da versão a ser utilizada, conforme recomendação da própria empresa, deve ser baseada na taxa de eficiência de transdução de cada um desses tipos de partículas lentivirais no tipo celular alvo.

Sendo assim, a fim de se eleger o tipo de partícula lentiviral que possui melhor capacidade de transdução em nosso tipo celular alvo (fibroblastos dérmicos humanos primários) realizamos um experimento piloto com partículas lentivirais contendo o gene repórter que codifica a proteína copGFP de ambas as versões, FIV e HIV (*pSIF1- and pSIH1-H1-siLuc-copGFP*, System Biosciences®), em diferentes linhagens de fibroblastos dérmicos humanos primários, sendo uma linhagem de fibroblastos humanos fetais (HDF51) e duas linhagens de fibroblastos humanos adultos (HDF48 e HDF49). As duas versões de partículas lentivirais foram testadas em diferentes MOI para cada uma das linhagens celulares avaliadas e sob diferentes porcentagens de confluência celular para as linhagens HDF48 e HDF549. Esses experimentos foram realizados durante meu estágio no laboratório da Dr Jeanne Loring, no *The Scripps Research Institute*, em La Jolla, Califórnia, EUA, e as três linhagens celulares utilizadas no experimento foram cedidas pelo seu laboratório.

As partículas lentivirais baseadas em HIV (*pSIH-H1-siLuc-copGFP*) apresentaram taxas consideravelmente maiores de eficiência de transdução viral em todas as três linhagens de fibroblastos primários humanos testadas, se comparadas às taxas de eficiência de transdução viral obtidas com as partículas baseadas em FIV (*pSIF1-H1-siLuc-copGFP*), conforme observado tanto por microscopia de fluorescência como por citometria de fluxo (Tabela 11). Não foi observado, em nenhuma das amostras de células transduzidas, aumento de mortalidade celular decorrente da infecção lentiviral para nenhuma das MOI testadas. Embora um pequeno aumento na eficiência de transdução tenha sido observado quando lentivirions baseados em FIV em todas as MOI avaliadas e lentivirions baseados em HIV em MOI de 1 e 5 foram transduzidos em placas mais confluentes do que em placas menos confluentes para ambas as linhagens HDF48 e HDF49, não houve

diferença significativa nas taxas de eficiência de transdução lentiviral entre as diferentes porcentagens de confluência celular testadas para essas duas linhagens após infecção com partículas lentivirais baseadas em HIV em MOI mais altas (MOI 10 e 20) (Tabela 11).

Com base nos resultados obtidos, pudemos concluir que as partículas virais baseadas em HIV (*pSIH-H1-siLuc-copGFP*) são mais apropriadas para transdução de fibroblastos primários dérmicos humanos do que as partículas virais baseadas em FIV (*pSIF1-H1-siLuc-copGFP*) e que ambas as porcentagens de confluência celular testadas são adequadas para a transdução de fibroblastos dérmicos humanos primários com lentivírons baseados em HIV. Além disso, foi observado que a MOI a ser utilizada para a transdução de fibroblastos dérmicos humanos primários com a biblioteca lentiviral de shRNAs baseada em HIV deve ser em torno de 5 a 7 quando o polibreno for utilizado como agente facilitador da transdução (de acordo com a Tabela 12, descrita no próximo item).

Tabela 11. Eficiência de Transdução Lentiviral em Células das Linhagens HDF48, HDF49 e HDF51 em Diferentes MOI e Confluências Celulares

Tipo de Lentivíron	Linhagem Celular/ Área de cultivo*	Porcentagem de células copGFP positivas			
		MOI=1	MOI=5	MOI=10	MOI=20
Baseado em FIV	HDF48 (24-well)	5,81%	10,15%	17,56%	37,26%
	HDF48 (6-well)	3,98%	7,78%	13,33%	30,99%
	HDF49 (24-well)	15,22%	18,95%	26,64%	38,55%
	HDF49 (6-well)	11,71%	15,57%	23,71%	34,32%
	HDF51 (24-well)	2,56%	7,50%	16,05%	40,66%
Baseado em HIV	HDF48 (24-well)	16,49%	48,46%	73,76%	90,78%
	HDF48 (6-well)	13,40%	40,51%	69,18%	91,08%
	HDF49 (24-well)	22,65%	49,27%	68,69%	81,43%
	HDF49 (6-well)	19,36%	46,77%	69,63%	83,99%
	HDF51 (24-well)	12,12%	43,06%	66,88%	83,90%

* 5×10^4 células foram plaqueadas em um poço de placa de 24-well ou em um poço de placa de 6-well.

4.8. Otimização da Transdução de Fibroblastos Primários Humanos da linhagem GM1661 com Partículas Lentivirais copGFP

Uma vez estabelecido que os fibroblastos primários humanos são mais eficientemente transduzidos por partículas lentivirais baseadas em HIV do que por partículas lentivirais baseadas em FIV, realizamos experimentos de titulação viral com o objetivo de determinar qual seria o número adequado de partículas virais por célula (MOI) para a transdução de nossas células-alvo com a biblioteca de shRNA, de maneira a garantir que cada célula presente na placa de cultura não fosse infectada por mais de um vírion. Esse procedimento é de fundamental importância para o sucesso da triagem funcional, já que a presença de mais de um tipo de shRNA lentiviral uma única célula resulta no *knockdown* simultâneo de mais de um gene, o que pode acarretar na obtenção de resultados falso-positivos ou falso-negativos.

O número de cópias de insertos lentivirais contendo o gene repórter coGFP presentes em uma determinada população celular pode ser estimado através do número de células copGFP positivas obtidas após a transdução (Tabela 12, extraída do manual que acompanha a biblioteca de shRNAs fornecido pela empresa System Biosciences®). Sendo assim, diferentes concentrações de partículas lentivirais (IFU/mL) foram testadas em nossas células-alvo a fim de se estabelecer o número MOI necessário para a obtenção de um número de cópias de shRNA integrados em nossa população de células menor ou igual a 1 cópia/célula (idealmente entre ~0,5 e 1,0).

Tabela 12. Correspondência Entre o Número de Células copGFP Positivas e o Número de Cópias do Gene Réporter por Célula

% Células copGFP-Positivas	10	20	30	40	50	60	70	80	90	>90
Número de Cópias por Célula	0.10	0.23	0.36	0.51	0.70	0.93	1.22	1.64	2.30	>2.50

Em verde: intervalo ideal no qual o número de cópias presentes por célula é entre 0,5 e 1,0.

Para esse experimento, utilizamos lentivírus controle que possuem a mesma estrutura dos plasmídeos utilizados na construção da biblioteca de shRNA e o gene repórter copGFP (partículas virais baseadas em HIV copGFP controle fornecidas juntamente com a biblioteca de shRNA [(pSIH1-H1-siLuc-copGFP Packaged Positive Transduction Control, System Biosciences®)] em fibroblastos primários humanos da linhagem GM1661/HPRT⁻. Além disso, utilizamos o reagente *TransDux*TM (System Biosciences®) em substituição ao polibreno como agente facilitador da transdução viral, uma vez que esse reagente demonstra ser muito mais eficiente e muito menos tóxico para as células do que o polibreno.

Foram testadas diferentes proporções de número de vírus (IFU) por número de células (MOI), em diferentes porcentagens de confluência celular, dentro do intervalo de confluência recomendado pela empresa System Biosciences® de 50% a 70% no momento da transdução.

O experimento permitiu verificar a eficiência de transdução especificamente da linhagem GM1661/HPRT⁻ pelas partículas virais baseadas em HIV pSIH-H1-siLuc-copGFP em diferentes MOI e taxas de confluência celular (Figura 19, Tabela 13), similar às obtidas com as linhagens HDF48, HDF49 e HDF51. Além disso, a eficiência de transdução demonstrou ser até duas vezes maior na presença do reagente *TransDux*TM do que na presença de polibreno (Tabelas 11 e 13).

Tabela 13. Eficiência de Transdução por Lentivírus Baseados em HIV e Número de Cópias do Gene Repórter em Células da Linhagem GM1661/HPRT⁻ em Diferentes MOI e Confluências Celulares

Nº de Células por Poço*	MOI	Porcentagem de células copGFP-positivas	Nº de Cópias do Gene Repórter por Célula**
4x10 ⁴	4	86,60%	2,08
5x10 ⁴	2	54,40%	0,80
5x10 ⁴	3	70,83%	1,25
5x10 ⁴	4	84,36%	1,93
5x10 ⁴	5	89,26%	2,25
5x10 ⁴	10	73,06%	1,22

*Refere-se ao momento do plaqueamento. As células foram plaqueadas em um poço de placa de 12-well. ** Calculado por interpolação linear usando como referência os valores associados aos dois pontos mais próximos descritos na Tabela 12.

Figura 19. Visualização por microscopia de fluorescência de fibroblastos primários humanos da linhagem GM1661 copGFP-positivos cerca de 48h após a transdução com partículas lentivirais pSIH1-H1-siLuc-copGFP em diferentes confluências e MOI. Aumento de 100X.

Com base nos dados obtidos por citometria de fluxo (Tabela 13) e na tabela de correspondência entre o número de células copGFP positivas e o número de cópias gênicas por célula (Tabela 12), o número MOI adequado para o experimento de transdução com a biblioteca de shRNAs em células da linhagem GM1661 previamente selecionadas em meio 6TG foi estabelecido em 1,8. Esse MOI corresponde a um número estimado de cópias do gene exógeno por célula igual a 0,52, conforme calculado por extrapolação e interpolação linear (disponível em <http://easycalculation.com/analytical/linear-interpolation.php>) dos dados de

porcentagem de células copGFP positivas obtidos para as MOI entre 2 e 5 nas amostras de 5×10^4 células plaqueadas (Tabela 13), seguido por interpolação linear do valor resultante (MOI 1,8 = 52,20% células GFP positivas) com os valores de referência dos dois pontos mais próximos (50% e 60%) expressos na Tabela 12.

4.9. Transdução de Fibroblastos Primários Humanos com a Biblioteca de shRNAs

Uma vez que nosso modelo celular de fibroblastos dérmicos primários humanos 46,XX, *HPRT*^{+/-} já havia sido estabelecido pela seleção em meio contendo 6TG e devidamente caracterizado, e que a transdução dessas células com as partículas lentivirais baseadas em HIV já havia sido otimizada, procedemos para o experimento de transdução com a biblioteca de shRNAs.

Dez milhões de células GM1661/*HPRT*⁻ de passagem 21 foram transduzidas com a biblioteca de shRNAs, na presença do reagente *TransDux*TM. As células transduzidas foram mantidas durante 72h em meio não seletivo. Durante esse período, não foi observado aumento na taxa de morte celular evidente decorrente da infecção viral. Após esse período, as células transduzidas foram tripsinizadas e contadas. Uma pequena porcentagem das células tripsinizadas (total de $2,1 \times 10^4$ células) foi dividida em poços de placa 12-well e tratadas por 10 dias em meio contendo 400ng/mL de puomicina, com o objetivo de se determinar a eficiência da transdução viral com a biblioteca de shRNAs, já que os plasmídeos contendo shRNAs possuem também um gene de resistência a esse antibiótico.

O tratamento com puomicina resultou claramente em uma redução considerável do número de células nas amostras transduzidas com a biblioteca de shRNA, conforme evidenciado por comparação visual dos poços controle (não tratado com puomicina) e experimental (tratado com puomicina) nos quais 5×10^3 células foram inicialmente plaqueadas (Figura 20).

5×10^3 Células em Puromicina

5×10^3 Células (Controle)

Figura 20. Visualização por microscopia de luz de fibroblastos primários humanos da linhagem GM1661 transduzidos com partículas lentivirais da biblioteca de shRNAs e submetidos a 10 dias de seleção em meio contendo puromicina (exceto controle não tratado). O número de células especificado está relacionado ao momento de plaqueamento. Aumento de 40X.

No entanto, o número de colônias presentes em cada poço não pode ser estimado devido à alta confluência celular após os 10 dias de tratamento em cultura. Por essa razão, as células presentes nos poços controle e experimental nos quais 5×10^3 células foram plaqueadas foram coletadas e contadas em câmara de Neubauer na presença de *trypan blue*. Enquanto $1,35 \times 10^5$ células estavam presentes no poço controle, apenas $8,45 \times 10^4$ células foram contadas no poço experimental, o que sugere que a taxa de transdução viral com a biblioteca de shRNAs foi de 62,59%. No entanto, esse número é, muito provavelmente, superestimado, uma vez que, no poço controle, não tratado com puromicina, as

células encontravam-se mais confluentes, o que implica em uma menor taxa de proliferação celular devido à maior taxa de inibição por contato entre as células ali presentes.

Em conclusão, com base nos resultados obtidos, a taxa de eficiência de transdução viral pela biblioteca de shRNAs pode ser estimada como sendo menor ou igual a 62,59%, o que representa um número de cópias do inserto de shRNA por célula eficientemente transduzida menor ou igual a 1 (calculado por interpolação linear desse dado de porcentagem de células coGFP positivas entre os dados de referência descritos para os dois pontos mais próximos na Tabela 12). Esse valor está dentro do intervalo de segurança recomendado pelo fabricante da biblioteca de shRNAs.

As células transduzidas com a biblioteca lentiviral de shRNAs foram divididas em duas populações 72 horas após a transdução: uma população controle, que foi mantida em meio de cultura seletivo contendo 6TG, e uma população experimental que foi selecionada em meio de cultura contendo HAT. O tempo de tratamento em meio seletivo contendo HAT ou 6TG foi de oito dias e foi baseado nas curvas de crescimento celular em meio HAT obtidas para essa linhagem.

Conforme o esperado, o número de células nas placas com meio HAT diminuiu progressivamente durante o curso do tratamento quando comparado ao número de células nas placas controle mantidas em 6TG. Ao final do período de tratamento, as células foram coletadas e contadas na presença de *trypan blue*. O número de células obtidas após seleção em meio 6TG foi $3,3 \times 10^7$ e em meio HAT foi $1,1 \times 10^7$, ou seja, uma diferença condizente com a diferença constatada para o mesmo tempo de tratamento durante a geração das curvas de crescimento celular em meio HAT para essa mesma linhagem.

4.10. Multiplex PCR para amplificação dos insertos de shRNAs

As células transduzidas com a biblioteca de shRNAs e selecionadas por 8 dias em meio contendo 6TG (amostra controle) ou HAT (amostra experimental) foram coletadas por tripsinização e submetidas à extração de DNA genômico total. Foram obtidos 400,4ug de DNA genômico das células mantidas em 6TG e 126,9ug das células mantidas em HAT.

Os DNAs foram utilizados em reações de *multiplex* PCR para amplificação dos insertos de shRNAs incorporados no genoma das células transduzidas pela biblioteca de shRNAs. Foram realizadas duas rodadas de amplificação, sendo que a segunda rodada foi constituída por *nested* PCR na qual foram utilizados primers localizados dentro da região amplificada na primeira rodada (Figura 10). Diversas combinações de número de ciclos de amplificação foram testadas a fim de se pudesse determinar qual o menor número de ciclos necessário em cada uma das rodadas de PCR para que os amplicons de tamanho de interesse pudessem ser detectados em gel de agarose ao final da segunda rodada de amplificação. Tal procedimento foi adotado a fim de evitar com que as reações de amplificação atingissem a fase de *plateau*, de maneira a garantir que a representação numérica de cada inserto de shRNA contido nas amostras estudadas fosse mantida. O número de mínimo de ciclos foi estabelecido em 18 tanto para a primeira como para a segunda rodada de PCR. Os produtos das reações de PCR foram corridos em gel de agarose e os padrões de bandas obtidos foram comparados com os padrões demonstrados no manual que acompanha a biblioteca de shRNAs (Figura 21).

Figura 21. Fotografias de géis de agarose mostrando os amplicons esperados após amplificação dos insertos de shRNA da biblioteca de shRNAs. (A) Bandas observadas após a primeira e segunda rodadas de PCR. Referência fornecida pelo fabricante da biblioteca de shRNAs. Modificado do manual que acompanha o produto fornecido pela empresa System Biosciences®. (B) Bandas observadas após a segunda rodada de PCR em nossas amostras. CP = Controle Positivo (DNAs plasmídeos contendo shRNAs), E = Amostra Experimental (cDNA de células da linhagem H1299 transduzidas com a biblioteca), CN = Controle Negativo (em (A): células H1299 transduzidas sem síntese de cDNA, em (B): células GM1661 não transduzidas), 6TG = DNA de células GM1661 transduzidas com a biblioteca e selecionadas em 6TG, HAT = DNA de células GM1661 transduzidas com a biblioteca e selecionadas em HAT, B = PCR “branco” (sem DNA).

Como o número mínimo de ciclagens de PCR foi utilizado em nossos experimentos, nenhum amplicon foi detectado em gel de agarose após a primeira rodada de amplificação. No entanto, após a segunda rodada de PCR, o mesmo padrão de bandas demonstrado no manual da biblioteca de shRNAs foi observado em nossas amostras, ou seja, a presença de duas bandas na região entre os marcadores de 100 e 200pb. Vale notar que, embora o tamanho esperado da banda correspondente aos insertos de shRNAs após a segunda rodada de PCR seja de 83pb, após eletroforese em gel de agarose 2,5% tais amplicons são observados na altura de ~150pb, o que ocorre devido a alteração da mobilidade eletroforética desses amplicons causada pela presença do *primer forward* biotinilado (desenhado pela empresa System Biosciences® para permitir detecção após hibridização em *chip* de *microarray*). Além disso, conforme mencionado pela própria empresa, após eletroforese em gel de agarose 2,5% dos amplicons gerados a partir do controle positivo (DNAs plasmidiais contendo os shRNAs), apenas uma banda pode ser observada (Figura 21). Entretanto, as rodadas de amplificação realizadas a partir de material proveniente de células transduzidas com a biblioteca de shRNAs gera não apenas uma banda, mas sim duas bandas bem próximas uma da outra na região de interesse do gel de agarose (Figura 21). Ainda segundo a empresa que comercializa a biblioteca de shRNAs, a banda inferior, menos proeminente, provavelmente é constituída em sua maioria por primers soltos, embora a presença de alguns tipos de shRNAs nessa região não possa ser excluída. De acordo com a equipe de suporte técnico da empresa, a presença dessas duas bandas não constitui um problema quando as amostras são posteriormente submetidas à hibridização em *microarray*. Para NGS, no entanto, a equipe de suporte recomendou a excisão e purificação dos amplicons contidos apenas na banda superior, mais proeminente, localizada entre 100 e 200pb no gel de agarose.

4.11. Purificação dos amplicons contendo shRNAs

As amostras de amplicons (controle 6TG e experimental HAT) provenientes da segunda rodada de amplificação por PCR (*nested* PCR) foram corridas em gel de agarose, as bandas de interesse foram rapidamente excisadas do gel e os fragmentos de agarose contendo amplicons foram submetidos ao procedimento de

purificação de DNA. Apenas a banda superior, mais proeminente, localizada entre os marcadores moleculares de 100 e 200bp, foi excisada do gel de agarose e submetida à purificação de DNA (Figura 22).

Figura 22. Fotografia do gel de agarose após primeiro corte para excisão das regiões contendo os amplicons desejados nas amostras controle (6TG) e experimental (HAT). Apenas a banda mais proeminente localizada entre os marcadores de 100 e 200pb foi utilizada.

Uma pequena alíquota das amostras de DNA purificadas (1,5 uL do volume total de cerca de 60uL) foi novamente corrida em gel de agarose 2%, ao lado de pequenas alíquotas dos produtos de PCR originais, não purificados, das amostras controle e experimental (2,0uL do volume total de 100uL), a fim de se verificar se o procedimento de purificação havia sido bem sucedido e de se acessar a qualidade dos amplicons purificados (Figura 23).

Figura 23. Fotografia do gel de agarose mostrando os amplicons presentes nas amostras controle (6TG) e experimental (HAT) antes e depois da purificação dos amplicons de interesse contidos nos fragmentos de agarose excisados.

A purificação dos amplicons de interesse, correspondentes à banda superior de tamanho entre 100 e 200pb, mostrou ter sido bem sucedida, uma vez que tais amplicons foram detectados de maneira proeminente após eletroforese do DNA isolado. A banda inferior de tamanho entre 100 e 200bp não foi detectada nas amostras purificadas. Além da presença da banda de interesse, foi observada também a presença de duas bandas fracas na região entre os marcadores moleculares de 200 e 300pb. Entretanto, tais bandas provavelmente não correspondem a produtos de PCR inespecíficos, uma vez que o fragmento de agarose excisado e submetido à purificação não compreendia essa região do gel de agarose. Sendo assim, é possível que essas bandas fracas correspondam a amplicons de interesse cuja mobilidade eletroforética foi alterada durante o procedimento de purificação.

4.12. Seqüenciamento de Próxima Geração (NGS)

A concentração de DNA das duas amostras eluídas do gel, controle 6TG e experimental HAT, foi medida em fluorômetro e 180ng de DNA da amostra controle 6TG e 140ng de DNA da amostra experimental HAT, em volume total de 50uL cada, foram enviados para a empresa Fasteris, na Suíça, para preparação da biblioteca de amplicons e corrida em aparelho de seqüenciamento de próxima geração HiSeq 2500 (Illumina®). A empresa Fasteris renomeou a amostra controle 6TG como “GZR-1” e a amostra experimental HAT como “GZR-2”. O resumo do plano experimental adotado pela empresa Fasteris está descrito no Anexo 1.

Ao receber as amostras, a empresa realizou um teste inicial na qual a qualidade dos DNAs enviados foi analisada. Ambas as amostras enviadas passaram esse primeiro controle de qualidade e foram então utilizadas para construção das bibliotecas de amplicons para NGS, que incluiu ligação de adaptadores (sequências de 120pb) aos amplicons e nova rodada de amplificação por PCR. Uma vez contruídas as bibliotecas, novos testes de qualidade foram realizados para análise do tamanho dos fragmentos e pureza das bibliotecas de amplicons geradas: corrida em aparelho LabChip® GX-LST (Perkin-Elmer®, Caliper Life Sciences) e corrida de NGS teste em coluna 1x50pb em aparelho HiSeq 2000 (Illumina®).

Conforme evidenciado pela corrida em aparelho LabChip® GX-LST (Perkin-Elmer®, Caliper Life Sciences), as duas bibliotecas de amplicons geradas eram constituídas por fragmentos de cerca de 200pb, o que estava de acordo com o esperado já que o tamanho dos amplicons enviados para a empresa era de 83pb e o tamanho dos adaptadores ligados à esses amplicons era de 120pb (Figura 24). Embora, duas bandas muito próximas tenham sido observadas na região de 200bp após corrida em aparelho LabChip® GX-LST (Figura 24), muito provavelmente essas duas bandas não correspondem às duas bandas observadas após a amplificação dos insertos de shRNAs por *nested* PCR, uma vez que ficou claro pelos experimentos de purificação a partir de gel de agarose que só a banda superior proveniente da *nested* PCR foi isolada. Outro fator que contribui para essa conclusão é que, na amostra 6TG corrida em aparelho LabChip® GX-LST a banda inferior apresentou maior intensidade, enquanto que a banda superior apresentou intensidade extremamente fraca, sendo de difícil visualização. Na amostra HAT, as duas bandas observadas após corrida em aparelho LabChip® GX-LST apresentaram intensidade similares. Tais observações não condizem com as intensidades observadas para as duas bandas observadas após *nested* PCR, na qual a banda superior, isolada da agarose, era de muito maior intensidade que a banda inferior, que foi descartada. Sendo assim, é provável que as duas bandas observadas após corrida em aparelho LabChip® GX-LST estivessem contidas na mesma banda isolada da agarose.

Além das duas bandas observadas na região de 200pb, nenhuma outra foi detectada após corrida em aparelho LabChip® GX-LST (Figura 24), nem mesmo as duas bandas fracas de tamanho entre 200 e 300pb que haviam sido observadas após purificação dos amplicons de interesse a partir dos fragmentos excisados do gel de agarose nas duas amostras submetidas para NGS (Figura 23).

Figura 24. Resultado obtido após corrida das bibliotecas de amplicons em aparelho LabChip® GX-LST (Perkin-Elmer, Caliper Life Sciences) realizado na empresa Fasteris. GZR-1: amostra controle 6TG, GZR-2: amostra experimental HAT.

Para a corrida teste de seqüenciamento de próxima geração, uma pequena alíquota de cada amostra das bibliotecas de amplicons geradas (6TG e HAT) foi submetida à corrida de NGS em coluna 1x50pb de modo que 50 nucleotídeos presentes em cada amplicon foram seqüenciados. Após a corrida, a empresa nos disponibilizou arquivos contendo 50 seqüências em formato fasta e arquivos contendo 20 mil seqüências em formato fastq para cada amostra para que pudéssemos analisar os dados obtidos. O resultado dessa análise nos permitiu comprovar que os amplicons submetidos ao NGS de fato continham em suas seqüências os insertos de shRNAs, e portanto, a qualidade das bibliotecas de amplicons geradas era satisfatória.

Sendo assim, autorizamos a empresa a dar continuidade ao experimento e submeter as duas amostras de bibliotecas de amplicons à corrida final em aparelho de NGS. As amostras foram corridas em canal 1x100pb de aparelho HiSeq 2500 (Illumina®) de modo que 100 nucleotídeos presentes em cada amplicon foram seqüenciados. Foram geradas 8.733.500 seqüências para a amostra controle 6TG (GZR-1) e 12.399.284 seqüências para a amostra experimental HAT (GZR-2). Os arquivos contendo as seqüências em formato fastq foram disponibilizados pela

empresa Fasteris para que pudéssemos realizar a análise dos dados. Os resultados obtidos pela análise dos dados provenientes da corrida teste de NGS e da corrida final de NGS estão descritos a seguir.

4.13. Análise dos Dados de NGS

4.13.1. Corrida Teste

Primeiramente, com o objetivo de determinar se os dados de NGS possuíam qualidade adequada, as 20 mil seqüências contidas nos arquivos fastq obtidas para cada amostra foram analisadas pelo programa FastQC. Essa análise possibilitou avaliar a composição e a qualidade das seqüências obtidas para cada amostra. Resultados similares foram observados para as amostras controle 6TG (GZR-1) e experimental HAT (GZR-2) (Figuras 25 e 26). Os parâmetros avaliados pelo programa e respectivos resultados obtidos foram:

a) Estatística básica – avalia o número total de sequencias contidas no arquivo, número de sequencias filtradas (sequencias de baixa qualidade marcadas para exclusão quando o programa é acionado no modo CASAVA), tamanho das sequencias (reporta o tamanho das sequencias de menor e maior tamanho) e porcentagem geral de GC (em relação à todas as bases de todas as sequencias).

Resultados: cada um dos arquivos continha, de fato, 20 mil sequencias. Nenhuma sequencia contida no arquivo foi excluída (filtrada) das amostras, o que demonstra que nenhuma sequencia foi sinalizada pelo software CASAVA como sendo de baixa qualidade. Todas as sequencias presentes nos dois arquivos possuíam 50pb de tamanho, o que demonstra que nenhuma base foi cortada durante o *pipeline* de *base-calling* devido à baixa qualidade. A porcentagem geral de GC das sequencias na amostra controle 6TG foi de 40% e na amostra experimental HAT foi de 41%.

b) Qualidade das seqüências por base – visão geral dos intervalos de valores de qualidade em todas as bases em cada posição nas sequencias de nucleotídeos. Quanto maior o score melhor o *base-call*. Como a qualidade das bases na maioria

das plataformas tende a decair com a progressão da corrida, é comum observar uma queda de qualidade na região final das sequencias. Vale notar que, como a região referente ao inserto de shRNA em nossas amostras abrange apenas os nucleotídeos 13 a 39, a presença de uma leve queda de qualidade referente aos nucleotídeos finais das sequencias não constitui uma preocupação para a nossa análise;

Resultados: para ambas as amostras, a qualidade geral das bases em cada posição uma das posições ao longo dos fragmentos sequenciados mostrou estar dentro do intervalo considerado ótimo, com média e mediana acima de 30 (o escore máximo para Casava versão 1.8 é de 41).

c) Escores de qualidade por sequencia – permite identificar se algum subgrupo das sequencias possui valores de qualidade baixos;

Resultados: os escores de qualidade por sequencia foram considerados ótimos para ambas as amostras, sendo que a grande maioria das sequencias apresentou escores entre 27 e 39.

d) Conteúdo das sequencias por base – avalia a porcentagem de cada base em cada posição ao longo das sequencias contidas no arquivo;

Resultados: em ambas as amostras a relação entre as porcentagens de cada nucleotídeo em cada posição dos fragmentos sequenciados demonstrou ser satisfatória para os nucleotídeos presentes nas posições 13 a 39 (região referente à segunda metade dos insertos de shRNAs). Os 12 primeiros e os 11 últimos nucleotídeos se referiam, respectivamente, às sequencias do *primer forward* e sequencias plasmidiais e, portanto, as porcentagens de cada nucleotídeo em cada uma das posições que compunham essas regiões encontravam-se enviesadas em ambas as amostras.

e) Conteúdo de GC por base – avalia o conteúdo de GC em cada posição ao longo das sequencias contidas no arquivo;

Resultados: o conteúdo de GC em cada posição das sequências referentes aos insertos de shRNAs mostrou variar entre 39,01 e 54,10% (média = 46,01%) para a amostra controle 6TG e entre 40,99 e 56,14% (média = 46,43%) para a amostra HAT. Como os valores observados variaram menos do que 10% em todas as posições relativas aos insertos de shRNAs e menos que 5% na grande maioria dessas posições, os resultados obtidos para esse parâmetro foram considerados satisfatórios.

f) Conteúdo de GC por sequência – avalia o conteúdo de GC através de todo o comprimento de cada sequência contida no arquivo e compara os resultados obtidos com o modelo de distribuição normal gerado pelo módulo;

Resultados: a distribuição do conteúdo médio de GC para as sequências contidas no arquivo mostrou acompanhar o modelo de distribuição normal gerado pelo programa, com pico condizente com o conteúdo médio de GC observado no modelo normal gerado em ambas as amostras. A soma dos desvios da distribuição normal representou menos do que 15% das sequências contidas no arquivo e, portanto, as amostras passaram esse controle de qualidade. Esses resultados indicam não haver contaminação, ou presença de outro tipo de subgrupo enviesado de sequências em nenhuma das duas bibliotecas de amplicons analisadas.

g) Conteúdo de N por base – denota a porcentagem de *base-calls* lidas como N (nucleotídeo não especificado) em cada posição ao longo das sequências presentes no arquivo;

Resultados: a quantidade de nucleotídeos não especificados durante o procedimento de *base-calling* demonstrou ser extremamente pequena tanto na região correspondente aos insertos de shRNAs ($\leq 0,015\%$ para amostra controle 6TG e $\leq 0,025\%$ para amostra experimental HAT) como na região correspondente ao primer forward e sequências plasmidiais ($\leq 0,155\%$ para amostra controle 6TG e $\leq 0,130$ para amostra experimental HAT). Como este valor está abaixo de 5% para todas as posições que compõem as sequências, pode-se considerar que o *pipeline* de análise foi capaz de interpretar os dados obtidos durante o NGS de maneira fortemente adequada e, portanto, gerou *base-calls* válidos.

h) Distribuição dos tamanhos das sequencias – distribuição dos tamanhos dos fragmentos contidos no arquivo;

Resultados: todas as 20 mil sequencias contidas em cada arquivo mostraram apresentar 50 nucleotídeos de comprimento, o que está de acordo com o esperado uma vez que as amostras foram sequenciadas em canal 1x50pb.

i) Sequencias duplicadas – conta o grau de duplicação de sequencias presentes no arquivo. Mostra o número relativo de sequencias com diferentes graus de duplicação. O número de sequencias únicas é sempre definido em 100. Nesse módulo, apenas as primeiras 200 mil sequencias contidas no arquivo são avaliadas (cada uma dessas 200 mil sequencias é comparada com todas as sequencias restantes presentes no arquivo) e as sequencias com mais de 75pb são truncadas de forma que apenas os 50pb iniciais são alinhados;

Resultados: O nível de duplicação de sequencias obtido para a amostra controle 6TG foi de $\geq 20,87\%$ e o nível de duplicação obtido para a amostra experimental HAT foi de $\geq 20,78\%$. No entanto, como apenas 20 mil sequencias estavam disponíveis para cada amostra, e também devido à natureza do nosso experimento (no qual a diversidade dos amplicons gerados é limitada e no qual se espera que células contendo shRNAs específicos sejam selecionadas em cultura) a análise desse parâmetro torna-se pouco informativa.

j) Sequencias super-representadas – lista todas as sequencias que constituem mais de 0,1% do total de sequencias contidas no arquivo. Neste módulo apenas as primeira 200 mil sequencias são avaliadas (cada uma dessas 200 mil sequencias é comparada com todas as sequencias restantes presentes no arquivo);

Resultados: nenhuma sequencia especifica mostrou estar representada mais do que 0,1% do total de sequencias contidas nos arquivos de ambas as amostras. Isso indica que as bibliotecas de amplicons geradas não estavam contaminadas.

k) Kmers super-representados – conta a presença de cada combinação de 5 nucleotídeos (kmers) nas sequencias contidas no arquivo para avaliar se alguma

combinação está enriquecida na biblioteca de amplicons. O módulo calcula o nível no qual cada kmer seria esperado de acordo com o conteúdo de bases observado para o total de sequências do arquivo e, então, utiliza o número realmente observado para calcular a razão esperado/observado para cada kmer. O módulo sinaliza caso algum kmer específico esteja representado, de maneira geral, mais do que 3 vezes do que o esperado ou caso um algum kmer seja detectado mais de 5 vezes do que esperado em uma posição específica. Nesse módulo, apenas 20% das sequências contidas no arquivo são analisadas.

Resultados: A grande maioria dos kmers mais amplamente representados nas amostras analisadas foi encontrada na região que compreende os 12 primeiros ou 11 últimos nucleotídeos das sequências de amplicons (respectivamente, *primer forward* e sequência plasmidial). Devido ao número reduzido de sequências presentes nos arquivos da corrida teste, a análise desse parâmetro torna-se pouco informativa.

Figura 25. Representação gráfica dos parâmetros avaliados pelo programa FastQC para as sequências geradas na corrida teste de NGS da amostra controle 6TG. (A) Qualidade das sequências por base. (B) Escores de qualidade por sequência. (C) Conteúdo das sequências por base. (D) Conteúdo de GC por base. (E) Conteúdo de GC por sequência. (F) Conteúdo de N por base. (G) Distribuição do tamanho das sequências. (H) Sequências duplicadas. (I) KMERs super-representados (apenas os seis KMERs mais frequentemente observados são representados).

Figura 26. Representação gráfica dos parâmetros avaliados pelo programa FastQC para as sequencias geradas na corrida teste de NGS da amostra experimental HAT. (A) Qualidade das sequencias por base. (B) Escores de qualidade por sequencia. (C) Conteúdo das sequencias por base. (D) Conteúdo de GC por base. (E) Conteúdo de GC por sequencia. (F) Conteúdo de N por base. (G) Distribuição do tamanho das sequencias. (H) Sequencias duplicadas. (I) KMERs super-representados (apenas os seis KMERs mais frequentemente observados são representados).

Após a análise em programa FastQC, as sequências de amplicons contidas nos dois arquivos foram submetidas a excisão das porções correspondentes ao primer forward localizadas na extremidade 5' e às sequências plasmidiais localizadas na extremidade 3' através do uso do programa Cutadapt, de forma que apenas a região central, contendo os 27 nucleotídeos correspondentes a segunda metade do inserto de shRNA foram preservadas. A sequência correspondente ao primer *forward* foi detectada e excisada em 19213 (96,1%) sequências de amplicons presentes no arquivo correspondente a amostra 6TG, sendo que em 19174 (95,9%) vezes a sequência relativa ao primer *forward* foi encontrada na porção 5' das sequências de amplicons. No arquivo correspondente à amostra HAT, essa mesma sequência foi detectada e excisada em 19880 (99,4%) sequências de amplicons, sendo que em 19850 (99,3%) vezes foi encontrada na extremidade 5' das mesmas. A sequência correspondente ao DNA plasmidial foi detectada e excisada em 19096 (95,5%) sequências de amplicons presentes no arquivo correspondente a amostra 6TG, sendo que em 19070 (95,4%) vezes estava presente na extremidade 3' dessas sequências. No arquivo correspondente a amostra HAT, essa mesma sequência foi detectada e excisada em 19754 (98,8%) sequências de amplicons, sendo que em 19737 (98,7%) vezes foi encontrada na extremidade 3' dessas sequências. Esses dados demonstram que a grande maioria das sequências presentes nos dois arquivos possuíam em suas extremidades as sequências esperadas.

Após o processamento pelo programa Cutadapt, as sequências de amplicons contidas em cada arquivo foram alinhadas ao banco de sequências-alvo dos shRNAs previamente preparado em programa Bowtie-built. O programa Bowtie foi utilizado para o alinhamento das sequências permitindo-se zero ou um *mismatch*. O número de sequências de amplicons processadas para a amostra 6TG foi de 19982 e para a amostra HAT 19987. O resumo do alinhamento está descrito na Tabela 14.

Tabela 14. Resumo dos Dados Obtidos nos Alinhamentos pelo Programa Bowtie Para a Corrida Teste de NGS

Amostra	Número de <i>Mismatches</i> Permitidos	Número de Sequências Alinhadas*
Controle (6TG)	0	9871 (49,40%)
	1	13901 (69,57%)
Experimental (HAT)	0	10204 (51,05%)
	1	14383 (71,96%)

* Número de sequências com pelo menos um alinhamento reportado.

Os resultados obtidos permitiram concluir que as amostras submetidas à corrida teste de NGS apresentavam excelente qualidade e correspondiam, pelo menos em grande parte, aos amplicons contendo insertos de shRNAs. Podemos observar que o cerca de 50% das sequências presentes em cada amostra avaliada foi alinhada com sucesso contra o banco de sequências-alvo e que essa porcentagem aumentou de maneira significativa quando permitimos a presença de um *mismatch* durante o alinhamento passando para cerca de 70%. Isso pode ser explicado pelo próprio modo com que os shRNAs contidos na biblioteca foram desenhados, pois os mesmos podem conter *mismatches* introduzidos intencionalmente em suas sequências a fim de melhorar a eficiência de *knockdown*, conforme nos foi informado pela equipe de suporte técnico da empresa System Biosciences®.

4.13.2. Corrida Final

O arquivo “Data Report” fornecido pela empresa Fasteris contém especificações técnicas referentes à corrida final de NGS (Anexo 3). A corrida final de NGS em canal 1x100pb gerou 8733500 sequências para a amostra controle 6TG (GZR-1) e 12399284 sequências para a amostra experimental HAT (GZR-2) em formato fastq. As sequências referentes a cada amostra foram primeiramente analisadas com o auxílio do programa FastQC. Resultados similares foram

observados para ambas as amostras (Figuras 27 e 28). Todas as sequencias obtidas na corrida final de NGS para as amostras 6TG e HAT mostraram apresentar 100 nucleotídeos de comprimento, o que está de acordo com o esperado uma vez que as amostras foram sequenciadas em canal 1x100pb. Os resultados obtidos para os outros diferentes parâmetros analisados pelo programa foram em sua maioria equivalentes aos obtidos para a corrida teste, especialmente para nucleotídeos presentes nas posições 13 a 39 (região referente à segunda metade dos insertos de shRNAs). As pequenas divergências observadas em relação aos resultados obtidos na corrida teste estão descritas a seguir:

a) Sequencias duplicadas

Resultados: O nível de duplicação de sequencias obtido para a amostra controle 6TG foi de $\geq 93,58\%$ e o nível de duplicação obtido para a amostra experimental HAT foi de $\geq 93,54\%$. Sequencias repetidas dez ou mais vezes foram 337 vezes mais predominantes na amostra 6TG e 422 vezes mais predominantes na amostra HAT do que as sequencias de cópia única presentes nas respectivas amostras. Embora padrões de duplicação como o observado (nível geral de duplicação $\geq 90\%$, e níveis mais baixos de sequencias com graus de duplicação de dois a nove com aumento abrupto do número de sequencias com dez ou mais cópias) possam estar associados a algum tipo de artefato de PCR (como por exemplo, número excessivo de ciclos de amplificação que podem gerar enriquecimento preferencial de um determinado subgrupo de amplicons), esse mesmo padrão também é condizente com o esperado após NGS de bibliotecas de amplicons que apresentam baixa diversidade. Isso está de acordo com o esperado para nossas bibliotecas de amplicons, geradas a partir de um modelo experimental que visou a depleção (na amostra controle 6TG) ou enriquecimento (na amostra experimental HAT) seletivos de sequencias de shRNAs (incluindo os shRNAs associados com a reativação do cromossomo X) a partir de um grupo limitado de sequencias de shRNAs específicas.

b) Kmers super-representados

Resultados: A grande maioria dos kmers mais amplamente representados nas amostras analisadas foi encontrada na região que compreende os 12 primeiros (*primer forward*) ou os 61 últimos nucleotídeos das sequências de amplicons (sequências plasmidiais, *primer reverse* e adaptadores). Devido às características particulares de nosso modelo experimental, a análise mais aprofundada desse parâmetro se torna difícil e pouco informativa. Isso porque, como o número esperado de cada kmer é calculado de acordo com o conteúdo de bases observado para as sequências contidas no arquivo, a presença dos *primer forward* e *reverse* ricos em pirimidinas e dos adaptadores ricos em purinas em nossas sequências de amplicons pode afetar o número esperado de diferentes kmers e gerar viés no número observado de alguns kmers principalmente nessas regiões. Uma vez que o número esperado de kmers ricos em purinas é diminuído pela alta proporção de pirimidinas presentes nos *primers*, e que a região final das sequências de amplicons, referente aos adaptadores, é extremamente rica em purinas, os seis kmers mais amplamente super-representados em nossas amostras (evidenciados no gráfico relacionado a esse parâmetro) encontram-se nessa região. O leve aumento de frequência desses mesmos kmers observado ao longo da região referente aos insertos de shRNAs pode ser explicado pelo mesmo raciocínio.

Figura 27. Representação gráfica dos parâmetros avaliados pelo programa FastQC para as sequencias geradas na corrida final de NGS da amostra controle 6TG. (A) Qualidade das sequencias por base. (B) Escores de qualidade por sequencia. (C) Conteúdo das sequencias por base. (D) Conteúdo de GC por base. (E) Conteúdo de GC por sequencia. (F) Conteúdo de N por base. (G) Distribuição do tamanho das sequencias. (H) Sequencias duplicadas. (I) KMERs super-representados (apenas os seis KMERs mais frequentemente observados são representados).

Figura 28. Representação gráfica dos parâmetros avaliados pelo programa FastQC para as sequencias geradas na corrida final de NGS da amostra experimental HAT. (A) Qualidade das sequencias por base. (B) Escores de qualidade por sequencia. (C) Conteúdo das sequencias por base. (D) Conteúdo de GC por base. (E) Conteúdo de GC por sequencia. (F) Conteúdo de N por base. (G) Distribuição do tamanho das sequencias. (H) Sequencias duplicadas. (I) KMERs super-representados (apenas os seis KMERs mais frequentemente observados são representados).

Antes do alinhamento, as sequências contidas no banco de dados de sequências de shRNAs propriamente ditos foram cortadas a fim de remover a primeira porção da sequência dos shRNAs (metade presente antes do *hairpin*) e alguns nucleotídeos da extremidade 3' que não eram equivalentes aos nucleotídeos correspondentes dos amplicons sequenciados por NGS. Além do alinhamento contra esse banco de dados, as sequências de amplicons obtidas na corrida final de NGS foram também alinhadas contra as sequências contidas no banco de dados de sequências-alvo e contra as sequências presentes no banco de dados RefMrna. O procedimento adotado para alinhamento e análise foi o mesmo para os três diferentes bancos de dados. Primeiramente, os bancos de dados foram preparados para alinhamento com o uso do programa Bowtie-built. As sequências de amplicons contidas em cada arquivo foram então alinhadas contra as sequências contidas em cada bancos de dados com o auxílio do programa Bowtie permitindo-se zero *mismatches*. O número de sequências alinhadas contra cada banco de dados e para cada amostra está descrito na Tabela 15.

Tabela 15. Resumo dos Dados Obtidos nos Alinhamentos pelo Programa Bowtie Para a Corrida Final de NGS

Banco de Dados	Amostra	Número de Sequências Alinhadas*
Sequências-alvo	Controle (6TG)	4179189 (47,85%)
	Experimental (HAT)	5952637 (48,01%)
Sequências de shRNAs	Controle (6TG)	4659028 (53,35%)
	Experimental (HAT)	6624772 (53,43%)
RefMrna	Controle (6TG)	3946513 (45,19%)
	Experimental (HAT)	5613911 (45,28%)

* Número de *mismatches* permitidos igual a zero. Número de sequências com pelo menos um alinhamento reportado.

Conforme o esperado, em ambas as amostras (controle 6TG e experimental HAT), o número de sequências alinhadas contra o banco de dados de sequências de shRNAs foi maior do que o número de sequências alinhadas contra o banco de sequências-alvo e contra o banco RefMrna. Isso provavelmente ocorreu devido à

presença de *mismatches* intencionalmente introduzidos nas sequências de alguns shRNAs durante o desenho da biblioteca lentiviral.

Após os alinhamentos terem sido completados, o número de amplicons alinhados contra cada sequência presente no banco de dados foi contado através do programa “script bowtie_parser.pl” para cada amostra. As duas tabelas de alinhamento obtidas para as amostras controle e experimental foram combinadas com auxílio do programa “script combine_tables.pl” e os dados foram analisados através da utilização do programa “DESeq” (Anders & Huber, 2010). O mesmo procedimento foi realizado para cada tipo de banco de dados. Essas análises resultaram na construção de tabelas nas quais as sequências (sequências-alvo, sequências de shRNA ou RefSeq dos genes) diferencialmente expressas nas amostras controle 6TG e experimental HAT, de maneira estatisticamente significativa (com p ajustado <0.1), estão listadas (Tabelas 16, 17 e 18). Vale notar que tanto candidatos cuja expressão estava aumentada na amostra controle 6TG como candidatos cuja expressão estava aumentada na amostra experimental HAT foram identificados (Tabelas 16, 17 e 18).

As sequências-alvo e sequências de shRNAs propriamente ditos identificadas como candidatas foram recuperadas dos bancos de dados correspondentes a partir de suas identificações e alinhadas contra o genoma e transcriptoma humano. Os transcritos e regiões cromossômicas que apresentaram alta similaridade (*query coverage* 100% e *identity* $>90\%$ para transcritos, *query coverage* 100% e *identity* 100% para regiões genômicas) com as sequências candidatas obtidas em nossa triagem funcional estão listados na Tabela 19.

Tabela 16. Lista de Candidatos Obtidos por Análise em Programa DESeq para o Banco de Dados de Sequências-Alvo

Identificação Alvo	Sonda Correspondente	Base Mean	Base Mean A	Base Mean B	Fold Change	Log2 Fold Change	p Value	Adjusted p Value
g8923671_160	219572_at	99,43689	0	198,873782	Infinito	Infinito	1,17E-05	7,03E-02
AL022724_278	215100_at	120,3319	1,131371	239,532422	2,12E+02	7,726005	5,52E-06	4,81E-02
g13569344_533	211918_x_at	148,29797	2,262742	294,333198	1,30E+02	7,023235	1,02E-06	1,00E-02
g12803794_456	201556_s_at	184,46648	5,656854	363,276109	6,42E+01	6,004923	4,42E-07	4,96E-03
g7106869_265	210075_at	360,65981	12,445079	708,874548	5,70E+01	5,831883	1,73E-11	1,35E-06
g8922889_795	219086_at	253,51546	14,707821	492,323096	3,35E+01	5,06495	1,25E-07	1,64E-03
g12056474_1102	206095_s_at	382,50941	38,466609	726,552218	1,89E+01	4,239388	2,52E-08	4,05E-04
g12275865_448	211001_at	288,8708	42,992092	534,749503	1,24E+01	3,63672	1,11E-05	7,03E-02
g3811344_154	209785_s_at	656,47794	195,727157	1117,22871	5,71E+00	2,513009	8,68E-06	6,32E-02
Z98200_598	215963_x_at	103,96237	207,040866	0,8838835	4,27E-03	-7,871844	1,38E-05	7,73E-02
BE217923_134	228686_at	107,92217	214,960461	0,8838835	4,11E-03	-7,925999	8,86E-06	6,32E-02
g4506984_326	208177_at	172,976	345,068109	0,8838835	2,56E-03	-8,608809	7,29E-09	2,86E-04
NM_033519_1040	1553061_at	152,16938	304,338759	0	0,00E+00	Infinito negativo	1,19E-08	3,12E-04
AF218012_831	222245_s_at	145,38115	290,762308	0	0,00E+00	Infinito negativo	2,58E-08	4,05E-04

Base Mean: número médio de sequências presentes em ambas as amostras (normalizado). Base Mean A: número médio de sequências presentes na amostra controle 6TG (normalizado). Base Mean B: número médio de sequências presentes na amostra experimental HAT (normalizado). Em letras pretas: sequências mais predominantes na amostra experimental HAT. Em letras vermelhas: sequências mais predominantes na amostra controle 6TG.

Tabela 17. Lista de Candidatos Obtidos por Análise em Programa DESeq para o Banco de Dados de Sequências de shRNAs

Identificação shRNA	Sonda Correspondente	Base Mean	Base Mean A	Base Mean B	Fold Change	Log2 Fold Change	p Value	Adjusted p Value
N(1)219572_atfr261	219572_at	96,86937	0	193,73873	Infinito	Infinito	9,97E-06	6,69E-02
N(0)215100_atfr130	215100_at	116,45457	1,130388	231,778755	2,05E+02	7,679786	5,11E-06	5,01E-02
N(2)211918_x_atfr482	211918_x_at	144,44397	2,260777	286,627163	1,27E+02	6,986213	8,68E-07	9,56E-03
N1_201556_s_at_30_f	201556_s_at	180,64097	5,651942	355,629998	6,29E+01	5,975487	3,50E-07	4,40E-03
N(0)210075_atfr346	210075_at	346,24286	11,303883	681,181837	6,03E+01	5,91315	1,18E-11	1,04E-06
N(1)219086_atfr527	219086_at	246,84241	12,434272	481,250545	3,87E+01	5,274394	4,51E-08	6,62E-04
N3_206095_s_at_93_f	206095_s_at	369,85815	37,302815	702,413479	1,88E+01	4,234964	2,26E-08	3,99E-04
N(1)211001_atfr19	211001_at	280,55747	41,824368	519,29057	1,24E+01	3,634126	9,63E-06	6,69E-02
N(1)211998_atfr158	211998_at	266,7225	40,69398	492,751018	1,21E+01	3,597972	1,75E-05	9,62E-02
N0_209785_s_at_0_f	209785_s_at	639,08716	186,514075	1091,66024	5,85E+00	2,549168	6,11E-06	5,38E-02
N(2)213183_s_atfr76	213183_s_at	107,14115	212,513006	1,7693035	8,33E-03	-6,908226	1,67E-05	9,62E-02
N(3)215963_x_atfr252	215963_x_at	102,17728	203,469899	0,8846517	4,35E-03	-7,84549	1,06E-05	6,69E-02
N(0)228686_atfr94	228686_at	106,13363	211,382618	0,8846517	4,19E-03	-7,900531	6,73E-06	5,39E-02
N3_208177_at_152_f	208177_at	170,00058	339,116499	0,8846517	2,61E-03	-8,582456	4,96E-09	2,18E-04
N(2)1553061_atfr0	1553061_at	148,08087	296,161743	0	0,00E+00	Infinito negativo	9,72E-09	2,85E-04
N(1)222245_s_atfr288	222245_s_at	142,42893	284,857859	0	0,00E+00	Infinito negativo	1,89E-08	3,99E-04

Sonda correspondente: refere-se a identificação da sequência correspondente no banco de dados de sequências-alvo. Base Mean: número médio de sequências presentes em ambas as amostras (normalizado). Base Mean A: número médio de sequências presentes na amostra controle 6TG (normalizado). Base Mean B: número médio de sequências presentes na amostra experimental HAT (normalizado). Em letras pretas: sequências mais predominantes na amostra experimental HAT. Em letras vermelhas: sequências mais predominantes na amostra controle 6TG.

Tabela 18. Lista de Genes Candidatos Obtidos por Análise em Programa DESeq para o Banco de Dados RefMrna

Identificação Alvo	Base Mean	Base Mean A	Base Mean B	Fold Change	Log2 Fold Change	p Value	Adjusted p Value	Gene Correspondente
NM_017665	72.26069	0	144.5213776	Infinito Positivo	Infinito Positivo	2.47E-07	9.39E-04	<i>Homo sapiens zinc finger, CCHC domain containing 10 (ZCCHC10), mRNA</i>
NM_032411	54.93503	0	109.8700531	Infinito Positivo	Infinito Positivo	8.82E-06	2.06E-02	<i>Homo sapiens chromosome 2 open reading frame 40 (C2orf40), mRNA (ou ECRG4)</i>
NM_032744*	54.51245	0	109.0248989	Infinito Positivo	Infinito Positivo	9.63E-06	2.07E-02	<i>Homo sapiens androgen-dependent TFPI-regulating protein (ADTRP), transcript variant 2, mRNA</i>
NM_145304	51.13183	0	102.2636648	Infinito Positivo	Infinito Positivo	1.96E-05	3.13E-02	<i>Homo sapiens chromosome 7 open reading frame 33 (C7orf33), mRNA</i>
NM_001010852	87.21992	1.183216	173.2566222	1.46E+02	7.194053	9.12E-08	6.93E-04	<i>Homo sapiens clavesin 2 (CLVS2), mRNA</i>
NM_032554*	70.73941	1.183216	140.2956063	1.19E+02	6.889613	2.15E-06	6.53E-03	<i>Homo sapiens hydroxycarboxylic acid receptor 1 (HCAR1), mRNA</i>
NM_001143948*	60.59756	1.183216	120.0119042	1.01E+02	6.66432	1.54E-05	2.73E-02	<i>Homo sapiens androgen-dependent TFPI-regulating protein (ADTRP), transcript variant 1, mRNA</i>
NR_026684	59.49886	2.366432	116.6312872	4.93E+01	5.623098	6.23E-05	7.90E-02	<i>Homo sapiens Helicobacter pylori responsive 1 (non-protein coding) (HPYR1), long non-coding RNA</i>
NM_001039840	73.61294	3.549648	143.6762233	4.05E+01	5.339002	1.32E-05	2.51E-02	<i>Homo sapiens cysteine-rich hydrophobic domain 1 (CHIC1), mRNA</i>
NM_006742	129.64666	26.030751	233.2625743	8.96E+00	3.163666	5.92E-05	7.82E-02	<i>Homo sapiens protein serine kinase H1 (PSKH1), mRNA</i>
NM_033195	241.79863	80.458685	403.1385795	5.01E+00	2.324956	5.24E-05	7.25E-02	<i>Homo sapiens lactate dehydrogenase A-like 6B (LDHAL6B), mRNA</i>
NM_014232	289.71888	99.39014	480.0476167	4.83E+00	2.272003	2.09E-05	3.17E-02	<i>Homo sapiens vesicle-associated membrane protein 2 (synaptobrevin 2) (VAMP2), mRNA</i>
NM_006122	454.94654	170.383098	739.5099729	4.34E+00	2.117787	2.58E-06	7.14E-03	<i>Homo sapiens mannosidase, alpha, class 2A, member 2 (MAN2A2), mRNA</i>
NM_014034	570.39461	268.590022	872.1991909	3.25E+00	1.699252	3.75E-05	5.43E-02	<i>Homo sapiens anti-silencing function 1A histone chaperone (ASF1A), mRNA</i>

Tabela 18. Lista de Genes Candidatos Obtidos por Análise em Programa DESeq para o Banco de Dados RefMrna (Continuação)

Identificação Alvo	Base Mean	Base Mean A	Base Mean B	Fold Change	Log2 Fold Change	p Value	Adjusted p Value	Gene Correspondente
NM_017868	275.77383	460.271007	91.2766595	1.98E-01	-2.334166	1.62E-05	2.73E-02	<i>Homo sapiens tetra tricopeptide repeat domain 12 (TTC12), mRNA</i>
NM_198576	203.09057	353.781571	52.3995638	1.48E-01	-2.755232	1.02E-05	2.07E-02	<i>Homo sapiens agrin (AGRN), mRNA</i>
NM_003052	267.49132	473.286383	61.6962606	1.30E-01	-2.939458	1.85E-07	8.01E-04	<i>Homo sapiens solute carrier family 34 (type II sodium/phosphate cotransporter), member 1 (SLC34A1), transcript variant 1, mRNA</i>
NM_017842	271.21	493.401054	49.0189468	9.93E-02	-3.331349	6.00E-09	6.08E-05	<i>Homo sapiens solute carrier family 48 (heme transporter), member 1 (SLC48A1), mRNA</i>
NM_207445	143.59171	262.673942	24.5094734	9.33E-02	-3.421862	5.40E-06	1.37E-02	<i>Homo sapiens chromosome 15 open reading frame 54 (C15orf54), mRNA</i>
NM_181644	150.26843	276.872534	23.6643191	8.55E-02	-3.548437	1.78E-06	6.01E-03	<i>Homo sapiens major facilitator superfamily domain containing 4 (MFS4), mRNA</i>
NM_018430	76.1484	151.451642	0.8451543	5.58E-03	-7.485427	1.26E-07	7.66E-04	<i>Homo sapiens translin-associated factor X interacting protein 1 (TSNAXIP1), mRNA</i>
NR_015360	112.82809	224.811032	0.8451543	3.76E-03	-8.055282	5.78E-11	8.78E-07	<i>Homo sapiens uncharacterized LOC644873 (FLJ33630), long non-coding RNA</i>
NR_002140	159.14255	318.285092	0	0.00E+00	Infinito Negativo	4.74E-16	1.44E-11	<i>Homo sapiens olfactory receptor, family 6, subfamily W, member 1 pseudogene (OR6W1P), non-coding RNA</i>
NM_004589	66.8517	133.703403	0	0.00E+00	Infinito Negativo	1.51E-07	7.66E-04	<i>Homo sapiens SCO1 cytochrome c oxidase assembly protein (SCO1), mRNA</i>

* Candidatos com pelo menos parte dos shRNAs mapeados alinhando também com outros genes-alvo (ambíguos). Base Mean: número médio de sequências presentes em ambas as amostras (normalizado). Base Mean A: número médio de sequências presentes na amostra controle 6TG (normalizado). Base Mean B: número médio de sequências presentes na amostra experimental HAT (normalizado). Em letras pretas: sequências mais predominantes na amostra experimental HAT. Em letras vermelhas: sequências mais predominantes na amostra controle 6TG.

Tabela 19. Identificação dos Transcritos e Regiões Genômicas Associados às Sequências Candidatas Obtidas nos Alinhamentos Contra os Bancos de Dados de Sequências-Alvo e Sequências de shRNAs

Identificação shRNA Identificação Sequência -Alvo Sequências Correspondentes	Transcriptoma*	Nº(s) de Acesso NCBI (Nucleotide)	Localização Cromossômica e Coordenadas Genômicas	Genoma**
N(1)219572_atfr261 TGCAGTGGTCCGAGCTGACTTCTCACA g8923671_160 TGTGAGAAGTCAGCTCGGACCACTGCA	<i>Homo sapiens</i> Ca ⁺⁺ -dependent secretion activator 2 (CADPS2), mRNA. Transcript variants 1, 2, 3. 100% query coverage e 96% identity.	NM_001167940.1 // NM_001009571.3 // NM_017954.10	7q31.32 chr7:121,958,477 - 122,526,812	Cromossomos 4, 5, 8, 10, 16, 17
N(0)215100_atfr130 TAGAGCCATAGGATGCGGCTGATGTAA AL022724_278 TTACATCAGCCGCATCCTATGGCTCTA	<i>Homo sapiens</i> androgen-dependent TFPI-regulating protein (ADTRP), mRNA. Transcript variants 1, 2. 100% query coverage e 100% identity.	NM_001143948.1 // NM_032744.3	6p24.1 chr6:11,713,887 - 11,779,279	----
N(2)211918_x_atfr482 TGCTCCAGAGTGTGAGCAGTGATATTC g13569344_533 GAATATCACTGCTGACACTCTGGAGCA	<i>Homo sapiens</i> pappalysin 2 (PAPPA2), mRNA. Transcript variant 1. 100% query coverage e 100% identity.	NM_020318.2	1q23-q25 chr1:176,432,307- 176,811,970	Cromossomos 1, 2, 6, 8, 14
N1_201556_s_at_30_f ATGATGGCGCAAATCACTCCCAAGATG g12803794_456 CATCTTGGGAGTGATTGCGCCATCAT	<i>Homo sapiens</i> vesicle-associated membrane protein 2 (synaptobrevin 2) (VAMP2), mRNA. 100% query coverage e 100% identity.	NM_014232.2	17p13.1 chr17:8,062,464 - 8,066,292	Cromossomos 2, 17
N(0)210075_atfr346 GGTCCGGAGAGCCCAGCATTCACTG g7106869_265 CAGTATGAATGCTGGGCTCTCCGACC	<i>Homo sapiens</i> membrane-associated ring finger (C3HC4) 2, E3 ubiquitin protein ligase (MARCH2), mRNA. Transcript variants 1, 2, 3. 100% query coverage e 100% identity.	NM_001005416.1 // NM_001005415.1 // NM_016496.4	19p13.2 chr19:8,478,186 - 8,503,900	Cromossomo 19
N(1)219086_atfr527 CAGACGCGCCACAGGGATTCTGATCA g8922889_795 TGATCAGGAATCCCTGTGGCGCGTCTG	<i>Homo sapiens</i> zinc finger protein 839 (ZNF839), mRNA. Transcript variants 1, 2 and 3. 100% query coverage e 100% identity.	NM_001267828.1 // NM_001267827.1 // NM_018335.4	14q32.31 chr14:102,786,088- 102,809,511	Cromossomo 14
N3_206095_s_at_93_f CAATGAGGCCATACTGTAATGATTCAC g12056474_1102 GTGAATCATTACAGTATGGCCTCATTG	<i>Homo sapiens</i> serine/arginine-rich splicing factor 10 (SRSF10), mRNA. Transcript variants 1, 4, 5, 6, 7. 100% query coverage e 100% identity.	NR_034035.1 // NM_001191009.1 // NM_001191007.1 // NM_001191006.1 // NM_006625.4	1p36.11 chr1:24,292,934 - 24,306,952	Cromossomos 1, 2, 4, 8, 10

Tabela 19. Identificação dos Transcritos e Regiões Genômicas Associados às Sequências Candidatas Obtidas nos Alinhamentos Contra os Bancos de Dados de Sequências-Alvo e Sequências de shRNAs (Continuação)

Identificação shRNA Identificação Sequência -Alvo Sequências Correspondentes	Transcriptoma*	Nº(s) de Acesso NCBI (Nucleotide)	Localização Cromossômica e Coordenadas Genômicas	Genoma**
N(1)211001_atfr19 CTTGAGGGATGCTCAGACCTTGCTTAA g12275865_448 TTAAGCAAGGTCTGAGCATCCCTCAAG	Sem correspondência de alta similaridade ***	***	***	Cromossomos 1, 7, 11, 14
N(1)211998_atfr158 CCAGACACGCTACAGCATAGCACGTTC	<i>Homo sapiens H3 histone, family 3B (H3.3B) (H3F3B), mRNA.</i> 100% query coverage e 100% identity.	NM_005324.3	17q25.1 chr17:73,772,514 - 73,775,859	Cromossomos 11, 17
N0_209785_s_at_0_f CCAAGCAGCAACTTCGGGCACTATCCT g3811344_154 AGGATAGTGCCCGAAGTTGCTGCTTGG	<i>Homo sapiens phospholipase A2, group IVC (cytosolic, calcium-independent) (PLA2G4C), mRNA. Transcript variants 1, 2, 3.</i> 100% query coverage e 100% identity.	NM_001159323.1 // NM_001159322.1 // NM_003706.2	19q13.33 chr19:48,551,099 - 48,614,108	Cromossomos 2, 19
N(2)213183_s_atfr76 CGAGAAGGTACACTGGTCCCAAGGTGT	<i>Homo sapiens cyclin-dependent kinase inhibitor 1C (p57, Kip2) (CDKN1C), mRNA. Transcript variants 1, 2, 3.</i> 100% query coverage e 100% identity.	NM_001122631.1 // NM_001122630.1 // NM_000076.2	11p15.4 – 11p15.5 chr11:2,904,447 - 2,906,994	Cromossomo 11
N(3)215963_x_atfr252 CAAAGCCACCCAGAGGATTGGTGCTCT Z98200_598 AGAGCACCAATCCTCTGGGTGGCTTTG	<i>Homo sapiens ribosomal protein L3 (RPL3), mRNA. Transcript variants 1, 2.</i> 100% query coverage e 93% identity.	NM_001033853.1 // NM_000967.3	22q13 Chr22:39,708,887- 39,715,670	Cromossomos 1, 3, 6, 9, 15, 20
N(0)228686_atfr94 CATGGTTGGCAGTGTCCAAGTGTTAA BE217923_134 TTAACCACTTGACACTGCCAACCATG	<i>Homo sapiens uncharacterized LOC644873 (FLJ33630), non- coding RNA.</i> 100% query coverage e 100% identity.	NR_015360.1	5q23.3 Chr5:127,357,243- 127,418,766	Cromossomos 5, 6, 10, 12, 14
N3_208177_at_152_f CACATGCACACACCGCACTCTCAC g4506984_326 GTGAGAGTGTGCGGTGTGTGTCATGTG	<i>Homo sapiens solute carrier family 34 (type II sodium/phosphate cotransporter), member 1 (SLC34A1), mRNA. Transcript variant 1.</i> 100% query coverage e 100% identity.	NM_003052.4	5q35.3 chr5:176,811,431 - 176,825,848	Cromossomos 4, 5, 7, 16, 21

Tabela 19. Identificação dos Transcritos e Regiões Genômicas Associados às Sequências Candidatas Obtidas nos Alinhamentos Contra os Bancos de Dados de Sequências-Alvo e Sequências de shRNAs (Continuação)

Identificação shRNA Identificação Sequência -Alvo Sequências Correspondentes	Transcriptoma*	Nº(s) de Acesso NCBI (Nucleotide)	Localização Cromossômica e Coordenadas Genômicas	Genoma**
N(2)1553061_atfr0 CAGCACTGCAGCCACAATGAGGATGTA	<i>Homo sapiens olfactory receptor, family 6, subfamily W, member 1 pseudogene (OR6W1P), non-coding RNA.</i>	NR_002140.1	7q34	Cromossomos 2, 4, 6, 7, 9, 11, 12, 14, 16, 21, 22, X
NM_033519_1040 TACATCCTCATTGTGGCTGCAGTGCTG	100% query coverage e 100% identity.		Chr7:142,759,381-142,760,882	
N(1)222245_s_atfr288 CCACCTTGCCCGTGAGGATGTACACAT	<i>Homo sapiens fer-1-like 4 (C. elegans) pseudogene (FER1L4), non-coding RNA.</i>	NR_024377.1	20q11.22	Cromossomos 12, 17, 19, 20
AF218012_831 ATGTGTACATCCTCACGGGCAAGGTGG	100% query coverage e 100% identity.		Chr20:34,146,507-34,195,818	

* As variantes de transcritos preditas não estão listadas. ** Estão listados somente os cromossomos para os quais a similaridade foi de 100% query coverage e 100% identity. *** Identificada no banco de dados de sequências-alvo como referente ao gene *Homo sapiens tripartite motif containing 29 (TRIM29)*, mRNA, NM_012101, localização cromossômica 11q23.3, coordenada genômica chr11:119,981,994-120,008,863. Em letras pretas: sequências mais predominantes na amostra experimental HAT. Em letras vermelhas: sequências mais predominantes na amostra controle 6TG.

Uma das sequencias candidatas, identificada tanto a partir do alinhamento contra o banco de sequencias-alvo quanto a partir do alinhamento contra o banco de sequencias de shRNAs propriamente ditos, de denominação g12275865_448 e N(1)211001_atfr19 respectivamente, não apresentou nenhum transcrito correspondente com alta similaridade quando blastada contra o transcriptoma humano no site do NCBI, embora no banco de dados de sequencias-alvo essa sequencia candidata estivesse relacionada ao gene TRIM29 (*tripartite motif containing 29*). De fato, o alinhamento dessa sequencia contra a sequencia do gene TRIM39 demonstrou não haver similaridade entre elas. Sendo assim, apenas regiões genômicas apresentaram alta similaridade com essa sequencia candidata (Tabela 19).

Em relação aos dados obtidos na alinhamento contra o banco RefMrna, as sequencias candidatas correspondentes às variantes de transcrito 1 e 2 do gene *androgen-dependent TFPI-regulating protein (ADTRP)* mostraram ser ambíguas. No entanto tal ambiguidade referia-se apenas à essas duas variantes de transcrito. Outro resultado ambíguo foi notado para o gene *hydroxycarboxylic acid receptor 1 (HCAR1)*. Pelo menos parte das sequencias alinhadas contra esse gene alinharam também com quatro outros transcritos (NM_003343, NM_182688 e NM_001202489: *ubiquitin-conjugating enzyme E2G 2 (UBE2G2)*, *transcript variant 1, 2 and 3, mRNA*; NM_005955: *metal-regulatory transcription factor 1 (MTF1)*, *mRNA*).

Conforme observado nas tabelas 18 e 19, todos os candidatos identificados a partir do alinhamento contra o banco de sequencias-alvo foram também identificados a partir do alinhamento contra o banco de sequencias de shRNAs propriamente ditas. Além disso, no alinhamento contra esse segundo banco de dados, mais duas sequencias diferenciamente representadas entre nossas duas amostras foram identificadas – uma significativamente mais representada na amostra experimental HAT (correspondente ao gene *H3 histone, family 3B (H3.3B) (H3F3B)*) e outra significativamente mais representada na amostra controle 6TG (correspondente ao gene *cyclin-dependent kinase inhibitor 1C (p57, Kip2) (CDKN1C)*). Foi posteriormente constatado que as sequencias candidatas relacionadas a esses dois genes de fato não estavam presentes no banco de sequencias-alvo.

As identidades dos genes candidatos obtidos a partir dos alinhamentos contra os bancos de sequencias-alvo/sequencias de shRNAs propriamente ditas

diferiram substancialmente da identidade dos genes candidatos obtidos a partir do alinhamento contra o banco RefMrna (Tabelas 18 e 19). De fato, somente cinco genes candidatos eram comuns aos três bancos de dados dos quais dois estavam mais representados na amostra experimental HAT (genes *androgen-dependent TFPI-regulating protein (ADTRP)* e *vesicle-associated membrane protein 2 (Synaptobrevin 2) (VAMP2)*) e três estavam mais representados na amostra controle 6TG (genes *solute carrier family 34 (type II sodium/phosphate cotransporter), member 1 (SLC34A1)*, *Homo sapiens uncharacterized LOC644873 (FLJ33630) non-coding RNA* e *olfactory receptor, family 6, subfamily W, member 1 pseudogene (OR6W1P) non-coding RNA*). Na análise contra os bancos de sequencias-alvo e sequencias de shRNAs propriamente ditos as sequencias de amplicons alinhadas contra cada uma das sequencias curtas de 27pb que compõem esses bancos são contadas. Desta maneira, diferentes shRNAs para um mesmo transcrito são avaliados isoladamente. Diferentemente, como o banco de RefMrna é constituído pelas sequencias completas de cada transcrito, na análise realizada contra esse banco de dados as sequencias de amplicons referentes a diferentes shRNAs para um mesmo transcrito são contadas conjuntamente.

Em conclusão, nossa triagem funcional permitiu identificar diversas sequencias/genes diferencialmente representadas entre nossas amostras controle 6TG e experimental HAT. Coletivamente, as análises realizadas contra os três bancos de dados utilizados permitiram identificar sequencias correspondentes a 20 diferentes genes de representação significativamente mais elevada na amostra experimental HAT. Uma outra sequencia identificada no alinhamento contra o banco de sequencias de shRNAs não apresentou correspondência significativa com nenhum gene quando alinhada contra o transcriptoma humano. Além disso, foram identificadas também sequencias correspondentes a 13 genes de representação significativamente mais elevada na amostra controle 6TG.

O significado desses achados bem como as características e funções de cada gene e sua possível relação com o processo de manutenção da XCI ou com outros mecanismos biológicos serão discutidos na seção seguinte.

Discussão

5. Discussão

5.1. Metodologia Utilizada

Durante o período que compreendeu este trabalho de doutorado, uma biblioteca de shRNAs com cobertura genômica foi utilizada como ferramenta para triagem funcional de genes candidatos potencialmente envolvidos no processo de manutenção da inativação do cromossomo X em humanos. O trabalho foi realizado a partir da premissa de que o *knockdown* de um determinado gene envolvido na manutenção da XCI resultaria na reativação de genes que estão normalmente silenciados no cromossomo Xi.

Para que fosse possível selecionar a população de células na qual houve reativação do cromossomo X após transdução com a biblioteca de shRNAs, a triagem funcional foi realizada em células da linhagem GM1661, heterozigota para uma mutação no gene submetido à XCI *HPRT*, que apresentavam desvio total da XCI de forma a só expressar o alelo mutante de *HPRT*, sendo assim *HPRT*⁻. A reativação do cromossomo X levaria essas células a expressar o alelo normal de *HPRT*, se tornando *HPRT*⁺ e permitindo com que fossem selecionadas em meio HAT. Sendo assim, as células *HPRT*⁻ foram transduzidas com a biblioteca de shRNAs e posteriormente divididas em duas populações – uma que foi mantida em meio 6TG para a seleção de células *HPRT*⁻, e outra que foi cultivada em meio HAT para a seleção de células *HPRT*⁺ (população que inclui células que sofreram reativação do cromossomo X, pelo menos em alguma extensão, como consequência da expressão de shRNAs).

Os shRNAs integrados no DNA de cada uma dessas duas populações de células foram amplificados por *multiplex* PCR e as amostras de amplicons purificadas de gel de agarose foram submetidas ao sequenciamento de próxima geração. As sequências obtidas para cada amostra foram alinhadas contra três diferentes bancos de dados e os resultados dos alinhamentos obtidos para as amostras controle 6TG e experimental HAT foram comparados através do uso do programa DESeq a partir do qual foram geradas listas de sequências/genes candidatos diferencialmente representados entre as duas amostras. Os resultados

dessas análises nos permitiram identificar 21 candidatos em potencial para o processo de manutenção da XCI em humanos.

5.2. Seleção em 6TG e HAT

Antes da transdução com a biblioteca de shRNAs, as células foram selecionadas em meio contendo 6TG na concentração de 60uM durante cerca de quatro semanas com o objetivo de isolar uma população HPRT⁻, na qual o mesmo cromossomo X estaria ativo em todas as células.

Entretanto, o sequenciamento convencional de Sanger para análise de SNPs heterozigotos presentes em genes submetidos à XCI realizado a partir de cDNA de células não selecionadas e de células selecionadas em 6TG, demonstrou que nossa amostra inicial de células já possuía um desvio praticamente total de inativação do cromossomo X, a favor de células HPRT⁻, antes mesmo da seleção em meio contendo 6TG. Essa observação foi corroborada pelos resultados obtidos para as curvas de crescimento geradas para as células dessa linhagem, uma vez que não foi observada diferença no número de células presentes nas amostras de células não selecionadas e células selecionadas em 6TG durante todo o período de tratamento em cultura.

Por já haver um desvio de inativação do cromossomo X em nossa amostra inicial de células da linhagem GM1661, esses experimentos, isoladamente, não permitiram avaliar a eficácia do tratamento em meio contendo a droga 6TG. Porém, vale notar que a concentração de 6TG utilizada em nossos experimentos (60uM) é o dobro da concentração recomendada para a seleção de células HPRT⁻ em cultura pela empresa que comercializa a droga, e está também de acordo com a concentração utilizada na literatura (Mekhoubad *et al.*, 2012). Além disso, os experimentos de seleção em meio 6TG realizados em fibroblastos humanos primários normais para o gene *HPRT* (*HPRT*^{+/+}) das linhagens FibN1 e Hs27 demonstram que a concentração de 6TG utilizada é suficiente para causar intensa depleção de células em cultura após poucos dias de tratamento.

A concentração de HAT utilizada seguiu a recomendação da empresa que comercializa o produto e também estava de acordo com o utilizado na literatura (Mekhoubad *et al.*, 2012). O meio de cultura utilizado durante a seleção continha

100uM de hipoxantina sódica, 400nM de aminopterina e 16uM de timidina. As curvas de crescimento geradas para células GM1661 não selecionadas ou previamente selecionadas em meio 6TG submetidas à seleção em meio contendo HAT demonstraram que o uso de HAT nas concentrações especificadas é eficiente para causar depleção acentuada de células HPRT⁻ após poucos dias de tratamento em cultura.

5.3. Representatividade da Biblioteca de shRNAs

A triagem funcional a partir de bibliotecas de shRNAs é uma ferramenta valiosa para identificação de reguladores genéticos envolvidos em uma extensa gama de processos biológicos. Entretanto, para que a triagem funcional seja bem sucedida, é imperativo que as condições experimentais sejam cuidadosamente otimizadas. Diversos fatores que exercem influência sob cada um dos passos experimentais adotados podem afetar a qualidade da triagem e a significância biológica dos resultados obtidos (Sims *et al.*, 2011; Strezoska *et al.*, 2012). Sendo assim, cada um dos passos experimentais envolvidos em nossa triagem funcional foi cuidadosamente otimizado antes que nossas amostras pudessem ser submetidas ao NGS.

Nossos primeiros esforços nesse sentido visaram assegurar uma ampla representação dos diferentes insertos de shRNAs que compõem a biblioteca lentiviral em nossas amostras de células transduzidas. Isso porque, quanto maior a representatividade da biblioteca, maior a reproducibilidade do experimento de triagem e menor a quantidade de candidatos falsos negativos e falsos positivos obtidos (Sims *et al.*, 2011; Strezoska *et al.*, 2012). A escolha da uma linhagem celular altamente permissiva à transdução por partículas lentivirais, a seleção da melhor versão da biblioteca de shRNAs para uso no modelo celular escolhido e a determinação da quantidade de partículas lentivirais a ser utilizada durante a transdução das células-alvo com a biblioteca de shRNAs foram aspectos cuidadosamente avaliados em nosso trabalho através de uma série de experimentos preliminares de transdução realizados em fibroblastos humanos primários que incluiu também a própria linhagem celular alvo GM1661.

O número de células submetidas à transdução com a biblioteca de shRNAs também afeta a taxa de representação dos diferentes tipos de shRNAs durante a triagem funcional e é provavelmente um dos fatores mais críticos para que uma boa representatividade da biblioteca seja adquirida (Sims *et al.*, 2011; Strezoska *et al.*, 2012). O número de células a serem transduzidas em nosso trabalho foi baseado na recomendação da empresa que comercializa a biblioteca de shRNAs e foi estabelecido em 10^7 . Sendo assim, como a biblioteca de shRNAs contém cerca de 200 mil tipos diferentes de shRNAs, que foram transduzidos de forma a obter uma taxa de cópias do inserto exógeno por célula entre 0,5 (determinada através dos experimentos de otimização da transdução de células GM1661 com o vírus controle pSIH-H1-siLuc-copGFP) e 1,0 (valor máximo esperado de acordo com os resultados obtidos pelo experimento de seleção de células transduzidas com a biblioteca de shRNAs em puromicina), podemos concluir que, logo após o procedimento de transdução, cada tipo de shRNAs estava representado de 25 a 50 vezes em nossa amostra.

Na literatura, o número relativo à representação de cada shRNA contido nas bibliotecas virais em experimentos de triagem funcional varia consideravelmente. Alguns trabalhos foram desenvolvidos de forma a obter dez cópias de cada inserto na amostra inicial, enquanto outros almejavam obter mil cópias (Huang *et al.*, 2008; Schlabachet *et al.*, 2008; Silva *et al.*, 2008; Hurov *et al.*, 2010; Mendes-Pereira *et al.*, 2011). Entretanto, a complexidade da biblioteca de shRNAs também influencia na determinação do número de células a serem utilizadas. Bibliotecas mais complexas exigem um maior número de células para que cada shRNA esteja representado na amostra e a quantidade de células necessárias para manter a representação da biblioteca aumenta conforme maiores números de cópias por inserto são esperados, o que torna a manipulação em cultura e a produção das reações de PCR progressivamente mais difíceis. Além disso, o número de cópias por inserto a ser utilizado depende também da estratégia experimental. Experimentos que, como o descrito nessa tese, visam identificar shRNAs enriquecidos após a aplicação de uma pressão seletiva tendem a ser menos estridentes do que experimentos que visam identificar shRNAs que foram depletados da amostra (Strezoska *et al.*, 2012).

Após aplicação da pressão seletiva nas células transduzidas (no nosso caso a seleção em meio contendo HAT ou 6TG), os fatores que influenciam a representação numérica dos insertos de shRNAs presentes em cada amostra estão

relacionados principalmente à amplificação dos insertos de shRNAs durante o procedimento de *multiplex* PCR. Primeiramente, é extremamente importante assegurar também que o número de ciclos utilizados durante a amplificação dos insertos de shRNAs não ultrapasse a fase exponencial da PCR, na qual os componentes da reação estão ainda em excesso e existe duplicação praticamente exata dos produtos de PCR a cada ciclo. Durante as fases posteriores da PCR, fase linear e fase de *plateau*, a taxa de eficiência da reação de amplificação é fortemente afetada devido ao consumo progressivo dos reagentes necessários para a geração dos amplicons e decaimento da atividade da DNA polimerase (inibida quando o número de amplicons presentes é alto), o que pode alterar a representatividade numérica das sequências de shRNA integradas na amostra (Kainz, 2000; Strezoska, 2012). O uso de nested primers em uma segunda rodada de amplificação não só permite que a primeira PCR seja conduzida com um menor número de ciclos, mas também aumenta a especificidade da amplificação, eliminando consideravelmente o número de produtos inespecíficos formados. A *nested* PCR melhora o limite inferior de detecção de produtos de amplificação e torna a reação mais robusta. A relação matemática entre o número de sequências-molde e o número de produtos amplificados é mantida, desde que a primeira rodada de amplificação seja mantida na fase exponencial e que concentrações adequadas de *primers* sejam utilizadas nas duas rodadas de amplificação (Haff, 1994).

A amplificação dos insertos de shRNAs contidos em nossas amostras foi realizada por *nested* PCR e as reações de amplificação foram baseadas no manual que acompanha a biblioteca. O número de ciclagens nas duas rodadas de amplificação realizadas para amplificação dos insertos de shRNAs em nossas amostras foi extensivamente testado antes que pudéssemos determinar o número mínimo de ciclagens necessárias para obtenção das bandas de interesse com tamanho adequado em nossas amostras. O número de ciclagens estabelecido para a primeira rodada de PCR foi inferior ao recomendado pelo manual da empresa que comercializa a biblioteca (18 ciclos ao invés dos 20 ciclos recomendados) e o número de ciclos utilizados durante a segunda rodada de amplificação foi condizente com o especificado no manual (18 ciclos).

Além disso, é importante que a reação de amplificação seja realizada a partir de uma quantidade de DNA genômico suficiente para garantir a representatividade dos insertos de shRNAs contidos em cada amostra. Em nosso trabalho, a

quantidade de DNA utilizada durante a primeira rodada de amplificação foi o dobro da quantidade recomendada pela empresa que comercializa a biblioteca. Como 10ug de DNA genômico foram utilizados na primeira rodada de PCR e cada célula possui cerca de 6pg de DNA (transduzidas de forma com que o número de cópias de inserto exógeno por célula esperado estava entre 0,5 e 1,0), o número esperado de cópias para cada shRNA que compõe a biblioteca contidas no DNA submetido à primeira PCR estava entre quatro e oito. Embora alguns trabalhos sugiram que o número de cópias de cada inserto de shRNA submetido à PCR seja o mesmo esperado logo após a transdução das células com a biblioteca (Strezoska, 2012), é importante considerar que diversos tipos de shRNAs são eliminados ou consideravelmente reduzidos da amostra devido à aplicação da pressão seletiva ou ao *knockdown* de genes necessários para a manutenção da vida e proliferação das células em cultura. Por outro lado, os diferentes tipos de shRNAs que conferem vantagem seletiva às células em cultura são aumentados em número durante o período de cultivo. Além disso, esses trabalhos consideraram a utilização de apenas um ciclo de PCR. O uso de uma segunda rodada de *nested* PCR melhora o limite inferior de detecção de produtos de amplificação e torna a reação significativamente mais robusta (Haff, 1994).

Outros pontos devem também ser considerados quando experimentos de triagem funcional são realizados. Mesmo com todos os cuidados tomados durante a amplificação dos insertos de shRNAs por multiplex PCR, é importante ter em mente a possibilidade de que nem todos os diferentes tipos de insertos contidos na amostra sejam amplificados com a mesma eficiência, o que pode ocorrer em decorrência da composição de bases de determinados insertos. Por exemplo, a produção de amplicons contendo as duas metades complementares do shRNA pode levar a formação de estruturas secundárias que afetam a taxa de eficiência de amplificação de acordo com a composição de bases da molécula. O uso de primers desenhados para amplificação de apenas a segunda metade do *hairpin*, como os utilizados em nosso trabalho, previne esse problema.

Finalmente, a produção de réplicas biológicas diminui o número de falsos positivos e falsos negativos gerados em decorrência de pequenas variações nas condições experimentais entre as amostras controle e experimental e, portanto, é fortemente indicada. Entretanto, a produção de réplicas biológicas nem sempre é possível devido à baixa disponibilidade de material, ao alto custo financeiro

associado aos experimentos de triagem funcional e ao tempo de trabalho necessário para execução do experimento. Em nosso trabalho, réplicas biológicas não foram geradas devido à ausência de tempo hábil para repetição do experimento dentro do prazo compreendido pelo programa de doutorado.

Todos esses fatores contribuem para reafirmar a importante necessidade de validação dos candidatos obtidos a partir de experimentos de triagem funcional que envolvam bibliotecas formadas por misturas de diferentes shRNAs e multiplex PCR, incluindo os candidatos identificados em no nosso trabalho.

Embora não tenha sido possível avaliar a representatividade da biblioteca de shRNAs nas células infectadas pelas partículas lentivirais logo após a transdução, os alinhamentos das sequências obtidas por NGS contra os bancos de dados de sequências-alvo e sequências de shRNAs propriamente ditas nos permitiram determinar a representatividade da biblioteca no momento em que as células foram coletadas. No total, 78465 sequências diferentes presentes no banco de dados de sequências-alvo foram detectadas após alinhamento pelo menos uma vez em alguma de nossas amostras, sendo que 60299 sequências-alvo diferentes foram detectadas pela menos uma vez na amostra controle 6TG e 74020 sequências-alvo diferentes foram detectadas pelo menos uma vez na amostra experimental HAT. Quando as sequências de amplicons foram alinhadas contra o banco de dados de sequências de shRNAs propriamente ditos, 88112 sequências de shRNAs diferentes foram detectadas pelo menos uma vez em nossas amostras, sendo que 74892 sequências foram detectadas pelo menos uma vez na amostra controle 6TG e 83234 sequências foram detectadas pelo menos uma vez na amostra experimental HAT. O alto número de diferentes tipos de shRNAs presentes em ambas as amostras suporta a teoria formulada a partir dos experimentos de geração de curvas de crescimento celular de que a principal consequência dos tratamentos com as drogas 6TG e HAT, pelo menos a curto prazo, seja a diminuição da taxa de proliferação de células com fenótipo HPRT desfavorável e não depleção completa dessas células na placa de cultura.

Como a complexidade da biblioteca de shRNAs é de cerca de 200 mil tipos de shRNAs, podemos concluir que nem todos os diferentes tipos de shRNAs que constituem a biblioteca lentiviral estavam representados em nossas amostras após os tratamentos nas drogas de seleção. Isso é esperado, uma vez que as células de ambas as amostras foram submetidas à pressão seletiva. Além disso, é esperado

que parte dos diferentes topos de shRNAs levem à inviabilidade, e consequente depleção de células em cultura, como consequência, por exemplo, do *knockdown* de genes envolvidos em processos metabólicos essenciais para a vida e proliferação celular. Vale notar ainda que o número de diferentes tipos de shRNAs presentes na biblioteca relatado pela empresa parece não ser acurado. Observamos que existem sequências ambíguas no banco de dados de sequências de shRNAs propriamente ditos (aproximadamente 5% do número total de sequências), ou seja, uma mesma sequência pode aparecer no banco de dados repetidamente, associada a diferentes identificações (genes). Sendo assim, a representatividade da biblioteca observada ao final de nossa triagem funcional parece ter sido satisfatória.

5.4. Escolha do Método de Análise dos Dados Gerados por NGS

O primeiro passo da análise de dados gerados por NGS consiste no alinhamento (ou mapeamento) de cada uma das curtas sequências geradas pelo sequenciador a um banco de sequências de referência a fim de inferir suas identidades e/ou localizações genômicas correspondentes. Dependendo do volume de sequências geradas por NGS e do tamanho do banco de dados utilizado como referência, esse processo pode se tornar complicado em termos computacionais e vários programas foram desenvolvidos com o intuito de resolver esse problema, incluindo o programa Bowtie, amplamente utilizado por laboratórios de sequenciamento *high-throughput*. Os algoritmos utilizados pelo Bowtie são baseados no método de Burrows-Wheeler e em estruturação de dados associados e permitem a obtenção rápida de alinhamentos exatos ou aproximados. A metodologia empregada pelo programa é desenhada para alinhar sequências geradas por NGS curtas de maneira exata ou com poucos *mismatches* e seu desempenho decai à medida que essas sequências se tornam mais longas e o número de *mismatches* permitidos aumenta (Langmead *et al.*, 2009). Optamos em utilizar esse programa em nossas análises já que nossas sequências eram curtas e nossos bancos de dados relativamente pequenos. Além disso, os programas utilizados para contagem e construção das tabelas dos alinhamentos, baseados nos dados obtidos pelo Bowtie e desenvolvidos pela Profa. Dra. Tatiana Teixeira Torres, estavam prontamente disponíveis.

A escolha do programa DESeq para nossa análise foi baseada em diversos fatores. O programa DESeq é um programa originalmente desenvolvido para detecção de genes diferencialmente expressos entre duas condições experimentais a partir de dados de RNA-seq com tamanho amostral pequeno e com super dispersão obtidos por NGS (Anders & Huber 2010; Dillies *et al.*, 2012). O programa utiliza um modelo de distribuição binomial e regressão local para estimar a relação entre a média e a variância das contagens obtidas para cada sequência/gene nas diferentes amostras e estima a significância do *fold-change* entre elas. O programa também aplica o teste de correção múltiplo de Benjamini-Hochberg aos p-valores calculados (Anders & Huber 2010; Dillies *et al.*, 2012; Strezoska *et al.*, 2012).

Outra característica importante do programa é que a estratégia de normalização empregada se baseia na premissa de que a maioria dos genes/sequências contidos nas amostras não são diferencialmente expressos e de que, entre os genes/sequências diferencialmente expressos, exista uma proporção aproximadamente balanceada de sequências sub e super representadas (Anders & Huber 2010; Dillies *et al.*, 2012). Assumindo que a maioria dos genes não possui expressão diferencial, um fator de correção é gerado para ajustar o tamanho das bibliotecas comparadas. Além disso, o teste estatístico empregado não assume que exista uma distribuição comparável das diferentes contagens de sequências geradas por NGS nas diferentes amostras, sendo que o método de normalização de dados utilizado pelo programa não inclui ajuste de distribuição das contagens entre elas. Isso permite com que as diferentes amostras comparadas exibam diferenças na composição das bibliotecas (Dillies *et al.*, 2012).

A metodologia empregada pelo programa mostrou ser a mais consistente quando diferentes números de réplicas biológicas são analisadas em comparação a outros programas desenvolvidos para a mesma finalidade, como por exemplo, edgeR e NBPSseq (Robles *et al.*, 2012) e permite, inclusive, que se trabalhe sem nenhuma réplica biológica (Anders & Huber 2010), o que o torna adequado ao nosso modelo experimental no qual nenhuma réplica biológica estava presente.

Finalmente, o DESeq é um programa amplamente utilizado (Dillies *et al.*, 2012) e já foi inclusive empregado por diversos grupos na análise de triagens genômicas funcionais baseadas em bibliotecas de shRNAs e NGS realizados com ou sem réplicas biológicas (Strezoska *et al.*, 2012).

5.5. Comparação dos Dados Obtidos a Partir de Alinhamentos Contra Diferentes Bancos de Dados

Os alinhamentos das sequências geradas na corrida final de NGS contra os bancos de dados foram realizados permitindo-se sempre zero *mismatches*. Isso porque, devido ao pequeno tamanho dos insertos de shRNAs (27pb), a ocorrência de apenas um *mismatch* poderia resultar em alinhamentos inespecíficos, o que confundiria a análise e poderia implicar em maior quantidade de candidatos falso-positivos.

Três bancos de dados diferentes foram utilizados nos alinhamentos contra as sequências obtidas na corrida final de NGS para nossas amostras. O primeiro deles é constituído pelas sequências-alvo de shRNAs e refere-se às sequências de sondas presentes no chip de microarray *GeneChip® Human Genome U133 Plus 2.0 Array* (Affymetrix®) que serviu de base para o desenho dos shRNAs que compõem a biblioteca lentiviral pela empresa que comercializa a biblioteca. Sendo assim, essas sequências-alvo não se referem especificamente às sequências dos insertos de shRNAs propriamente ditos. De fato, alguns dos shRNAs presentes na biblioteca não correspondem totalmente às sequências-alvo contidas nesse banco de dados, podendo conter mismatches, intencionalmente introduzidos com o objetivo de melhorar a eficiência de *knockdown* dos genes quando expressos na célula transduzida. Parte das sequências-alvo presentes nesse banco de dados também não possui nenhuma correspondência com nenhuma sequência de inserto de shRNA contida na biblioteca lentiviral. De maneira inversa, alguns insertos de shRNAs não possuem nenhuma sequência-alvo correspondente nesse banco de dados. Vale notar que, na época em que a corrida teste de NGS foi analisada, contávamos apenas com esse banco de dados, que é fornecido juntamente com a biblioteca de shRNAs.

A análise dos alinhamentos das sequências obtidas na corrida final de NGS para nossas amostras contra o banco de dados de sequências-alvo nos permitiu identificar nove sequências mais significativamente representadas na amostra experimental HAT do que na amostra experimental 6TG. Também foram observadas cinco sequências significativamente mais representadas na amostra controle 6TG do que na amostra experimental HAT.

O segundo banco de dados é constituído pelas sequências dos shRNAs propriamente ditos e, portanto, contém as sequências específicas de cada inserto de shRNA desenhado pela empresa System Biosciences e presente na biblioteca. Os alinhamentos das sequências de amplicons obtidas de nossas amostras contra esse banco de dados nos permitiu identificar dez sequências significativamente mais representadas na amostra experimental HAT do que na amostra controle 6TG. Seis sequências estavam mais representadas na amostra controle 6TG do que na amostra experimental HAT. Todas as sequências diferencialmente representadas a partir do alinhamento contra o banco de sequências-alvo estavam também presentes na lista de candidatos gerados a partir do alinhamento contra o banco de sequências de shRNAs propriamente ditas. As duas sequências detectadas apenas no alinhamento contra o banco de sequências de shRNAs propriamente ditas (relacionadas aos genes *H3 histone, family 3B (H3.3B) (H3F3B)* mais representada na amostra HAT e *cyclin-dependent kinase inhibitor 1C (p57, Kip2) (CDKN1C)* mais representada na amostra 6TG) não possuíam correspondência total com nenhuma das sequências contidas no banco de sequências-alvo e, por isso, não foram detectadas no alinhamento contra esse banco de dados.

Embora três a cinco diferentes shRNAs para cada transcrito-alvo existam na biblioteca lentiviral, somente um tipo de shRNA para cada transcrito-alvo mostrou estar diferencialmente representado nas amostras controle 6TG e experimental HAT após os alinhamentos contra os bancos de sequências-alvo e sequências dos shRNAs propriamente ditos. Os demais shRNAs que tinham como alvo os mesmos transcritos candidatos não estavam representados em nossas amostras ou não estavam diferencialmente representados entre elas.

O terceiro banco de dados utilizado nos alinhamentos foi o banco de sequências de RNAs anotadas RefMrna do NCBI. Nesse último alinhamento, shRNAs que tinham como alvos 13 genes estavam significativamente mais representados na amostra experimental HAT do que na amostra controle 6TG e shRNAs que tinham como alvos 10 genes estavam significativamente mais representados na amostra controle 6TG do que na amostra experimental HAT.

Entretanto, a identidade dos genes candidatos encontrados no alinhamento contra o banco de RefMrna diferiu substancialmente da identidade dos genes candidatos encontrados nos alinhamentos contra os dois outros bancos de dados. Dos 13 genes candidatos mais representados na amostra experimental HAT após

alinhamento contra o banco de RefMrna, apenas dois estavam também presentes na lista gerada a partir dos alinhamentos contra os bancos de dados de sequencias-alvo e sequencias de shRNAs propriamente ditos, os genes *androgen-dependent TFPI-regulating protein* (ADTRP) e o gene *vesicle-associated membrane protein 2* (*synaptobrevin 2*) (VAMP2). Dos 10 genes candidatos mais representados na amostra controle 6TG após alinhamento contra o banco de RefMrna, apenas três estavam presentes também na lista gerada a partir dos alinhamentos contra os bancos de dados de sequencias-alvo e sequencias de shRNAs propriamente ditos, os genes *solute carrier family 34 (type II sodium/phosphate cotransporter), member 1* (SLC34A1), o gene *Homo sapiens uncharacterized LOC644873 (FLJ33630) non-coding RNA* e o gene *olfactory receptor, family 6, subfamily W, member 1 pseudogene (OR6W1P) non-coding RNA*.

A diferença nos resultados obtidos entre a análise realizada com os bancos de sequencias-alvo/sequencias de shRNAs propriamente ditos e a análise realizada com o banco de RefMrna poderia ser explicada pela presença dos *mismatches* introduzidos propositadamente durante o desenho dos shRNAs, já que isso poderia ocasionar o alinhamento inespecífico dos mesmos com as sequencias presentes no banco de RefMrna. No entanto, quando os shRNAs candidatos identificados nos primeiros alinhamentos foram alinhados contra o transcriptoma humano, a grande maioria deles foi mapeada contra seus respectivos transcritos com *query coverage* e *identity* de 100% (Tabela 19). Isso demonstra que esses shRNAs candidatos não possuíam mismatches intencionalmente introduzidos em suas sequencias e torna essa explicação não satisfatória.

Sendo assim, a explicação mais plausível para a diferença observada entre os resultados desses alinhamentos refere-se ao mecanismo pelo qual as sequencias de amplicons são alinhadas em cada tipo de análise. Os bancos de sequencias-alvo e sequencias de shRNAs propriamente ditos são constituídos por sequencias curtas de 27bp referentes a cada um dos shRNAs (ou suas sequencias-alvo) que compõem a biblioteca. Sendo assim, nos alinhamentos realizados contra esses bancos de dados, cada um dos shRNAs referentes à um mesmo transcrito-alvo é analisado isoladamente.

Já o banco de dados RefMrna é constituído pelas sequencias completas de cada transcrito. Nos alinhamentos contra esse banco de dados, as sequencias de amplicons são mapeadas ao longo de todo o comprimento de cada transcrito que

compõe o banco de dados e então todas as sequências de amplicons que possuem correspondência com qualquer região de um determinado transcrito são coletivamente contadas. Isso significa que todos os diferentes shRNAs desenhados para um mesmo transcrito são avaliados em conjunto. Esse tipo de análise precisa ser cuidadosamente avaliada. É preciso ter em mente que diferentes shRNAs para um mesmo transcrito podem apresentar diferentes taxas de eficiência de *knockdown*. Alguns deles podem inclusive ser completamente ineficientes. Essa é, aliás, a razão pela qual nesse tipo de análise a significância de um determinado shRNA pode ser diluída pela presença de shRNA que não-significantes. Seguindo esse raciocínio, podemos explicar a ausência da grande maioria dos candidatos identificados nos dois primeiros alinhamentos na terceira análise, realizada com o banco de RefMrna. Por outro lado, diferentes shRNAs para um mesmo transcrito que apresentem individualmente uma significância baixa terão seus efeitos somados durante esse tipo de alinhamento. Isso explica o aparecimento dos novos candidatos identificados após alinhamento contra o banco de RefMrna.

5.6. Genes Candidatos

Coletivamente, os resultados das análises dos dados provenientes do NGS nos permitiram identificar 21 sequências/genes de representação significativamente mais elevada na amostra experimental HAT do que na amostra controle 6TG. Essas sequências/genes constituem candidatos potencialmente envolvidos no processo de manutenção da inativação do cromossomo X em humanos. No entanto, é preciso considerar também que, devido ao nosso desenho experimental, é possível que parte das sequências/genes diferencialmente representadas entre nossas duas amostras estejam relacionadas a mecanismos celulares que não estão envolvidos na manutenção da XCI e sim a mecanismos envolvidos na resistência celular ao tratamento com HAT. Da mesma maneira, as 13 sequências/genes de representação significativamente mais elevadas na amostra controle 6TG constituem candidatos relacionados a possíveis mecanismos de resistência celular à droga 6TG e, portanto, estão fora do escopo desse trabalho e não serão discutidos nesse tese de doutorado.

A existência de um papel na manutenção do processo de inativação do cromossomo X para cada candidato identificado no trabalho descrito nessa tese de doutorado somente poderá ser confirmada após experimentos de validação que serão futuramente realizados em nosso laboratório. Entretanto, o estudo da estrutura e função das proteínas codificadas por cada gene candidato pode fornecer pistas sobre um possível papel dos mesmos na manutenção da XCI ou em mecanismos de resistência ao tratamento com HAT. Esse estudo foi baseado em dados descritos na literatura e/ou relacionados nos bancos de dados GeneCards, NCBI AceView e OMIM.

A Tabela 20 apresenta um resumo das principais informações coletadas para cada gene candidato. Os genes evidenciados em vermelho foram considerados como os mais prováveis de ter um papel relacionado à manutenção da XCI já que conhecidamente desempenham papel em mecanismos de controle epigenético. Genes candidatos de função desconhecida ou pouco conhecida também foram considerados como candidatos em potencial para envolvimento nesse processo e na tabela estão marcados em azul. Os demais genes possuem localização celular e funções conhecidas que os tornam os candidatos menos prováveis para envolvimento no processo de XCI.

Tabela 20. Principais Características e Funções dos Genes Candidatos de Representação mais Elevada na Amostra Experimental Tratada em HAT

Nome do Gene, Número(s) de Acesso (NCBI)	Características e Funções Conhecidas	Referências Seleccionadas
<p><i>Homo sapiens phospholipase A2, group IVC (cytosolic, calcium-independent) (PLA2G4C), mRNA</i></p> <p>NM_001159323.1 NM_001159322.1 NM_003706.2</p>	<p>Enzima ligada ao metabolismo de lipídeos de função independente do cálcio associada à membrana e citossólica. Presente no retículo endoplasmático e mitocôndria. Hidrolisa glicerofosfolipídeos para produção de ácidos graxos livres e lisofosfolipídeos (ambos servem de precursores de moléculas sinalizadoras). Possui várias isoformas. Expressa preferencialmente em cérebro coração e músculo esquelético e em níveis menores em outros tecidos. Pode ser que a função seja modulada por acetilação e fosforilação.</p>	<p>Yamashita <i>et al.</i>, 2009; Xu <i>et al.</i>, 2012; Bickford <i>et al.</i>, 2012; Bickford <i>et al.</i>, 2013.</p>
<p><i>Homo sapiens mannosidase, alpha, class 2A, member 2 (MAN2A2), mRNA</i></p> <p>NM_006122</p>	<p>Envolvida na formação de N-glicanas terminadas em N-acetilglucosamina. Porém sua atividade catalítica é baixa. Hidrolisa dois resíduos de manose de M6Gn2 convertendo-a em M4Gn2, reação que pode ocorrer no início da via alternativa de biosíntese de N-glicanas complexas. Expresso em uma grande variedade de tecidos. Presente no complexo de Golgi, integral da membrana. Camundongo <i>knockout</i> macho é infértil, pois as células germinativas falham em aderir às células de Sertoli e são precocemente liberadas do testículo para o epidídimo. Camundongo duplo-<i>knockout</i> MAN2A1 e MAN2A2 morre no final da gestação ou logo após o nascimento, nesse último caso por falência respiratória. Dentre proteínas com as quais possivelmente interaja encontra-se C7orf55-LUC7L2.</p>	<p>Oh-Eda <i>et al.</i>, 2001; Fukuda & Akama, 2002; Akama <i>et al.</i>, 2006; Hato <i>et al.</i>, 2006; Bi <i>et al.</i>, 2013.</p>
<p><i>Homo sapiens zinc finger, CCHC domain containing 10 (ZCCHC10), mRNA</i></p> <p>NM_017665</p>	<p>Função não conhecida. Liga-se a ácidos nucleicos e íons metálicos. Expresso em diversos tecidos e células tronco-embrionárias. Localização provável no núcleo e mitocôndrias. 32 proteínas diferentes supostamente interagem com ZCCHC10, incluindo C7orf55-LUC7L2 (proteína de possível ligação ao RNA), DPPA2 (pode ser que a proteína se ligue a cromatina e ative a transcrição gênica, pode ter papel na manutenção da pluripotência em ESCs), EEF1A1 (fator de alongamento da tradução), PIN1 (se liga a substratos fosforilados mudando sua conformação), PSME3 (regulador da atividade do proteossomo), R3HDM2-STAC3, THAP1 (proteína pró-apoptótica) e TP53 (supressão tumoral). Possui diversas variantes de <i>splicing</i>. Proteína pode ser precipitada juntamente com spliceossomos.</p>	<p>Rual <i>et al.</i>, 2005; Stelzl <i>et al.</i>, 2005; Hegele <i>et al.</i>, 2012.</p>
<p><i>Homo sapiens chromosome 2 open reading frame 40 (C2orf40), mRNA</i></p> <p>NM_032411</p>	<p>Função pouco conhecida. Gene de supressão tumoral. Hipermetilado em alguns tipos de câncer. Sua super-expressão está relacionada à diminuição da proliferação e viabilidade celular em células de linhagem tumoral. Localiza-se no complexo de Golgi e possivelmente no núcleo e citoplasma. Expresso em diversos tecidos, mas parece ser principalmente expresso em cartilagem, fígado, testículo e cérebro. Suposta atividade como hormônio-peptídeo envolvido em sinalização celular. Pode ser que induza senescência de células neurais. Possível interação com TP53 (supressão tumoral, regulação da transcrição gênica incluindo ncRNAs, apoptose, senescência, induz necrose relacionada a stress oxidativo).</p>	<p>Yue <i>et al.</i>, 2003; Mirabeau <i>et al.</i>, 2007; Mori <i>et al.</i>, 2007; Götze <i>et al.</i>, 2009.</p>

Tabela 20. Principais Características e Funções dos Genes Candidatos de Representação mais Elevada na Amostra Experimental Tratada em HAT (Continuação)

Nome do Gene, Número(s) de Acesso (NCBI)	Características e Funções Conhecidas	Referências Seleccionadas
<p><i>Homo sapiens lactate dehydrogenase A-like 6B (LDHAL6B), mRNA</i></p> <p>NM_033195</p>	<p>Função pouco conhecida. Expresso em diversos tecidos, mas principalmente no testículo. Possui todos os principais domínios presentes em outros membros da família de lactato-desidrogenases (lactato/malato desidrogenase, domínio de ligação a NAD e domínio lactato/malato desidrogenase α/β C-terminal) e domínios transmembrana. Por isso, espera-se que tenha função relacionada com atividade de L-lactato desidrogenase (convertendo reversivelmente piruvato em lactato e, concomitantemente, NADH em NAD⁺, e vice-versa) e, portanto, obtenção de energia para a célula. Pode ser que esteja envolvida em processos como metabolismo de propanoato, cisteína e metionina, glicólise/gluconeogênese, metabolismo de carboidratos, reações de oxi-redução. Possível que a proteína esteja localizada também no citoplasma e membrana mitocondrial. Possível que o domínio de ligação a NAD também se ligue a RNA presente em hnRNP (como sugerido também para outras lactato-desidrogenases) e que tenha, portanto, um papel na regulação da expressão gênica. Possível interação com HDAC6, HNRPD e diversas outras proteínas.</p>	<p>Pioli <i>et al.</i>, 2002; Wang <i>et al.</i>, 2005; Weimann <i>et al.</i>, 2013.</p>
<p><i>Homo sapiens protein serine kinase H1 (PSKH1), mRNA</i></p> <p>NM_006742</p>	<p>Proteína pode estar associada à fosforilação de serinas. Se associa a compartimentos contendo fatores de splicing onde tem papel no tráfego intranuclear de proteínas SR (fatores de splicing não-snRNP contendo domínios ricos em serina e arginina) e processamento de pré-mRNA. Se auto-fosforila. Nenhum outro substrato para fosforilação (além dela mesma) foi ainda identificado. Parece não fosforilar histonas. Sofre miristoilação e palmitoilação e possivelmente glicosilação. Possível que se ligue ao ATP. Função pode depender da concentração de cálcio. Localizado no núcleo, corpos nucleares, centrossomos, citoplasma, membrana plasmática, complexo de Golgi, retículo endoplasmático e possivelmente citoesqueleto. Expresso em diversos tecidos. Possível interação com uma diversa gama de proteínas.</p>	<p>Brede <i>et al.</i>, 2000; Brede <i>et al.</i>, 2002; Brede <i>et al.</i>, 2003.</p>
<p><i>Homo sapiens serine/arginine-rich splicing factor 10 (SRSF10), mRNA</i> Transcript variants 1, 4, 5, 6, 7.*</p> <p>NR_034035.1 NM_001191009.1 NM_001191007.1 NM_001191006.1 NM_006625.4</p>	<p>Proteína SR (fator de splicing não-snRNP contendo domínios ricos em serina e arginina). Participa da regulação de <i>splicing</i> constitutivo e alternativo. Possui motivo RNP N-terminal necessário para ligação a RNAs e múltiplas repetições SR/RS C-terminais que são importantes para associação com outras proteínas. Geralmente, atua como repressor geral de <i>splicing</i>, em sua forma desfosforilada. Necessário para repressão de <i>splicing</i> na mitose e após <i>heat-shock</i> levando ao silenciamento gênico. Na forma fosforilada pode ativar <i>splicing</i> quando na presença de um co-fator não identificado. Pode influenciar a escolha de sítios de <i>splicing</i>. Parece interferir no reconhecimento do sítio de <i>splicing</i> 5' por U1 snRNP do SNRNP70. Possui várias isoformas. Pode ser que sua função seja modulada também por acetilação. Expresso em diversos tecidos. Presente no núcleo, corpos nucleares e citoplasma. Interege com diversas proteínas, incluindo YWHAB, YWHAB.1, YWHAG e YWHAQ que previnem sua desfosforilação. Parece ser induzido a desfosforilação na presença de andrógenos. Camundongos <i>knockouts</i> morrem durante a segunda metade do período gestacional e apresentam múltiplos defeitos cardíacos, heterozigotos são aparentemente normais. Quando super-expresso antagoniza a atividade das proteínas SR SF2/ASF e SC35.</p>	<p>Shin <i>et al.</i>, 2004; Fushimi <i>et al.</i>, 2005; Shin <i>et al.</i>, 2005; Shi & Manley, 2007; Biamonti & Caceres, 2009; Feng <i>et al.</i>, 2009; Xiao <i>et al.</i>, 2011.</p>

Tabela 20. Principais Características e Funções dos Genes Candidatos de Representação mais Elevada na Amostra Experimental Tratada em HAT (Continuação)

<p><i>Homo sapiens zinc finger protein 839 (ZNF839), mRNA.*</i></p> <p>NM_001267828.1 NM_001267827.1 NM_018335.4</p>	<p>Função não conhecida. Ligação a íons metálicos e íons de zinco. Pode estar localizada no citoplasma e núcleo. Expresso em diversos tecidos. É provável que a função da proteína seja modulada por acetilação e fosforilação. Possível interação com as proteínas YWHAE, YWHAZ e YWHAG (modulam a atividade de diversas proteínas incluindo quinases, envolvidas no metabolismo, tráfego de proteínas, transdução de sinal, apoptose e regulação do ciclo celular, reconhecem motivos de serina/treonina fosforilados, se ligam também a fatores de transcrição, podem interagir com HDACs e todos os tipos de histonas, podem formar ancoramento para outras proteínas, aumentam atividade de proteínas com que estimulam proliferação e sobrevivência celular) e WINK1 (serina/treonina quinase importante para manutenção da homeostase e sinalização, proliferação e sobrevivência celular, ativa co-transportadores de cloro associados ao sódio e inibem co-transportadores de cloro associados ao potássio, participa da reorganização do citoesqueleto). Identificado como antígeno em carcinoma renal.</p>	<p>Scanlan <i>et al.</i>, 1999; Healy <i>et al.</i>, 2011.</p>
<p><i>Homo sapiens vesicle-associated membrane protein 2 (synaptobrevin 2) (VAMP2), mRNA. **</i></p> <p>NM_014232</p>	<p>Envolvido em exocitose. Participa da regulação da localização e fusão de vesículas sinápticas através da montagem de complexos multiméricos. VAMP2 interage diretamente com as proteínas Sintaxina-1A e SNAP25 (<i>synaptosome associated protein of 25 kDa</i>) formando um complexo (SNARE) que serve como receptor de fatores de tráfego da família SNAPs (<i>soluble N-ethylmaleimide sensitive factors attachment proteins</i>). A fusão do SNARE à membrana plasmática é modulada pelos níveis de cálcio através do sensor Sinaptotagmina 1 que se liga ao complexo. Participa do mecanismo de absorção de glucose através da formação de um complexo com Sintaxina-4 e SNAP23 presente em vesículas que fazem a translocação de GLUT4 (proteína transportadora de glucose 4) de membranas intracelulares para a membrana plasmática mediante sinalização de insulina. Forma um outro complexo distinto com sinaptofisina envolvido no endereçamento de VAMP2 para compartimentos intracelulares. Pode ser que VAMP2 esteja presente também no complexo de Golgi e que faça ligação com calmodulina. Possível que interaja com outras VAMPs, sintaxinas, HDAC4 e proteínas YWHAB, YWHAE, YWHAG, YWHAH, YWHAQ e YWHAZ, entre outros. Expresso em diversos tecidos, mas possui expressão aumentada no cérebro. Camundongos <i>knockout</i> apresentam alteração da morfologia corporal (acúmulo de gordura no dorso superior), morrem logo após o nascimento e apresentam diminuição de taxa de fusão de vesículas sinápticas.</p>	<p>Hao <i>et al.</i>, 1997; Kawanishi <i>et al.</i>, 2000; Schoch <i>et al.</i>, 2001; Bonanomi <i>et al.</i>, 2007; Prescott <i>et al.</i>, 2009; Lin & Sun, 2011; Bogan, 2012.</p>

Tabela 20. Principais Características e Funções dos Genes Candidatos de Representação mais Elevada na Amostra Experimental Tratada em HAT (Continuação)

<p><i>Homo sapiens</i> Ca^{++}-dependent secretion activator 2 (CADPS2), mRNA. Transcript variants 1, 2, 3*</p> <p>NM_001167940.1 NM_001009571.3 NM_017954.10</p>	<p>Envolvido na regulação da exocitose. Expresso em uma ampla gama de tecidos, no cérebro é maior no cerebelo. Funciona como um sensor de cálcio que regula tráfego e secreção constitutivos de vesículas do tipo LDCV (large dense core vesicles). O domínio C2 media a ligação ao cálcio e a fosfoinositídeos (fosfatidil inositol 4,5-bifosfato). Possui também domínio PH (presentes em proteínas envolvidas em sinalização celular ou associadas ao citoesqueleto), dois domínios <i>coiled-coiled</i> (permite interação com outras proteínas) e um domínio de ligação a dinactina-1. Encontrado em vesículas citoplasmáticas, sinápticas e membrana plasmática. Possível localização em junções celulares. Pode ser que sua atividade seja modulada por fosforilação. Os camundongos <i>knockout</i> apresentam déficit de desenvolvimento neuronal e características de comportamento autista. Também tem papel no comportamento materno e circadiano, interação social e ansiedade. Diversas variantes de transcrito foram identificadas. Pode ser que interaja com VAMP-4 e Sintaxina-6, entre outros.</p>	<p>Cisternas <i>et al.</i>, 2003; Sadakata <i>et al.</i>, 2007; Corpet <i>et al.</i>, 2011; Sadakata <i>et al.</i>, 2012.</p>
<p><i>Homo sapiens</i> membrane-associated ring finger (C3HC4) 2, E3 ubiquitin protein ligase (MARCH2), mRNA Transcript variants 1, 2, 3.*</p> <p>NM_001005416.1 NM_001005415.1 NM_016496.4</p>	<p>Enzima ubiquitina-E3-ligase associada à membrana. Contém um domínio N-terminal RING-CH, um domínio PDZ de ligação a outras proteínas e domínios transmembrana C-terminais. Interage com enzimas ubiquitina-E2-de conjugação (principalmente ubcH6 e ubcH9) através de um domínio PHD. Presente na membrana de endossomos, complexo de Golgi e membrana plasmática. Regula tráfego entre os endossomos e o Complexo de Golgi. Associa-se diretamente a Sintaxina-6. Ancorado pela DLG1 (<i>Disc Large Gene 1</i>, supressão tumoral, confere polaridade celular) em locais de contato entre células. Pode ubiquitinar DLG1 diminuindo a expressão de DLG1 em locais de contato entre células. Super-expressão inibe o transporte de transferrina absorvida para o endossomo. Interage com Sintaxina-6 e CAL (<i>cystic fibrosis transmembrane conductance regulator (CFTR)-associated ligand</i>) e promove ubiquitinação de CFTR (fator de transcrição, também delimita região da cromatina protegendo regiões teloméricas de silenciamento). Associa-se também à MARCH3. Expressão em uma grande gama de tecidos. Possível localização também em retículo endoplasmático e lisossomos.</p>	<p>Bartee <i>et al.</i>, 2004; Nakamura <i>et al.</i>, 2005; Fukuda <i>et al.</i>, 2006; Cao <i>et al.</i>, 2008; Cheng & Guggino, 2013.</p>
<p><i>Homo sapiens</i> clavesin 2 (CLVS2), mRNA NM_001010852</p>	<p>Presente em endossomos tardios e lisossomos de neurônios e vesículas cobertas por clatrin. Pode ser que esteja presente também em endossomos precoces, no complexo de Golgi e citoplasma. Se liga a lipídeos (fosfatidilinositol-3,5-bifosfato). Expressão em diversos tecidos mas preferencialmente cerebral. Pode ser que sua atividade seja modulada por fosforilação.</p>	<p>Katoh <i>et al.</i>, 2009.</p>

Tabela 20. Principais Características e Funções dos Genes Candidatos de Representação mais Elevada na Amostra Experimental Tratada em HAT (Continuação)

<p><i>Homo sapiens pappalysin 2 (PAPPA2), mRNA. Transcript variant 1.*</i></p> <p>NM_020318.2</p>	<p>Membro da família de metaloproteases metzincinas. Regula localmente a biodisponibilidade de IGFs (IGF1 e IGF2, mitógenos). Cliva a proteína IGFBP5 (<i>insulin-like growth factor-binding protein 5</i>) associada a IGF ou livre aumentando a biodisponibilidade de IGF. Apresenta atividade proteolítica mais baixa contra IGFBP3. O <i>splicing</i> alternativo resulta em múltiplas variantes. Expressão em diversos tecidos, mas principalmente na placenta. Hipóxia, PGE2 e TNF-alfa causam aumento da expressão de PAPPA2. Também expresso no feto sendo importante para seu crescimento e desenvolvimento e para crescimento pós-natal normal. Localização provável na membrana plasmática, intracelular e extracelular. Possui domínios sushi (presentes em proteínas complemento ou de adesão), domínio Notch (presente em proteínas transmembrana), fibronectina 3 (ligação a receptores celulares ou DNA) e similar à concanavalina-A (ligação a carboidratos complexos).</p>	<p>Overgaard <i>et al.</i>, 2001; Wang <i>et al.</i>, 2009; Yan <i>et al.</i>, 2010; Wagner <i>et al.</i>, 2011.</p>
<p><i>Homo sapiens anti-silencing function 1A histone chaperone (ASF1A), mRNA</i></p> <p>NM_014034</p>	<p>Apesar do nome, ASF1A está principalmente envolvida na montagem de nucleossomos e silenciamento da expressão gênica. Interage diretamente com a região em hélice C-terminal de H3 envolvida no dobramento da mesma e com a cauda C-terminal de H4 envolvida na ligação com H2A. Participa do primeiro passo da formação de nucleossomos incorporando H3/H4 ao DNA. Asf1a interage com CAF1 e HIRA (<i>histone repression A fator</i>, não-replicativa). Asf1a é a principal forma relacionada ao silenciamento gênico e em complexo com HIRA foi implicado também na formação de heterocromatina facultativa. Asf1a interage com HIRA também para deposição de H3.3 mediada por transcrição e interage com HIRA e H3 para a formação de SAHF (<i>senescence-associated heterochromatic foci</i>, ricos em MacroH2A, associados ao silenciamento de genes envolvidos na proliferação celular). Asf1a é expressa durante todas as fases do ciclo celular e é necessária para a deposição de H3.3 em células quiescentes. Se associam indiretamente com helicase MCM2-7 na forquilha de replicação. Asf1 interage com o domínio de histona acetiltransferase de TFIID (transcription factor IID) e com quinases envolvidas em <i>checkpoints</i> do ciclo celular, incluindo Rad53 e <i>Tousled-like</i> quinases. Mutantes são incapazes de promover silenciamento gênico e de fornecer resistência a danos no DNA. Também envolvido na desmontagem de nucleossomos (de-repressão da cromatina) e ativação da transcrição gênica. É necessária para que ocorra acetilação de H3K56 (PTM necessária para conferir estabilidade genômica e resistência ao <i>stress</i> genotóxico). Expresso em diversos tecidos. Atividade protéica modulada por fosforilação e possivelmente ubiquitinação.</p>	<p>Munakata <i>et al.</i>, 2000; Chimura <i>et al.</i>, 2002; Mello <i>et al.</i>, 2002; Mousson <i>et al.</i>, 2005; Zhang <i>et al.</i>, 2005; Ozdemir <i>et al.</i>, 2006; Agez <i>et al.</i>, 2007; Gazin <i>et al.</i>, 2007; Groth <i>et al.</i>, 2007; Hayashi <i>et al.</i>, 2007; Natsume <i>et al.</i>, 2007; Galvani <i>et al.</i>, 2008; Banumathy <i>et al.</i>, 2009; Pilyugin <i>et al.</i>, 2009; Akai <i>et al.</i>, 2010; Ishikawa <i>et al.</i>, 2011.</p>

Tabela 20. Principais Características e Funções dos Genes Candidatos de Representação mais Elevada na Amostra Experimental Tratada em HAT (Continuação)

<p><i>Homo sapiens H3 histone, family 3B (H3.3B) (H3F3B), mRNA *</i></p> <p>NM_005324.3</p>	<p>Histonas H3.3 são a forma predominante de H3 presente em células quiescentes e células em G1 e G2, embora transcritos sejam produzidos durante todas as fases do ciclo celular. Substituem a histona canônica H3.1 em uma grande gama de nucleossomos associados a genes ou regiões intergênicas. H3.3 é associada preferencialmente a genes com alto conteúdo GC na região promotora. Enriquecida em PTMs associadas à ativação gênica. São incorporadas de maneira independente da síntese de DNA embora possam ser incorporadas também durante a replicação. Deposição independente da replicação pode ser mediada por HIRA em complexo com ASF1a ou por HIRA em complexo com o fator remodelador de cromatina CDH1. Deposição de H3.3 dependente de HIRA geralmente está associada com ativação da transcrição através de incorporação após desmontagem de nucleossomos causada pelo alongamento da transcrição, mas também podem estar envolvidas em silenciamento gênico. Em genes ativos H3.3 é encontrada tanto ao redor de sítios de ligação a fatores de transcrição como ao longo do gene e em genes reprimidos H3.3 é encontrada preferencialmente ao redor de sítios de ligação a fatores de transcrição. Deposição ao longo do gene é dependente de HIRA e está diretamente relacionada ao nível de expressão. Deposição ao redor de alguns sítios de ligação a fatores de transcrição é dependente de HIRA e ao redor de outros sítios de de ligação a fatores de transcrição independe dessa chaperone. A histona H3.3 está relacionada a regiões pericentroméricas e teloméricas (deposição nessas regiões não depende de HIRA). A histona chaperone DAXX (<i>death domain associated protein</i>) e a helicase/fator remodelador de cromatina ATRX (<i>alpha-thalassemia/ mental retardation syndrome protein</i>) depositam H3.3 em regiões teloméricas e pericentroméricas e em corpúsculos PLM (<i>promyelocytic leukemia protein nuclear bodies</i>), ricos em heterocromatina. DAXX também pode interagir com DEK (<i>chromatin-bound oncogene product</i>), outra histona chaperone envolvida na deposição de H3.3 em elementos regulatórios de genes ativos. H3.3 tem papel importante na diferenciação e sua transcrição é mantida durante esse período (ao contrário de H3.1 e H3.2). De fato, é a forma de H3 mais abundante em tipo celulares de longa-vida em camundongos. Embora existam diferenças entre os genes H3F3A e H3F3B, inclusive na região promotora, não há diferenças na sequência de aminoácidos das proteínas codificadas por essas duas isoformas. A inativação de cromossomos sexuais X e Y durante a meiose masculina envolve incorporação massiva de H3.3 e subsequente silenciamento gênico.</p>	<p>Witt <i>et al.</i>, 1997; Frank <i>et al.</i>, 2003; Tagami <i>et al.</i>, 2004; Groth <i>et al.</i>, 2005; Daury <i>et al.</i>, 2006; Loyola & Almouzni, 2007; Delbarre <i>et al.</i>, 2010; Drané <i>et al.</i>, 2010; Elsaesser <i>et al.</i>, 2010; Goldberg <i>et al.</i>, 2010; Wong <i>et al.</i>, 2010; Szenker <i>et al.</i>, 2011; Harada <i>et al.</i>, 2012; Behjati <i>et al.</i>, 2013.</p>
---	--	--

Tabela 20. Principais Características e Funções dos Genes Candidatos de Representação mais Elevada na Amostra Experimental Tratada em HAT (Continuação)

<p><i>Homo sapiens androgen-dependent TFPI-regulating protein (ADTRP), mRNA Transcript variants 1,2. **</i></p> <p>NM_001143948 NM_032744</p>	<p>Em células endoteliais, regula a expressão (provavelmente indiretamente já que parece não estar no núcleo), localização e a atividade anticoagulante da isoforma presente nas cavéolas e <i>lipids-rafts</i> da proteína <i>Tissue Factor Pathway Inhibitor</i> (TFPI, anticoagulante natural que neutraliza FXa e inibe produção de FXa dependente de TF-FVIIa. A disfunção de células endoteliais relacionadas a problemas vasculares está relacionada a baixos níveis de TFPI). Expressão induzida por andrógenos. Função provavelmente regulada por palmitoilação. Localização na superfície celular e no citoplasma. Possível localização também no retículo endoplasmático. Em células endoteliais está colocalizada com TFPI e caveolina-1. Expresso em diversos tecidos. Outras supostas interações protéicas incluem o fator de transcrição CREB-3 (presente também no retículo endoplasmático) e vimentina.</p>	<p>Lupu <i>et al.</i>, 2011.</p>
<p><i>Homo sapiens chromosome 7 open reading frame 33 (C7orf33), mRNA</i></p> <p>NM_145304</p>	<p>Função desconhecida. Possui domínios que inferem localização nuclear. Parece ser expresso em diversos tecidos. Pode ser que tenha alguma relação com Diabetes tipo 1.</p>	<p>Lewis & Schlitt; 2011.</p>
<p><i>Homo sapiens hydroxycarboxylic acid receptor 1 (HCAR1), mRNA</i></p> <p>NM_032554*</p>	<p>Receptor acoplado à proteína G. Possui sete domínios transmembrana e transduz sinais extracelulares através de trimeros de proteína G. Expressão predominante em tecido adiposo e baixa em diversos tecidos. Se liga ao lactato provocando diminuindo a resposta lipolítica em células adiposas. Camundongos <i>knockout</i> são viáveis e férteis e não possuem fenótipo evidente. Proteína está localizada na membrana plasmática.</p>	<p>Lee <i>et al.</i>, 2001; Cai <i>et al.</i>, 2008; Liu <i>et al.</i>, 2009; Kuei <i>et al.</i>, 2011; Offermanns <i>et al.</i>, 2011.</p>
<p><i>Homo sapiens Helicobacter pylori responsive 1 (non-protein coding) (HPYR1), long non-coding RNA</i></p> <p>NR_026684</p>	<p>Função não conhecida. Expressão aumentada frente à infecção por <i>Helicobacter pylori</i>. É um lincRNA.</p>	<p>Cox <i>et al.</i>, 2001.</p>

Tabela 20. Principais Características e Funções dos Genes Candidatos de Representação mais Elevada na Amostra Experimental Tratada em HAT (Continuação)

<i>Homo sapiens cysteine-rich hydrophobic domain 1 (CHIC1), mRNA</i> NM_001039840	Função não conhecida. Localizado no XIC, próximo ao XIST. Expressão de Chic1 se inicia cedo na embriogênese. Em camundongos, é preferencialmente expresso no cérebro e submetido à XCI. Expresso também em outros tecidos e células-tronco embrionárias. Pode estar localizado na membrana plasmática e vesículas citoplasmáticas, de acordo com os domínios preditos (Golgin subfamily A member 7/ERF4 family, domínios transmembrana e coiled-coiled). Sofre ubiquitinação e pode ser que sofra fosforilação e palmitoilação.	Simmler <i>et al.</i> , 1997; Cools <i>et al.</i> , 2001; Povlsen <i>et al.</i> , 2012.
--	---	---

* Genes candidatos identificados nos alinhamentos contra os bancos de dados de sequências-alvo e/ou de shRNAs propriamente ditos. ** Genes candidatos identificados nos alinhamentos contra os três bancos de dados utilizados. Os demais genes candidatos foram identificados somente no alinhamento contra o banco de dados RefMrna. Em vermelho: candidatos mais prováveis para envolvimento no processo de manutenção da XCI por função conhecida relacionada a mecanismos de controle epigenético. Em azul: candidatos de função pouco conhecida. Os demais candidatos possuem menor probabilidade de estarem envolvidos na manutenção da XCI.

5.6.1. H3.3B e ASF1A

Interessantemente, diversas correlações podem ser observadas entre os diferentes candidatos identificados em nossa triagem funcional. A mais clara delas refere-se às proteínas ASF1A (*anti-silencing function 1A histone chaperone*) e H3.3B (codificada pelo gene *H3F3B*, *H3 histone family 3B*). As histonas H3.3 são variantes das formas canônicas, replicativas, H3.1 e H3.2 e são depositadas nos nucleossomos de maneira independente da replicação celular (Khare *et al.*, 2012; revisto por Law & Cheung, 2012). As formas mais conhecidas de H3.3 são as formas H3.3A e H3.3B, codificadas por genes distintos que diferem na composição de nucleotídeos inclusive na região promotora. Apesar disso, os produtos finais da expressão de ambos os genes são proteínas histonas que apresentam sequências idênticas de aminoácidos. Por esse motivo, H3.3A e H3.3B são geralmente referidas coletivamente como H3.3. No entanto, foi observado que alguns tipos de tumor estão associados preferencialmente a mutações no gene H3F3A enquanto outros tipos estão associados mais frequentemente a mutações em H3F3B. Isso sugere que esses dois genes possam apresentar algumas funções distintas. Outra explicação é de que essas duas formas apresentem perfis de expressão distintos em diferentes tipos celulares e/ou diferentes momentos da vida do organismo. Mais estudos são necessários para resolver essa questão (Behjati *et al.*, 2013).

H3.3 pode ser encontrada de maneira enriquecida tanto em regiões gênicas como intergênicas. Embora a deposição de histonas H3.3 tenha sido primeiramente relacionada com eucromatina e ativação da transcrição gênica, estudos recentes demonstram que essa variante tem também um papel no silenciamento gênico. De fato, estudos compreensivos que visaram mapear as regiões genômicas enriquecidas em H3.3, demonstraram que essa variante de histona está associada também a sítios de início da transcrição de genes reprimidos (Elssaesser *et al.*, 2010; Goldberg *et al.*, 2010). Mais ainda, esses estudos mostraram que o padrão de deposição de H3.3 em genes ativos difere do padrão encontrado em genes silenciados. Embora a deposição em torno de sítios de início de transcrição esteja relacionada tanto a genes ativos como a genes reprimidos, a deposição de H3.3 ao longo de regiões codificadoras e de regiões localizadas além do sítio de término da transcrição está associada a genes transcricionalmente ativos. A histona H3.3 é encontrada também em promotores silenciados bivalentes de células pluripotentes,

caracterizados pelas PTMs H3K4me3 (geralmente associada a ativação da transcrição gênica) e H3K27me3 (geralmente associada a repressão da transcrição gênica). Além disso, está presente em regiões de heterocromatina facultativa pericentroméricas e teloméricas (Goldberg *et al.*, 2010; Szenker *et al.*, 2011).

ASF1 é uma histona chaperone envolvida na deposição de dímeros de H3/H4 ao DNA de maneira tanto dependente como independente da replicação (English *et al.*, 2006; Agez *et al.*, 2007). Interage diretamente com H3 e H4 e está envolvida preferencialmente na montagem de nucleossomos e repressão da atividade gênica, embora possa também estar envolvida na demontagem de nucleossomos e ativação da transcrição. Duas isoformas são encontradas em mamíferos, ASF1A e ASF1B. Um trabalho voltado para o estudo da deposição de histonas H3.1 e H3.3 demonstrou que a forma H3.1 é preferencialmente encontrada em associação com ASF1B durante a divisão celular, enquanto que a forma H3.3 é preferencialmente encontrada em associação com ASF1A em células quiescentes (Tagami *et al.*, 2004). De fato, ASF1B é expressa preferencialmente durante a divisão celular. ASF1B interage diretamente com CAF1, uma histona chaperone multimérica formada pelas subunidades polipeptídicas RabAp48, p60 e p150 também associada a deposição de histonas H3.1 ao DNA durante a divisão celular (Tagami *et al.*, 2004; Groth *et al.*, 2005).

Embora ASF1A seja expressa durante todas as fases do ciclo celular e também possa se associar a CAF1, é a isoforma necessária para deposição de H3.3 em células quiescentes. Outro parceiro que interage diretamente com ASF1 é a histona chaperone HIRA, extensamente descrita na literatura como tendo função na deposição da variante de H3.3 em células que não estão em divisão (Tagami *et al.*, 2004). A deposição de H3.3 ao longo de região codificadoras de genes parece ser dependente de HIRA e os níveis de H3.3 observados nessas regiões são diretamente proporcionais a atividade gênica. Entretanto, a deposição em sítios de início da transcrição, tanto em genes ativos como silenciado, em algumas regiões é dependente e, em outras, independe dessa chaperone (Elsaesser *et al.*, 2010; Goldberg *et al.*, 2010).

Embora a interação entre ASF1A e HIRA possa ser observada durante a deposição de H3.3 mediada por transcrição (processo geralmente associado a genes ativos), ASF1a é a principal isoforma relacionada ao silenciamento gênico e em complexo com HIRA foi implicada também na formação de heterocromatina

facultativa. Além disso, ASF1A interage com HIRA e H3 para a formação de SAHF (*senescence-associated heterochromatic foci*), regiões presentes em células senescentes ricas em MacroH2A e associadas ao silenciamento de genes envolvidos na proliferação celular (Zhang *et al.*, 2005; Banumathy *et al.*, 2009).

Finalmente, em *Schizosaccharomyces pombe*, Asf1 e HIRA se associam a Swi6/HP1 formando um complexo que se espalha em regiões silenciadas da cromatina reforçando o silenciamento transcricional. Asf1 se associa e atua em sincronia com o complexo HDAC/Clr6 para promover intensa desacetilação de histonas, silenciar elementos repetitivos na heterocromatina e suprimir inclusive a expressão de transcritos intragênicos antisense, tanto na hetero como na eucromatina. Além disso, Asf1 e SHREC agem em vias sobrepostas facilitando a deposição de nucleossomos na heterocromatina (Yamane *et al.*, 2011).

A deposição de H3.3 em regiões teloméricas e pericentroméricas parece ser independente de HIRA e mediada pela histona chaperone DAXX e pelo fator remodelador de cromatina ATRX (Drané *et al.*, 2010; Goldberg *et al.*, 2010; Lewis *et al.*, 2010; Wong *et al.*, 2010). Foi demonstrado que DAXX é capaz de interagir diretamente com ATRX (Szenker *et al.*, 2011). Embora a presença de H3.3 nessas regiões tenha sido relacionada com repressão da expressão de elementos teloméricos repetitivos, o papel dessa variante nos telômeros e regiões pericentroméricas ainda precisa ser elucidado (Goldberg *et al.*, 2010).

Embora nenhuma associação específica entre H3.3 ou ASF1 e o processo de XCI em fêmeas ou formação de heterocromatina constitutiva tenha sido encontrada na literatura, é sabido que o fator remodelador de cromatina ATRX exerce um papel importante na XCI. Em células murinas da camada granulosa de ovário, fibroblastos embrionários e células-tronco trofoblásticas, a proteína Atrx permanece estavelmente associada ao Xi durante a intérfase e metáfase onde exibe um padrão de bandamento ao longo de todo Xi. Além disso, foi demonstrado que Atrx está preferencialmente enriquecido em regiões de H3K9me, similar a marcadores de Xi H3K9me2 and H3K27me3. Além disso, Atrx está co-localizado com macroH2A (um outro marcador de Xi) no cromossomo X de replicação tardia. De fato, foi demonstrado que Atrx é um fator crítico para promover formação de heterocromatina facultativa no Xi de camundongos. Além disso, trofoblastos murinos deficientes em Atrx exibem um padrão anormal de *imprinting* relacionado a XCI não-aleatória (Baumann & De La Fuente 2009; De La Fuente *et al.*, 2011).

O papel específico de ATRX na XCI ainda não é claro, mas evidências indicam que ATRX deva estar envolvida na manutenção do Xi ao invés de estabelecimento do silenciamento em fases iniciais da XCI. Isso porque, o enriquecimento significativo nos níveis desse fator remodelador de cromatina, assim como o aumento da distribuição do mesmo ao longo do Xi, é observado apenas após o oitavo dia de diferenciação espontânea de mESCs em cultura (Baumann & De La Fuente, 2009; De La Fuente *et al.*, 2011).

Poderia se especular que a detecção de ASF1 e H3.3 em nossa triagem funcional tenha sido decorrente de um mecanismo que permitiu o escape de células que continham shRNAs para um desses dois genes ao processo de senescência celular. Entretanto, nesse caso não se esperaria que estes shRNAs estivessem significativamente mais representados em nossa amostra HAT do que em nossa amostra 6TG, uma vez que seria esperado que células tratadas com qualquer uma dessas drogas se beneficiariam igualmente do escape à senescência. Sendo assim, a investigação do papel dessas duas proteínas na manutenção da XCI merece e deve ser investigada, tanto de maneira isolada quanto conjuntamente.

5.6.2. Candidatos Envolvidos em Mecanismos de *Splicing*

Outro tipo de correlação encontrada entre os candidatos identificados em nossa triagem funcional refere-se ao envolvimento em mecanismos de *splicing*. Três proteínas identificadas em nosso trabalho parecem estar envolvidas nesse tipo de mecanismo, as proteínas codificadas pelos genes SRSF10, PSKH1 e ZCCH10.

A proteína SRSF10 é uma proteína do tipo SR (fator de *splicing* não-snRNP que contém domínios ricos em serina e arginina) que participa da regulação de *splicing* constitutivo e alternativo (Shin *et al.*, 2004; Shi *et al.*, 2005; Feng *et al.*, 2009). Sua atividade é regulada principalmente por fosforilação. Embora geralmente atue em sua forma desfosforilada como repressor geral de *splicing* (Shin & Manley, 2002; Shin *et al.*, 2004), foi demonstrado que SRSF10 também pode induzir o processamento de RNAs na presença de um co-fator ainda não identificado (Feng *et al.*, 2008). Ainda não se conhece a quinase capaz de fosforilar SRSF10, e a proteína parece ser induzida a desfosforilação na presença de andrógenos (testosterona) (Xiao *et al.*, 2011).

A presença de SRSF10 pode influenciar a escolha de sítios de *splicing*. De fato, foi demonstrado que essa proteína interfere no reconhecimento do sítio de *splicing* 5' por U1 snRNP presente em SNRNP70 (Shin *et al.*, 2004).

Acredita-se que a proteína PSKH1 possa ter papel relacionado à fosforilação de serinas. Entretanto, além de ser capaz de promover autofosforilação (Brede *et al.*, 2000), nenhum outro substrato foi ainda identificado para essa proteína. Além de estar presente em outros compartimentos celulares, onde pode estar associada a funções distintas, PSKH1 é encontrada em compartimentos contendo fatores de *splicing* onde exerce papel no tráfego intranuclear de proteínas SR e processamento de pré-mRNA (Brede *et al.*, 2002).

ZCCHC10 foi frequentemente detectada em associação com spliceossomos (Rual *et al.*, 2005; Stelzl *et al.*, 2005; Hegele *et al.*, 2012) e é possível que LDHAL6B interaja com a HNRPD (*heterogeneous nuclear ribonucleoprotein D*), outra proteína envolvida no processamento de pré-mRNAs de transcritos relacionados à proto-oncogenes e citocinas.

Genes envolvidos no processamento de moléculas de RNA encontrados em um estudo de triagem funcional de genes envolvidos na manutenção da XCI em camundongos (Chan *et al.*, 2011) mostraram ter algum papel na manutenção da XCI. Sendo assim, o possível envolvimento das proteínas envolvidas em mecanismos de *splicing* identificados em nossa triagem funcional precisa ser experimentalmente investigado.

5.6.3. Outras Possíveis Correlações entre Genes Candidatos

Diferentes candidatos identificados em nosso trabalho possuem relação com a formação ou funcionamento de vesículas intracelulares associadas ao processo de endo ou exocitose. Esse é o caso das proteínas CAPSD2, VAMP2, MARCH2 e CLVS2 (Cisternas *et al.*, 2003; Sadakata *et al.*, 2007; . Além disso, tanto VAMP2 como MARCH2 são capazes de interagir com sintaxinas. VAMP2 interage com Sintaxina-1 e MARCH2 com Sintaxina-6 (Hao *et al.*, 1997; Cao *et al.*, 2008; Cheng & Guggino, 2013). Além disso, é possível que CAPSD2 também interaja com Sintaxina-6 e com um outro tipo de VAMP, a proteína VAMP4 (Genecards, www.genecards.org). O significado dessas correlações são aspectos que ainda

precisam ser elucidados. Entretanto, devido à localização celular e às funções biológicas descritas para essas proteínas, é improvável que esses candidatos desempenhem algum papel importante na manutenção da XCI. Sendo assim, a predominância de shRNAs relacionados a esses genes em nossa amostra experimental pode estar relacionada a mecanismos que conferem algum tipo de resistência celular ao tratamento com HAT.

Outras correlações observadas entre os genes candidatos identificados em nosso trabalho dizem respeito a possíveis interações de diferentes proteínas candidatas com proteínas em comum. Entre as proteínas preditas para interagir com ZCCHC10 e C2orf40 inclui-se a proteína TP53 (*Tumor Protein p53*), uma proteína de supressão tumoral que age em resposta a diversas situações relacionadas ao estresse celular regulando a expressão de genes que induzem retenção no ciclo celular, apoptose, senescência, reparo de DNA ou alterações no metabolismo (Genecards, www.genecards.org).

ZCCHC10 e MAN2A2 possivelmente interagem com Corf55-LUC2L, uma proteína de função desconhecida que possivelmente se liga a moléculas de RNA. É possível também que LDHAL6B e VAMP2 interajam com HDACs (LDHAL6B com HDAC6 e VAMP2 com HDAC4) (Genecards, www.genecards.org).

Finalmente, SRSF10 demonstrou interagir, em sua forma fosforilada, com proteínas da família YWHA que previnem a desfosforilação de SRSF10 (Shi & Manley, 2007). As proteínas YWHA possuem funções variadas. Reconhecem motivos de serina/treonina fosforilados e modulam a atividade de diversas outras proteínas incluindo quinases, proteínas envolvidas em tráfego de proteínas, transdução de sinais, apoptose e regulação do ciclo celular. Além disso, podem se ligar a fatores de transcrição e interagir com HDACs e com todos os tipos de histonas. Formam regiões de ancoramento para outras proteínas e aumentam a atividade de proteínas que estimulam a proliferação e sobrevivência celular. É possível que VAMP2 e ZNF839 também interajam com membros da família YWHA (Genecards, www.genecards.org).

Entretanto, a existência da maioria dessas interações ainda precisa ser confirmada. Sendo assim, o significado dessas correlações, assim como o significado da identificação desses genes candidatos em nossa triagem funcional, são aspectos que ainda precisam ser experimentalmente elucidados.

5.7. Genes Envolvidos na Manutenção da XCI

Conforme mencionado na introdução desse trabalho, *Xist*, *Dnmt1*, *SmcHD1* e *Mbd2* são fatores relacionados ao processo de manutenção da XCI. No entanto, nenhum desses genes foi identificado como candidato para envolvimento no processo de manutenção da XCI em nossa triagem funcional.

Vários fatores podem explicar a ausência desses genes em nossa lista de candidatos. Primeiramente vale lembrar que, embora os genes *Xist*, *Dnmt1*, *SmcHD1* e *Mbd2* exerçam um papel importante no processo de manutenção da XCI em camundongos (Sado *et al.*, 2000; Barr *et al.*, 2007; Blewitt *et al.*, 2008; Gendrel *et al.*, 2012), pouco se conhece sobre o papel de SMCHD1 e MBD2 na manutenção da XCI em humanos e o grau de importância de XIST e DNMT1 para a manutenção da XCI em humanos parece ser relativamente menor (Brown & Willard, 1999; Rhee *et al.*, 2002; Noozawa *et al.*, 2013; Vasquez *et al.*, 2005).

Além disso, a ausência de DNMT1, SMCHD1 e MBD2 em nossa lista de candidatos, pode ser explicada com base nas importantes funções desempenhadas por essas proteínas. DNMT1 possui um importante papel na manutenção da metilação do DNA não apenas associada ao cromossomo Xi, mas em todo o genoma da célula (Sado *et al.*, 2000; Rhee *et al.*, 2002; Vasquez *et al.*, 2005). SMCHD1 também está relacionada à metilação do DNA (Blewitt *et al.*, 2008; Gendrel *et al.*, 2012) e MBD2 se liga ao DNA metilado mediando o efeito da metilação em células de mamíferos (Hendrich *et al.*, 1998). Além disso, embora camundongos *knockouts* para MBD2 sejam viáveis e férteis (Sansom *et al.*, 2003), camundongos *knockouts* para *Dnmt1* ou *SmcHD1* morrem antes do nascimento (Blewitt *et al.*, 2008). Sendo assim, é possível que o *knockdown* desses genes cause um grande impacto no funcionamento das células de forma a conferir uma desvantagem seletiva dessas células em relação as demais células transduzidas com a biblioteca de shRNAs presentes na mesma placa de cultura, tanto para nossa amostra 6TG como para nossa amostra HAT, o que poderia explicar a ausência desses genes em nossa lista de candidatos.

Em relação ao XIST, é preciso levar em conta que ele é um lincRNA de localização nuclear, fortemente associado ao DNA e outras proteínas, e o próprio método de triagem utilizado em nosso trabalho. Embora em alguns estudos tenha

sido demonstrado que shRNAs e siRNAs são capazes de promover o *knockdown* de alguns tipos de ncRNAs (Ramos *et al.*, 2013), é preciso considerar que a eficiência de *knockdown* de ncRNAs mediado por shRNAs e siRNAs pode ser fortemente impactada por diferentes fatores. O primeiro deles está relacionado com a localização dos ncRNAs dentro da célula. Embora tenha sido demonstrado que pelo menos parte da maquinaria de RNAi também possa ser encontrada no núcleo, ela é predominantemente citoplasmática, o que pode dificultar o *knockdown* de ncRNAs nucleares. O segundo deles refere-se à disponibilidade do sítio de reconhecimento ao RNAi presente no ncRNA. Como os ncRNAs possuem fita simples, a formação de estruturas secundárias pode ocultar a região de reconhecimento ao RNAi, impedindo o silenciamento da expressão dos mesmos. Além disso, os ncRNAs estão frequentemente associados ao DNA e/ou a proteínas, o que também pode restringir seu reconhecimento pelos RNAi. É preciso considerar também que alguns ncRNAs possuem altos níveis de expressão e, nesses casos, o *knockdown* pode não ser suficientemente efetivo em revelar sua função biológica. Finalmente, é preciso considerar que a utilização de apenas um tipo de siRNA ou shRNA por célula pode não fornecer um nível de degradação que seja suficiente para inibir totalmente o funcionamento do ncRNA-alvo. Quando níveis de degradação insuficientes são obtidos, é possível que os domínios funcionais do ncRNAs permaneçam intactos e continuem a desempenhar alguma função (Ploner *et al.*, 2009; Gutschner *et al.* 2011). Todos esses fatores podem explicar a ausência de gene *XIST* em nossa lista de candidatos.

Chan e colaboradores (2011), utilizaram uma biblioteca de shRNA e *microarray* para triar genes envolvidos na manutenção da XCI em fibroblastos embrionários de camundongos. Noventa e seis genes candidatos foram identificados nesse estudo, dos quais 46 estão relacionados no artigo científico publicado e 32 foram validados com sucesso. *Xist*, *Dnmt1*, *ShcMD1* e *Mbd2* não estão presentes na lista contendo 46 candidatos.

Nenhum dos genes candidatos relacionados na publicação de Chan e colaboradores (2001) foi identificado em nosso trabalho. No entanto, diversos fatores podem explicar as diferenças observadas. Primeiramente, o estudo de Chan e colaboradores (2011) foi realizado em fibroblastos embrionários de camundongos imortalizados por antígeno T de SV40 que continham um transgene GFP no Xi, enquanto que nossa triagem funcional foi realizada em fibroblastos humanos

primários e adultos. Não se sabe qual o efeito do uso de células imortalizadas nos resultados obtidos. Enquanto Chan e colaboradores (2011) utilizaram a expressão direta de um transgene contido no cromossomo X como estratégia de detecção da reativação do Xi, a nossa estratégia baseou-se na reversão do fenótipo XPRT⁻. Por outro lado, o marcador de reativação do Xi usado por Chan e colaboradores (2011) é um gene exógeno, enquanto o marcador utilizado em nosso trabalho é um gene endógeno do cromossomo X.

Além disso, conforme já mencionado, o processo de XCI parece apresentar diferenças significantes entre camundongos e humanos. Outra diferença está relacionada ao tempo de cultivo decorrido após transdução com a biblioteca de shRNAs. No trabalho de Chan e colaboradores (2011) as células foram coletadas 72 horas após a transdução enquanto que em nosso trabalho as células foram coletadas após 11 dias. É possível que células contendo dois cromossomos X ativos sejam afetadas negativamente por um tempo de cultura mais prolongado, assim como é possível que tempos de cultura muito curtos não sejam suficientes para se detectar o efeito de alguns shRNAs no processo de XCI.

Finalmente, a técnica empregada para detecção de shRNAs diferencialmente representados entre as amostras controle e experimental divergiu entre os dois trabalhos. Enquanto em nosso trabalho avaliamos a presença de insertos de shRNAs incorporados ao genoma das células transduzidas a partir da extração de DNA de células das amostras controle e experimental e NGS, Chan e colaboradores (2011) avaliaram a presença dos transcritos de shRNAs nas amostras controle e experimental através de *microarray*, um método menos sensível do que o NGS.

Sendo assim, a ausência de genes previamente associados ao processo de manutenção da XCI em nossa lista de candidatos não deve ser considerada como um indicativo de falta de especificidade de nossa triagem funcional. Entretanto, o papel dos genes candidatos identificados em nosso trabalho na manutenção da XCI em humanos somente poderá ser confirmado através de experimentos de validação. Esses experimentos serão realizados em trabalho de continuação ao descrito nessa tese.

5.8. Direções Futuras

Os experimentos de validação e estudo dos genes candidatos obtidos através do trabalho descrito nessa tese de doutorado já estão sendo programados e devem compreender os seguintes passos:

- a. Células HPRT⁻ serão transduzidas com shRNAs individuais para cada um dos genes candidatos selecionados (Tabela 20) submetidas à extração de RNA total, síntese de cDNA e qPCR a fim de conformar o *knockdown* do gene alvo. A diminuição na taxa de expressão de cada gene alvo também será confirmada através de experimentos de western blot, quando aplicável (para genes candidatos codificadores de proteínas);
- b. O cDNA obtido a partir de cada população de células submetidas ao *knockdown* de um gene candidato específico será submetido ao sequenciamento direto para determinação dos genótipos referentes aos SNPs heterozigotos identificados nesse trabalho de doutorado. A detecção de expressão gênica proveniente de ambos os alelos relacionados a cada SNP será considerada como um forte indicativo de reativação do cromossomo X;
- c. As células HPRT⁻ transduzidas com os shRNAs individuais também serão selecionadas em meio HAT para que se possa averiguar se o *knockdown* de cada gene candidato torna as células HPRT⁺, um indicativo de reativação do Xi.
- d. A presença de reativação do cromossomo X também será investigada através de RNA-FISH para o gene *XIST* e imunofluorescência para H3K27me3 – ambos conhecidos marcadores moleculares associados ao Xi.
- e. O papel dos genes candidatos validados com sucesso pelos métodos descritos acima no processo de manutenção da XCI será mais amplamente investigado através de outros métodos de biologia molecular que poderão incluir imunofluorescência, co-imunoprecipitação (co-IP), imunoprecipitação de cromatina (ChIP) ou outras metodologias que consideráremos relevantes;

- f. Os shRNAs relacionados aos genes candidatos aprovados no processo de validação poderão ser utilizados em conjunto, em diferentes combinações, para induzir o *knockdown* simultâneo de diferentes genes candidatos em uma mesma população de células. Isso permitirá avaliar possíveis efeitos sinérgicos entre eles no processo de manutenção da XCI em humanos.

Conclusões

6. Conclusões

O presente trabalho teve como principal objetivo a realização de uma triagem funcional em células somáticas femininas para identificação de genes envolvidos no processo de manutenção da XCI em humanos. Nossa estratégia experimental consistiu na transdução de fibroblastos primários *HPRT*^{+/-} e *HPRT*⁻ com uma biblioteca de shRNAs de cobertura genômica e subsequentemente transferência de parte da população de células transduzidas para meio de cultura contendo HAT. Com isso, esperávamos selecionar células que tiveram seu cromossomo Xi reativado (células *HPRT*⁺). Esperava-se que a presença de células *HPRT*⁺ na placa de cultura tratada em HAT estivesse fortemente relacionada ao *knockdown* de genes envolvidos na manutenção do estado de inativação do cromossomo X. A outra parte da população de células transduzidas foi mantida em meio 6TG já que nessas condições esperava-se que houvesse contra-seleção de células que reativaram seu Xi. A população de células mantida em 6TG foi utilizada como controle.

Utilizamos a técnica de NGS seguida de alinhamento dos insertos de shRNAs sequenciados contra diferentes banco de dados de referência para identificar e comparar a frequência relativa a cada inserto de shRNA observada nas amostras experimental tratada com HAT e controle mantida em 6TG. Através dessa metodologia, conseguimos identificar sequencias relacionadas a 20 genes com representação significativamente mais elevada na amostra tratada em HAT do que na amostra controle 6TG. Nenhum dos candidatos identificados em nossa triagem funcional foi ainda relacionado ao processo de manutenção da XCI na literatura.

Em posse da lista de candidatos, investigamos as características e funções de cada gene na literatura e em bancos de dados disponíveis na internet a fim de determinar quais candidatos melhor se adequavam a um possível papel na manutenção da XCI. Diversos candidatos identificados em nosso estudo possuem atualmente função pouco ou não conhecida. Esse é o caso dos genes *ZCCHC10*, *C2orf40*, *LDHAL6B*, *PSKH1*, *ZNF839*, *ADTRP*, *C7orf33*, *HPYR1* e *CHIC1*. Além disso, espera-se que muitos deles possam ser encontrados no núcleo celular (genes *ZCCHC10*, *C2orf40*, *PSKH1*, *ZNF839*, *C7orf33* e *HPYR1*). Esses fatores colocam

esses candidatos em um grupo cuja função na manutenção da XCI merece ser investigada.

Outros candidatos identificados em nossa análise já possuem funções bastante conhecidas e foram relacionados a processos celulares que nunca foram associados ao processo de XCI. Esse é o caso das proteínas MAN2A2, VAMP2, CADPS2, MARCH2, CLVS2, HCAR1, PAPPA2 e PLA2G4C. Além disso, as proteínas codificadas por esses genes parecem ser encontradas somente fora do compartimento nuclear, incluindo vesículas citoplasmáticas associadas aos processos de endocitose ou exocitose (VAMP2, CADPS2, MARCH2, CLVS2) e membrana plasmática (PAPPA2, HCAR1 e PLA2G4C). Sendo assim, um papel na manutenção da XCI se torna menos provável e, portanto, a identificação desses candidatos em nossa triagem funcional provavelmente deva estar relacionada a mecanismos que conferem resistência a droga HAT em cultura.

Nossa triagem funcional identificou também candidatos cuja função relatada na literatura está diretamente relacionada a mecanismos de controle epigenético. Embora esses candidatos não tenham sido ainda relacionados com o processo de XCI, o envolvimento em mecanismos de controle epigenético os enquadra em um grupo de candidatos de maior probabilidade de envolvimento na manutenção da XCI. Esse é o caso da proteína SRSF10 (envolvida na regulação de *splicing*), da histona chaperone ASF1 e da variante de histona H3.3B.

Nosso estudo identificou ainda algumas correlações funcionais entre diferentes genes candidatos. Esse é o caso das proteínas ASF1 e H3.3B que, conforme já demonstrado, são capazes de interagir diretamente; das proteínas envolvidas em *splicing* SRSF10 e PSKH1 (e possivelmente ZCCH10 e LDHAL6B) e das proteínas CAPSD2, VAMP2, MARCH2 e CLVS2, presentes em vesículas citoplasmáticas e relacionadas a processos de endo ou exocitose.

Experimentos de validação devem ser conduzidos a fim de que o envolvimento de cada um dos genes candidatos no processo de manutenção da XCI possa ser confirmado.

Resumo

7. Resumo

A compensação da dosagem gênica entre fêmeas XX e machos XY em mamíferos é adquirida através de um complexo mecanismo epigenético que resulta na inativação de grande parte de um dos cromossomos X nas células femininas. O processo de inativação do cromossomo X (XCI) se inicia cedo durante a embriogênese, concomitantemente à diferenciação celular, e envolve a aquisição de modificações epigenéticas características do cromossomo X inativo (Xi). Uma estabelecido, o padrão de inativação é estavelmente mantido através de todas as mitoses celulares subsequentes e por toda a vida do organismo (exceto para células germinativas que sofrem reativação do Xi).

Os mecanismos envolvidos na iniciação e estabelecimento da XCI foram extensivamente estudados, especialmente em camundongos. Embora algumas características epigenéticas associadas à manutenção da XCI tenham sido descritas, a identidade e modo específico de ação de fatores envolvidos durante essa fase da XCI são aspectos ainda não bem compreendidos. Além disso, o processo de XCI apresenta diferenças importantes entre humanos e camundongos e estudos direcionados para a identificação de novos componentes envolvidos na manutenção da XCI em humanos tornam-se de fundamental importância.

Triagens funcionais genômicas por bibliotecas de shRNAs constituem uma ferramenta poderosa para a identificação de genes envolvidos em diferentes mecanismos celulares e vias bioquímicas. Sendo assim, utilizamos essa ferramenta para triar genes envolvidos na manutenção da XCI em humanos. Células somáticas femininas primárias $HPRT^{+/-}/HPRT^{-}$ foram transduzidas com uma biblioteca lentiviral de shRNAs e posteriormente tratadas em meio de cultura contendo a droga HAT para seleção de células $HPRT^{+}$ nas quais esperava-se que o cromossomo Xi presente tivesse sofrido reativação em decorrência do *knockdown* de genes envolvidos na manutenção da XCI. Essa estratégia nos permitiu identificar 20 novos genes candidatos a estarem envolvidos na manutenção da XCI. Esses candidatos deverão ser avaliados individualmente para confirmar seu papel no processo de controle epigenético do cromossomo X.

Abstract

8. Abstract

Transcriptional dosage compensation between mammalian XX females and XY males is acquired through a complex epigenetic mechanism that leads to the inactivation of most part of one of the X chromosomes in the female cells. The X chromosome inactivation (XCI) process takes place early during embryogenesis and involves the acquisition of epigenetic modifications that are characteristic of the inactive X chromosome (Xi). Once silencing is established, the inactivation pattern is maintained in through all the subsequent mitosis and the same X chromosome remains stably silenced in all the descendant cells and throughout the life of the organism (except for the germ line cells that undergo X chromosome reactivation).

The initiation of XCI has been studied extensively, especially in mice. Although some epigenetic features associated with the maintenance of XCI have already been described, the identity and specific mode of action of the factors involved in this phase of XCI are largely unknown. Moreover, the XCI process presents important differences between mice and humans, and studies directed to the identification of new players involved in the maintenance of human XCI are fundamentally important.

Functional genome-wide screens using multiplex shRNA libraries are a powerful tool for the identification of genes involved in different cellular mechanisms and biochemical pathways. In order to screen for genes involved in the maintenance of XCI in humans, a population of $HPRT^{+/-}/HPRT^{-}$ primary somatic female cells were transduced with a multiplex lentiviral shRNA library and subsequently treated in HAT medium to select for $HPRT^{+}$ cells in which we expected that the Xi would undergo reactivation as a result of the *knockdown* of genes involved in the maintenance of XCI. As a result, we identified 20 new candidate genes that could potentially be involved in the maintenance of XCI. These candidates should be individually evaluated in order to confirm their role in the epigenetic control of the X chromosome.

Referências Bibliográficas

9. Referências Bibliográficas

Agez M, Chen J, Guerois R, van Heijenoort C, Thuret JY, Mann C, Ochsenbein F. Structure of the histone chaperone ASF1 bound to the histone H3 C-terminal helix and functional insights. *Structure*, 15(2):191-199, 2007.

Akai Y, Adachi N, Hayashi Y, Eitoku M, Sano N, Natsume R, Kudo N, Tanokura M, Senda T, Horikoshi M. Structure of the histone chaperone CIA/ASF1-double bromodomain complex linking histone modifications and site-specific histone eviction. *Proc Natl Acad Sci U S A*, 107(18):8153-8158, 2010.

Akama TO, Nakagawa H, Wong NK, Sutton-Smith M, Dell A, Morris HR, Nakayama J, Nishimura S, Pai A, Moremen KW, Marth JD, Fukuda MN. Essential and mutually compensatory roles of α -mannosidase II and α -mannosidase IIx in N-glycan processing in vivo in mice. *Proc Natl Acad Sci U S A*, 103(24):8983-8988, 2006.

Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Chromosomal DNA and Its Packaging in the Chromatin Fiber. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK26834/>

Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*, 11(10):R106, 2010.

Auclair G, Weber M. Mechanisms of DNA methylation and demethylation in mammals. *Biochimie*, 94(11):2202-2211, 2012.

Augui S, Filion GJ, Huart S, Nora E, Guggiari M, Maresca M, Stewart AF, Heard E. Sensing X Chromosome Pairs Before X Inactivation Via a Novel X-Pairing Region of the Xic. *Science*, 318: 1632-1636, 2007.

Augui S, Nora EP, Heard E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet*, 12(6):429-442, 2011.

Ausió J. Histone variants - the structure behind the function. *Brief Funct Genomic Proteomic*, 5(3):228-243, 2006.

Bailey JA, Carrel L, Chakravarti A, Eichler EE. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. *Proc Natl Acad Sci U S A*, 97:6634-6639, 2000.

Banaszynski LA, Allis CD, Lewis PW. Histone variants in metazoan development. *Dev Cell*, 19(5):662-674, 2010.

Banerjee T, Chakravarti D. A peek into the complex realm of histone phosphorylation. *Mol Cell Biol*, 31(24):4858-4873, 2011.

Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res*, 21(3):381-95, 2011.

Banumathy G, Somaiah N, Zhang R, Tang Y, Hoffmann J, Andrade M, Ceulemans H, Schultz D, Marmorstein R, Adams PD. Human UBN1 is an ortholog of yeast Hpc2p and has an essential role in the HIRA/ASF1a chromatin-remodeling pathway in senescent cells. *Mol Cell Biol*, 29(3):758-770, 2009.

Barakat TS, Gribnau J. X Chromosome Inactivation and Embryonic Stem Cells. Em: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000-. Publicado em 2010. Disponível em <http://www.ncbi.nlm.nih.gov/books/NBK45037/>

Barakat TS, Gunhanlar N, Pardo CG, Achame EM, Ghazvini M, Boers R, Kenter A, Rentmeester E, Grootegoed JA, Gribnau J. RNF12 activates Xist and is essential for X chromosome inactivation. *PLoS Genet*, 7(1):e1002001, 2011.

Barr ML, Bertram EG. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature*, 163(4148):676, 1949.

Barr ML, Bertram EG. The behaviour of nuclear structures during depletion and restoration of Nissl material in motor neurons. *J Anat*, 85(2):171-181, 1951.

Barr H, Hermann A, Berger J, Tsai HH, Adie K, Prokhortchouk A, Hendrich B, Bird A. Mbd2 contributes to DNA methylation-directed repression of the Xist gene. *Mol Cell Biol*, 27(10):3750–3757, 2007.

Bartee E, Mansouri M, Hovey Nerenberg BT, Gouveia K, Früh K. Downregulation of major histocompatibility complex class I by human ubiquitin ligases related to viral immune evasion proteins. *J Virol*, 78(3):1109-1120, 2004.

Batista PJ, Chang HY. Long noncoding RNAs: cellular address codes in development and disease. *Cell*, 152(6):1298-1307, 2013.

Baumann C, De La Fuente R. ATRX marks the inactive X chromosome (Xi) in somatic cells and during imprinted X chromosome inactivation in trophoblast stem cells. *Chromosoma*, 118(2):209-222, 2009.

Beard C, Li E, Jaenisch R. Loss of methylation activates Xist in somatic but not in embryonic cells. *Genes Dev*, 9:2325–2334, 1995.

Behjati S, Tarpey PS, Presneau N, Scheipl S, Pillay N, Van Loo P, Wedge DC, Cooke SL, Gundem G, Davies H, Nik-Zainal S, Martin S, McLaren S, Goodie V, Robinson B, Butler A, Teague JW, Halai D, Khatri B, Myklebost O, Baumhoer D, Jundt G, Hamoudi R, Tirabosco R, Amary MF, Futreal PA, Stratton MR, Campbell PJ, Flanagan AM. Distinct H3F3A and H3F3B driver mutations define chondroblastoma and giant cell tumor of bone. *Nat Genet*, 45(12):1479-1482, 2013.

Bergman Y, Cedar H. DNA methylation dynamics in health and disease. *Nat Struct Mol Biol*, 20(3):274-281, 2013.

Berletch JB, Yang F, Disteche CM. Escape from X inactivation in mice and humans. *Genome Biol*, 11(6):213, 2010.

Bi J, Li Y, Sun F, Saalbach A, Klein C, Miller DJ, Hess R, Nowak RA. Basigin null mutant male mice are sterile and exhibit impaired interactions between germ cells and Sertoli cells. *Dev Biol*, 380(2):145-156, 2013.

Biamonti G, Caceres JF. Cellular stress and RNA splicing. *Trends Biochem Sci*, 34(3):146-153, 2009.

Bickford JS, Newsom KJ, Herlihy JD, Mueller C, Keeler B, Qiu X, Walters JN, Su N, Wallet SM, Flotte TR, Nick HS. Induction of group IVC phospholipase A2 in allergic asthma: transcriptional regulation by TNF α in bronchoepithelial cells. *Biochem J*, 442(1):127-137, 2012.

Bickford JS, Mueller C, Newsom KJ, Barilovits SJ, Beachy DE, Herlihy JD, Keeler B, Flotte TR, Nick HS. Effect of allergy and inflammation on eicosanoid gene expression in CFTR deficiency. *J Cyst Fibros*, 12(3):258-265, 2013.

Blewitt ME, Gendrel AV, Pang Z, Sparrow DB, Whitelaw N, Craig JM, Apedaile A, Hilton DJ, Dunwoodie SL, Brockdorff N, Kay GF, Whitelaw E. SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat Genet*, 40(5):663-669, 2008.

Bogan JS. Regulation of glucose transporter translocation in health and diabetes. *Annu Rev Biochem*, 81:507-532, 2012.

Bonanomi D, Rusconi L, Colombo CA, Benfenati F, Valtorta F. Synaptophysin I selectively specifies the exocytic pathway of synaptobrevin 2/VAMP2. *Biochem J*, 404(3):525-534, 2007.

Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, Grompe M, Pizzuti A, Muzny D, Lawrence C, Willard HF, Avner P, Ballabio A. Characterization of a murine gene expressed from the inactive X chromosome. *Nature*, 351(6324):325-329, 1991.

Brede G, Solheim J, Tröen G, Prydz H. Characterization of PSKH1, a novel human protein serine kinase with centrosomal, golgi, and nuclear localization. *Genomics*, 70(1):82-92, 2000.

Brede G, Solheim J, Prydz H. PSKH1, a novel splice factor compartment-associated serine kinase. *Nucleic Acids Res*, 30(23):5301-5309, 2002.

Brede G, Solheim J, Stang E, Prydz H. Mutants of the protein serine kinase PSKH1 disassemble the Golgi apparatus. *Exp Cell Res*, 291(2):299-312, 2003.

Brinkman AB, Roelofsen T, Pennings SW, Martens JH, Jenuwein T, Stunnenberg HG. Histone modification patterns associated with the human X chromosome. *EMBO Rep*, 7(6): 628-634, 2006.

Brockdorff N, Ashworth A, Kay GF, Cooper P, Smith S, McCabe VM, Norris DP, Penny GD, Patel D, Rastan S. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature*, 351(6324):329-331, 1991.

Brown CJ, Willard HF. The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature*, 368(6467):154-156, 1999.

Brown CJ, Lafreniere RG, Powers VE, Sebastio G, Ballabio A, Pettigrew AL, Ledbetter DH, Levy E, Craig IW, Willard HF. Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature*, 349(6304):82-84, 1991.

Cai TQ, Ren N, Jin L, Cheng K, Kash S, Chen R, Wright SD, Taggart AK, Waters MG. Role of GPR81 in lactate-mediated reduction of adipose lipolysis. *Biochem Biophys Res Commun*, 377(3):987-991, 2008.

Cao Z, Huett A, Kuballa P, Giallourakis C, Xavier RJ. DLG1 is an anchor for the E3 ligase MARCH2 at sites of cell-cell contact. *Cell Signal*, 20(1):73-82, 2008.

Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, 434(7031):400-404, 2005.

Carrel L, Park C, Tyekucheva S, Dunn J, Chiaromonte F, Makova KD. Genomic environment predicts expression patterns on the human inactive X chromosome. *PLoS Genet*, 2:e151, 2006.

Chadwick BP, Willard HF. Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc Natl Acad Sci U S A*, 101:17450–17455, 2004.

Chadwick, B.P. Variation in Xi chromatin organization and correlation of the H3K27me3 chromatin territories to transcribed sequences by microarray analysis. *Chromosoma*, 116:147–157, 2007.

Chan KM, Zhang H, Malureanu L, van Deursen J, Zhang Z. Diverse factors are involved in maintaining X chromosome inactivation. *Proc Natl Acad Sci U S A*, 108(40):16699-16704, 2011.

Chang B, Chen Y, Zhao Y, Bruick RK. JMJD6 is a histone arginine demethylase. *Science*, 318(5849):444-447, 2007.

Cheng J, Guggino W. Ubiquitination and degradation of CFTR by the E3 ubiquitin ligase MARCH2 through its association with adaptor proteins CAL and STX6. *PLoS One*, 8(6):e68001, 2013.

Chimura T, Kuzuhara T, Horikoshi M. Identification and characterization of CIA/ASF1 as an interactor of bromodomains associated with TFIID. *Proc Natl Acad Sci U S A*, 99(14):9334-9339, 2002.

Chow JC, Heard E. Nuclear organization and dosage compensation. *Cold Spring Harb Perspect Biol*, 2(11):a000604, 2010.

Cisternas FA, Vincent JB, Scherer SW, Ray PN. Cloning and characterization of human CADPS and CADPS2, new members of the Ca²⁺-dependent activator for secretion protein family. *Genomics*, 81(3):279-291, 2003.

Corpet A, De Koning L, Toedling J, Savignoni A, Berger F, Lemaître C, O'Sullivan RJ, Karlseder J, Barillot E, Asselain B, Sastre-Garau X, Almouzni G. Asf1b, the necessary Asf1 isoform for proliferation, is predictive of outcome in breast cancer. *EMBO J*, 30(3):480-493, 2011.

Cools J, Mentens N, Marynen P. A new family of small, palmitoylated, membrane-associated proteins, characterized by the presence of a cysteine-rich hydrophobic motif. *FEBS Lett*, 492(3):204-209, 2001.

Cox JM, Clayton CL, Tomita T, Wallace DM, Robinson PA, Crabtree JE. cDNA array analysis of cag pathogenicity island-associated *Helicobacter pylori* epithelial cell response genes. *Infect Immun*, 69(11):6970-6980, 2001.

Csankovszki G, Panning B, Bates B, Pehrson JR, Jaenisch R. Conditional deletion of Xist disrupts histone macroH2A localization but not maintenance of X-inactivation. *Nat Genet*, 22:323–24, 1999.

Csankovszki G, Nagy A, Jaenisch R. Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *Cell Biol*, 153:773–84, 2001.

Daury L, Chailleux C, Bonvallet J, Trouche D. Histone H3.3 deposition at E2F-regulated genes is linked to transcription. *EMBO Rep*, 7(1):66-71, 2006.

De La Fuente R, Baumann C, Viveiros MM. Role of ATRX in chromatin structure and function: implications for chromosome instability and human disease. *Reproduction*, 142(2):221-234, 2011.

Debrand E, Chureau C, Arnaud D, Avner P, Heard E. Functional analysis of the DXPas34 locus, a 3' regulator of Xist expression. *Mol Cell Biol*, 19:8513–8525, 1999.

Delbarre E, Jacobsen BM, Reiner AH, Sørensen AL, Küntziger T, Collas P. Chromatin environment of histone variant H3.3 revealed by quantitative imaging and genome-scale chromatin and DNA immunoprecipitation. *Mol Biol Cell*, 21(11):1872-1884, 2010.

Dhanasekaran K, Kumari S, Kanduri C. Noncoding RNAs in Chromatin Organization and Transcription Regulation: An Epigenetic View. *Subcell Biochem*, 61:343-372, 2012.

Dickie MM. The tortoiseshell house mouse. *J Hered*, 45:158–159, 1954.

Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, Guernec G, Jagla B, Jouneau L, Laloë D, Le Gall C, Schaëffer B, Le Crom S, Guedj M, Jaffrézic F; French StatOmique Consortium. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform*, 14(6):671-683, 2013.

Disteche CM, Filippova GN, Tsuchiya KD. Escape from X inactivation. *Cytogenet Genome Res*, 99(1-4):36-43, 2002.

Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, Tanzer A, Lagarde J, Lin W, Schlesinger F, Xue C, Marinov GK, Khatun J, Williams BA, Zaleski C, Rozowsky J, Röder M, Kokocinski F, Abdelhamid RF, Alioto T, Antoshechkin I, Baer MT, Bar NS, Batut P, Bell K, Bell I, Chakraborty S, Chen X, Chrast J, Curado J, Derrien T, Drenkow J, Dumais E, Dumais J, Duttagupta R, Falconnet E, Fastuca M, Fejes-Toth K, Ferreira P, Foissac S, Fullwood MJ, Gao H, Gonzalez D, Gordon A, Gunawardena H, Howald C, Jha S, Johnson R, Kapranov P, King B, Kingswood C, Luo OJ, Park E, Persaud K, Preall JB, Ribeca P, Risk B, Robyr D, Sammeth M, Schaffer L, See LH, Shahab A, Skancke J, Suzuki AM, Takahashi H, Tilgner H, Trout D, Walters N, Wang H, Wrobel J, Yu Y, Ruan X, Hayashizaki Y, Harrow J, Gerstein M, Hubbard T, Reymond A, Antonarakis SE, Hannon G, Giddings MC, Ruan Y, Wold B, Carninci P, Guigó R, Gingeras TR. Landscape of transcription in human cells. *Nature*, 489(7414):101-108, 2012.

Donohoe ME, Silva SS, Pinter SF, Xu N, Lee JT. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature*, 460(7251):128-132, 2009.

Drané P, Ouarrhni K, Depaux A, Shuaib M, Hamiche A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev*, 24(12):1253-1265, 2010.

Dvash T, Fan G. Epigenetic Regulation of X-Inactivation in Human Embryonic Stem Cells. *Epigenetics*, 4(1): 19-22, 2009.

Elsaesser SJ, Goldberg AD, Allis CD. New functions for an old variant: no substitute for histone H3.3. *Curr Opin Genet Dev*, 20(2):110-117, 2010.

ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146):799-816, 2007.

ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57-74, 2012.

English CM, Adkins MW, Carson JJ, Churchill ME, Tyler JK. Structural basis for the histone chaperone activity of Asf1. *Cell*, 127(3):495-508, 2006.

Erwin JA, Lee JT. New Twists in X-Chromosome Inactivation. *Curr Opin Cell Biol*, 20: 1-7, 2008.

Esteller M. Non-coding RNAs in human disease. *Nat Rev Genet*, 12(12):861-874, 2011.

FANTOM Consortium, RIKEN Genome Exploration Research Group, Genome Science Group (Genome Network Project Core Group). The transcriptional landscape of the mammalian genome. *Science*, 309(5740):1559-1563, 2005.

Feng Y, Chen M, Manley JL. Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. *Nat Struct Mol Biol*, 15(10):1040-1048, 2008.

Feng Y, Valley MT, Lazar J, Yang AL, Bronson RT, Firestein S, Coetzee WA, Manley JL. SRp38 regulates alternative splicing and is required for Ca(2+) handling in the embryonic heart. *Dev Cell*, 16(4):528-538, 2009.

Flaquer A, Rappold GA, Wienker TF, Fischer C. The human pseudoautosomal regions: a review for genetic epidemiologists. *Eur J Hum Genet*, 16(7):771-779, 2008.

Fong JJ, Nguyen BL, Bridger R, Medrano EE, Wells L, Pan S, Sifers RN. β -N-Acetylglucosamine (O-GlcNAc) is a novel regulator of mitosis-specific phosphorylations on histone H3. *J Biol Chem*, 287(15):12195-203, 2012.

Ford CE, Jones KW, Polani PE, de Almeida JC, Briggs JH. A sex chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet*, 1:711-713, 1959.

Frank D, Doenecke D, Albig W. Differential expression of human replacement and cell cycle dependent H3 histone genes. *Gene*, 312:135-143, 2003.

Fujimoto WY, Subak-Sharpe JH, Seegmiller JE. Hypoxanthine-guanine phosphoribosyltransferase deficiency: chemical agents selective for mutant or normal cultured fibroblasts in mixed and heterozygote cultures. *Proc Natl Acad Sci U S A*, 68(7):1516-1519, 1971.

Fukuda MN, Akama TO. In vivo role of alpha-mannosidase IIx: ineffective spermatogenesis resulting from targeted disruption of the Man2a2 in the mouse. *Biochim Biophys Acta*, 1573(3):382-387, 2002.

Fukuda H, Nakamura N, Hirose S. MARCH-III Is a novel component of endosomes with properties similar to those of MARCH-II. *J Biochem*, 139(1):137-145, 2006.

Fushimi K, Osumi N, Tsukahara T. NSSRs/TASRs/SRp38s function as splicing modulators via binding to pre-mRNAs. *Genes Cells*, 10(6):531-541, 2005.

Galvani A, Courbeyrette R, Agez M, Ochsenbein F, Mann C, Thuret JY. In vivo study of the nucleosome assembly functions of ASF1 histone chaperones in human cells. *Mol Cell Biol*, 28(11):3672-3685, 2008.

Gartler SM, Goldman MA. Biology of the X chromosome. *Curr Opin Pediatr*, 13(4):340-345, 2001.

Gazin C, Wajapeyee N, Gobeil S, Virbasius CM, Green MR. An Elaborate Pathway Required for Ras-Mediated Epigenetic Silencing. *Nature*, 449(7165):1073-1077, 2007.

Gendrel AV, Apedaile A, Coker H, Termanis A, Zvetkova I, Godwin J, Tang YA, Huntley D, Montana G, Taylor S, Giannoulatou E, Heard E, Stancheva I, Brockdorff. Smchd1-dependent and -independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome. *Dev Cell*, 23(2):265-79, 2012.

Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, Wen D, Chappier A, DeKolver RC, Miller JC, Lee YL, Boydston EA, Holmes MC, Gregory PD, Greally JM, Rafii S, Yang C, Scambler PJ, Garrick D, Gibbons RJ, Higgs DR, Cristea IM, Urnov FD, Zheng D, Allis CD. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell*, 140(5):678-691, 2010.

Gontan C, Achame EM, Demmers J, Barakat TS, Rentmeester E, van IJcken W, Grootegeed JA, Gribnau J. RNF12 initiates X-chromosome inactivation by targeting REX1 for degradation. *Nature*, 485(7398):386-390, 2012.

González-González E, López-Casas PP, Del Mazo J. Gene silencing by RNAi in mouse Sertoli cells. *Reprod Biol Endocrinol*, 6:29, 2008.

González-Romero R, Rivera-Casas C, Ausió J, Méndez J, Eirín-López JM. Birth-and-death long-term evolution promotes histone H2B variant diversification in the male germinal cell line. *Mol Biol Evol*, 27(8):1802–1812, 2010.

Götze S, Feldhaus V, Traska T, Wolter M, Reifemberger G, Tannapfel A, Kuhnen C, Martin D, Müller O, Sievers S. ECRG4 is a candidate tumor suppressor gene frequently hypermethylated in colorectal carcinoma and glioma. *BMC Cancer*, 9:447, 2009.

Groth A, Ray-Gallet D, Quivy JP, Lukas J, Bartek J, Almouzni G. Human Asf1 regulates the flow of S phase histones during replicational stress. *Mol Cell*, 17(2):301-311, 2005.

Groth A, Corpet A, Cook AJ, Roche D, Bartek J, Lukas J, Almouzni G. Regulation of replication fork progression through histone supply and demand. *Science*, 318(5858):1928-1931, 2007.

Gutschner T, Baas M, Diederichs S. Noncoding RNA gene silencing through genomic integration of RNA destabilizing elements using zinc finger nucleases. *Genome Res*, 21(11):1944-1954, 2011.

Haff LA. Improved quantitative PCR using nested primers. *PCR Methods Appl*, 3(6):332-337, 1994.

Hall LL, Byron M, Butler J, Becker KA, Neleson A, Amit M, Itskovitz-Eldor J, Stein J, Stein G, Ware C, Lawrence JB. X-Inactivation Reveals Epigenetic Anomalies in Most hESC but Identifies Sublines That Initiate as Expected. *J Cell Physiol*, 216: 445-452, 2008.

Han G, Li J, Wang Y, Li X, Mao H, Liu Y, Chen CD. The hydroxylation activity of Jmjd6 is required for its homo-oligomerization. *J Cell Biochem*, 113(5):1663-1670, 2012.

Hangauer MJ, Vaughn IW, McManus MT. Pervasive transcription of the human genome produces thousands of previously unidentified long intergenic noncoding RNAs. *PLoS Genet*, 9(6):e1003569, 2013.

Hao JC, Salem N, Peng XR, Kelly RB, Bennett MK. Effect of mutations in vesicle-associated membrane protein (VAMP) on the assembly of multimeric protein complexes. *J Neurosci*, 17(5):1596-1603, 1997.

Harada A, Okada S, Konno D, Odawara J, Yoshimi T, Yoshimura S, Kumamaru H, Saiwai H, Tsubota T, Kurumizaka H, Akashi K, Tachibana T, Imbalzano AN, Ohkawa Y. Chd2 interacts with H3.3 to determine myogenic cell fate. *EMBO J*, 31(13):2994-3007, 2012.

Harper PS. Mary Lyon and the hypothesis of random X chromosome inactivation. *Hum Genet*, 130(2):169-174, 2011.

Harrison KB. X-Chromosome Inactivation in the Human Cytotrophoblast. *Cytogenet Cell Genet*, 52:37-41, 1989.

Hato M, Nakagawa H, Kuroguchi M, Akama TO, Marth JD, Fukuda MN, Nishimura S. Unusual N-glycan structures in alpha-mannosidase II/IIx double null embryos identified by a systematic glycomics approach based on two-dimensional LC mapping and matrix-dependent selective fragmentation method in MALDI-TOF/TOF mass spectrometry. *Mol Cell Proteomics*, 5(11):2146-2157, 2006.

Hayashi R, Goto Y, Tanaka R, Oonogi K, Hisasue M, Yoshida K. Transcriptional regulation of human chromatin assembly factor ASF1. *DNA Cell Biol*, 26(2):91-99, 2007.

Heard E, Clerc P, Avner P. X-Chromosome Inactivation in Mammals. *Annu Rev Genet*, 31: 571-610, 1997.

Heard E. Recent advances in X-chromosome inactivation. *Curr Opin Cell Biol*, 16:247-255, 2004.

Heard E. Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Curr Opin Genet Dev*, 15(5):482-489, 2005.

Healy S, Khan DH, Davie JR. Gene expression regulation through 14-3-3 interactions with histones and HDACs. *Discov Med*, 11(59):349-358, 2011.

Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, Will CL, Pena V, Lührmann R, Stelzl U. Dynamic protein-protein interaction wiring of the human spliceosome. *Mol Cell*, 45(4):567-80, 2012.

Hendrich B, Bird A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol*, 18(11):6538-6547, 1998.

Hoffman L, Hall L, Batten JL, Young H, Pardasani D, Baetge EE, Lawrence J, Carpenter MK. X-Inactivation Status Varies in Human Embryonic Stem Cell Lines. *Stem Cells*, 23: 1468-1478, 2005.

Hu G, Luo J. A primer on using pooled shRNA libraries for functional genomic screens. *Acta Biochim Biophys Sin (Shanghai)*, 44(2):103-112, 2012.

Huang X, Wang JY, Lu X. Systems analysis of quantitative shRNA-library screens identifies regulators of cell adhesion. *BMC Syst Biol*, 2:49, 2008.

Huang Y, Zhang JL, Yu XL, Xu TS, Wang ZB, Cheng XC. Molecular functions of small regulatory noncoding RNA. *Biochemistry (Mosc)*, 78(3):221-330, 2013.

Hurov KE, Cotta-Ramusino C, Elledge SJ. A genetic screen identifies the Triple T complex required for DNA damage signaling and ATM and ATR stability. *Genes Dev*, 24(17):1939-1950, 2010.

Inbar-Feigenberg M, Choufani S, Butcher DT, Roifman M, Weksberg R. Basic concepts of epigenetics. *Fertil Steril*, 99(3):607-615, 2013.

International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931-945, 2004.

Ishikawa K, Ohsumi T, Tada S, Natsume R, Kundu LR, Nozaki N, Senda T, Enomoto T, Horikoshi M, Seki M. Roles of histone chaperone CIA/Asf1 in nascent DNA elongation during nucleosome replication. *Genes Cells*, 16(10):1050-1062, 2011.

Jacobs PA, Strong JA. A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature*, 183:302–303, 1959.

Jonkers I, Barakat TS, Achame EM, Monkhorst K, Kenter A, Rentmeester E, Grosveld F, Grootegoed JA, Gribnau J. RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. *Cell*, 139(5):999-1011, 2009.

Jufvas A, Strålfors P, Vener AV. Histone variants and their post-translational modifications in primary human fat cells. *PLoS One*, 6(1):e15960, 2011.

Junqueira LC, Carneiro J. O núcleo da célula. In: *Biologia celular e molecular*. 7^o ed, Guanabara-Koogan SA, pp.147-151, 2000.

Katoh Y, Ritter B, Gaffry T, Blondeau F, Höning S, McPherson PS. The clavesin family, neuron-specific lipid- and clathrin-binding Sec14 proteins regulating lysosomal morphology. *J Biol Chem*, 284(40):27646-27654, 2009.

Kawanishi M, Tamori Y, Okazawa H, Araki S, Shinoda H, Kasuga M. Role of SNAP23 in insulin-induced translocation of GLUT4 in 3T3-L1 adipocytes. Mediation of complex formation between syntaxin4 and VAMP2. *J Biol Chem*, 275(11):8240-8247, 2000.

Khare SP, Habib F, Sharma R, Gadewal N, Gupta S, Galande S. Histome--a relational knowledgebase of human histone proteins and histone modifying enzymes. *Nucleic Acids Res*, 40(Database issue):D337-342, 2012. Banco de dados disponível em: <http://www.actrec.gov.in/histome/>

Kothapalli N, Camporeale G, Kueh A, Chew YC, Oommen AM, Griffin JB, Zemleni J. Biological functions of biotinylated histones. *J Nutr Biochem*, 16(7):446-448, 2005.

Kouzarides T. Chromatin modifications and their function. *Cell*, 128(4):693-705, 2007.

Kugel JF, Goodrich JA. Non-coding RNAs: key regulators of mammalian transcription. *Trends Biochem Sci*, ;37(4):144-151, 2012.

Kuei C, Yu J, Zhu J, Wu J, Zhang L, Shih A, Mirzadegan T, Lovenberg T, Liu C. Study of GPR81, the lactate receptor, from distant species identifies residues and motifs critical for GPR81 functions. *Mol Pharmacol*, 80(5):848-858, 2011.

Kumari S, Swaminathan A, Chatterjee S, Senapati P, Boopathi R, Kundu TK. Chromatin organization, epigenetics and differentiation: an evolutionary perspective. *Subcell Biochem*, 61:3-35, 2012.

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*, 10(3):R25, 2009.

Law C, Cheung P. Histone variants and transcription regulation. *Subcell Biochem*, 61:319-341, 2012.

Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet*, 21(4):400-404, 1999.

Lee DK, Nguyen T, Lynch KR, Cheng R, Vanti WB, Arkhitko O, Lewis T, Evans JF, George SR, O'Dowd BF. Discovery and mapping of ten novel G protein-coupled receptor genes. *Gene*, 275(1):83-91, 2001.

Lehne B, Lewis CM, Schlitt T. PLoS One. From SNPs to genes: disease association at the gene level. *PLoS One*, 6(6): e20133, 2011.

Lewis PW, Elsaesser SJ, Noh KM, Stadler SC, Allis CD. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proc Natl Acad Sci U S A*, 107(32):14075-14080, 2010.

Lin Y, Sun Z. Thyroid hormone promotes insulin-induced glucose uptake by enhancing Akt phosphorylation and VAMP2 translocation in 3T3-L1 adipocytes. *J Cell Physiol*, 226(10):2625-2632, 2011.

Liu C, Wu J, Zhu J, Kuei C, Yu J, Shelton J, Sutton SW, Li X, Yun SJ, Mirzadegan T, Mazur C, Kamme F, Lovenberg TW. Lactate inhibits lipolysis in fat cells through activation of an orphan G-protein-coupled receptor, GPR81. *J Biol Chem*, 284(5):2811-2822, 2009.

Loyola A, Almouzni G. Marking histone H3 variants: how, when and why? *Trends Biochem Sci*, 32(9):425-433, 2007.

Lupu C, Zhu H, Popescu NI, Wren JD, Lupu F. Novel protein ADTRP regulates TFPI expression and function in human endothelial cells in normal conditions and in response to androgen. *Blood*, 118(16):4463-4471, 2011.

Lyon MF. A further mutation of the mottled type in the house mouse. *J Hered*, 51:116–121, 1960.

Lyon MF. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature*, 190:372-3, 1961.

Lyon MF. Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet*, 14:135-148, 1962.

Lyon MF. Possible mechanisms of X chromosome inactivation. *Nat New Biol*, 232(34):229-232, 1971.

Lyon MF. X-chromosome inactivation: a repeat hypothesis. *Cytogenet Cell Genet*, 80:133-137, 1998.

Mekhoubad S, Bock C, de Boer AS, Kiskinis E, Meissner A, Eggan K. Erosion of dosage compensation impacts human iPSC disease modeling. *Cell Stem Cell*, 10(5):595-609, 2012.

Mello JA, Silljé HH, Roche DM, Kirschner DB, Nigg EA, Almouzni G. Human Asf1 and CAF-1 interact and synergize in a repair-coupled nucleosome assembly pathway. *EMBO Rep*, 3(4):329-334, 2002.

Mendes-Pereira AM, Sims D, Dexter T, Fenwick K, Assiotis I, Kozarewa I, Mitsopoulos C, Hakas J, Zvelebil M, Lord CJ, Ashworth A. Genome-wide functional screen identifies a compendium of genes affecting sensitivity to tamoxifen. *Proc Natl Acad Sci U S A*, 109(8):2730-2735, 2012.

Messner S, Hottiger MO. Histone ADP-ribosylation in DNA repair, replication and transcription. *Trends Cell Biol*, 21(9):534-42, 2011.

Migeon BR. X-linked hypoxanthine-guanine phosphoribosyl transferase deficiency: detection of heterozygotes by selective medium. *Biochem Genet*, 4:377-383, 1970.

Migeon BR, Chowdhury AK, Dunston JA, McIntosh I. Identification of TSIX, encoding an RNA antisense to human XIST, reveals differences from its murine counterpart: implications for X inactivation. *Am J Hum Genet*, 69(5):951-960, 2001.

Migeon BR, Lee CH, Chowdhury AK, Carpenter H. Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation. *Am J Hum Genet*, 71(2):286-293, 2002.

Miller JL, Grant PA. The role of DNA methylation and histone modifications in transcriptional regulation in humans. *Subcell Biochem*, 61:289-317, 2012.

Mirabeau O, Perlas E, Severini C, Audero E, Gascuel O, Possenti R, Birney E, Rosenthal N, Gross C. Identification of novel peptide hormones in the human proteome by hidden Markov model screening. *Genome Res*, 17(3):320-327, 2007.

Mise N, Goto Y, Nakajima N, Takagi N. Molecular cloning of antisense transcripts of the mouse Xist gene. *Biochem Biophys Res Commun*, 258(3):537-541, 1999.

Monkhorst K, Jonkers I, Rentmeester E, Grosveld F, Gribnau J. X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator. *Cell*, 132(3):410-421, 2008.

Moreira de Mello JC, de Araújo ES, Stabellini R, Fraga AM, de Souza JE, Sumita DR, Camargo AA, Pereira LV. Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. *PLoS One*, 5(6):e10947, 2010.

Morey C, Avner P. Employment opportunities for non-coding RNAs. *FEBS Lett.*, 567(1):27-34, 2004.

Mori Y, Ishiguro H, Kuwabara Y, Kimura M, Mitsui A, Kurehara H, Mori R, Tomoda K, Ogawa R, Katada T, Harata K, Fujii Y. Expression of ECRG4 is an independent prognostic factor for poor survival in patients with esophageal squamous cell carcinoma. *Oncol Rep*, 18(4):981-985, 2007.

Mousson F, Lautrette A, Thuret JY, Agez M, Courbeyrette R, Amigues B, Becker E, Neumann JM, Guerois R, Mann C, Ochsenbein F. Structural basis for the interaction of Asf1 with histone H3 and its functional implications. *Proc Natl Acad Sci U S A*, 102(17):5975-5980, 2005.

Munakata T, Adachi N, Yokoyama N, Kuzuhara T, Horikoshi M. A human homologue of yeast anti-silencing factor has histone chaperone activity. *Genes Cells*, 5(3):221-233, 2000.

Nakamura N, Fukuda H, Kato A, Hirose S. MARCH-II is a syntaxin-6-binding protein involved in endosomal trafficking. *Mol Biol Cell*, 16(4):1696-1710, 2005.

Natsume R, Eitoku M, Akai Y, Sano N, Horikoshi M, Senda T. Structure and function of the histone chaperone CIA/ASF1 complexed with histones H3 and H4. *Nature*, 446(7133):338-341, 2007.

Navarro P, Chambers I, Karwacki-Neisius V, Chureau C, Morey C, Rougeulle C, Avner P. Molecular coupling of Xist regulation and pluripotency. *Science*, 321(5896):1693-1695, 2008.

Ngo VN, Davis RE, Lamy L, Yu X, Zhao H, Lenz G, Lam LT, Dave S, Yang L, Powell J, Staudt LM. A loss-of-function RNA interference Screen for Molecular Targets in Cancer. *Nature*, 441(7089):106-110, 2006.

Nicodemi M, Prisco A. Self-Assembly and DNA binding of the Blocking Factor in X Chromosome Inactivation. *Plos*, 3(11): 2135-2142, 2007.

Nozawa RS, Nagao K, Igami KT, Shibata S, Shirai N, Nozaki N, Sado T, Kimura H, Obuse C. Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway. *Nat Struct Mol Biol*, 20(5):566-73, 2013.

Offermanns S, Colletti SL, Lovenberg TW, Semple G, Wise A, IJzerman AP. International Union of Basic and Clinical Pharmacology. LXXXII: Nomenclature and Classification of Hydroxy-carboxylic Acid Receptors (GPR81, GPR109A, and GPR109B). *Pharmacol Rev*, 63(2):269-290, 2011.

Oh-Eda M, Nakagawa H, Akama TO, Lowitz K, Misago M, Moremen KW, Fukuda MN. Overexpression of the Golgi-localized enzyme alpha-mannosidase IIx in Chinese hamster ovary cells results in the conversion of hexamannosyl-N-acetylchitobiose to tetramannosyl-N-acetylchitobiose in the N-glycan-processing pathway. *Eur J Biochem*, 268(5):1280-1288, 2001.

Ohno S, Kaplan WD, Kinosita R. Formation of the sex chromatin by a single X-chromosome in liver cells of *Rattus norvegicus*. *Exp Cell Res*, 18:415-418, 1959.

Ohno S, Hauschka TS. Allocycly of the X-chromosome in tumors and normal tissues. *Cancer Res*, 20:541-545, 1960.

Otto SP, Pannell JR, Peichel CL, Ashman TL, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet*, 27(9):358-367, 2011.

Overgaard MT, Boldt HB, Laursen LS, Sottrup-Jensen L, Conover CA, Oxvig C. Pregnancy-associated plasma protein-A2 (PAPP-A2), a novel insulin-like growth factor-binding protein-5 proteinase. *J Biol Chem*, 276(24):21849-21853, 2001.

Ozdemir A, Masumoto H, Fitzjohn P, Verreault A, Logie C. Histone H3 lysine 56 acetylation: a new twist in the chromosome cycle. *Cell Cycle*, 5(22):2602-2608, 2006.

Panning B, Jaenisch R. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes Dev*, 10:1991–2002, 1996.

Payer B, Lee JT. X Chromosome Dosage Compensation: How Mammals Keep the Balance. *Ann Rev Genet*, 42: 733-772, 2008.

Pereira LV, Vasques LR. X-Chromosome Inactivation: Lessons from Transgenic Mice. *Gene*, 255: 363-371, 2000.

Perry J, Palmer S, Gabriel A, Ashworth A. A short pseudoautosomal region in laboratory mice. *Genome Res*, 11(11):1826-1832, 2001.

Pierce BA. Estrutura dos cromossomos e elementos de transposição. In: Genética um enfoque conceitual. 7º ed, Guanabara-Koogan SA, pp.281-287, 2004.

Pioli PA, Hamilton BJ, Connolly JE, Brewer G, Rigby WF. Lactate dehydrogenase is an AU-rich element-binding protein that directly interacts with AUF1. *J Biol Chem*, 277(38):35738-35745, 2002.

Plath K, Mlynarczyk-Evans S, Nusinow DA, Panning B. Xist RNA and the Mechanism of X Chromosome Inactivation. *Annu Rev Genet*, 36:233-278, 2002.

Plath K, Fang J, Mlynarczyk-Evans SK, Cao R, Worringer KA, Wang H, Cruz CCC, Otte AP, Panning B, Zhang Y. Role of Histone H3 Lysine 27 Methylation in X Inactivation. *Science*, 300: 131-135, 2003.

Ploner A, Ploner C, Lukasser M, Niederegger H, Hüttenhofer A. Methodological obstacles in knocking down small noncoding RNAs. *RNA*, 15(10):1797-1804, 2009.

Polani PE, Hunter JF, Lennox B. Chromosomal sex in Turner's syndrome with coarctation of the aorta. *Lancet*, 2:120–121, 1954.

Polani PE, Lessof MH, Bishop PMF. Colour blindness in ovarian agenesis (gonadal dysplasia). *Lancet*, 271:118–120, 1956.

Povlsen LK, Beli P, Wagner SA, Poulsen SL, Sylvestersen KB, Poulsen JW, Nielsen ML, Bekker-Jensen S, Mailand N, Choudhary C. Systems-wide analysis of ubiquitylation dynamics reveals a key role for PAF15 ubiquitylation in DNA-damage bypass. *Nat Cell Biol*, 14(10):1089-1098, 2012.

Prescott GR, Gorleku OA, Greaves J, Chamberlain LH. Palmitoylation of the synaptic vesicle fusion machinery. *J Neurochem*, 110(4):1135-1149, 2009.

Pilyugin M, Demmers J, Verrijzer CP, Karch F, Moshkin YM. Phosphorylation-mediated control of histone chaperone ASF1 levels by Tousled-like kinases. *PLoS One*, 4(12):e8328, 2009.

Rao DD, Vorhies JS, Senzer N, Nemunaitis J. siRNA vs. shRNA: similarities and differences. *Adv Drug Deliv Rev*, 61(9):746-759, 2009.

Rastan S. Non-random X-chromosome inactivation in mouse X-autosome translocation embryos--location of the inactivation centre. *J Embryol Exp Morphol*, 78:1-22, 1983.

Rastan S, Robertson EJ. X-chromosome deletions in embryo-derived (EK) cell lines associated with lack of X-chromosome inactivation. *J Embryol Exp, Morphol*, 90, 379–388, 1985.

Rhee I, Bachman KE, Park BH, Jair KW, Yen RW, Schuebel KE, Cui H, Feinberg AP, Lengauer C, Kinzler KW, Baylin SB, Vogelstein B. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*, 416(6880):552-556, 2002.

Robles JA, Qureshi SE, Stephen SJ, Wilson SR, Burden CJ, Taylor JM. Efficient experimental design and analysis strategies for the detection of differential expression using RNA-Sequencing. *BMC Genomics*, 13:484, 2012.

Romito A, Rougeulle C. Origin and evolution of the long non-coding genes in the X-inactivation center. *Biochimie*, 93(11):1935-1942, 2011.

Ropers HH, Wolff G, Nitzeroth HW. Preferential X-Inactivation in Human Placenta Membranes: Is the Paternal X Inactive in Early Embryonic Development of Female Mammals? *Hum Genet*, 43: 265-173, 1978.

Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, Frankish A, Lovell FL, Howe KL, Ashurst JL, Fulton RS, Sudbrak R, Wen G, Jones MC, Hurles ME, Andrews TD, Scott CE, Searle S, Ramser J, Whittaker A, Deadman R, Carter NP, Hunt SE, Chen R, Cree A, Gunaratne P, *et al.*. The DNA sequence of the human X chromosome. *Nature*, 434:325-337, 2005.

Rossetto D, Avvakumov N, Côté J. Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics*, 7(10):1098-1108, 2012.

Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S, Albala JS, Lim J, Fraughton C, Llamas E, Cevik S, Bex C, Lamesch P, Sikorski RS, Vandenhaute J, Zoghbi HY, Smolyar A, Bosak S, Sequerra R, Doucette-Stamm L, Cusick ME, Hill DE, Roth FP, Vidal M. Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, 437(7062):1173-1178, 2005.

Russell LB. Mammalian X-chromosome action: inactivation limited in spread and region of origin. *Science*, 140(3570):976-978, 1963.

Russell LB. *Another look at the single-active-X hypothesis*. Trans NY Acad Sci, 26:726–736, 1964.

Sadakata T, Washida M, Iwayama Y, Shoji S, Sato Y, Ohkura T, Katoh-Semba R, Nakajima M, Sekine Y, Tanaka M, Nakamura K, Iwata Y, Tsuchiya KJ, Mori N, Detera-Wadleigh SD, Ichikawa H, Itohara S, Yoshikawa T, Furuichi T. Autistic-like phenotypes in Cadps2-knockout mice and aberrant CADPS2 splicing in autistic patients. *J Clin Invest*, 117(4):931-943, 2007.

Sadakata T, Shinoda Y, Oka M, Sekine Y, Sato Y, Saruta C, Miwa H, Tanaka M, Itohara S, Furuichi T. Reduced axonal localization of a Caps2 splice variant impairs axonal release of BDNF and causes autistic-like behavior in mice. *Proc Natl Acad Sci U S A*, 109(51):21104-21109, 2012.

Sado T, Fenner MH, Tan SS, Tam P, Shioda T, Li E. X-inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X-inactivation. *Dev Biol*, 225:294–303, 2000.

Sakabe K, Wang Z, Hart GW. Beta-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proc Natl Acad Sci USA*, 107(46):19915-19920, 2010.

Sansom OJ, Berger J, Bishop SM, Hendrich B, Bird A, Clarke AR. Deficiency of Mbd2 suppresses intestinal tumorigenesis. *Nat Genet*, 34(2):145-147, 2003.

Scanlan MJ, Gordan JD, Williamson B, Stockert E, Bander NH, Jongeneel V, Gure AO, Jäger D, Jäger E, Knuth A, Chen YT, Old LJ. Antigens recognized by autologous antibody in patients with renal-cell carcinoma. *Int J Cancer*, 83(4):456-464, 1999.

Schlabach MR, Luo J, Solimini NL, Hu G, Xu Q, Li MZ, Zhao Z, Smogorzewska A, Sowa ME, Ang XL, Westbrook TF, Liang AC, Chang K, Hackett JA, Harper JW, Hannon GJ, Elledge SJ. Cancer proliferation gene discovery through functional genomics. *Science*, 319(5863):620-624, 2008.

Schoch S, Deák F, Königstorfer A, Mozhayeva M, Sara Y, Südhof TC, Kavalali ET. SNARE function analyzed in synaptobrevin/VAMP knockout mice. *Science*, 294(5544):1117-1122, 2001.

Sharman GB. Late DNA replication in the paternally derived X chromosome of female kangaroos. *Nature*, 230(5291):231-232, 1971.

Shen Y, Matsuno Y, Fouse SD, Rao N, Root S, Xu R, Pellegrini M, Riggs AD, Fan G. X-Inactivation in Female Human Embryonic Stem Cells is in a Nonrandom Pattern and Prone to Epigenetic Alterations. *PNAS*, 105(12): 4709-4714, 2008.

Shen L, Zhang Y. 5-Hydroxymethylcytosine: generation, fate, and genomic distribution. *Curr Opin Cell Biol*, 25(3):289-296, 2013.

Shi Y, Manley JL. A complex signaling pathway regulates SRp38 phosphorylation and pre-mRNA splicing in response to heat shock. *Mol Cell*, 28(1):79-90, 2007.

Shiio Y, Eisenman RN. Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A*, 100(23):13225-13230, 2003.

Shin C, Manley JL. The SR protein SRp38 represses splicing in M phase cells. *Cell*, 111(3):407-417, 2002.

Shin C, Feng Y, Manley JL. Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature*, 427(6974):553-558, 2004.

Shin C, Kleiman FE, Manley JL. Multiple properties of the splicing repressor SRp38 distinguish it from typical SR proteins. *Mol Cell Biol*, 25(18):8334-8343, 2005.

Silva SS, Rowntree RK, Mekhoubad S, Lee JT. X-Chromosome Inactivation and Epigenetic Fluidity in Human Embryonic Stem Cells. *PNAS*, 105(12): 4820-4825, 2008.

Silva JM, Marran K, Parker JS, Silva J, Golding M, Schlabach MR, Elledge SJ, Hannon GJ, Chang K. Profiling essential genes in human mammary cells by multiplex RNAi screening. *Science*, 319(5863):617-620, 2008.

Simmler MC, Heard E, Rougeulle C, Cruaud C, Weissenbach J, Avner P. Localization and expression analysis of a novel conserved brain expressed transcript, Brx/BRX, lying within the Xic/XIC candidate region. *Mamm Genome*, 8(10):760-766, 1997.

Sims D, Mendes-Pereira AM, Frankum J, Burgess D, Cerone MA, Lombardelli C, Mitsopoulos C, Hakas J, Murugaesu N, Isacke CM, Fenwick K, Assiotis I, Kozarewa I, Zvelebil M, Ashworth A, Lord CJ. High-throughput RNA interference screening using pooled shRNA libraries and next generation sequencing. *Genome Biol*, 12(10):R104, 2011.

Smith BC, Denu JM. Chemical mechanisms of histone lysine and arginine modifications. *Biochim Biophys Acta*, 1789(1):45-57, 2009.

Stabellini R, de Mello JC, Hernandez LM, Pereira LV. MAOA and GYG2 are submitted to X chromosome inactivation in human fibroblasts. *Epigenetics*, 4(6):388-393, 2009.

Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S, Timm J, Mintzlaff S, Abraham C, Bock N, Kietzmann S, Goedde A, Toksöz E, Droege A, Krobitsch S, Korn B, Birchmeier W, Lehrach H, Wanker EE. A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, 122(6):957-968, 2005.

Stout JT, Caskey CT. HPRT: gene structure, expression, and mutation. *Annu Rev Genet*, 19:127-48, 1985.

Strezoska Ž, Licon A, Haimes J, Spayd KJ, Patel KM, Sullivan K, Jastrzebski K, Simpson KJ, Leake D, van Brabant Smith A, Vermeulen A. Optimized PCR conditions and increased shRNA fold representation improve reproducibility of pooled shRNA screens. *PLoS One*, 7(8):e42341, 2012.

Sun S, Del Rosario BC, Szanto A, Ogawa Y, Jeon Y, Lee JT. Jpx RNA activates Xist by evicting CTCF. *Cell*, 153(7):1537-1551, 2013.

Szenker E, Ray-Gallet D, Almouzni G. The double face of the histone variant H3.3. *Cell Res*, 21(3):421-434, 2011.

Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell*, 116(1):51-61, 2004.

Takagi N, Sasaki M. Preferential Inactivation of the Paternally Derived X Chromosome in the Extraembryonic Membranes of the Mouse. *Nature*, 256: 640-642, 1975.

Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A*, 99(6):3740-3745, 2002.

Tian D, Sun S, Lee JT. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell*, 143(3):390-403, 2010.

Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol*, 7 Suppl 1:S12.1-14, 2006.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. Embryonic stem cell lines derived from human blastocysts. *Science*, 282(5391):1145-1147, 1998.

Tsuchiya KD, Gready JM, Yi Y, Noel KP, Truong JP, Distech CM. Comparative sequence and x-inactivation analyses of a domain of escape in human Xp11.2 and the conserved segment in mouse. *Genome Res*, 14:1275-1284, 2004.

Vasques LR, Pereira LV. Allele-specific X-linked gene activity in normal human cells assayed by expressed single nucleotide polymorphisms (cSNPs). *DNA Res*, 8(4):173-177, 2001.

Vasques LR, Stabellini R, Xue F, Tian XC, Soukoyan M, Pereira LV. XIST repression in the absence of DNMT1 and DNMT3B. *DNA Res*, 12(5):373-378, 2005.

Wagner PK, Otomo A, Christians JK. Regulation of pregnancy-associated plasma protein A2 (PAPPA2) in a human placental trophoblast cell line (BeWo). *Reprod Biol Endocrinol*, 9:48, 2011.

Wang H, Zhou Z, Lu L, Xu Z, Sha J. Cloning and characterization of a novel intronless lactate dehydrogenase gene in human testis. *Int J Mol Med*, 15(6):949-953, 2005.

Wang J, Qiu Q, Haider M, Bell M, Gruslin A, Christians JK. Expression of pregnancy-associated plasma protein A2 during pregnancy in human and mouse. *J Endocrinol*, 202(3):337-345, 2009.

Watson JM, McKay LM, Marshall Graves JA. Genome instability in interspecific cell hybrids II. Aminopterin resistance and gene amplification in lines arising from fusions of cells from divergent mammalian species. *J Genet*, 67(2):75-86, 1988.

Webby CJ, Wolf A, Gromak N, Dreger M, Kramer H, Kessler B, Nielsen ML, Schmitz C, Butler DS, Yates JR 3rd, Delahunty CM, Hahn P, Lengeling A, Mann M, Proudfoot NJ, Schofield CJ, Böttger A. Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science*, 325(5936):90-93, 2009.

Weimann M, Grossmann A, Woodsmith J, Özkan Z, Birth P, Meierhofer D, Benlasfer N, Valovka T, Timmermann B, Wanker EE, Sauer S, Stelzl U. A Y2H-seq approach defines the human protein methyltransferase interactome. *Nat Methods*, 10(4):339-342, 2013.

Welshons WJ, Russell LB. The Y-chromosome as the bearer of male determining factors in the mouse. *Proc Natl Acad Sci USA*, 45:560–566, 1959.

Westbrook TF, Martin ES, Schlabach MR, Leng Y, Liang AC, Feng B, Zhao JJ, Roberts TM, Mandel G, Hannon GJ, Depinho RA, Chin L, Elledge SJ. A Genetic Screen for Candidate Tumor Suppressors Identifies REST. *Cell*, 121(6):837-848, 2005.

White MA, Ikeda A, Payseur BA. A pronounced evolutionary shift of the pseudoautosomal region boundary in house mice. *Mamm Genome*, 23(7-8):454-466, 2012.

Witt O, Albig W, Doenecke D. Transcriptional regulation of the human replacement histone gene H3.3B. *FEBS Lett*, 408(3):255-260, 1997.

Wolf SF, Migeon BR. Studies of X chromosome DNA methylation in normal human cells. *Nature*, 295(5851):667-671, 1982.

Wong LH, McGhie JD, Sim M, Anderson MA, Ahn S, Hannan RD, George AJ, Morgan KA, Mann JR, Choo KH. ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome Res*, 20(3):351-360, 2010.

Wright MW, Bruford EA. Naming 'junk': human non-protein coding RNA (ncRNA) gene nomenclature. *Hum Genomics*, 5(2):90-98, 2011.

Wright DE, Wang CY, Kao CF. Histone ubiquitylation and chromatin dynamics. *Front Biosci (Landmark Ed)*, 17:1051-1078, 2012.

Xiao PJ, Peng ZY, Huang L, Li Y, Chen XH. Dephosphorylated NSSR1 is induced by androgen in mouse epididymis and phosphorylated NSSR1 is increased during sperm maturation. *PLoS One*, 6(9):e25667, 2011.

Xu N, Tsai CL, Lee JT. Transient Homologous Chromosome Pairing Marks the Onset of X Inactivation. *Science*, 311: 1149-1152, 2006.

Xu N, Donohoe ME, Silva SS, Lee JT. Evidence that homologous X-chromosome pairing requires transcription and Ctf protein. *Nat Genet*, 39(11):1390-1396, 2007.

Xu S, Pei R, Guo M, Han Q, Lai J, Wang Y, Wu C, Zhou Y, Lu M, Chen X. Cytosolic phospholipase A2 gamma is involved in hepatitis C virus replication and assembly. *J Virol*, 86(23):13025-13037, 2012.

Xue F, Tian XC, Du F, Kubota C, Taneja M, Dinnyes A, Dai Y, Levine H, Pereira LV, Yang X. Aberrant patterns of X chromosome inactivation in bovine clones. *Nat Genet*. 2002 Jun;31(2):216-20. Epub 2002 May 28.

Yamashita A, Tanaka K, Kamata R, Kumazawa T, Suzuki N, Koga H, Waku K, Sugiura T. Subcellular localization and lysophospholipase/transacylation activities of human group IVC phospholipase A2 (cPLA2gamma). *Biochim Biophys Acta*, 1791(10):1011-1022, 2009.

Yan X, Baxter RC, Firth SM. Involvement of pregnancy-associated plasma protein-A2 in insulin-like growth factor (IGF) binding protein-5 proteolysis during pregnancy: a potential mechanism for increasing IGF bioavailability. *J Clin Endocrinol Metab*, 95(3):1412-1420, 2010.

Yang C, Chapman AG, Kelsey AD, Minks J, Cotton AM, Brown CJ. X-chromosome inactivation: molecular mechanisms from the human perspective. *Hum Genet*, 130(2):175-185, 2011.

Yue CM, Deng DJ, Bi MX, Guo LP, Lu SH. Expression of ECRG4, a novel esophageal cancer-related gene, downregulated by CpG island hypermethylation in human esophageal squamous cell carcinoma. *World J Gastroenterol*, 9(6):1174-1178, 2003.

Zhang R, Poustovoitov MV, Ye X, Santos HA, Chen W, Daganzo SM, Erzberger JP, Serebriiskii IG, Canutescu AA, Dunbrack RL, Pehrson JR, Berger JM, Kaufman PD, Adams PD. Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. *Dev Cell*, 8(1):19-30, 2005.

Zhang LF, Huynh KD, Lee JT. Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell*, 129:693–706, 2007.

Zhang S, Roche K, Nasheuer HP, Lowndes NF. Modification of histones by sugar β -N-acetylglucosamine (GlcNAc) occurs on multiple residues, including histone H3 serine 10, and is cell cycle-regulated. *J Biol Chem*, 286(43):37483-37495, 2011.

Zhang R, Zhang L, Yu W. Genome-wide expression of non-coding RNA and global chromatin modification. *Acta Biochim Biophys Sin (Shanghai)*, 44(1):40-47, 2012.

Zempleni J, Chew YC, Bao B, Pestinger V, Wijeratne SS. Repression of transposable elements by histone biotinylation. *J Nutr*, 139(12):2389-2392, 2009.

Anexo 1.

Resumo do Plano Experimental de NGS Adotado pela Empresa Fasteris

Date / ref.: CP-2012-06-29_Ligation-on-PCR

HiSeq service

Custom Protocol: Ligation on PCR

Introduction:

High-throughput DNA sequencing using Illumina technology:

Samples are processed *in vitro* to generate a library of short inserts, the "DNA Colonies Template Library".

Having passed quality control, the library is sequenced on the Illumina HiSeq 2000.

For each lane we guarantee 130-150 million DNA Colonies producing pass filter sequences.

The read lengths are 1x50 or 1x100 bp for **single-reads** runs using the "forward" sequencing primer.

The inserts can also be sequenced from both ends using a "forward" and a "reverse" sequencing primer, generating **paired-reads** of 2x100 bp.

The data is then processed using bioinformatics tools to extract biologically useful information.

NB: Although the HiSeq system delivers high-quality data, it is a young technology and we recommend independent validation of the data obtained.

Fasteris is the first facility certified by Illumina for sequencing applications (2008)

Requirements:

Please provide:

- Purified double stranded genomic DNA PCR product: **100-200ng** in a maximum of 50 uL.
- Sample concentration measured with Qubit (Nanodrop often overestimates concentrations);
- Volume (**max 50 uL**);
- Copy of quotation and sample-sheet with the samples.

Samples can be sent at room temperature in a bubble envelope.

Shipping address:

FASTERIS SA
HiSeq Services
Ch. du Pont-du-Centenaire 109
CH-1228 Plan-les-Ouates
Switzerland

Tel: +41-22-794 22 23
ht_seq@fasteris.com

Sample preparation processing steps – Custom Protocol:

General comment: We may be using a modified protocol, e.g. Fasteris-designed indexed adapters.

1. Initial Quality Control;
*If the sample fails the initial QC, another one is done free of charge.
If no library is done, CHF 100.- will be charged for the initial QC.*
2. Ends repair;
3. 3' A addition;
4. Ligation of adapters;
5. PCR amplification to generate the DNA Colonies Template Library,
library purification;
6. Library quantification and dilution at 10 nM.
7. Library Quality Control, done by titration run on the HiSeq (see note below).
For large orders, Library Quality Control is only performed on selected libraries.
=> We will only proceed with sequencing on the HiSeq if you are satisfied with the quality control results.

Note on titration run: All the libraries to be quality controlled are multiplexed in one lane of a 1x50 bp run on the HiSeq. After demultiplexing of the reads according to the indexes, we typically provide 50 fasta reads and 20'000 fast-q reads from each library. Customers can then blast the fasta sequences to check the quality of the library. In addition, the titration run helps us for quantification of the libraries.

NB: For some libraries, e.g. of low complexity, the sequence quality of the titration run is not always the same as the quality of the full HiSeq run.

HiSeq processing steps:

1. Flow-cell preparation on the cBot;
2. High-throughput DNA sequencing on the illumina HiSeq 2000;
A PhiX control lane normally run on each flow-cell;
A PhiX control library is spiked in each lane.
3. Real-time base-calling on the HiSeq computer.

Time lines:

- The libraries are prepared in 2-3 weeks after the samples have passed the initial quality control.
For large orders, this applies only to the first samples.

- Run on the HiSeq is booked on next available run when libraries have passed quality control. Runs normally start within 2 weeks for 1x50 or 1x100 and 3-4 weeks for 2x100.
NB: runs of 2x50 bases are now abandoned, unless a full flow-cell is ordered (7 lanes).
- Delivery may be delayed, e.g. if quality control is not satisfactory, if the HiSeq has to undergo maintenance or if illumina has difficulties delivering the kits.
- Optional: for urgent project, additional fees (+20%) will be added to the whole price

We deliver:

- *If applicable:* The Library Quality Control results.
- The millions of reads generated on the HiSeq.
 - The sequences are provided under illumina fast-Q format (i.e. with quality information).
 - These very large files are delivered on Internet. If shipping of external hard drive is requested, a fee will be charged.
 - The reads are sorted according to the indexes. For non-multiplexed libraries, this permits removing the reads from the spiked PhiX control library.

NB: Should you need the intensity files, please inform us as soon as possible, as the intensity files will be **deleted** at the end of the run.

Bioinformatics analyses:

Example of typical analyses, details to be defined with our bioinformaticians.

Option A. Standard analyses: reference genome available:

- Mapping the reads on **standard** reference genomes, e.g. human, mouse, etc.;
- BAM files for graphical display of the mapped reads;
- Coverage information;
- List of SNPs, filtered according to Fasteris-defined criteria;
- List of short InDels (not available for 1x50 bp reads);
- Comparisons between samples.

Option B. Standard analyses: de novo assembly of small genomes:

De novo assembly of the reads into contigs using VELVET or EDENA (developed by Dr. David Hernandez from Prof. Schrenzel's group at the University of Geneva, in collaboration with Fasteris).

Contigs extension is expected to stop if the coverage is not sufficient or if there are ambiguities, e.g. repeated regions.

Minimum coverage required: 40-60x.

Option C. Custom analyses:

- SNPs or short InDels filtered according to user-defined criteria;
- Advanced comparison between samples;
- Finding annotations linked to SNPs or InDels.
- Statistical significance of SNPs.
- Custom analyses.

Option D. Custom analyses: Genomes smaller than 10 MB

Mapping and targeted *de novo* assembly using DNA STAR's SeqMan NGen.

Data is provided as SeqMan file that can be analyzed on a Windows desktop computer, e.g. to select SNPs.

Anexo 2.

Descrição do Formato fastq das Sequências Obtidas por NGS Fornecida pela Empresa Fasteris

Sequence format and quality values

(pipeline CASAVA.1.8)

fastq format

4 lines for each sequence:

- 1- Unique identifier, with the following format:
`@<machine_id>:<Run number>:<Flow-cell ID>:<lane>:<tile>:<x_coord>:<y_coord>
<read>:<filtered>:<control number>:<index sequence>`
- 2- Sequence (A, T, C, G or N) (N stands for undetermined)
- 3- Orientation (always forward without mapping)
- 4- Quality value for each base, corresponding to a Phred-like score encoded in ASCII format. With CASAVA 1.8, the offset is of 33 (e.g. "J" gives a value of 41) and is in accordance with Sanger FASTQ format.

example:

```
@HWI-ST365:196:C03F2ABXX:2:1101:14893:2232 1:N:0:TCCAGA
TGAAATGCAAGTGAAATTAGGCCTGTAGGCACCATCAATAGANCGGAAGA
+
CCCCFFFFFHGHHHIJJJJJJJJJJIDHJIHGJIIJJJJIIHG#0??DD@H
```

N.B.: For previous CASAVA pipeline versions, until CASAVA.1.7, illumina had its own fastq format (e.g. ASCII minus 64). With CASAVA.1.8, illumina has adopted Sanger FASTQ format.

PhiX quality score

In fastq file format, a Phred-like quality score is associated to each base of the sequence. It is calculated to reflect the probability that the base is called erroneously.

In illumina's quality scoring, the quality score is typically quoted as QXX, where the "XX" is the score, and it means that a particular call has a probability of error (P_e) of $P_e = 10^{-(XX/10)}$. In other terms, the quality is $QXX = -10 \times \log_{10} P_e$.

e.g. Q10 is associated to an expected error rate of 1 in 10 (10%) when
Q30 is associated to an expected error rate of 1 in 1000 (0.1%)

The highest base quality value is determined according to the algorithms that generate the matrices for the base-calling, and has been modified with the different releases and updates of the base-calling pipelines (see the table below summarizing the main changes). Currently, the highest base quality value is 41 and is encoded by the "J" character.

Illumina promotes the use of the quality of the reads according to the percentage of bases having a base quality greater or equal than Q30.

At FASTERIS, we use an indexed PhiX library for spiking in every lane of the flow-cell. This allows the computation of the sample library independent quality scoring for each lane (sample libraries are also indexed which permits the removal of PhiX reads from the sample libraries).

Our base-calling quality specifications is based on the quality of the spiked PhiX (see note below)

HiSeq: The current FASTERIS specifications are based on the PhiX spiked on the lane:

<u>Read Length</u>	<u>Bases having a quality score greater or equal to Q30</u>
50 bases	≥85%
100 bases	≥80%

MiSeq: The current FASTERIS specifications are based on the PhiX spiked on the lane:

<u>Read Length</u>	<u>Bases having a quality score greater or equal to Q30</u>
up to 100 bases	≥85%
150 bases	≥80%
250 bases	≥70%

N.B.: In some cases, the calculated quality score of the spiked PhiX might be different from the quality score of other libraries in the same lane. A difference in the size of the DNA colonies (insert length) and/or a difference in GC-content and/or the presence of secondary structures can influence the base-calling and possibly lead to a difference between PhiX quality score and other libraries' quality scores.

Table: main changes on quality score values over the pipeline versions

Quality scoring type	Solexa	Illumina				Sanger	
RTA pipeline version				1.10	1.12	1.12	1.17
CASAVA pipeline version	previous to 1.3	1.3 – 1.5	1.6 – 1.7	1.7	1.7	1.8	1.8
Highest quality value	40	35	38	40	41	41	41
Character associated	h	b	f	h	i	J	J

Control lane

HiSeq high output run:

For most runs, a lane is loaded with the PhiX control library. The control lane can be set on the PhiX lane (by default) or on a lane loaded with a homogeneous genomic library.

The control lane's base-calling parameters are used to guide the base-calling of other lanes.

In addition, the quality values of this lane are used to assess the run overall sequencing quality (i.e. library independent sequencing control).

PhiX error rate

At FASTERIS, we use an indexed PhiX library for spiking in every lane of the flow-cell. This allows the computation of the error rate for each lane by mapping the de-multiplexed PhiX reads (sample libraries are also indexed which permits the removal of PhiX reads from the sample libraries).

The PhiX reads are mapped on the PhiX reference genome. The error rate is then estimated by the number of mismatches over the total number of bases of mapped PhiX reads.

The current FASTERIS specifications apply to the PhiX spiked library (internal control) if present, or to the PhiX control lane (in case of HiSeq run).

HiSeq (based on the PhiX spike)

<u>Error rate below</u>	<u>Read length</u>
0.5%	50 bases
1.5%	100 bases

MiSeq (based on the PhiX spike)

<u>Error rate below</u>	<u>Read length</u>
0.5%	50 bases
1.0%	100 bases
1.5%	150 bases
2.5%	250 bases

N.B.: In some cases, the calculated error rate of the spiked PhiX might be different from the other libraries in the same lane. A difference in the size of the DNA colonies (insert length) and/or a difference in GC-content and/or the presence of secondary structures can influence the base-calling and possibly lead to a difference between PhiX error rate and other libraries' error rates.

We observe improvement on the error rates with each new version of the base-calling pipeline and sequencing kits.

Anexo 3.

***Data Report* referente à corrida final de NGS fornecido pela empresa Fasteris**

HiSeq Services
Data Report
GZR-1-2

Customer		Fasteris	
Institution	IB-USP	Project	GZR-1-2
		Date / ref.	DR_2013-06-24_GZR-1-2
Contact	Naja Vergani	Order ref.	SO-2013-03-04-561-SO-2013-03-04-574_Vergani_Custom-LigationonPCR
e-mail	nvergani@gmail.com	Author	Loïc Baerlocher

Summary	
Sequenced libraries	GZR-1-2
Number of libraries	2
Sample type	PCR product
Run ID	130515_SN365_A
Number of cycle	1x100+7
Instrument	HiSeq
Yield(MB)	2'113
Quality control	Within specifications
Author's comments: The number of reads obtained is not very high. The ratio between the barcode are good.	

1 Introduction

High-throughput DNA sequencing using Illumina technology:

Samples are processed *in vitro* to generate a library of short inserts, the DNA Colonies Template Library.

Having passed quality control, the library is sequenced on the Illumina **HiSeq 2000**.

For each lane we guarantee 130-150 million DNA Colonies producing pass filter sequences.

The read lengths are 1x50 or 1x100 bp for **single-reads** runs using the forward sequencing primer.

The inserts can also be sequenced from both ends using a forward and a reverse sequencing primer, generating **paired-reads** of 2x100 bp.

The data is then processed using bioinformatics tools to extract biologically useful information.

NB: Although the HiSeq system delivers high-quality data, it is a young technology and we recommend independent validation of the data obtained.

Fasteris is the first facility certified by Illumina for sequencing applications (2008)

2 Samples

Sample name	Sample type	Library ID	Library type	Index
6TG	PCR product	GZR-1	Custom	CTGAAC (Fasteris)
HAT	PCR product	GZR-2	Custom	GCGGTC (Fasteris)

Table 1: Description of the samples

3 Sequencing specifications

3.1 Instrument	
Serial number	SN365
Manufacturer	Illumina
Version	HiSeq 2500
Slot used	A
Basecalling pipeline	- HiSeq Control Software 2.0.5 - RTA 1.17.20.0 - CASAVA-1.8.2
3.2 Run	
Run ID	130515_SN365_A
Number of cycles	1x100+7
Number of lanes	1
Flow cell ID	C20U1ACXX
Flow cell version	HiSeq Flow Cell v3
Kit version	TruSeq SBS Kit v3

4 Basecalling summary

4.1 Parameters

Lane	Expected lane %	Number of mismatch in the index selection	Phix spiked*	Other multiplexed libraries**
8	20	1	Yes	Yes

Table 2: Basecalling parameters

* Indicates if a PhiX reference is spiked (i.e. a low concentration of a PhiX library is added before sequencing and selectively retrieved through its related index) in your lane to estimate the error rate for your sequences.

** Indicates whether libraries from other projects are present in the lane

The base calling pipeline proceeds to the demultiplexing prior to the generation of fast-q sequence files, i.e. by separating the libraries according to their indexes.

4.2 Results

The sequences are attributed according to their index code (6 bases).

The data set provided contains only sequences attributed without ambiguity to only one library.

Lane	Expected lane %	Library ID	Sample name	Yield (Mb)	%PF*	Cluster (PF)	Q30**	Mean qual. (PF)
8	10	GZR-1	6TG	873	86.47	8'733'500	93.81	36.61
8	10	GZR-2	HAT	1'240	87.48	12'399'284	94.07	36.70

* PF stands for 'passed filter' i.e. clusters that fulfill the default Illumina quality criteria

** % of bases (PF) with a quality score greater or equal to 30

4.3 Quality control

In order to evaluate the quality of the sequencing, we perform a quality control for each lane independently.

In addition to the error of the spiked PhiX, Illumina evaluates the quality of the reads according to the percentage of bases having a base quality greater or equal to 30 (Q30).

Lane	Error rate %	Q30 %	Within specifications
8	0.27	91.29	Yes

N.B.: We do not perform quality filters on sequence files that are within our quality specifications, besides the default 'failed chastity' filter done by the pipeline itself. We have observed that additional filters can alter the representation/variability of the sequences (e.g. specific removal of sequences with secondary structures).

The base-calling pipeline can sometimes call bases as blanks ('N'). Blanks correspond to unattributed bases. If the blank rate is lower than 0.05%, no further information is provided.