Joana Carvalho Moreira de Mello

Estudos da inativação do cromossomo X em
humanos: iniciação e *imprinting*

X-chromosome inactivation in humans: initiation
and imprinting

São Paulo

2015

Joana Carvalho Moreira de Mello

Estudos da inativação do cromossomo X em
humanos: iniciação e *imprinting*

X-chromosome inactivation in humans: initiation
and imprinting

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

Orientadora: Dra. Lygia da Veiga Pereira
Carramaschi

São Paulo

2015

Este trabalho foi realizado com auxílio financeiro das agências de fomento:

Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP),
Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq),
Fundação Coordenação de Pessoal de Nível Superior (CAPES-PROEX)

Ficha Catalográfica

Moreira de Mello, Joana Carvalho
Estudos da inativação do
cromossomo X em humanos: iniciação e
imprinting
142 páginas

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

1. Inativação do cromossomo X 2.
epigenética 3. *imprinting* 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.

Orientadora

À minha família, por estarem sempre ao meu lado.

Agradecimentos

Gostaria de agradecer a todos que contribuíram com o desenvolvimento desse trabalho e com o meu desenvolvimento científico, intelectual e emocional. Isso inclui colaboradores, colegas, amigos e familiares, obrigada!

Agradeço à minha orientadora, Profa. Dra. Lygia Pereira por enfrentar comigo o desafio que este projeto nos trouxe, confiando em mim por todos esses anos, me dando toda a liberdade para trabalhar e me ajudando a buscar soluções. Obrigada, Lygia.

Agradeço à Profa. Dra. Maria Vibranovski, que aceitou me co-orientar, colaborando com extrema dedicação para a realização deste projeto. Sem sua ajuda não teríamos chegado tão longe. Obrigada, Maria.

Agradeço a Profa. Dra. Wendy Robinson que me entregou as chaves de seu laboratório e à toda a sua equipe. Em especial à Maria Peñaherrera e à Ruby Jiang.

Agradeço aos Drs. Noboru Sakabe e Gustavo Ribeiro pela infinita paciência que tiveram em me ajudar ao explicar detalhes da Bioinformática e da estatística.

Um agradecimento muito especial à profa. Dra. Iscia Teresinha Lopes Cendes e à pesquisadora Dra. Danyella Dogini, ambas do laboratório de genética molecular da FCM-UNICAMP, que muito gentilmente cederam amostras de cérebros humanos. Sei como são preciosas, obrigada.

Agradeço à Profa. Dra. Estela Bevilacqua e sua equipe que me ajudou com as amostras de placentas e me acolheu com tanto carinho.

Sou muito grata também à profa. Dra. Tatiana Torres e ao Dr. Jorge E. S. Souza pelas discussões sugestões e opiniões tão pertinentes.

Agradeço também aos fóruns de bioinformática e à toda comunidade que deles participam fazendo e respondendo perguntas, dando dicas e explicando minuciosamente os detalhes ocultos da bioinformática.

Aos professores do departamento de Genética e Biologia Evolutiva deste Instituto, em especial à prfa. Dra. Angela Morgante, ao prof. Dr. Paulo Otto, e ao prof, Dr. Luis Netto, prfa. Dra. Regina Mingroni e prof. Dr. Eduardo Gorab.

Aos funcionários e colegas deste mesmo departamento, Fátima, Mara, Maraísa, Deyse e Helenice, Lilian, Fernando, Juliana, Débora.

Aos meus colegas de laboratório, passados e presentes, Giovana, Luis, Max, Gustavo, Patrícia, Juliana, Elisa, Lucas, Fabiano, Mari, Gisele, Fernando, Sofia, Fernanda, Simone, Roberta, Luciana, Claudia e Amanda. Em especial minhas amigas e mentoras, Raquel, Ana, Erica e Rafa.

Aos colegas de trabalho em outras terras Ilse, Cindy, Marjon e claro Dra Hikke.

E finalmente à minha família, sobretudo ao Ruggero por todo o seu amor e sua paciência. Mas agradeço especialmente aos meus pais, aos quais tenho grande admiração. Amantes do fascinante mundo da epigética, desde o meu mestrado, sempre interessados em minhas descobertas e me apoiando em todas as minhas decisões.

Índice

Lista de Tabelas	5
Lista de figuras.....	6
Capítulo I	8
ESTUDOS SOBRE A INATIVAÇÃO DO CROMOSSOMO X: INÍCIAÇÃO E <i>IMPRINTING</i>	8
1. INTRODUÇÃO GERAL.....	9
2. JUSTIFICATIVAS.....	14
3. OBJETIVOS GERAIS	15
Capítulo II	16
ATIVIDADE TRANSCRICIONAL DO CROMOSSOMO X HUMANO EM ESTÁGIOS PRÉ-IMPLANTACIONAIS, UMA ANÁLISE <i>IN SILICO</i>	16
1. INTRODUÇÃO	17
2. OBJETIVOS ESPECÍFICOS	21
3. MATERIAL E MÉTODOS	22
3.1. Obtenção dos dados	22
3.2. Bancos de dados:	23
3.3. Plano geral do processo de alinhamento e identificação das variantes	25
3.4. Seleção dos SNPs	34
3.5. Determinação do sexo cromossômico das amostras.....	34
3.6. Expressão de <i>XIST</i>	35
3.7. Genotipagem	36
3.8. Determinação do padrão de inativação do cromossomo X.....	37
3.9. Razão entre o nível de expressão gênica dos genes do cromossomo X e aquela dos autossomos	39
4. RESULTADOS E DISCUSSÃO.....	40
4.1. Qualidade, alinhamento das reads e variantes anotadas	40
4.2. Determinação do sexo cromossômico das amostras.....	46
4.3. Expressão de <i>XIST</i>	53
4.4. Desvios de alinhamento em favor do alelo de referência	56
4.5. Inativação do cromossomo X pela análise de expressão alelo específica.....	57
5. Razão da expressão global entre genes do cromossomo X e autossômicos.....	70
5. CONCLUSÕES	79
Capítulo III	82
GENES REGULADOS POR <i>IMPRINTING</i> NO CROMOSSMO X HUMANO	82
1. INTRODUÇÃO.....	83
2. Objetivos específicos	88
3. Material e métodos	89
3.1. Laboratórios.....	89

3.2. Seleção dos polimorfismos e síntese dos iniciadores	89
3.3. Obtenção das amostras	91
3.4. Obtenção dos ácidos nucleicos	92
3.5. Genotipagem do DNA genômico e verificação do alelo expresso	94
4. RESULTADOS E DISCUSSÃO.....	97
4.1. Análise de expressão alelo-específica em placenta humana a termo	97
4.1.1. Genotipagem	97
4.2. Análise de expressão alelo-específica em cérebro humano:	107
5. CONCLUSÃO	111
Conclusões gerais	113
Resumo	115
Abstract	117
Referências bibliográficas.....	119
Endereços na internet	128
Anexos	129

Lista de Tabelas

Tabela 1 Número de embriões e células utilizados nos trabalhos de Xue <i>et al.</i> e Yan <i>et al.</i> (2013).....	23
Tabela 2 Média por estágio embrionário dos resultados de sequenciamento, de filtragem, dos alinhamentos e dos SNPs anotados.....	40
Tabela 3 Total de variantes anotadas por tipo cromossômico.....	44
Tabela 4 Resultado do processo de determinação do sexo cromossômico.....	51
Tabela 5 Genotipagem dos embriões femininos por cromossomo.....	60
Tabela 6 Comparação entre o número de genes com expressão monoalélica no cromossomo X e nos autossomos.....	62
Tabela 7 Genes e polimorfismos selecionados.....	90
Tabela 8 Amostras informativas após genotipagem de material genômico extra-embrionário e parental.....	97
Tabela 9 Alelo expresso determinado por sequenciamento direto tipo Sanger das amostras com um fragmento coletado.....	100
Tabela 10 Alelo expresso determinado por sequenciamento direto tipo Sanger das amostras com mais de um segmento coletado.....	101
Tabela 11 Alelo expresso determinado por sequenciamento direto tipo Sanger das amostras de cérebro humano.....	107

Lista de figuras

Figura 1 Diagrama do plano de trabalho.....	26
Figura 2 Gene e três possíveis transcritos hipotéticos com SNP identificado no exon 3.....	32
Figura 3 Gene hipotético e resultados de dois alinhamentos independentes tendo como base diferentes referências.....	33
Figura 4 Resultados hipotéticos representando os padrões esperados de inativação do cromossomo X ao longo do desenvolvimento extra-embrionário.....	38
Figura 5 Número de <i>reads</i> em cada fase do processo de alinhamento.....	41
Figura 6 Número de variantes anotadas.....	45
Figura 7 Proporção de genes autossômicos expressos.....	48
Figura 8 Determinação do sexo cromossômico dos embriões através da contagem de genes expressos mapeados no cromossomo Y.....	50
Figura 9 Expressão de <i>XIST</i> ao longo do desenvolvimento embrionário humano.....	53
Figura 10 Genotipagem através da identificação de variantes em regiões expressas.....	59
Figura 11 Genes informativos de expressão monoalélica e bialélica.....	61
Figura 12 Análise do estabelecimento e do padrão de inativação do cromossomo X durante o desenvolvimento pré-implantacional humano.....	65
Figura 13 Número total de <i>reads</i> alinhadas em relação ao estágio do desenvolvimento.....	69
Figura 14 Razão da expressão global dos genes do cromossomo X pela dos autossômicos por embrião.....	71
Figura 15 Razão da expressão global dos genes do cromossomo X pela dos autossômicos para cada célula.....	72
Figura 16 Verificação da presença de células aneuploides nos embriões em estágio de mórula.....	76
Figura 17 Eletroferogramas representativos do caso pl.40.....	102
Figura 18 Curva padrão construída com os valores obtidos por pirosequenciamento.....	104

Figura 19 Origem parental do alelo expresso do gene <i>MAOA</i> em placenta a termo.....	105
Figura 20 Eletroferogramas representativos do sequenciamento de cDNA das amostras de cérebro.....	108

Capítulo I

ESTUDOS SOBRE A INATIVAÇÃO DO CROMOSSOMO X: INÍCIAÇÃO E *IMPRINTING*

A definição contemporânea de epigenética faz referência a mecanismos bioquímicos de modificações no DNA que são transmitidas mitoticamente sem, no entanto, alterar sua sequência original de nucleotídeos. Todas essas mudanças bioquímicas têm uma substancial influência na estrutura da cromatina e função gênica, que difere de acordo com o tipo e localização da modificação. A reprogramação epigenética é o processo pelo qual o genótipo do organismo é modulado pelo ambiente para produzir seu fenótipo. Este fenômeno também ajuda a explicar as variações individuais e a singularidade de células, tecidos e órgãos apesar de possuírem uma informação genética idêntica.

As modificações epigenéticas melhor caracterizadas hoje são a metilação do DNA, as modificações das histonas que compõem os nucleossomos (Jenuwein and Allis, 2001) e ainda a ação de RNAs não-codificadores (ncRNA - do inglês, *non-coding RNA*) (revisto por Richards, 2006)).

Os processos epigenéticos são hoje reconhecidos como fundamentais durante o desenvolvimento e envelhecimento (revisto por Calvanese *at al.*, 2009), e cada vez mais estão associados a mecanismos de doenças como o câncer, doenças auto-imunes e síndromes como a de Prader-Willi e de Angelmann (Ballestar *at al.*, 2006; Butler, 2009; Tang and Ho, 2007). Já foi também demonstrado que mudanças ambientais são capazes de interferir e modificar o padrão de metilação em regiões específicas do DNA, caracterizando o complexo sistema de comunicação entre fatores externos e a epigenética (Postovit *at al.*, 2007).

Os fenômenos epigenéticos mais reconhecidos são:

- aqueles que levam à diferenciação celular durante o desenvolvimento, responsáveis pelo padrão coordenado de controle da expressão de um zigoto indiferenciado a um adulto com diferentes tecidos e tipos celulares (Surani *at al.*, 2007). Uma vez diferenciada, a identidade celular é mantida e cada tipo terá seu padrão único de expressão gênica derivado de seu perfil epigenético, possuindo assim um epigenoma próprio;
- aqueles que determinam o padrão de expressão monoalélico de um gene como a expressão monoalélica aleatória e o imprinting genômico, onde o alelo expresso é determinado de acordo com a origem parental;
- a inativação do cromossomo X, que também resulta em expressão monoalélica dos genes deste cromossomo.

A inativação do cromossomo X

A inativação do cromossomo X (ICX) é um impressionante exemplo de modificação e regulação epigenética nos mamíferos. O evento é responsável por garantir a compensação de dose em relação à transcrição gênica nos cromossomos sexuais entre fêmeas (XX) e machos (XY) (LYON, 1961; OHNO *at al.*, 1959).

O processo tem início nas primeiras etapas do desenvolvimento do embrião feminino, com a transcrição do gene XIST, a partir do futuro X inativo (Xi). Este gene exclusivo de mamíferos eutérios transcreve para um ncRNA e suas moléculas se depositam em *cis* e então recrutam todas as modificações epigenéticas associadas a heterocromatina, incluindo modificações de histona e metilação de DNA deixando o cromossomo num estado transcricional inativo (revisto em (Okamoto *at al.*, 2004).

Um *locus* chave para a ICX é o Centro de Inativação do Cromossomo X (XIC – do inglês, *X Inactivation Centre*) (Brown *at al.*, 1991; Rastan and Cattanach, 1983). Antes de se

estabelecer a inativação, em cada célula do embrião, ocorre uma íntima interação entre os XICs homólogos (Bacher *at al.*, 2006; Xu *at al.*, 2006). Essa interação está diretamente relacionada aos processos de contagem, escolha e iniciação do silenciamento, sendo que cada célula irá inativar todos menos um cromossomo X por genoma diploide.

No XIC de camundongos existe um outro importante gene com atividade relacionada à supressão de *Xist*, mantendo o X ativo (*Xa*), o *Tsix* (do inglês, *X (inactive)-specific transcript, antisense*) (Lee *at al.*, 1999). Este gene também produz um ncRNA e sua sequência apresenta sobreposição à toda a extensão de *Xist* (Lee and Lu, 1999). Apesar de o homólogo humano *TSIX* ter sido descrito em 2001 (Migeon *at al.*, 2001), estudos sobre sua existência e sua função continuam gerando controvérsias (Lee, 2011; Migeon, 2003; Payer and Lee, 2008). Em extensão, o *TSIX* humano não cobre a região promotora de *XIST* (fundamental para sua repressão em camundongos), pois possui uma extremidade 3' truncada, e ainda, suas ilhas CpG (essenciais na expressão de *Tsix*) são incompletas se comparadas às de camundongos (Migeon *at al.*, 2001).

Embora ainda não se tenha evidências claras sobre alguma função envolvendo o *TSIX* humano no processo de ICX, um estudo de 2013 identificou outro ncRNA, sem homólogo em camundongos: o gene *XACT* (do inglês, *X active coating transcript*). Em linhagens de células tronco pluripotentes este gene é expresso a partir do *Xa* e seu RNA também se deposita em *cis*, entretanto ao contrário do que ocorre com *XIST*, a expressão e o acúmulo de *XACT* só pode ser observado em células indiferenciadas em estágio pré-ICX (Vallot *at al.*, 2013).

Como característica de um fenômeno epigenético, o estado inativo do cromossomo X é firmemente mantido através das mitoses, e assim as células pertencentes a uma população clonal devem possuir sempre o mesmo Xi. A estabilidade do silenciamento observado durante as mitoses é resultado de um sinergismo entre múltiplos fatores epigenéticos que caracterizam o estado inativo do cromossomo X (Csankovszki *at al.*, 2001).

A ICX pode ocorrer basicamente de duas maneiras: aleatória, onde cada célula escolhe ao acaso qual será o cromossomo Xi, paterno ou materno; ou de forma dependente da origem parental, “imprintada” (adaptado do inglês, *imprinted*) (Boumil and Lee, 2001). As células somáticas das fêmeas dos mamíferos eutérios são exemplos inativação aleatória, onde qualquer um dos seus dois cromossomos X pode ser inativado (Lyon, 1961). Já as células somáticas de fêmeas marsupiais são um clássico exemplo de inativação “imprintada”, nas quais o X de origem paterna é sempre inativado (Sharman, 1971).

Durante o desenvolvimento embrionário de camundongos a ICX ocorre em duas etapas: a primeira tem início no embrião em estágio de duas células e o X de origem paterna é preferencialmente inativado (ICX “imprintada”). Na fase de blastocisto as células da massa celular interna reativam o cromossomo X paterno e o processo de inativação se inicia novamente naquelas células, desta vez com escolha aleatória do Xi (Mak *at al.*, 2004; Okamoto *at al.*, 2004). Por consequência deste fenômeno, o que se observa em camundongos é a ICX “imprintada” nos tecidos extra-embrionários e aleatória no embrião propriamente dito (Takagi e Sasaki, 1975; West *at al.*, 1977). Em humanos até o momento foi observado apenas o padrão de ICX aleatória tanto em tecidos extra-embrionários quanto no organismo adulto (Amos-Landgraf *at al.*, 2006; Moreira de Mello *at al.*, 2010).

Dados relacionados aos mecanismos de ICX no início do desenvolvimento humano são difíceis de serem obtidos não apenas por causa das restrições éticas feitas a estudos que envolvem embriões humanos, mas também pela dificuldade em se obter e manipular o material. Entretanto dois importantes estudos buscaram determinar em que momento a ICX teria início em embriões em estágio pré-implantacional. Através de experimentos com hibridação fluorescente *in situ* (FISH, sigla em inglês para *fluorescent in situ hybridization*) de RNA van den Berg e colaboradores (2009) e Okamoto e colaboradores (2011) detectaram a presença do RNA de XIST e seu acúmulo se intensificando em embriões a partir do estágio de oito células. Além disso, van den Berg e colaboradores (2009) foram

capazes de detectar expressão monoalélica do gene *CHIC1* no X ativo, enquanto Okamoto e colaboradores (2009) observaram expressão bialélica para os genes *ATRX*, *FGD1* e *HUWE1*. Assim, a extensão do silenciamento transcricional dos genes no cromossomo X no estágio de blastocisto não foi conclusiva.

Em células adultas humanas, nas quais a ICX já se estabeleceu, cerca de 15% dos genes se mantêm transcricionalmente ativos a partir do Xi, mostrando expressão bialélica e outros 10% apresentam padrões de atividade que podem variar entre tecidos e até mesmo entre os indivíduos (Anderson e Brown, 1999; Carrel e Willard, 2005). Muitos desses genes que escapam à ICX estão presentes nas regiões pseudoautosômicas ou apresentam homólogos no cromossomo Y, e assim são expressos de forma bialélica em ambos os gêneros. Já àqueles genes do X que escapam à ICX sem equivalentes funcionais no cromossomo Y, foram atribuídas funções relacionadas com as diferenças entre os gêneros, e podem também explicar as anormalidades clínicas reconhecidas em pacientes com aneuploidias do cromossomo X. Assim, acredita-se que os produtos transcricionais dos genes que escapam à ICX devem ter papel nos fenótipos encontrados em indivíduos com a síndrome de Klinefelter (47,XXY) e a síndrome de Turner (45,X) (Berletch *et al.*, 2011).

Além disso, diferenças observadas no fenótipo de mulheres com síndrome de Turner com diferentes origens parentais do cromossomo X herdado sugerem a existência de regulação por imprinting em algum(ns) dos gene(s) do cromossomo X (revisto por Lee and Bartolomei, 2013). Entretanto, até o momento, não foi descrita a ocorrência de nenhum gene assim regulado no cromossomo X humano.

Existem muitos estudos sobre a inativação do cromossomo X (ICX) durante o início do desenvolvimento em camundongos - em nenhum outro modelo animal o mecanismo é melhor compreendido. As principais razões para tal são a facilidade em se obter embriões em estágio pré-implantacional, e a capacidade que as células tronco embrionárias desses animais têm de recapitular *in vitro* as etapas que envolvem o processo de ICX. Em humanos entretanto, o pouco conhecimento que se tem está baseado em análises de imagem por ensaios de hibridação de RNA *in situ* fluorescente em embriões a partir do estágio de oito células. Estudos com técnicas de sequenciamento em larga escala do RNA total (RNA-Seq) de células únicas são cada vez mais comuns, e em 2013 foram publicados os dados do sequenciamento de células isoladas de embriões humanos desde o início do desenvolvimento pré-implantacional (Xue *et al.*, 2013; Yan *et al.*, 2013). Esses dados, agora públicos, propiciam uma oportunidade única de se estudar o início da ICX em diferentes estágios do desenvolvimento.

O *imprinting* genômico é um fenômeno bastante conhecido que regula a expressão de genes de acordo com a origem parental, sendo a maioria dos genes assim regulados expressos na placenta e no cérebro. Todavia, devido ao fato de que tanto machos (XY) quanto fêmeas (XX, com ICX) apresentam um único X ativo por célula, a caracterização de genes “imprintados” no cromossomo X é desafiadora. Diferenças fenotípicas entre mulheres portadoras da síndrome de Turner (45,X), foram relacionadas à origem parental do cromossomo X herdado apontando para a existência de genes assim regulados (Daniels *et al.*, 1997; Ray *et al.*, 1997). Em particular os genes do cromossomo X *MAOA*, *MAOB* e *USP9X* foram sugeridos como candidatos a serem regulados por *imprinting*. Até o momento, entretanto, não há nenhuma publicação que avalie o padrão de expressão desses genes em placenta e cérebro humanos.

- Verificar quando se dá o início do processo de inativação do cromossomo X no desenvolvimento embrionário humano a partir de dados gerados por sequenciamento em larga escala do RNA total obtido de células únicas;
- Verificar se a escolha do X inativo no início do desenvolvimento humano é aleatória ou dependente da origem parental;
- Testar a hipótese de que os genes MAOA, MAOB e USP9X sejam regulados por imprinting em cérebro e placenta humanos.

Capítulo II

ATIVIDADE TRANSCRICIONAL DO CROMOSSOMO X HUMANO EM
ESTÁGIOS PRÉ-IMPLANTACIONAIS, UMA ANÁLISE *IN SILICO*

Após mais de meio século desde a proposição de Mary Lyon sobre a existência de compensação de dose através da inativação transcricional global de um dos cromossomos sexuais de mamíferos eutérios (Lyon, 1961 e 1962) podemos dizer que muito se avançou no conhecimento das etapas e mecanismos que envolvem o silenciamento transcricional de um cromossomo X em todas as células das fêmeas. Entretanto, a maioria do que se sabe hoje sobre o início do processo de ICX está baseada em estudos envolvendo camundongos, e cada vez mais evidências surgem mostrando que humanos e camundongos diferem substancialmente quanto ao início do processo de inativação e à escolha do X inativo em certas fases do desenvolvimento (Okamoto *at al.*, 2011, Moreira de Mello *at al.*, 2010).

Tanto em humanos como em camundongos, sabe-se que o início do processo de ICX depende da expressão e super-regulação de um gene em especial, o *XIST/Xist* localizado no centro de inativação do cromossomo X (XIC) (Payer and Lee, 2008). Nas duas espécies, este gene transcreve um lncRNA (do inglês, *long non-coding RNA*) que é depositado em *cis* no Xi mantendo esse padrão ao longo de toda a vida da célula (Brockdorff *at al.*, 1992; Brown *at al.*, 1992; Clemson *at al.*, 1996). Este fenômeno pode ser observado através de técnicas de FISH de RNA como um forte acúmulo de suas moléculas sobre o Xi em células adultas. Genes importantes relacionados ao Xa foram descritos em camundongos e humanos, *Tsix* e *XACT* respectivamente (Lee *at al.*, 1999; Stavropoulos *at al.*, 2001; Vallot *at al.*, 2013). Apesar de terem atividade funcional exclusiva em cada uma das espécies onde foram descritos, ambos foram relacionados ao Xa no início do processo de ICX, como apresentado na introdução geral.

Uma vez que o cromossomo X é silenciado, ele é mantido neste estado nas células somáticas durante toda a vida do organismo de maneira estável, mesmo após as sucessivas mitoses. Em termos gerais, o estado inativo do X é mantido por uma ação conjunta de

mecanismos como metilação de suas ilhas CpG, a constante transcrição de *XIST/Xist*, modificações das histonas (metilação e hipoacetilação), e a presença da variante de histona macroH2A (Csankovszki *at al.*, 2001; Miettton *at al.*, 2009).

Em camundongos, o gene *Xist* tem sua transcrição iniciada no embrião em estágio pré-implantacional ainda na fase de duas células, o que coincide com a ativação do genoma embrionário (EGA - do inglês, *Embryonic genome activation*). Nesta fase, o gene é expresso a partir de ambos os cromossomos. Entretanto, sua expressão é rapidamente estabilizada e seu RNA se acumula apenas sobre o cromossomo X paterno (Payer and Lee, 2008). No estágio de blastocisto, nas células da massa celular interna, o que se observa é a reativação do X paterno e uma segunda etapa de ICX - desta vez a escolha do Xi é aleatória, e se mantém nas células do embrião até a fase adulta (Deng *at al.*, 2014; Patrat *at al.*, 2009). Em humanos pouco se sabe sobre os estágios iniciais de ICX: em que momento se inicia a transcrição de *XIST*, se no início a ICX ocorre de maneira “imprintada”, e se existem dois momentos de ICX como observado em camundongos.

Dois estudos entretanto, tiveram acesso a embriões humanos em estágio pré-implantacional e por técnica de FISH de RNA buscaram responder às questões pendentes no processo de ICX no início do desenvolvimento: van den Berg e colaboradores (2009) apresentam resultados de embriões a partir do estágio de oito células e Okamoto e colaboradores (2011) investigaram a expressão de *XIST* em embriões em estágio de mórula e blastocisto. van den Berg e colaboradores (2009) concluíram em seu estudo que um leve acúmulo de *XIST* pode ser detectado já no embrião de oito células e que em blastocisto o X coberto pelo RNA de *XIST* já apresenta o gene *CHIC1* inativado, sugerindo que neste estágio do desenvolvimento a ICX já tenha acontecido. Okamoto e colaboradores (2011) entretanto reportam que os genes *ATRX*, *FGD1* e *HUWE1* se mantêm ativos mesmo no X com acúmulo de *XIST* em fase de blastocisto. Especula-se que a disseminação dos sinais para o silenciamento transcricional dos genes no Xi ocorra de maneira paulatina e que os genes próximos ao XIC sejam os primeiros a serem inativados (Pinter *at al.*, 2012). Assim, o

fato de *CHIC1* encontrar-se na mesma banda cromossômica que *XIST* (Xq 13.2, montagem genômica de referência GRCh38) e os genes estudados por Okamoto e colaboradores (2011) estarem mais distantes (*FGD1* (Xp11.21, GRCh38), *HUEW1* (Xp11.22, GRCh38) e *ATRX* (Xq21.1, GRCh38), pode explicar a divergência entre os resultados dos dois estudos – em blastocistos humanos a ICX estaria se estabelecendo e os sinais para inativação transcricional se disseminando, logo *CHIC1*, mais próximo de *XIST* já estaria inativado, enquanto que aqueles genes mais distantes ainda estariam ativos.

Em camundongos é possível reproduzir *in vitro* o processo de ICX ao longo do desenvolvimento ao se diferenciar *in vitro* células de linhagens de células tronco embrionárias (CTEs) femininas (que apresentam dois X ativos, XaXa) (revisto em Bermejo-Alvarez *at al.*, 2012). Em humanos porém, o estado de inativação do X em CTEs indiferenciadas parece ser muito mais complicado: por FISH de RNA foram descritos três tipos de linhagens de acordo com o padrão de acúmulo de *XIST* antes e depois da diferenciação *in vitro*: classe I, sem sinal de *XIST* em células indiferenciadas e como sinal após diferenciação; classe II com sinal de *XIST* antes e depois da diferenciação; e classe III sem sinal de *XIST* antes ou depois da diferenciação (Silva *at al.*, 2008). Sob a ótica do estado transcricional dos genes do cromossomo X em CTEs humanas, um outro trabalho também definiu três tipos de linhagens: classe I, dois X ativos; classe II, inativação parcial; e classe III, inativação total de um dos homólogos (Bruck and Benvenisty, 2011). Assim, ao contrário do observado em camundongos, organismo utilizado universalmente como modelo, em humanos ainda não é possível reproduzir *in vitro* fielmente o processo de inativação do cromossomo X.

Em agosto de 2013, foram publicados dois estudos envolvendo sequenciamento de última geração de RNA total (RNA-Seq) de embriões humanos em diferentes estágios pré-implantacionais (Xue *at al.*, 2013 e Yan *at al.*, 2013). Estes trabalhos tornaram públicos dados de RNA-Seq de células únicas isoladas de embriões em estágio de 2, 4 e 8 células, e

em estágio de mórula e de blastocisto, bem como de oócitos e zigotos humanos gerados a partir de técnicas de fertilização *in vitro*.

Ambos os estudos apresentam como objetivo comum estudar a variação no transcriptoma ao longo do desenvolvimento humano. Para tanto, eles realizaram análises similares que os levaram às mesmas conclusões, corroborando de maneira independente os resultados dos dois trabalhos. Entre os principais resultados estão a identificação de genes chave no controle do desenvolvimento, novos lncRNAs com expressão específica em determinados estágios, início de ativação do genoma embrionário, degradação dos mRNAs maternos estocados durante a gametogênese, entre outros. Embora muito detalhados, nenhum dos dois trabalhos se interessou em estudar especificamente a atividade transcricional do cromossomo X durante o desenvolvimento pré-implantacional humano. Assim, esses dados públicos gerados propiciam uma oportunidade única de se estudar o início da ICX em diferentes estágios do desenvolvimento. O fato de os dois estudos supramencionados terem realizado os experimentos em células únicas permite análises ainda mais detalhadas do que aquelas propostas originalmente, sendo possível determinar se todas as células de um determinado embrião dão início à ICX em um mesmo momento, discriminando individualmente as células da massa celular interna e as da trofoectoderme no estágio de blastocisto; além disso, pode-se avaliar ainda se a inativação do cromossomo X ocorre de maneira “imprintada” ou aleatória nesta fase do desenvolvimento.

2. OBJETIVOS ESPECÍFICOS

A partir de dados gerados por sequenciamento em larga escala do RNA total obtido de células únicas:

- Identificar em que momento do desenvolvimento humano tem início o processo de inativação do cromossomo X;
- Verificar se há sincronia entre as células de um mesmo embrião em relação ao início da inativação do cromossomo X;
- Constatar se a escolha do X inativo é aleatória ou dependente de origem parental.

3.1. Obtenção dos dados

Os dados envolvendo sequenciamento de última geração de RNA total (RNA-Seq) de embriões humanos em diferentes estágios pré-implantacionais estão disponíveis no banco de dados do NCBI (<http://www.ncbi.nlm.nih.gov/geo/>) sob o número de acesso GSE44183 (Xue *et al.*, 2013) e GSE36552 (Yan *et al.*, 2013). Ambos os experimentos foram realizados em plataforma Illumina Hiseq 2000, as bibliotecas do estudo GSE44183 foram geradas para sequenciamento do tipo *paired-end* e as do estudo GSE36552 tipo *single-end*.

Os dados foram depositados separadamente para cada célula sequenciada, sendo que alguns foram obtidos de embriões íntegros. Com as informações na publicação de Yan e colaboradores (2013, GSE36552) foi possível reagrupar os dados de células individuais de acordo com o embrião de origem; entretanto, no trabalho de Xue e colaboradores (2013, GSE44183), as informações em relação à origem de cada célula não está clara, por esta razão nem todas puderam ser reagrupadas de acordo com sua origem. O trabalho de Xue e colaboradores (2013) publicou ainda o sequenciamento do exoma paterno, neste estudo todos os embriões foram produzidos *in vitro* com o espermatozoides do mesmo genitor. No portal do GEO podem ser encontrados os arquivos com informações sobre o nível de expressão gênica atribuído à cada amostra ou célula individual bem como os arquivos em formato SRA com os dados brutos de RNA-Seq de cada experimento; cada arquivo representa o sequenciamento de uma única célula isolada ou de um embrião inteiro, quando não houve dissociação, e do exoma paterno.

Na tabela 1 está o número de embriões utilizados e células isoladas de cada trabalho.

Tabela 1 | Número de embriões e células utilizadas nos trabalhos de Xue et al. e Yan et al. (2013)

Série	Estágio	Embriões	Células sequenciadas
GSE44183*	Oócito	-	3
	Pronúcleo	-	3
	Zigoto	-	2
	2-células	3	3
	4-células	4	4
	8-células	4	10
	Mórula	3	-
GSE36552**	Oócito	-	3
	Zigoto	-	3
	2-células	3	6
	4-células	3	12
	8-células [§]	3	20
	Mórula [§]	2	16
	Blastocisto	3	30
	hESp0	-	8

* Xue et al., 2013; **Yan et al., 2013

[§] embriões íntegros sequenciados

3.2. Bancos de dados:

Alguns bancos de dados públicos fornecem informações sobre o genoma e o transcriptoma humano e todos os arquivos e dados relacionados às sequências, transcritos, nome, identidade, posição, cromossomo de origem e fita de leitura dos genes humanos dos cromossomos sexuais e autossomos foram adquiridas da montagem Hg19 (GgrCh37.p13).

3.2.1. Genoma humano de referência

As sequências em formato fasta foram adquiridas de cada cromossomo individualmente no depósito de dados públicos do UCSC (<http://hgdownload.cse.ucsc.edu/goldenpath/Hg19/chromosomes/>

montagem Hg19/GRCh37 Genome Reference Consortium Human Reference 37 (GCA_000001405.1)).

Estas foram então concatenadas gerando um novo arquivo multifasta. Com o programa `fasta-grep`, opção `-w 0 ''` removeu-se possíveis espaços em branco que poderiam atrapalhar os procedimentos seguintes (programa desenvolvido por Tim Booth, disponível em <http://nebc.nerc.ac.uk/>).

3.2.2. Transcritos humanos de referência

As sequências de todos os transcritos humanos já anotados foram adquiridas em formato multifasta no depósito de dados públicos do UCSC (<http://hgdownload.cse.ucsc.edu/goldenpath/Hg19/bigZips/> montagem Hg19/GRCh37 Genome Reference Consortium Human Reference 37 (GCA_000001405.1)).

Novamente com o programa `fasta-grep`, opção `-w 0 ''`, removeu-se possíveis espaços em branco que poderiam atrapalhar os procedimentos seguintes.

3.2.3. Polimorfismos de base única (SNPs)

Devido a erros de anotação encontrados no banco disponível de SNPs humanos (dbSNP138), foi necessária a construção de um novo banco personalizado para as necessidades deste projeto.

Para a construção foram utilizados:

- Arquivo tipo `.bed` com a posição no cromossomo e identidade dos SNPs, disponível em: ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/BED/.

- Arquivo tipo `.gtf` (`gencode.v19.annotation.gtf`) com diversas informações sobre os genes humanos conhecidos, como: cromossomo, posição de início e término, nome e identidade do gene, nome e identidade dos possíveis transcritos e classificação (e.g. `lincRNA`, `miRNA`, `processed_transcript`, `protein_coding`, `pseudogene`, `rRNA` etc); disponível em:

`ftp://ftp.sanger.ac.uk/pub/gencode/Gencode_human/release_19`).

- Lista de genes presentes na região do centro de inativação do cromossomo X (XIC, Xq12-q13, de acordo com HUGO Gene Nomenclature Committee). Esta foi obtida do banco <http://www.ensembl.org/biomart/martview>.

Todas as listas foram filtradas com comandos básicos de Unix para a remoção de genes e polimorfismos que eventualmente estivessem anotados em mais de uma posição no genoma. Também foi necessário o desenvolvimento de alguns programas em linguagem *perl*, pela profa. Dra. Maria Vibranovski, capazes de cruzar as informações das listas adquiridas e gerar o banco de SNPs adequado para a realização deste trabalho.

3.3. Plano geral do processo de alinhamento e identificação das variantes

Ensaio de RNA-Seq geram números muito grandes de sequências curtas (cerca de 40 a 100 bases dependendo da técnica utilizada), as quais são comumente chamadas de *reads* e assim o serão ao longo desta tese. A figura 1 é um esquema simplificado ordenando os passos seguidos desde o processo de tratamento das amostras, alinhamento das *reads*, identificação das variantes até anotação dos SNPs. Esta será citada ao longo deste item para melhor compreensão do texto.

Figura 1| Diagrama do plano de trabalho. Estão apresentadas as etapas desde a formatação das *reads* até a obtenção do arquivo final com os SNPs anotados. Os números entre parênteses fazem referência ao subitem com a explicação mais detalhada.

3.3.1. Conversão do formato SRA

As *reads* geradas por experimentos de RNA-Seq disponíveis para download no NCBI estão em formato .sra e foram convertidos em .fastq com o programa fastq-dump (sraToolkit v2.3.5, disponível em <http://www.ncbi.nlm.nih.gov/>) em modo *default* quando a técnica de sequenciamento era *single-end*, ou com a opção `--split-3` quando em *paired-end* (Figura 1: 3.3.1).

3.3.2. Controle de qualidade das *reads*

SNPs são polimorfismos de uma única base, e o sequenciamento de nova geração não está imune a erros de incorporação de bases, ainda que em pequeno número. De acordo com Wang e colaboradores (2013), para se evitar falsos-positivos, devem ser utilizadas apenas *reads* que apresentem qualidade (*Q-score*) maior ou igual a 3. O programa Trimmomatic (Bolger *at al.*, 2014) é capaz de trabalhar eficientemente tanto *reads* geradas por *paired-end* quanto *single-end* e foi aqui utilizado com os seguintes parâmetros: limiar de corte da região 5' (LEADING) : 3; limiar de corte da região 3' (TRAILING) : 3; tamanho da janela deslizante e qualidade média mínima de corte (SLIDINGWINDOW) : 4 : 15; tamanho mínimo do fragmento resultante (MINLEN) : 36; adicionando-se as opções PE (*paired-end*) ou SE (*single-end*) e ainda -phred64 ou -phred63 dependendo da metodologia utilizada na geração das *reads* (Figura 1: 3.3.2).

3.3.3. Alinhamento propriamente dito

Aquelas *reads* que passaram no controle de qualidade puderam então ser alinhadas contra as sequências de referência humana (Figura 1: 3.3.3). Para tanto foi utilizado o software livre Bowtie, versão v1.1.0 (Langmead *at al.*, 2009).

–Alinhamento contra genoma de referência (Hg19) (Figura1: 3.3.3a):

```
bowtie -n 3 -e 150 --chunkmbs 1000 -k 1 -m 1 --best --strata --al <FILE_Hg19.G1> --un <FILE_Hg19.G0> --max <FILE_Hg19.Gm1> <genoma de referência:Hg19_ready_index> <arquivo de entrada:FILE_trimm.fq> -S <arquivo de saída:FILE_Hg19.sam>. Adequando-se os arquivos de entrada quando o sequenciamento foi realizado em paired-end e indicando o índice phred utilizado quando este era 64 (--phred64-quals).
```

O parâmetro `-n 3`, permite até 3 *mismatches* (trocas de base) no alinhamento de cada *read*; `-e 150` indica que a soma máxima permitida das qualidades das bases que apresentam *mismatch* seja 150. Esses parâmetros são fundamentais para não eliminar as *reads* com potenciais SNPs. As opções `-k 1`: limita a apenas um alinhamento válido por *read*; `-m 1`, suprime todos os alinhamentos de uma *read* em particular se existir mais de um alinhamento válido para a mesma *read*, ambos parâmetros importantes para que se obtenha apenas *reads* que alinhem unicamente no genoma eliminando regiões repetitivas e de alta homologia.

A opção `--al` gera um arquivo, aqui nomeado como `FILE_Hg19.G1`, com todas as *reads* que alinham uma única vez; `--un` gera um outro arquivo, nomeado como `FILE_Hg19.G0`, com todas as *reads* que não alinham; `--max`, gera um arquivo, nomeado como `FILE_Hg19.Gm1`, com aquelas *reads* que alinham em mais de uma região.

–Alinhamento contra as sequências dos possíveis transcritos (refmRNA) (Figura 1: 3.3.3b):

Aquelas *reads* que não encontraram identidade com a sequência genômica (presentes no arquivo `FILE_Hg19.G0`) podem ainda representar regiões de *splicing*, assim estas foram alinhadas novamente agora utilizando-se como referência as sequências dos possíveis transcritos humanos. A linha de comando é basicamente a mesma com uma

única modificação, trocou-se o parâmetro `-m` por `-a`, este permite que *reads* seja alinhadas em diferentes regiões, importante aqui para se recuperar eventuais transcritos alternativos.

```
bowtie -n 3 --chunkmbs 1000 -k 1 -a --best -strata --al
FILE_refmRNA1 --un FILE_refmRNA0 --max FILE_refmRNA1 <genoma
de      referência:refMrna_200714_ready_index>      <arquivo   de
entrada:FILE_Hg19.G0> -S <arquivo de saída:FILE_refmRNA.sam>.
Adequando-se os arquivos de entrada quando o sequenciamento foi realizado em paired-
end e indicando o índice phred utilizado quando este era 64 (--phred64-quals).
```

A exemplo do passo anterior (alinhamento contra o genoma) foram especificados três arquivos de saída: `--al` gerando o arquivo `FILE_refmRNA1` com todas as *reads* que alinham uma única vez; `--un FILE_refmRNA0`, com todas as *reads* que não alinham; `--max FILE_refmRNA1`, com aquelas *reads* que alinham mais de uma vez.

Ao permitir que uma *read* alinhe mais de uma vez ela pode tanto encontrar similaridade em transcritos alternativos de um mesmo gene como pode também alinhar em regiões parecidas de genes homólogos ou que simplesmente tenham sequência muito semelhante. Para se remover essas *reads* promíscuas a profa. Dra. Maria Vibranovski desenvolveu um programa em linguagem *perl* capaz de identificar e remover uma *read* que apresentasse identidade em genes diferentes mas manter aquelas que representassem transcritos diferentes de um mesmo gene (Figura 1: 3.3.3c).

Os nomes dos arquivos de entrada e saída aqui são genéricos (`FILE`) e variaram de acordo com a identificação da amostra.

Para todas as amostras à exceção do exoma paterno (arquivo SRR689250, estudo GSE44183) foram realizadas as duas etapas de alinhamento acima explicados. Por se tratar do sequenciamento de exoma a amostra de DNA paterna foi submetida apenas ao primeiro passo de alinhamento contra o genoma de referência.

3.3.4. Identificação de possíveis variantes (Figura 1: 3.3.4)

O pacote de comandos SAM Tools (Li *et al.*, 2009) apresenta diversas ferramentas de manipulação, visualização, formatação entre outras tantas para se trabalhar com arquivos resultantes de alinhamentos. Associados SAM Tools e o programa VarScan (Koboldt *et al.*, 2009, 2012) se tornam uma potente ferramenta de detecção de possíveis polimorfismos. Existem diversas ferramentas disponíveis para se identificar variantes mas o VarScan permite muitos ajustes importantes como a determinação do número mínimo de *reads* que cobrem a região do polimorfismo, a qualidade de sequenciamento dessas *reads*, a qualidade da base da *read* na posição do polimorfismo, a qualidade do genótipo encontrado, significância estatística entre outros.

O VarScan combinado com o SAM Tools reporta as variantes em potencial tendo como base o genoma de referência utilizado no processo de alinhamento, assim os arquivos gerados pelos alinhamentos contra o genoma de referência (FILE_Hg19) e as sequências dos transcritos, (FILE_G0refmRNA_filtered, filtrado para a remoção das *reads* promíscuas, subitem 3.3.3) foram submetidos separadamente mas seguindo as mesmas opções e parâmetros:

```
samtools mpileup -B -f <Hg19_ready.fa ou  
refMrna_200714_ready.fa> -q 1 <arquivo de entrada:FILE.bam> |  
java -jar VarScan.v2.3.7.jar mpileup2cns --output-vcf >  
<arquivo de saída:FILE_Hg19.vcf ou FILE_refmRNA.vcf>. Este  
programa é capaz de encontrar e mapear tanto genótipos iguais ao do genoma de  
referência quanto os alternativos desde que seja especificada a opção mpileup2cns.
```

3.3.5. Anotação de possíveis variantes

O VarScan é uma potente ferramenta que permite tanto a genotipagem quanto a identificação de novos polimorfismos, mas esses últimos devem ser posteriormente validados. Como o material utilizado neste trabalho é fruto da extração de RNA total, impossibilitando o acesso direto ao genótipo das amostras, uma importante decisão tomada aqui foi a de utilizar apenas SNPs conhecidos depositados em bancos de dados públicos. No subitem 3.2.3 deste capítulo foi apontada a necessidade de se gerar um banco de SNPs personalizado para este trabalho. Este banco contém a identidade do SNP, o cromossomo, a posição no cromossomo, a classificação e nome do gene que este se encontra. Dado que as variantes identificadas no processo de alinhamento contra as sequências dos transcritos (refmRNA) apresentam a posição no transcrito e não no genoma foi necessária fazer uma correção da posição dessas. Para tanto a profa. Dra. Maria Vibranovski desenvolveu um programa que reposiciona essas variantes de acordo com a posição no genoma. O programa toma como base um arquivo com dados da localização genômica de genes e seus transcritos e reposiciona a variante encontrada pelo VarScan para a posição genômica e assim é possível anotar os SNPs encontrados (Figura 1: 3.3.5a).

Feita esta correção os arquivos gerados pelo VarScan puderam então ter suas variantes anotadas tendo como base o banco de SNPs personalizado utilizando-se um outro programa também desenvolvido pela profa. Dra. Maria Vibranovski capaz de cruzar a informação da posição da variante identificada pelo VarScan com a posição de SNPs conhecidos eliminando-se assim possíveis falso-positivos (Figura 1: 3.3.5b).

Devido a real possibilidade de um único *read* ter sido alinhado em diferentes transcritos de um mesmo gene (Figura 2) o arquivo resultante da anotação dos SNPs teve também que ser corrigido (Figura 1: 3.3.5c). A figura 2 exemplifica um caso no qual um mesmo SNP está representado em transcritos diferentes de um gene hipotético, na porção 3' do exon 3. A figura ilustra também que uma mesma *read* foi mapeada em dois

transcritos (transcrito 1 e 2) na região de cobertura do SNP. É intenção deste trabalho observar expressão alelo-específica e não identificar diferença de expressão entre transcritos de um mesmo gene. A contagem de uma mesma *read* nos passos seguintes pode provocar um desvio nos valores observados de expressão alelo-específica logo, para se excluir esta possibilidade optou-se aqui por selecionar o transcrito que apresentasse maior número de *reads* cobrindo a posição de um SNP no caso deste estar representado em mais de um transcrito. De acordo com a figura 2 o transcrito 2 exibe apenas uma única *read* cobrindo a posição do SNP já o transcrito 1 apresenta quatro *reads* assim neste exemplo os resultados referentes ao transcrito dois seria mantido nas análises subsequentes.

Figura 2 | Gene e três possíveis transcritos hipotéticos com SNP identificado no exon 3. A figura exemplifica a possibilidade de uma mesma *read* (barra preta) ter sido alinhada em transcritos diferentes. As barras vermelhas representam diversas *reads* únicas que alinharam apenas em uma região.

Como foi explicado no subitem 3.3.3 aquelas *reads* que não encontraram similaridade no genoma de referência (FILE_Hg19.G0) são novamente testadas contra as sequências dos transcritos na tentativa de se resgatar *reads* que representem junções exon-exon

(Figura 1: 3.3.3b). Caso um SNP esteja presente próximo a uma das extremidades de um exon ele pode estar representado duas vezes no arquivo final. A figura 3 ilustra essa possibilidade, o SNP presente na extremidade 3' do exon 3 do gene hipotético, foi identificado tanto por *reads* do alinhamento com a sequência genômica (Hg19) quanto com a sequência dos transcritos (refmRNA). Foi necessária então fazer a soma daquelas *reads* que cobrem a região de um mesmo SNP (Figura 1: 3.3.5d), assim no exemplo da figura 3 a contagem final ficaria de oito *reads*, quatro encontradas no alinhamento com o genoma de referência (Hg19) e mais quatro do segundo alinhamento tendo como referência as sequências dos transcritos (refmRNA).

Figura 3| Gene hipotético e resultados de dois alinhamentos independentes tendo como base diferentes referências. **Gene**, exemplifica a estrutura geral de um gene qualquer; **Hg19**, *reads* alinhadas tendo como referência a sequência do genoma humano (passo descrito no subitem 3.3.3 (Figura 1: 3.3.3a); **refmRNA**, *reads* alinhadas tendo como referência as sequências dos transcritos (passo descrito no subitem 3.3.3 Figura 1: 3.3.3b). A barra vertical vermelha indica a posição de um SNP qualquer, barras horizontais pretas indicam *reads* que cobrem a posição do SNP; barras horizontais marrons indicam *reads* em outras posições.

3.4. Seleção dos SNPs

À semelhança do que foi descrito por Wang e colaboradores (2008) e aplicado e validado para se estudar padrão de ICX a partir de dados de RNA-Seq de placentas de equinos e muar (Wang *et al.*, 2012), foi feita a contagem direta das *reads* de cobertura de um determinado SNP, baseada na anotação de SNPs em regiões transcritas, para depois se calcular a razão relativa da expressão entre os alelos nas amostras de embriões humanos. Apenas foram consideradas variantes com no mínimo 20 de *reads* de cobertura. Como controle de padrão de expressão bialélico são utilizados os polimorfismos presentes nos genes autossômicos. Os dados das amostras de células-tronco masculinas foram utilizados como controle positivo para padrão de expressão monoalélica de genes do cromossomo X.

3.5. Determinação do sexo cromossômico das amostras

Nenhuma das amostras disponibilizadas por Xue e colaboradores (2013) ou Yan e colaboradores (2013) possuem o sexo cromossômico determinado, à exceção da linhagem de células tronco embrionárias (hESC, série GSE36552) que foi caracterizada como 46, XY. Entretanto podemos assumir que as amostras de oócitos não apresentam genes do cromossomo Y. Uma forma indireta de se determinar o sexo cromossômico das amostras embrionárias pode ser realizada pela verificação de genes expressos a partir do cromossomo Y. Os dados de expressão gênica para cada célula das séries GSE44183 e GSE36552 estão publicamente disponíveis no portal do GEO no NCBI (<http://www.ncbi.nlm.nih.gov/geo/>), já normalizados em rpkm (do inglês, *reads per kilobase per million*). Depois do processo de alinhamento contra as sequências de referência, ensaios de RNA-Seq podem fornecer aproximações quantitativas da abundância de transcritos. De forma simplificada: essas aproximações são estimadas através da contagem direta de *reads*, levando em conta seu tamanho, que alinham em cada gene, sendo assim normalizadas (rpkm). Esse processo ajuda a remover desvios técnicos

inerentes aos passos de preparação do material e da biblioteca de sequenciamento, resultando assim em valores quantitativos e não apenas qualitativos de expressão gênica.

Em suas análises tanto Xue e colaboradores (2013) quanto Yan e colaboradores (2013) consideram um gene expresso se este apresentar rpkms maior ou igual a 0,1, logo, foram aqui selecionados e contados apenas aqueles genes com $rpkms \geq 0,1$. Como os embriões tiveram suas células sequenciadas separadamente seus dados foram reunidos para a contagem de genes únicos por embrião a fim de se evitar a contagem repetida de um mesmo gene caso este estivesse expresso em diferentes células de um mesmo embrião. Utilizando-se ferramentas de busca de linguagem Unix como por exemplo awk foram contados quantos genes estavam expressos separados por tipo cromossômico: autossomo, cromossomo X ou cromossomo Y, evitando-se aqueles presentes na região pseudoautossômica.

Para se comparar o número absoluto de genes do cromossomo Y observado para cada embrião com aquele dos oócitos ou da linhagem de células tronco masculina foi utilizado o teste exato de Fisher com correção de Bonferroni (usando as análises estatísticas do programa R (v.3.1.0; R Development Core Team, 2008) e assim definir o sexo cromossômico da amostra. O número máximo de genes encontrado em qualquer das amostras foi utilizado como o número máximo que poderia ser encontrado em qualquer uma das amostras com o método empregado para o sequenciamento do transcriptoma.

As amostras de cada estudo foram tratadas separadamente e os genes autossômicos foram utilizados como controle pois não espera-se encontrar diferenças significativas.

3.6. Expressão de *XIST*

Como explicado no item anterior os dados sobre o nível de expressão gênica estão disponíveis normalizados em rpkms para todos os genes encontrados no processo de alinhamento realizado por Xue e colaboradores (2013) e Yan e colaboradores (2013). Através de ferramentas de busca de linguagem Unix (awk e grep) procurou-se pelas

células que apresentassem nível de expressão de $XIST \geq 0,1$. Os valores foram então diretamente utilizados na montagem do gráfico de barras para comparação.

3.7. Genotipagem

Como realizado no trabalho de Wang e colaboradores (2011), é possível utilizar as leituras de sequenciamento de todos os genes do cromossomo X com os dados de expressão gerados nos trabalhos de Xue e colaboradores (2013) e Yan e colaboradores (2013) para se verificar padrões de expressão alelo-específicos e indiretamente questionar o estágio de inativação do X em que a amostra se encontra. Ao se alinhar os resultados de RNA-Seq com o genoma de referência do cromossomo X humano, é possível identificar em regiões transcritas que contenham SNPs se há alguma tendência pela expressão de um alelo em particular, o que indicaria que a inativação do cromossomo X já se estabeleceu naquela célula.

Primeiramente foram concatenados os resultados do processo de anotação de variantes de todas as células de um mesmo embrião (a partir do estágio de duas células) ou aquelas da linhagem de células tronco. Através do uso de um programa desenvolvido para o programa R (pelo Dr. Gustavo Ribeiro Fernandes) foi possível então buscar por SNPs que estivessem representados em ao menos duas células de uma determinada amostra. Um outro programa, escrito pela profa. Dra. Maria Vibranosvski, foi utilizado para se filtrar daqueles SNPs em comum que apresentassem expressão bialélica ou expressão monoalélica diferencial *i.e.*, quando uma célula expressava um alelo e outra célula de um mesmo embrião expressava outro.

O processo de genotipagem se baseou naqueles SNPs que indicassem expressão bialélica e também naqueles que representassem expressão monoalélica diferencial. Somente se uma amostra for heterozigota para um determinado polimorfismo será observada expressão bialélica baseada em sua presença; da mesma forma duas células de

um mesmo embrião só podem expressar diferentes alelos se seu genótipo for heterozigoto. Assim: quando uma célula apresentou um alelo expresso (A) e outra o outro alelo (A') foi possível inferir que geneticamente aquele embrião era heterozigoto (A|A') para aquele determinado polimorfismo. SNPs que sugerem expressão bialélica (A/A') também indicam que a amostra é heterozigota.

Para se determinar o padrão de expressão a partir de um SNP foi realizada a contagem direta das *reads* representando o alelo de referência e o alternativo, a partir da qual calculou-se a razão entre os alelos expressos. Foram considerados genes de expressão bialélica aqueles com razão entre os alelos dentro do intervalo de 0,2 a 0,8 (Amos-Landgraf *at al.*, 2006; Borel *at al.*, 2015), por conseguinte foram classificados genes de expressão monoalélica aqueles com razão menor ou igual a 0,2 ou maior ou igual a 0,8. Quando todas as células de um mesmo embrião apresentaram o mesmo alelo expresso para um determinado polimorfismo este foi caracterizado como não informativo pois não é possível distinguir entre um caso de homozigose ou expressão monoalélica.

3.8. Determinação do padrão de inativação do cromossomo X

A presença de genes do cromossomo X com expressão bialélica pode indicar que o silenciamento transcricional ainda não se estabeleceu, foram então contados quantos genes por embrião apresentavam expressão bialélica e foi calculada a razão entre esses genes e o total de genes expressos no cromossomo X. Como controle foram utilizados os genes autossômicos.

De maneira complementar buscou-se por genes heterozigotos de expressão monoalélica diferencial, quando um determinado gene apresentou um alelo expresso em uma célula e outro alelo em outra célula de um mesmo embrião este foi considerado informativo. Calculou-se então a razão entre os genes de expressão monoalélica e os genes

heterozigotos de cada embrião, com a intenção de se determinar o padrão de ICX ao longo do desenvolvimento. Onde assumiu-se que:

- se a ICX ainda não tiver ocorrido espera-se encontrar uma proporção de genes bialélicos que não varie significativamente de acordo com o estágio do desenvolvimento; adicionalmente espera-se que a proporção de genes heterozigotos do cromossomo X com expressão monoalélica diferencial se mantenha baixa;
- se estivermos diante de um caso de ICX “imprintada” espera-se que a proporção de genes bialélicos diminua de forma dependente ao estágio do desenvolvimento; enquanto que a proporção de genes heterozigotos do cromossomo X com expressão monoalélica diferencial deve se manter baixa;
- se o caso for de ICX aleatória veremos que a proporção de genes bialélicos diminuirá de forma dependente ao estágio do desenvolvimento junto com um aumento significativo da proporção de genes heterozigotos do cromossomo X com expressão monoalélica diferencial.

Figura 4| Resultados hipotéticos representando os padrões esperados de inativação do cromossomo X ao longo do desenvolvimento extra-embriônico. Em **A** nenhuma fase embrionária apresenta ICX (dois X ativos); em **B** inativação dependente da origem parental; em **C** inativação aleatória sendo estabelecida ao longo do desenvolvimento. Em preto razão entre o número de genes com expressão bialélica pelo número total de genes com curva de regressão; em vermelho razão entre o número de genes com expressão monoalélica diferencial pelo número total de com a curva de regressão.

3.9. Razão entre o nível de expressão gênica dos genes do cromossomo X e aquela dos autossomos

Uma forma de se verificar ICX pode ser feita através do cálculo da razão entre o nível de expressão dos genes do cromossomo X por aqueles dos autossomos (Bruck and Benvenisty, 2011). Em 1997 Susumo Ohno sugeriu que o único X ativo tanto em fêmeas quanto em machos de mamíferos eutérios teria seu nível de expressão aumentado como uma forma de compensação de dose entre genes do cromossomo X e autossomos mais recentemente revisto em (Veitia *et al.*, 2015). Ainda hoje a hipótese de Ohno é muito debatida, existem trabalhos que suportam sua teoria observando razão $X:A = 1$ e outros que a refutam com razão $X:A = 0,5$ (revisto em Pessia *et al.*, 2014; Veitia *et al.*, 2015). Com o intuito de se calcular a razão $X:A$, no presente trabalho, foram utilizados os dados de expressão em rpk (rpk $\geq 0,1$), para tanto foi calculada a mediana do nível de expressão dos genes de cada cromossomo separadamente, e também de todos os autossomos em conjunto.

4. RESULTADOS E DISCUSSÃO

4.1. Qualidade, alinhamento das *reads* e variantes anotadas

Na tabela 2 encontram-se as médias de *reads* por estágio do desenvolvimento para as amostras de cada estudo GSE36552 (Yan *et al.*, 2013) e GSE44183 (Xue *et al.*, 2013). Com os valores desta tabela, e também na representação gráfica da figura 5, pode-se verificar que o processo de sequenciamento do estudo GSE36552 resultou em muito mais *reads* sequenciadas por amostra (em média 36.936.192 *reads*) do que o estudo GSE44183 (em média 20.400.987 *reads*), à exceção do exoma paterno (47.632.194 *reads*, tabela 2 e figura 5). As *reads* alinhadas com as sequências de referência dos transcritos (refmRNA) são parte daquelas que não encontraram similaridade no genoma de referência (Hg19). Os valores absolutos de cada amostra para o processo de remoção de *reads* com baixa qualidade, cada etapa do alinhamento e total de SNPs anotados encontram-se no anexo 1 para o estudo GSE36552 e anexo 2 para GSE44183.

Série	Estágio	<i>reads</i> iniciais	<i>reads</i> com boa qualidade (%)	<i>reads</i> alinhadas com Hg19 (%)	<i>reads</i> alinhadas com refmRNA (%)	Total de <i>reads</i> alinhadas (%)	SNPs anotados
GSE36552	Oócito	29729122	94.16	58.66	55.53	75.21	277004
	Zigoto	36816984	94.66	44.67	24.76	54.34	277241
	2-células	52072244	93.37	45.41	29.59	56.21	296127
	4-células	50317149	93.11	55.60	49.54	68.30	406237
	8-células	35014889	93.07	50.92	36.96	62.28	276769
	Mórula	33095667	91.13	50.73	32.22	58.58	212803
	Blastocisto	31972718	90.60	43.41	40.52	63.60	213703
	hESCp0	26470766	93.42	56.15	51.80	66.02	262392
	Média total	36936192	92.94	50.69	40.12	63.07	277785
GSE44183	Oócito	18737845	93.47	26.18	41.94	53.63	100343
	Pronúcleo	23878668	93.64	26.37	50.01	72.26	248262
	Zigoto	22346410	93.06	26.28	39.57	52.57	190949
	2-células	20260599	94.33	24.62	40.44	51.46	193133
	4-células	21245108	93.77	23.78	30.20	42.65	171418
	8-células	18306067	85.90	20.63	32.34	43.02	127879
	Mórula	18032213	94.65	24.25	44.19	50.49	133115
	Média total	20400987	92.69	24.59	39.81	52.30	166443
	DNA Paterno	47632194	93.88	48.52		48.52	1811877

Figura 5| Número de *reads* em cada fase do processo de alinhamento. Gráficos em *boxplot* gerados com o número de *reads* de cada amostra nas categorias: “**Total de *reads***”: número bruto de *reads* iniciais disponíveis para análise; “***reads* com qualidade**” número de *reads* que passaram no filtro de qualidade; “***reads* alinhadas**” *reads* que encontraram similaridade com o genoma de referência (Hg19 e refmRNA). *Outliers* da série GSE44183: genoma paterno representado pelos pontos superiores nas três categorias; 8cél.#C tem suas duas células representadas como os pontos inferiores na categoria da qualidade e apenas uma das células como o ponto inferior na categoria das *reads* alinhadas; pronúcleo #2 também aparece como *outlier* superior na categoria das *reads* alinhadas. Os dados estão agrupados por estudo, GSE36552 (Yan *et al.*, 2013) e GSE44183 (Xue *et al.*, 2013).

O exoma paterno é resultado do sequenciamento de amostra de DNA purificado de sangue periférico do doador. Como esse material é bastante abundante e em comparação às amostras embrionárias, é muito mais simples de trabalhar do que o RNA total purificado de células isoladas. Entretanto as diferenças em relação à quantidade total de *reads* por amostra observadas entre os estudos GSE36552 e GSE44183 (Tabela2 e figura 5) talvez se deva à técnica de preparação da biblioteca para o RNA-Seq. Yan e colaboradores (GSE36552) relatam em seu trabalho que a técnica utilizada por eles é mais eficiente, mas

infelizmente não foi possível compará-la ao publicado por Xue e colaboradores (GSE44183) por falta de informações mais detalhadas na descrição dos métodos deste último.

Do total de *reads* iniciais disponíveis foram removidas aquelas com baixa qualidade de sequenciamento. Este processo é fundamental para minimizar o efeito negativo de eventuais bases com baixa qualidade de sequenciamento na identificação das variantes. Estão também na tabela 2 as porcentagens médias das *reads* por estágio do desenvolvimento que passaram nessa filtragem.

De uma maneira geral, o filtro de qualidade manteve quase 93% das *reads* originais para ambos os estudos. Entretanto, do sequenciamento das células dos embriões no estágio de oito células da série GSE44183 restaram em média 85,9% das *reads* originais. Essa perda de cerca de 14% das *reads*, assim como o baixo o número inicial de *reads* das amostras da série GSE44183, se reflete nas etapas subsequentes: *reads* alinhadas e variantes anotadas como discutido a seguir.

Daquelas *reads* com qualidade adequada para análise de variantes, cerca de 50% alinharam uma única vez contra o genoma de referência (Hg19) para as amostras da série GSE36552 e 28% para as amostras da série GSE44183. Dentre estas últimas, apenas o exoma paterno teve quase 50% de suas *reads* alinhadas contra o genoma de referência (Hg19). Aquelas *reads* que não encontraram identidade no genoma de referência passaram por um novo processo de alinhamento, desta vez contra as sequências de referência dos transcritos humanos (refmRNA). Esta foi uma etapa importante, pois agregou de 8% a quase 46% de *reads* a mais, aumentando de maneira expressiva a quantidade de dados para busca por variantes (Tabela 2 e anexos 1 e 2). Uma visão geral da tabela 2 aponta para um expressivo número de *reads* não alinhadas, variando de aproximadamente de 25% entre os oócitos da série GSE36552 e um pouco menos do que 58% entre os embriões no estágio de quatro células da série GSE44183. Esses valores podem ser considerados altos entretanto, refletem a escolha dos parâmetros utilizados no processo de alinhamento

visando ponderar entre a um ótimo alinhamento da *read* com o genoma de referência e ainda assim permitir a presença de *mismatches* para a identificação das variantes.

Os valores do estudo GSE44183 tabulados na figura 5 apontam para alguns pontos fora da curva (*outliers*) nas três categorias avaliadas. Como discutido acima, o exoma paterno apresenta um número muito maior de *reads* sequenciadas do que os embriões, o que se reflete nas etapas subsequentes. Essa amostra aparece na figura 5 como os *outliers* superiores nas três categorias ('Total de *reads*', '*reads* com qualidade' e '*reads* alinhadas'). Por alguma razão a qual não foi possível detectar no trabalho original publicado por Xue e colaboradores (2013), o sequenciamento das duas células do embrião no estágio de oito células, identificado como embrião C (8-cél.#C), resultou em um número muito menor de *reads* de boa qualidade do que os outros (Anexo 2), ambos aparecendo como *outliers* inferiores de pontos sobrepostos na categoria '*reads* com qualidade' (Figura 5). A célula #5 do mesmo embrião C também apresentou número muito inferior de *reads* totais alinhadas quando comparado às outras amostras, *outlier* inferior na categoria '*reads* alinhadas'. Ainda observando o campo com o total de *reads* alinhadas, o pronúcleo #2 junto com o exoma paterno apresentaram muito mais *reads* alinhadas do que as outras amostras, em geral e se apresentam como os *outliers* superiores (Figura 5 e anexo 2). Ainda que as células do embrião C tenham apresentado um baixo número de *reads*, decidiu-se seguir com as análises para esta amostra.

Como consequência de se ter iniciado o processo de filtragem e alinhamento com um maior número de *reads*, de uma maneira geral as amostras do estudo GSE36552 apresentaram também um maior número de variantes anotadas quando comparadas ao estudo GSE44183 (Tabela 3), o mesmo acontece para o exoma paterno (GSE44183).

Na tabela 3 estão representadas as contagens das variantes anotadas nos autossomos e nos cromossomos sexuais para cada amostra, agrupadas por estudo. Para aqueles embriões dos quais foi possível sequenciar mais de uma célula, foi calculada a média dos valores totais.

Tabela 3 Total de variantes anotadas por tipo cromossômico					
Série	Estágio	Amostra	Autossomos	crom. X	crom. Y
GSE36552	Oócito	#1	274360	5089	2
		#2	283439	5331	2
		#3	258052	4734	2
	Zigoto	#1	271022	5162	2
		#2	237432	4512	2
		#3	307927	5663	2
	2-células	emb.#1*	296590	5533	2
		emb.#2*	290296	5525	2
		emb.#3*	291050	5379	2
	4-células	emb.#1*	401285	7731	1
		emb.#2*	393573	7983	2
		emb.#3*	401201	7629	2
	8-células	emb.#1*	212055	5272	1
		emb.#2*	229641	5571	1
		emb.#3*	271309	5437	23
	Mórula	emb. #1*	189457	2967	42
		emb. #2*	233192	2806	45
	Blastocisto	emb. #1*	207142	3565	65
		emb. #2*	203618	3569	75
		emb. #3*	221052	5054	44
	hESCp0	*	256731	5559	101
GSE44183	Oócito	#1	86044	1633	0
		#3	128369	2712	0
		#4	80585	1685	0
	Pronúcleo	#2	229129	4373	6
		#3	233572	4854	9
		#4	267515	5319	8
	Zigoto	#1	239568	4546	8
		#2	135030	2742	3
	2-células	#1	276174	5450	6
		#5	148585	3082	6
		#6	143392	2702	2
	4-células	#1	271943	5457	3
		#2	148591	2688	6
		#3	67854	1318	3
		#4	184412	3392	6
	8-células	emb.#a*	218272	3965	46
		emb.#b*	114844	2220	28
		emb.#c*	44991	997	11
		emb. íntegro	61285	1060	13
	Mórula (íntegra)	emb. #1	151048	2701	30
		emb. #4	111425	2162	37
		emb. #5	129293	2606	43
	DNA Paterno		1777679	33936	262

* valores resultantes da média de variantes encontradas nas respectivas células

Na figura 6 os gráficos foram construídos com o número total de variantes anotadas para todas as amostras agrupadas por estudo, à exceção do exoma paterno, e separadas por autossomos ou cromossomo X. Os valores de cada amostra podem ser observados nos anexos 3 e 4 para cada cromossomo individualmente (estudos GSE36552 e GSE44183 respectivamente). Novamente é possível notar a diferença entre os dois estudos em relação à quantidade de variantes finais anotadas por estágio do desenvolvimento (GSE36552: número mínimo de variantes anotadas 212.803 e máximo 406.237; GSE44183:

mínimo 100.434 e máximo 248.262) condizente com o número inicial de *reads* sequenciadas (Tabela 2).

Figura 6| Número de variantes anotadas. Gráficos em *boxplot* gerados a partir do número total de SNPs que puderam ser anotados em todos os cromossomos (A) e para os cromossomos autossomos (B) e cromossomo X (C) separadamente. Foram utilizados os dados referentes às amostras embrionárias e oócitos. *Outliers* da série GSE36552: embrião 4-cél.#1 célula#1 aparece como o ponto superior para todas as categorias; no cromossomo X (C) embrião 4-cél.#1 célula#2 é o segundo ponto mais alto. *Outliers* inferiores no cromossomo X (C) em ordem decrescente: mórula#2 células #3, #4, #6 e #7. Os dados estão agrupados por estudo, GSE36552 (Yan *et al.*, 2013) e GSE44183 (Xue *et al.*, 2013).

O *outlier* superior que aparece entre as amostras da série GSE36552 (Figura 6 A, B e C) é a célula #1 do embrião #1 em estágio de quatro células (4-cél.#1 célula#1) com 498.698 SNPs anotados, quase 55 mil a mais do que o máximo encontrado para as outras amostras desse grupo (Anexo 3). De fato, para as categorias ‘total de *reads*’, ‘*reads* com qualidade’ e ‘total de *reads* alinhadas’ (Figura 5, tabela 2 e anexo 1), os resultados dessa célula específica se apresentam a cima do percentil 75. Uma outra célula do mesmo embrião (4-cél.#1 célula #2) também se apresenta como *outlier* logo abaixo da célula#1 quando contados apenas os SNPs anotados no cromossomo X. Ainda neste cromossomo, para o estudo GSE36552 aparecem quatro *outliers* inferiores: são as células #3, #4, #6 e #7 do embrião #2 em estágio de mórula.

Apesar de as amostras de RNA-Seq da série GSE44183 apresentarem número inferior de *reads* em todas as categorias até aqui analisadas, e consequentemente de SNPs anotados, o estudo apresenta uma vantagem em relação as amostras da série GSE36552: o sequenciamento do exoma paterno. Como o doador de sêmen é o mesmo para todas as amostras, é possível identificar a origem parental do alelo expresso nas amostras embrionárias.

4.2. Determinação do sexo cromossômico das amostras

À exceção da linhagem de células tronco embrionárias desenvolvida no trabalho de Yan e colaboradores (GSE36552) que foi identificada como 46, XY, os trabalhos de Yan e de Xue (2013) não fazem menção ao sexo cromossômico das amostras embrionárias. Entretanto é possível se basear na expressão de genes conhecidos do cromossomo Y para assim se inferir de maneira indireta o gênero dos embriões como descrito a seguir.

Nos trabalhos publicados por Xue e colaboradores (2013) e Yan e colaboradores (2013) os autores consideraram um gene como expresso quando a rpkf fosse maior ou igual a 0,1. Assim, realizou-se aqui a contagem dos genes com rpkf $\geq 0,1$ para a verificação de quantos são expressos a partir de cada tipo cromossômico autossomos ou sexuais (X e Y separadamente) para cada amostra individualmente. Para tanto, quando as

células de um mesmo embrião (a partir do estágio de duas células) ou da mesma linhagem de células tronco foram sequenciadas individualmente, os dados de expressão de todas as células de uma mesma amostra foram concatenados gerando um arquivo único no intuito de se representar o embrião como um todo.

Os gráficos de barras da figura 7 foram construídos tendo como base a razão entre o número de genes autossômicos identificados em cada amostra pelo número máximo de genes autossômicos identificados. Em ambos os estudos, embriões no estágio de oito células foram os que apresentaram maior número de genes autossômicos expressos; no caso da série GSE36552, o embrião identificado como 8-cél.#2 apresentou 13.577 genes expressos e no da série GSE44183 o embrião identificado como 8-cél.#A apresentou 10.890 genes expressos, estes foram então utilizados no cálculo da razão como número máximo de genes autossômicos expressos encontrados para cada série respectivamente (Figura 7 **A** e **B**).

Figura 7 | Proporção de genes autossômicos expressos. Gráfico de barras mostrando a razão entre o número de genes expressos a partir dos cromossomos autossomos de cada amostra e o número máximo total de genes autossômicos encontrado em uma das amostras calculado para cada série em separado. Em **A** resultados da série GSE36552 (Yan *et al.*, 2013) e em **B** da série GSE44183 (Xue *et al.*, 2013).

Os gráficos da figura 7 mostram que, independente do estágio do desenvolvimento, as amostras não apresentam grande variação entre as razões do número absoluto de genes autossômicos expressos, o que pode ser observado em particular nas amostras da série GSE36552.

De maneira similar ao que foi feito com os genes autossômicos, realizou-se a contagem daqueles genes mapeados no cromossomo Y com rpkms maior ou igual a 0,1 mas que não estivessem na região pseudoautossômica, a fim de se evitar falsos positivos em células femininas. Como os cromossomos X e Y humanos apresentam homologia, há grande chance de que as *reads* geradas pelo processo de sequenciamento sejam alinhadas erroneamente. Assim, mesmo tomando-se o cuidado de remover os genes da região

pseudoautossômica das análises de sexagem, é possível que uma amostra XX apresente *reads* mapeadas no cromossomo Y. Todavia, espera-se que a quantidade de genes do cromossomo Y encontrada em células femininas seja numericamente inferior àquela encontrada em uma célula verdadeiramente XY e que ambas sejam estatisticamente diferentes. O número de genes do cromossomo Y encontrado para cada amostra foi comparado com aqueles encontrados nos oócitos, utilizados como controle de células femininas. Apenas para o estudo GSE36552 pôde-se usar também os resultados de expressão da linhagem de células tronco embrionárias indiferenciadas na passagem zero (hESCp0) como controle de célula masculina. Para se testar essa hipótese utilizou-se o teste exato de Fisher com correção de Bonferroni.

Durante as análises de expressão alelo específica, apresentadas e discutidas mais adiante no item 4.5, foram identificados alguns SNPs que indicavam expressão bialélica de dois genes do cromossomo Y (*UTY* e *TTY14*). Dado que em células normais existe apenas um único cromossomo Y e todos os seus polimorfismos devem se apresentar em hemizigose, um gene que apresente expressão bialélica neste cromossomo deve ser considerado um artefato de alinhamento. Assim tais genes foram removidos da análise de determinação do sexo cromossômico.

A figura 8 representa os resultados da sexagem baseada na expressão de genes do cromossomo Y para cada estudo em separado. A exemplo do que foi feito para os genes autossômicos, os gráficos de barras foram construídos tendo como base a razão entre o número de genes do cromossomo Y identificados em cada amostra pelo número máximo de genes do Y identificado em uma das amostras. No caso da série GSE36552, o embrião em estágio de mórula (identificado como Mórula #1) apresentou 29 genes expressos a partir do cromossomo Y, e no da série GSE44183 o embrião em estágio de oito células (identificado como 8-cél.#A) apresentou 11 genes mapeados no cromossomo Y estes foram então utilizados como número máximo de genes expressos do Y encontrados para cada série respectivamente. Estão também representados nas tabelas à direita de cada

gráfico da figura 8 os resultados dos testes exatos de Fisher que foram realizados separadamente para cada estudo.

Figura 8| Determinação do sexo cromossômico dos embriões através da contagem de genes expressos mapeados no cromossomo Y. Foram utilizados os valores de expressão gênica em rpkm disponibilizados por cada estudo: em A resultados da série GSE36552 (Yan *et al.*, 2013) e em B da série GSE44183 (Xue *et al.*, 2013). A esquerda gráficos de barras mostrando a razão entre o número de genes expressos a partir do cromossomo Y de cada amostra e o número máximo total de genes do Y encontrado; a direita resultado do teste exato de Fisher comparando cada amostra com aquelas nas quais o sexo cromossômico é conhecido, as amostras de oócitos foram utilizadas como controle de células femininas e as de células tronco embrionárias como masculina. Quando para GSE36552 $p < 0,0025$ ou para GSE44183 $p < 0,0023$ assume-se que as diferenças são significativas. As cores estão de acordo com os resultados do teste exato de Fisher, onde foram coloridas com rosa aquelas amostras classificadas como femininas e com azul aquelas masculinas de acordo com o valor de p ajustado.

Da comparação feita pelo teste exato de Fisher pôde-se então inferir o sexo cromossômico de cada embrião levando-se em conta o valor de p ajustado pelo número de testes realizados: quando $p < 0,0025$ para GSE36552 ou $p < 0,0023$ para GSE44183 rejeita-se a hipótese nula de que as amostras comparadas sejam iguais. Assim, quando uma amostra embrionária foi comparada a um oócito e apresentou valor de p menor do que o ajustado ela foi classificada como masculina; de maneira oposta, aquelas com p menor do que o ajustado quando comparada à linhagem de células tronco hESCp0 (46, XY) foram classificadas como feminina. É importante notar que para o estudo GSE36552 o teste controle da comparação entre o número de genes do cromossomo Y detectados nas células indiferenciadas da linhagem de células tronco embrionária com os oócitos calculou com sucesso $p < 0,0025$ em todos os casos ($p = 0,0000$). Quando comparadas aos controles tanto feminino quanto masculino aquelas amostras que apresentaram valor de $p < 0,05$ e maior do que o valor ajustado (0,0025, GSE36552 ou 0,0023, GSE44183), não puderam ter o sexo cromossômico definido já que o teste de Bonferroni é bastante conservador e por isto apresenta grande número de erro do tipo 2, ou seja aceitar a hipótese nula quando ela é falsa.

Durante a oogênese são estocados pela mãe transcritos e proteínas que serão utilizados nas primeiras etapas do desenvolvimento, antes que o embrião seja capaz de transcrever a partir de seu próprio genoma (Schultz, 2002). A literatura mostra que em humanos há dois momentos de ativação do genoma embrionário (EGA, do inglês, *Embryonic Genome Activation*): entre os estágios de duas a quatro células observa-se a primeira EGA, ainda que discreta, e ativação mais pronunciada se dá no embrião de oito células (Vassena *et al.*, 2011; Xue *et al.*, 2013; Yan *et al.*, 2013). É importante então ter em mente que a expressão gênica em células obtidas de embriões nos estágios mais precoces do desenvolvimento é representada em parte pelos transcritos maternos, o que pode dificultar a determinação do sexo cromossômico quando baseada na contagem dos genes expressos do cromossomo Y. Este é um cromossomo muito especializado, com menos de

80 genes codificadores de proteína que pertencem a basicamente 27 famílias gênicas, das quais 11 são expressas exclusivamente nos tecidos reprodutivos masculinos. Consequentemente, a expressão de grande parte de seus genes só será mais fortemente detectada em estágios mais tardios do desenvolvimento (Skaletsky *et al.*, 2003).

Ainda assim, a sexagem da forma como foi aqui realizada se mostra bastante conveniente para nossos propósitos, principalmente para os embriões a partir de estágio de oito células, com EGA completa.

Para facilitar a compreensão das análises seguintes, a tabela 4 apresenta um resumo do que foi resultado do processo de determinação do sexo cromossômico que pode ser visualizado também na figura 8.

Tabela 4 Resultado do processo de determinação do sexo cromossômico					
GSE36552	Sexo cromossômico	Células sequenciadas	GSE44183	Sexo cromossômico	Células sequenciadas
Oócito#1	Feminino*	1	Oócito#1	Feminino*	1
Oócito#2	Feminino*	1	Oócito#3	Feminino*	1
Oócito#3	Feminino*	1	Oócito#4	Feminino*	1
Zigoto#1	?	1	Pronúcleo#2	?	1
Zigoto#2	?	1	Pronúcleo#3	?	1
Zigoto#3	?	1	Pronúcleo#4	?	1
2-cél.#1	?	2	Zigoto#1	?	1
2-cél.#2	?	2	Zigoto#2	?	1
2-cél.#3	?	2	2-cél.#1	?	1
4-cél.#1	?	4	2-cél.#5	?	1
4-cél.#2	?	4	2-cél.#6	?	1
4-cél.#3	?	4	4-cél.#1	?	1
8-cél.#1	Feminino	4	4-cél.#2	?	1
8-cél.#2	Feminino	8	4-cél.#3	?	1
8-cél.#3	Masculino	8	4-cél.#4	?	1
Mórula#1	Masculino	8	8-cél.#A	Masculino	3
Mórula#2	Masculino	8	8-cél.#B	Masculino	5
Blastoc.#1	Masculino	12	8-cél.#C	?	2
Blastoc.#2	Masculino	10	8-cél.Int.	?	íntegro
Blastoc.#3	Feminino	8	Mórula#1	Masculino	íntegro
hESCp0	Masculino**	8	Mórula#4	Masculino	íntegro
			Mórula#5	Masculino	íntegro

*não apresentam cromossomo Y

** cariótipo 46, XY

4.3. Expressão de *XIST*

Ainda utilizando os resultados de expressão (rpkm) dos estudos de Yan e colaboradores (2013) e Xue e colaboradores (2013), buscamos então identificar em que fase do desenvolvimento humano tem início a expressão de *XIST* e se seu nível varia dependendo do gênero do embrião (Figura 9 A), traçando-se também um paralelo com o que se conhece dos resultados de FISH de RNA dos trabalhos de van den Berg e colaboradores (2009) e Okamoto e colaboradores (2011) (Figura 9 B).

Figura 9 | Expressão de *XIST* ao longo do desenvolvimento embrionário humano. Em A nível de expressão do gene *XIST* em rpkm para cada célula de cada embrião a partir do estágio de oito células e da linhagem de células tronco embrionária (hESCp0) (série GSE36552); individualmente cada embrião é identificado pela fase do desenvolvimento em que se encontrava e por um número; abaixo de cada identificação está o símbolo universal de gênero indicando o sexo

cromossômico da amostra. Em **B** compilação dos resultados de FISH de RNA com uma representação esquemática do observado em células individuais por microscopia confocal; entre parênteses a porcentagem de células observadas com respectivo padrão; resultados de van den Berg *at al.* (2009) em preto e de Okamoto *at al.* (2011) em vermelho pra embriões em estágio pré-implantacional; resultados de células tronco embrionárias humanas (Silva *at al.*, 2008).

Os embriões da série GSE44183 não estão representados no gráfico de barras da figura 9 pois não se detectou a expressão de *XIST* em qualquer das amostras em nenhuma das fases submetidas a sequenciamento. A partir dos dados das amostras da série GSE36552, observa-se que a expressão de *XIST* tem início em embriões no estágio de oito células, pois sua expressão não foi detectada em nenhuma amostra que representasse estágio anterior (zigoto, duas células, quatro células ou oócito). É interessante notar que a expressão de *XIST* nos embriões femininos no estágio de oito células é bastante baixa se comparada àquela do blastocisto feminino (8-cél#1 com rpkm máxima igual a 2,28 e Blastocisto#3 com rpkm máxima igual à 23,96), corroborando as observações de van den Berg e colaboradores (2009) de 68% das células em estágio de oito células com um acúmulo pontual de RNA de *XIST*.

O processo de sexagem apresentado anteriormente indica que os embriões em estágio de mórula da série GSE36552 sejam XY. Entretanto algumas de suas células expressam *XIST* com igual ou maior intensidade do que aquelas dos embriões em estágio de oito células femininos (8-cél.#1 e 8-cél.#2), corroborando o que foi descrito pelos trabalhos utilizando técnicas de FISH de RNA (Figura 9) (van den Berg *at al.*, 2009; Okamoto *at al.*, 2011). Comparáveis também são os resultados para blastocistos masculinos: van den Berg e colaboradores (2009) observaram que 7% das células de blastocistos masculinos podem acumular RNA de *XIST* em seu único cromossomo X, ainda que em pequena quantidade, e em nossas análises de nível de expressão, apenas uma das 12 células do blastocisto #1 apresentou expressão aumentada de *XIST*.

Apesar do pequeno número de amostras e células que temos disponível para as análises, podemos dizer que a expressão de *XIST* tem início no embrião em estágio de oito células, e que este gene é expresso também em algumas células de embriões masculinos nos estágios de mórula e blastocisto.

Sabe-se que em camundongos, no início do processo de ICX aleatória um sinal discreto de *Xist* é detectado em ambos os cromossomos X de células femininas. Ao longo da diferenciação celular, a expressão de *Xist* é estabilizada, e seu sinal duplo desaparece permanecendo apenas no futuro Xi das células femininas, onde o gene tem sua expressão regulada positivamente e seu RNA acumula-se *in cis* (Avner e Heard, 2001). Um paralelo ao que acontece com camundongos pode explicar os resultados aqui expostos para embriões humanos: a detecção de transcritos de *XIST* nas células dos embriões femininos em estágio de oito células é de menor intensidade se comparada à do blastocisto #3 (XX) (Figura 9). E ainda, os resultados de van den Berg mostraram que é possível detectar um leve acúmulo de *XIST* em ambos os cromossomos nos estágios de oito células e mórula, mas que depois sua expressão é estabilizada permanecendo o acúmulo em apenas um X por célula em cerca de 90% dos casos das amostras femininas.

Parece haver uma correlação entre o início da transcrição de *Xist/XIST* em diferentes animais e a ativação do genoma do próprio embrião (EGA): camundongos, no estágio de duas células; coelhos e bovinos no estágio de oito células (revisto por Dupont and Gribnau, 2013). Em humanos esse padrão parece não ser diferente: embriões apresentam EGA mais proeminente em estágio de oito células (Vassena *at al.*, 2011; Xue *at al.*, 2013; Yan *at al.*, 2013), e nesta mesma fase a expressão de *XIST* foi detectada tanto por FISH de RNA (van den Berg *at al.*, 2009) quanto pela expressão detectada por RNA-Seq apresentada no presente estudo. Infelizmente o trabalho de van den Berg e colaboradores (2009) não apresenta resultados com embriões em estágios anteriores ao de oito células. Entretanto, a análise dos dados de expressão de embriões desde o início do desenvolvimento da série GSE36552 aqui realizada revela com clareza que em humanos a

expressão de *XIST* tem início no embrião de oito células, na mesma fase que a EGA. Não se deve descartar a hipótese de que todos os embriões da série GSE36552 em estágios anteriores ao de oito células sejam XY entretanto, teríamos que supor que nenhum dos nove embriões sequenciados nestas fases é do sexo feminino.

Como já foi discutido no item 4.1 deste capítulo, os sequenciamentos das amostras da série GSE44183 parecem menos robusto, implicando em resultados mais pobres em todas as análises subsequentes. Outra diferença fundamental entre os estudos é que o trabalho de Yan e colaboradores (2013) (GSE36552) foi mais eficiente em isolar e sequenciar quase todas as células dos embriões em estágio de duas a oito células (Tabela 1 no item 3.1), enquanto que Xue e colaboradores (2013) (GSE44183) apresentam resultados de apenas uma célula por embrião, à exceção daqueles no estágio de oito células. Das quatro amostras em estágio de oito células da série GSE44183 duas são masculinas, 8-cél#A e 8-cél#B e tiveram apenas três e cinco células sequenciadas, respectivamente (Tabela 4). As outras duas amostras neste mesmo estágio não puderam ter o sexo cromossômico determinado com clareza (Figura 8, tabela 4 e discussão ao longo do item 4.2) e ainda ou tiveram apenas duas células sequenciadas (8-cél.#C), ou foram sequenciadas sem dissociação (8-cél. Int.). O estudo de Xue e colaboradores (2013) também não apresenta nenhum embrião feminino em estágio de mórula, ou blastocisto sequenciado.

4.4. Desvios de alinhamento em favor do alelo de referência

Estudos *in silico* envolvendo grandes quantidades de dados estão sujeitos a erros e desvios que podem ficar mascarados no montante de resultados. Uma crítica que se pode fazer ao presente trabalho é que este não teve acesso ao material genômico das amostras embrionárias e portanto tivemos que basear a genotipagem em evidências indiretas de genes com expressão bialélica e genes com expressão monoalélica aleatória. O processo de alinhamento das *reads* se baseia na identidade de suas sequências com aquela do genoma de referência, e este processo pode favorecer o alinhamento de *reads* que sejam idênticas

à sequência de referência, pois penaliza a qualidade de alinhamento quando os nucleotídeos não são os mesmos. Assim, dependendo dos parâmetros utilizados no passo de alinhamento pode-se perder informações sobre *reads* em posições de SNPs que representem o alelo alternativo (subitem 3.3.3). Mesmo que os parâmetros sejam modificados, Degner e colaboradores (2009) reportam que um desvio de alinhamento em favor do alelo de referência pode ainda assim ser observado.

Com a intenção de verificar se esse tipo de desvio incorre na presente análise, para cada SNP anotado que indicasse expressão bialélica nos genes autossômicos foi calculada a proporção entre *reads* representando o alelo de referência e o alelo alternativo. São considerados SNPs representando expressão bialélica aqueles com razão entre os alelos dentro do intervalo de 0,2 a 0,8 de acordo com o descrito na literatura (Amos-Landgraf *et al.*, 2006; Borel *et al.*, 2015).

Dentre todas as células sequenciadas da série GSE36552, em média 52,29% das *reads* alinharam no alelo de referência (Anexo 5 apresenta o valor de cada uma). Em teoria, genes de expressão bialélica devem ser expressos igualmente entre os alelos, mas biologicamente é difícil de calcular se há algum pequeno desvio (Degner *et al.*, 2009). Como os métodos de análises do presente trabalho se baseiam na identificação de genes de expressão bialélica ou mono alélica de maneira discreta, um pequeno desvio em favor do alelo de referência não apresenta grande influência nas conclusões.

4.5. Inativação do cromossomo X pela análise de expressão alelo específica

4.5.1. Genotipagem

Uma etapa fundamental deste trabalho foi a anotação de polimorfismos nos genes expressos de cada célula sequenciada nos trabalhos de Yan e colaboradores (2013) e Xue e colaboradores (2013). A anotação propriamente dita deu-se para cada célula de cada embrião individualmente, utilizando-se o banco de dados de SNPs personalizado para este

projeto (Tabela 3 e anexos 3 e 4). No presente trabalho, a presença de polimorfismos foi utilizada como ferramenta para a identificação de genes com expressão bialélica ou de expressão monoalélica diferencial ou seja, em um mesmo indivíduo algumas células expressam um alelo de um determinado gene e outras expressam o outro alelo para esse mesmo gene. Entretanto, para se buscar por genes de expressão bialélica ou com expressão monoalélica diferencial, é necessária a identificação de variantes em heterozigose. Para tanto, primeiramente identificou-se aqueles SNPs anotados que estivessem presentes nas diferentes células de um mesmo embrião. Quando uma célula apresentou um alelo expresso (A) e outra célula do mesmo embrião o outro alelo (A'), foi possível concluir que aquele embrião era heterozigoto (A|A') para aquele determinado polimorfismo. SNPs que sugerem expressão bialélica (A/A') também indicam que a amostra é heterozigota.

Na figura 10 está esquematizado um caso ilustrativo do processo de genotipagem baseado em três genes do cromossomo X. No exemplo, o padrão de expressão de cada uma das oito células do Blastocisto#3 foi determinado pela presença de SNPs para os genes *VCX*, *ATP6AP2* e *VBP1*. As variantes rs151212725 no gene *VCX*, rs192230843 no gene *ATP6AP2* e rs11544264 em *VBP1* não são informativas, pois todas as células expressam o mesmo alelo (A), sendo provavelmente homozigotas para essas variantes; já os SNPs rs5963816 e rs1060063 em *ATP6AP2* indicam que as células identificadas como #3, #5 e #7 tem expressão bialélica (A/A'; círculo cinza); as células #1 e #2 tem expressão monoalélica para um alelo (A'; círculo branco); e finalmente a célula #8 tem expressão monoalélica para o outro alelo (A; círculo preto). Assim, o blastocisto#3 é considerado heterozigoto para os SNPs rs80291436 em *VCX*, rs5963816 e rs1060063 em *ATP6AP2* e rs11887 em *VBP1*.

Figura 10| Genotipagem através da identificação de variantes em regiões expressas. Acima, **Blastocisto#3**, teve o genótipo determinado de acordo com o padrão de expressão de cada uma de suas células (numeradas de **#1** a **#8**). Nas células os círculos em preto e em branco indicam expressão monoalélica dos diferentes alelos: A e A' respectivamente; os círculos cinza indicam expressão bialélica. O genótipo do blastocisto #3 segue o mesmo padrão de cores, sendo que os círculos em preto e em branco indicam variante em homozigose dos diferentes alelos: A e A' respectivamente; os círculos em cinza indicam heterozigose. "X"; variante não detectada; "/" espaçamento arbitrário entre os genes. A posição dos genes é meramente ilustrativa.

Dessa maneira, foram genotipados os genes de cada cromossomo autossomo e do cromossomo X, fora da região pseudoautossômica, para cada embrião com sexo cromossômico determinado como XX, para então seguir com a investigação sobre a ICX (Tabela 5).

A tabela 5 apresenta os resultados da genotipagem para cada cromossomo dos embriões 8-cél.#1, 8-cél.#2 e Blastocisto#3. Nela é possível observar quantos genes foram considerados heterozigotos e homozigotos para cada cromossomo. No cromossomo X foram identificados 33, 70 e 45 genes informativos nos embriões 8-cél.#1, 8-cél.#2 e Blastocisto#3, respectivamente.

Tabela 5 Genotipagem dos embriões femininos por cromossomo						
Amostra	8-cél.#1		8-cél.#2		Blastocisto#3	
	HTR	HOM	HTR	HOM	HTR	HOM
crom.X	33	167	70	236	45	214
crom.1	163	458	309	548	253	577
crom.2	109	309	238	413	163	408
crom.3	83	257	186	320	144	315
crom.4	46	174	117	248	91	207
crom.5	68	200	147	290	112	253
crom.6	67	210	157	286	108	266
crom.7	77	192	168	255	114	248
crom.8	59	149	120	242	83	213
crom.9	56	167	123	211	102	226
crom.10	55	177	125	231	117	212
crom.11	67	233	153	306	118	329
crom.12	92	232	173	331	133	301
crom.13	19	97	68	108	44	95
crom.14	61	148	127	184	83	187
crom.15	39	156	109	192	64	186
crom.16	58	201	127	234	99	255
crom.17	96	241	165	292	134	343
crom.18	19	82	55	112	35	87
crom.19	122	246	196	312	151	358
crom.20	45	112	85	140	76	155
crom.21	9	41	27	58	26	51
crom.22	30	89	65	115	73	132

"HTR": heterozigoto; "HOM": homozigotos; "crom.": cromossomo

4.5.2. Padrão de expressão dos genes informativos

A partir dos resultados da genotipagem, buscou-se saber então, para cada célula individualmente, quantos genes do cromossomo X apresentavam expressão bialélica ou monoalélica diferencial. É reconhecido que o início da ICX é dependente da expressão de *XIST*, mas o processo como um todo depende de diversos fatores epigenéticos e a simples expressão de *XIST* não indica necessariamente que todos os genes do cromossomo X já

estejam inativados (Okamoto *at al.*, 2011). Em humanos tampouco há informações sobre a sincronia entre as células em relação à ICX, *i.e.*, se todas as células de um determinado embrião iniciam e estabilizam o processo de ICX ao mesmo tempo ou se é possível que a ICX já esteja estabilizada em algumas células e em outras não. Uma maneira simples de se averiguar a ICX em células individuais é através da contagem do número de genes do cromossomo X com expressão bialélica. Espera-se que para uma célula com silenciamento transcricional dos genes do cromossomo X a proporção de genes informativos com expressão monoalélica seja maior do que aquela dos genes autossômicos. Para melhor representar a proporção entre genes heterozigotos de expressão monoalélica e bialélica foram construídos os gráficos expostos na figura 11.

Figura 11| Genes informativos de expressão monoalélica e bialélica. Gráficos de pizza representando a proporção encontrada entre genes informativos de expressão bialélica (preto) e monoalélica (branco) para cada célula dos embriões femininos da série GSE36552 nos estágios de

oito células (**8-cél.#1**, **8-cél.#2**) e blastocisto (**Blastocisto#3**). À esquerda, genes do cromossomo X fora da região pseudoautossômica; à direita genes autossômicos.

Primeiramente verificou-se se para cada célula individualmente se a proporção entre genes informativos de expressão monoalélica e bialélica do cromossomo X era semelhante àquela dos autossômicos, aplicando-se o teste do χ^2 (Tabela 6) .

Tabela 6 Comparação entre o número de genes com expressão monoalélica no cromossomo X e nos autossomos						
Amostra	célula	cromossomo X		Autossomos		χ^2 valor de p
		Bilélicos	Monoalélicos	Bilélicos	Monoalélicos	
8cél.#1	#1	22	8	964	231	0,317
	#2	21	8	921	237	0,349
	#3	19	4	813	212	0,700
	#4	23	7	955	308	0,894
8cél.#2	#1	18	13	867	772	0,568
	#2	25	9	959	785	0,031
	#3	21	15	881	700	0,755
	#4	28	11	1103	864	0,050
	#5	19	32	1128	1057	0,042
	#6	21	30	1204	984	0,049
	#7	11	11	923	637	0,385
	#8	18	20	1029	772	0,229
Blastocisto#3	#1	5	13	335	601	0,482
	#2	11	13	741	712	0,616
	#3	13	12	878	748	0,842
	#4	9	16	497	755	0,708
	#5	11	12	690	723	0,924
	#6	5	13	464	656	0,243
	#7	15	16	875	715	0,462
	#8	12	14	701	819	0,997

Os resultados dos testes expostos na tabela 6 indicam que apenas as células #2, #5 e #6 do embrião #2 no estágio de oito células apresentam diferença entre a proporção de genes de expressão monoalélica no cromossomo X e nos autossomos ($p < 0,05$; destacado em vermelho), sendo a célula #2 a que menos apresentou genes com expressão monoalélica no cromossomo X (26%) e as células #5 e #6 as que mais apresentaram (63% e 59%, respectivamente). Entretanto, se aplicada a correção para testes múltiplos de Bonferroni essa diferença deixa de ser significativa (usaria-se $p < 0,0025$). Assim, a simples verificação da proporção entre genes informativos de expressão monoalélica e bialélica em

uma célula individual não foi capaz de indicar se os transcritos do cromossomo X já estão sendo inativados nos estágios de oito células e blastocisto.

Buscou-se ainda verificar se havia diferença estatística em relação ao número de genes com expressão monoalélica do cromossomo X entre as células de um mesmo embrião. Novamente apenas entre as células do embrião 8-cél.#2 o teste do χ^2 acusou diferença significativa ($p= 0,005$). Espera-se que com a ICX existam menos genes informativos de expressão bialélica no cromossomo X do que quando comparados aos autossomos, entretanto como não houve diferenças significativas entre genes de expressão monoalélica e bialélica entre o X e os autossomos não se pode afirmar com esses resultados que o silenciamento transcricional dos genes do cromossomo X já tenha se estabelecido.

Uma observação inesperada foi o elevado número de genes informativos com expressão monoalélica entre os autossômicos. É senso comum acreditar que estes genes sejam expressos a partir de ambos os alelos. Entretanto, recentemente foi publicado um estudo baseado no sequenciamento do transcriptoma de células adultas humanas individuais mostrando que cerca de 76% dos genes autossômicos apresentam expressão monoalélica (Borel *et al.*, 2015). Os níveis aqui apresentados não passam de 64% (Blastocisto#3 célula #3), mas isso pode ser reflexo da reestruturação epigenética que as células ainda embrionárias estão passando, como discutido mais adiante. O alto número de genes informativos com expressão monoalélica preferencial em autossomos demonstrado no presente estudo ajuda então a estender os recentes achados de Borel e colaboradores (2015) para o início do desenvolvimento embrionário humano.

4.5.3. Padrão de inativação do cromossomo X

No intuito de verificar se a ICX em humanos é dependente da origem parental, buscou-se evidências de **expressão monoalélica “imprintada”** ou **expressão monoalélica aleatória** dos genes do cromossomo X. A exemplo do que acontece com os genes autossômicos de expressão monoalélica, a inativação do X deve seguir um ou outro padrão (ICX aleatória ou ICX “imprintada”). Assim, a observação no embrião como um todo do padrão de expressão de cada um de seus genes deve indicar o padrão global de ICX. Parte-se da premissa de que o padrão esperado de ICX “imprintada” deve seguir o padrão do que ocorre com células masculinas, que por possuírem apenas o cromossomo X de origem materna, todas as variantes anotadas no cromossomo X indicam expressão de um único alelo.

Assim, de forma separada para genes autossômicos e do cromossomo X, foram calculados quantos genes exibiam expressão bialélica e, daqueles genes informativos, foram calculados quantos apresentavam expressão monoalélica. Os dados de expressão de todas as células de uma mesma amostra foram reunidos no intuito de se representar o espécime como um todo.

A figura 12 representa em gráficos de dispersão os resultados obtidos das análises do padrão de ICX para os embriões do estudo GSE36552.

Figura 12 | Análise do estabelecimento e do padrão de inativação do cromossomo X durante o desenvolvimento pré-implantacional humano. Resultados obtidos das variantes anotadas por amostra da série GSE36552. Gráficos na linha de cima: “**Todas as amostras**”, sem distinção entre machos e fêmeas; na do meio: “**Amostras femininas***”, aquelas classificadas como XX de acordo com o processo de genotipagem; em baixo: “**Amostras masculinas***”, aquelas classificadas como XY

pelo processo de genotipagem; *, embriões em estágio de duas a quatro células nos quais não é possível determinar o sexo cromossômico por padrão de expressão de genes do cromossomo Y foram utilizados em todas as análises. Na abscissa: amostras agrupadas por ordem de desenvolvimento; hESCp0, linhagem de célula tronco masculina na passagem zero. Em preto está razão entre o número de genes com expressão bialélica pelo número total de genes em cada categoria (autossomos ou cromossomo X) com curva de regressão e valores do teste de correlação; em vermelho, também por categoria está a razão entre o número de genes informativos com expressão monoalélica diferencial pelo número total de genes informativos com as respectivas curvas de regressão e valores do teste de correlação atribuídos.

Para os genes autossômicos, os resultados expostos nos gráficos da figura 12 mostram que não há correlação entre a proporção de genes bialélicos (em preto) e o estágio do desenvolvimento. Entretanto, o número de genes informativos com expressão monoalélica diferencial (em vermelho) parece aumentar de acordo a fase embrionária. Como já foi discutido, o RNA materno estocado durante a oogênese pode ser detectado em estágios iniciais do desenvolvimento, sendo ainda abundantes nas fases de duas a quatro células; recentemente, foi demonstrado que a expressão monoalélica de genes autossômicos é um fenômeno epigenético bastante comum podendo afetar cerca de 76% dos genes humanos em células adultas (Borel *et al.*, 2015). Isso pode explicar o aumento observado no número de genes autossômicos com expressão monoalélica diferencial, mais marcante a partir do estágio de oito células (Figura 12: categoria 'Autossomos'). Este fenômeno de regulação para expressão monoalélica dos genes autossômicos está presente tanto em células masculinas quanto femininas e pode ser igualmente detectado utilizando todos os embriões ou mesmo separando-os por sexo cromossômico (correlação $> 0,7407$ e valor de $p < 0,005$ para todas as análises).

Da mesma forma, foram avaliados os resultados para os genes do cromossomo X. Observando-se a figura 12, fica claro que para as células masculinas a proporção de genes com expressão bialélica diminui drasticamente com o desenvolvimento - o gráfico mostra uma forte correlação negativa entre a proporção de genes bialélicos do cromossomo X e o estágio do desenvolvimento (correlação = $-0,8754$, valor de $p = 0,0001$, em preto). Além

disso, não há correlação entre a proporção de genes de expressão monoalélica diferencial e a fase do desenvolvimento (correlação = 0,0271, valor de $p = 0,9333$, em vermelho). Esses padrões são esperados para células masculinas, pois com a degradação do RNA materno as células terão apenas um cromossomo X de onde seus genes serão expressos, fenômeno observado notadamente nos embriões a partir do estágio de oito células.

Para embriões femininos, consideramos três hipóteses possíveis e os respectivos padrões de expressão gênica:

–ICX ainda não se estabeleceu - a proporção de genes informativos do cromossomo X com expressão monoalélica diferencial deve permanecer baixa e sem correlação com as fases embrionárias; e a proporção de genes com expressão bialélica deve se manter alta, resultado da presença de dois X ativos (Figura 4 **A**, item 3.8);

–ICX “imprintada” - a proporção de genes com expressão bialélica deve diminuir acentuadamente; e a proporção de genes informativos do X com expressão monoalélica diferencial deve permanecer baixa e sem correlação com as fases embrionárias, ou seja, o padrão deve obedecer àquele observado em amostras masculinas onde todas as células expressam os genes de um mesmo cromossomo X (Figura 4 **B**, item 3.8);

–ICX aleatória - a proporção de genes com expressão bialélica deve diminuir; e a proporção de genes informativos do X com expressão monoalélica diferencial deve aumentar de acordo com as fases embrionárias, resultado da presença de um X ativo e um inativo (Figura 4 **C**, item 3.8).

O resultado obtido de genes do cromossomo X para células masculinas obedece ao padrão esperado mostrando o poder da técnica de detecção de ICX ao longo do desenvolvimento (Figuras 12 e 4 **B**, item 3.8). A proporção de genes do X com expressão bialélica (em preto) diminui consideravelmente à medida em que se esgotam os recursos de mRNA materno estocados na oogênese e com a EGA (estágio de oito células); e ainda, a

proporção de genes do X com expressão monoalélica diferencial (em vermelho) se mantém baixa e sem correlação com as fases embrionárias.

Já para as células femininas, os resultados observados para os genes do cromossomo X (Figura 12) não parecem se adequar a nenhuma das hipóteses descritas acima (Figura 4, item 3.8). O que se observa é que a proporção de genes com expressão bialélica (em preto) se mantém ao longo do desenvolvimento. Esse efeito pode ser explicado pela presença de dois X ativos principalmente a partir do estágio de oito células. Contudo, a proporção de genes informativos do X com expressão monoalélica diferencial aumenta significativamente (correlação = 0,8282 e $p = 0,0058$, em vermelho), indicando um possível início de inativação transcricional dos genes daquele cromossomo de maneira aleatória.

Pode-se supor que o número de genes com expressão bialélica diminui com o estágio do desenvolvimento devido a algum artefato gerado pelo método aplicado. Dado que a EGA é mais acentuada no estágio de oito células verificou-se se embriões mais tardios apresentavam um número maior de *reads* alinhadas, e por consequência um maior número de *reads* cobrindo a região de um SNP o que diminuiria a chance de considerarmos falso positivos - no caso genes de expressão bialélica. Foi averiguada então a possibilidade do número de *reads* variar de acordo com o estágio do desenvolvimento. Para tanto foi gerado um gráfico de dispersão relacionando-se o número de *reads* alinhadas com a fase em que o embrião se encontrava (Figura 13).

Figura 13| Número total de *reads* alinhadas em relação ao estágio do desenvolvimento. As células de cada embrião estão agrupadas por estágio, de oócito a blastocisto ou células da linhagem de células tronco embrionárias na passagem zero (hESCp0) do estudo GSE36552 (Yan *et al.*, 2013). A curva de regressão encontra-se em vermelho e os valores do teste de correlação no canto superior direito do gráfico, também destacado em vermelho.

Os valores apresentados na figura 13 apresentam uma correlação inversa (correlação = -0,4296) e o valor de p é muito baixo ($p = 0,00001$) indicando que o teste é significativo e as amostras apresentam menos *reads* alinhadas ao longo do desenvolvimento. Foi demonstrado que um oócito humano pode ter cerca de 330 pg de mRNA estocados durante a oogênese e uma típica célula adulta de mamífero tem cerca de 10 a 30 pg de mRNA (Kocabas *et al.*, 2006). Assim, embriões em estágios mais avançados não apresentam mais *reads* alinhadas e a causa para a número de genes com expressão bialélica diminuir ao longo do desenvolvimento é mais provavelmente uma característica biológica do que um artefato metodológico.

5. Razão da expressão global entre genes do cromossomo X e autossômicos

Foi demonstrada que a expressão dos genes do único X ativo em células adultas (sejam femininas já com ICX: XaXi, ou masculinas com um único cromossomo X: XaY) está elevada atingindo-se um equilíbrio transcricional entre o cromossomo X ativo e autossomos (hipótese proposta por Susumo Ohno em 1967). Consequentemente, a razão da transcrição global entre cromossomo X e autossomos (X:A) é próxima a 1 (Gupta *et al.*, 2006; Johnston *et al.*, 2008; Nguyen and Disteché, 2006; Talebizadeh *et al.*, 2006). De fato, linhagens de hESCs femininas que apresentam ICX possuem uma razão X:A próxima a 1, assim como células de linhagens masculinas. Todavia, aquelas linhagens que quando indiferenciadas possuem dois X ativos (XaXa) exibem uma razão X:A significativamente mais alta do que as XaXi (razão X:A próxima a 1,3) (Bruck and Benvenisty, 2011).

Com a intenção de se averiguar se o cromossomo X já estaria inativado em algum embrião sob essa perspectiva, foi calculada a razão entre o nível de expressão global dos genes do cromossomo X em relação aos autossômicos em cada embrião. Para tanto foram obtidos os valores de transcrição relativa (rpkm) das amostras da série GSE36552, e os valores resultantes da razão X:A foram utilizados na construção do gráfico da figura 14.

Figura 14| Razão da expressão global dos genes do cromossomo X pela dos autosossômicos por embrião. Cada ponto no gráfico representa um embrião individual; à exceção dos oócitos e dos zigotos que foram utilizados em conjunto. A partir dos dados de expressão (rpkm) foi calculada a mediana do nível de expressão dos genes do X e dos autosossômicos. Os pontos no gráfico foram obtidos da razão média dos genes do cromossomo X por aqueles dos autosossomos. O erro padrão está evidenciado em vermelho. hESCp0 representa a linhagem de células tronco embrionárias masculina (46, XY).

Nenhuma de nossas análises resultou em razão X:A próxima a 1,3 como sugerido por Bruck e Benvenisty (2011) para células femininas com dois X ativos. O maior valor aqui observado foi para os embriões femininos no estágio de oito células (Figura 14: 8-cél#1 e 8-cél#2; X:A próximo a 1,0). A linhagem masculina de células tronco embrionárias aqui utilizada apresentou valor para X:A de 0,86 (Figura 14: hESCp0) enquanto as de Bruck e Benvenisty (2011) apresentaram X:A entre 0,90 e 1,0. Não é adequado comparar nossos valores diretamente aos de Bruck e Benvenisty (2011) pois estes utilizam resultados de experimentos de *microarray* com mRNA obtido de várias células juntas, enquanto os dados aqui apresentados se baseiam em RNA-Seq de células únicas. Estudos de expressão por *microarray* são limitados aos genes presentes na lâmina de hibridação, enquanto que o sequenciamento do transcriptoma total tem o potencial de representar todos os transcritos presentes na amostra.

Optou-se então por verificar a possibilidade do cromossomo X estar inativado em alguma célula em particular, calculando-se a razão entre o nível de expressão global dos genes do cromossomo X em relação aos autossômicos para cada célula individualmente. Para tanto foram utilizados os valores de transcrição relativa (rpkm) das amostras da série GSE36552, e os valores resultantes foram utilizados na construção do gráfico da figura 15.

Oócito 2 células: emb1 4 células: emb1 8 células: emb1 Mórula: emb1 Blastocisto: emb1 hESC: p0
 Zigoto emb2 emb2 emb2 emb2 emb2 emb2
 emb3 emb3 emb3 emb3 emb3 emb3

Figura 15| Razão da expressão global dos genes do cromossomo X pela dos autossômicos para cada célula. Cada ponto no gráfico representa uma célula individual. A partir dos dados de expressão (rpkm) foi calculada a mediana do nível de expressão dos genes do cromossomo X e dos autossômicos e os valores plotados foram obtidos do cálculo da razão X:A. Os estágios estão separados pelas linhas tracejadas e as fases e a identificação de cada amostra são mostradas na parte de baixo. “hESC”, linhagem masculina de células tronco indiferenciadas na passagem zero (p0).

As amostras de oócitos e zigotos possuem apenas o RNA materno estocados durante a oogênese, e a razão X:A se apresenta perto de 0,75. Já os embriões em estágio de duas células identificados como #2 e #3 apresentam um tímido aumento da razão X:A que não é observado no embrião #1. Foi descrito que durante a embriogênese humana observa-se uma discreta ativação do genoma embrionário (EGA) entre os estágios de duas a quatro células (Vassena *at al.*, 2011; Xue *at al.*, 2013; Yan *at al.*, 2013). Com base nas figuras 14 e 15 é possível propor uma falta de sincronia de EGA para os embriões no estágio de duas células pois, a razão X:A é variável, isso pode ser resultado de uma sutil diferença da idade entre eles que não pode ser observada morfológicamente (Figura 14: 2cél#1 X:A = 0,76; 2cél#2 X:A = 0,87; 2cél#3 X:A = 0,90).

Apesar da falta de informação sobre o sexo cromossômico dos embriões em estágio de quatro células é coerente propor, baseado nos resultados das figuras 14 e 15, que as amostras 4-cél#1 e 4-cél#3 apresentem dois X ativos pois se comparadas ao embrião 4-cél#2 a razão X:A é maior (Figura 14: 4-cél#1 X:A = 0,93; 4-cél#3 X:A = 0,99; 4-cél#2 X:A = 0,76). Nos embriões em estágio de oito células classificados como XX (8-cél.#1 e 8-cél.#2), todas as células apresentam razão X:A maior do que 0,88 (Figura 15) reflexo da presença de dois X ativos (também observado na figura 14: 8-cél#1 X:A = 1,02; 8-cél#2 X:A = 0,97). Já o espécime 8-cél.#3 classificado como XY, tem cinco de suas oito células com razão média menor que 0,8 (Figura 15, também observado na figura 14: 8-cél#3 X:A = 0,84).

Mais interessante ainda são os resultados para os embriões em estágio de blastocisto quando comparados aos das células tronco indiferenciadas (hESCp0). Essa linhagem foi estabelecida a partir da massa celular interna de um blastocisto maduro (46,XY), e teve suas células sequenciadas ainda na passagem zero. Podemos assumir que esta deva ser a melhor representação do cenário transcricional de células desta fase do desenvolvimento embrionário. A média da razão X:A das células de hESCp0 é 0,86. Já as dos blastocistos masculinos (Blastocistos #1 e #2) estão próximas de 0,69 e aquela do blastocisto #3, feminino, é de 0,94 (Figura 15).

Neste momento faz-se necessária a retomada do que foi proposto por Susumo Ohno: a hipótese da super-regulação de genes do cromossomo Xa (Ohno, 1967). Esta hipótese propõe que a compensação de dose envolvendo os cromossomos sexuais de mamíferos eutérios ocorra em duas etapas: a inativação de um cromossomo X nas fêmeas e a regulação aumentada de genes do cromossomo X ativo tanto nos machos quanto nas fêmeas. Se comparados os níveis de expressão global entre genes do cromossomo X e autossômicos, a razão X:AA deveria ser um valor próximo a 1 em células adultas de ambos os gêneros. Caso contrário se a super-regulação dos genes no X não for observada espera-se uma razão X:AA próxima a 0,5.

De acordo com a hipótese de Ohno, a razão X:A observada para as células da hESCp0 (Figuras 14 e 15) é decorrência da regulação transcricional aumentada do seu único X. Entretanto, os valores observados para os outros blastocistos masculinos (Blastocistos #1 e #2) não acompanham o padrão da hESC. Já o Blastocisto#3 tem todas as suas células com razão acima de 0,80 e a razão média entre elas é de 0,94 (Figuras 14 e 15), o que seria esperado segundo Bruck e Benvenisty (2011) para células femininas com inativação do X completa. Entretanto, os dados do subitem 4.5.3 sugerem que a ICX no Blastocisto#3 já tenha tido início mas não se estabelecido (Figura 12). Assim, é coerente propor que o Blastocisto#3 apresenta razão X:A alta devido à presença dos dois X ativos (como os embriões femininos em estágio de oito células) e os Blastocistos #1 e #2 apresentem a razão X:A mais baixa na maioria de suas células pois ainda não estabeleceram a super-regulação de seu X ativo.

Trabalhos que refutam a hipótese de Ohno reportam que a razão X:A observada para uma célula adulta seja próxima de 0,5 pois não haveria compensação de dose entre o Xa e os cromossomos autossômicos (revisto em Pessia *et al.*, 2014). Toda via, os dados apresentados aqui não mostram nenhum dos embriões ou hESCs com razão X:A próxima a 0,5 à exceção da Mórula#2 - esta entretanto apresenta um grande erro padrão se comparada às outras amostras (Figura 14).

O padrão explicitamente heterogêneo observado nos embriões em estágio de mórula, em especial a Mórula#2 que apresenta quatro células com razão X:A menor do que 0,17 e grande erro padrão (Figuras 14 e 15), levantou a suspeita de que talvez esses embriões representassem casos de aneuploidias. A falta de acesso ao material genômico dos embriões dificulta a verificação direta de casos de aneuploidias em qualquer das células analisadas. Entretanto, é possível utilizar a razão entre expressão global de cada cromossomo individualmente pela expressão global dos genes autossômicos. Sabe-se que em casos de câncer genes apresentam maior nível de expressão quando duplicados (Lopez-Gines *et al.*, 2010) – assim, assumiu-se aqui que se algum cromossomo em particular apresentar uma razão expressivamente maior ou menor do que a média para os outros cromossomos, é razoável supor que a célula seja aneuploide.

A figura 16 apresenta o resultado desta análise para os embriões em estágio de mórula. O padrão observado para as células da hESCp0 foi utilizado como controle, já que esta teve seu cariótipo determinado como normal (46,XY) (Yan *et al.*, 2013).

Figura 16| Verificação da presença de células aneuploides nos embriões em estágio de mórula. Cada linha no gráfico representa uma célula individual. A partir dos dados de expressão (rpkm) foi calculada a mediana do nível de expressão dos genes de cada cromossomo separadamente e os valores plotados foram obtidos do cálculo da razão de cada um deles com a mediana da totalidade dos genes autossômicos. “hESCp0”, linhagem masculina de células tronco indiferenciadas na passagem zero com cariótipo normal (46,XY).

Por ser a única dentre as amostras utilizadas no presente trabalho com cariótipo determinado como normal (46,XY), o padrão observado para as hESC na passagem zero (p0) foi utilizado como controle do esperado para células normais. Apesar de apresentar alguma variação, todas as células sequenciadas da hESCp0 mantêm a razão no intervalo entre 0,5 e 1,5 para todos os cromossomos. Já a célula #8 da Mórula#1 possui razão maior que 1,5 ou menor que 0,5 para todos os seus cromossomos a exceção dos cromossomos 3, 9, 11 e 21; em especial os cromossomos 4, 6, 8, 13 e 18 com razão acima de 2,0. Ainda dentre as células da Mórula#1 a célula #1 apresentou razão igual a 0,18 para o cromossomo 10 e a célula #3 apresentou razão menor que 0,5 para os cromossomos 16, 19 e 20 e maior que 1,85 para os cromossomos 6, 13, 15 e 18 (Figura 16). Estas células são então fortes candidatas a apresentarem aneuploidia. Seguindo-se o mesmo tipo de análise pode-se especular alguma anomalia envolvendo o cromossomo X nas células #3, #4, #6 e #7 da Mórula#2.

Na busca por eventuais desvios de alinhamento em favor do alelo representado no genoma de referência (Hg19 e refmRNA) no item 4.4 deste capítulo, duas células apresentaram um desvio bastante expressivo: as células #3 e #8 da Mórula#1, com mais de 57% de representação do alelo de referência naqueles SNPs que indicavam expressão bialélica em genes autossômicos (Anexo 5). Porém, se consideradas todas as células em conjunto à exceção das células #3 e #8 da Mórula#1, a média de desvio para o alelo de referência é de 52,18%, corroborando a hipótese de que aquelas células tenham aneuploidias ou deleções.

O resultado de variantes anotadas no cromossomo X (Figura 6 **C** no item 4.1) aponta para quatro *outliers* inferiores, células da Mórula#2 (estudo GSE36552) que tiveram um número expressivamente menor de SNPs identificados naquele cromossomo: célula #3 com 451 SNPs anotados em genes do cromossomo X, célula #4 com 306, célula #6 com 302 e célula #7 com apenas 144 SNPs (Anexo 3). Estas mesmas células aparecem novamente na figura 16 (Mórula#2) com uma razão X:A menor que 0,17. De maneira independente, os resultados expostos nas figuras 6 **C** e 16 apontavam para alguma anomalia nessas quatro células da Mórula#2, e juntos corroboram a hipótese de deleções ou mesmo aneuploidia envolvendo o cromossomo X. Os resultados da análise de aneuploidia de todas as outras amostras embrionárias encontram-se no anexo 6. Não foi possível identificar com clareza nenhum padrão aberrante derivado dos resultados da razão dos cromossomos individuais.

Apesar de apenas um blastocisto ter tido o sexo cromossômico determinado como XX, os resultados apresentados neste capítulo contribuem significativamente para a compreensão do início e estabelecimento da inativação do cromossomo X em humanos. Alguns poucos estudos tentaram observar o início da transcrição de *XIST* em estágios pré-implantacionais: por PCR convencional, Daniels e colaboradores (1997) e Ray e colaboradores (1997) detectaram a expressão de *XIST* em embriões em diferentes fases pré-implantacionais tanto femininos quanto masculinos. Porém, na PCR convencional o grande número de ciclos utilizados na amplificação do material obtido pode mascarar a verdadeira intensidade de sinal resultado da expressão de *XIST* nos diferentes estágios e gêneros, o que pode justificar a detecção de RNA de *XIST* já em zigotos por Daniels e colaboradores (1997). Todavia, ambos os trabalhos relatam aumento de expressão de *XIST* dependente da fase do desenvolvimento (Daniels *et al.*, 1997; Ray *et al.*, 1997).

Uma técnica mais sensível é a de FISH de RNA, utilizada por van den Berg e colaboradores (2009) e Okamoto e colaboradores (2011). Apesar de concordarem em alguns aspectos, os resultados desses trabalhos divergem em outros. Em blastocistos femininos, van den Berg e colaboradores (2009) encontraram 90% das células com acúmulo de *XIST* em um único cromossomo X; já Okamoto e colaboradores (2011) observaram dois acúmulos de sinal moderado de *XIST* em 85% das células (Figura 9). Os resultados divergentes podem ser explicados por variações nas condições de cultura e qualidade dos embriões utilizados, bem como em diferenças nos protocolos dos ensaios de FISH de RNA (Payer *et al.*, 2011). Porém, desde então não houve novas publicações sobre a questão do início da ICX em humanos.

A tecnologia de RNA-Seq utilizada no presente trabalho se mostrou bastante sensível e corroborou os achados dos estudos de FISH de RNA de *XIST* de van den Berg e colaboradores (2009), pois sua expressão foi detectada apenas em embriões a partir do

estágio de oito células (Figura 9). Porém, não podemos afirmar se o gene é expresso a partir de um único ou dos dois cromossomos X, pois não foram detectados polimorfismos informativos em *XIST*.

A identificação de genes do cromossomo X que apresentam expressão monoalélica independente da origem parental foi de fundamental importância para se elucidar as questões ainda existentes sobre o estabelecimento e o padrão de ICX em estágios iniciais do desenvolvimento pré-implantacional humano. O gráfico construído com os embriões femininos mostra claramente que há genes já inativados no cromossomo X (Figura 12, em vermelho). Além disso, esta inativação é independente da origem parental, pois existem células que expressam um alelo enquanto outras, de um mesmo embrião, expressam o outro alelo, ao contrário do observado nas células masculinas, onde todas as células expressam predominantemente o mesmo alelo.

O sequenciamento em larga escala do transcriptoma de células únicas pode apresentar muita variação devido à escassa quantidade de material inicial. Mas os dados apresentados aqui são bastante consistentes principalmente entre as amostras do estudo GSE36552, se observarmos as diferentes células de um mesmo embrião ou ainda os oócitos e zigotos nas diversas análises feitas ao longo deste capítulo.

Podemos assim concluir que o gene responsável pelo início da inativação do cromossomo X em humanos, *XIST*, começa a ser expresso em embriões no estágio de oito células, coincidindo com a ativação do genoma do embrião; este gene pode ter a sua expressão detectada tanto em embriões femininos quanto masculinos, mas sua expressão é estabilizada na fase de blastocisto sendo altamente expresso principalmente em embriões femininos (Figura 9). Entretanto, apesar da expressão de *XIST* aumentada em blastocisto, o silenciamento transcricional de todo o cromossomo X ainda não pode ser observado nas células de embriões desta fase. Mesmo assim, a análise de expressão alelo-específica mostrou que alguns genes do X já apresentam expressão monoalélica, e que a ICX ocorre de forma aleatória (Figura 12).

Adicionalmente, as análises de expressão global, relacionando o nível de expressão dos genes do X com aqueles autossômicos, indicam que como proposto por Susumo Ohno (1967) células masculinas apresentam um aumento de expressão dos genes ligados ao X – resultado do mecanismo de compensação de dose entre o seu único X e os autossomos (Figuras 14 e 15: hESCp0). É interessante notar também que todas as amostras classificadas como masculinas apresentaram razão X:A menor do que aquelas femininas (Figuras 14 e 15: 8-cél#3, Blastocistos #1 e #2), reflexo dos dois X ativos nas femininas.

Infelizmente não podemos dizer nada a respeito da expressão de *XIST* em estágio de mórula pois os embriões disponíveis nessa fase são muito provavelmente XY (Figura 8 A e tabela 4) e ainda apresentam aneuploidias e/ou deleções envolvendo cromossomos autossomos (Figura 13: Mórula#1) e mesmo o cromossomo X (Figura 6: *boxplot* ‘cromossomo X’ e figura 16: Mórula#2, células #3, #4, #6 e #7).

O que se pôde observar aqui foi a presença de dois X ativos em embriões femininos em estágio de oito células e o início efetivo da inativação aleatória no estágio de blastocisto ao menos nas células do epiblasto e endoderme primitiva como foram classificadas as células sequenciadas do blastocisto#3 no estudo de Yan e colaboradores (2013).

Capítulo III

GENES REGULADOS POR *IMPRINTING* NO CROMOSSMO X HUMANO

Ainda que células diploides tenham dois alelos de cada gene em seu genoma nuclear, nem todos exibem expressão bialélica: cerca de 0,5 a 15% dos genes presentes nos autossomos são expressos a partir de um único alelo revisto em (Eckersley-Maslin and Spector, 2014). A exemplo do que ocorre com os cromossomos X de mamíferos eutérios, esses genes têm seus alelos inativados de maneira independente da origem parental; este padrão de expressão é comumente identificado como **expressão monoalélica aleatória**.

Classicamente os genes de expressão monoalélica aleatória foram estudados em grandes grupos gênicos como os genes olfatórios, das protocaderinas e das imunoglobulinas (revisto em Eckersley-Maslin and Spector, 2014). Mais recentemente, têm sido encontrados em humanos alguns genes isolados de expressão monoalélica aleatória de funções diversas, como *BCLAF1* (cromossomo 6) envolvido no processo de apoptose; *ABP1* (cromossomo 7) de ação metabólica, com propriedades deaminase oxidativa; *ZFAT* (cromossomo 8) também envolvido com apoptose e angiogênese na placenta; e *IFI30* (cromossomo 19) envolvido no controle intracelular de proteases (Metsalu *et al.*, 2014).

Outro fenômeno que resulta em expressão monoalélica é o *imprinting* genômico, resultado de marcas epigenéticas adicionadas durante a gametogênese masculina e feminina que ocasionam expressão **monoalélica dependente da origem parental**. No início dos anos 90 foram identificados os primeiros genes autossômicos regulados por *imprinting* (em inglês, *imprinted genes*). Atualmente, conhecemos cerca de 70 genes humanos assim regulados e outros 200 são candidatos aguardando validação (Wei *et al.*, 2014).

Genes “imprintados” podem ocorrer de forma isolada, mas em geral estão agrupados e compartilham elementos regulatórios em *cis*. O cromossomo 7 de camundongos, por exemplo, apresenta pelo menos cinco domínios de *imprinting* que concentram quase metade dos genes “imprintados” da espécie. Dentre os elementos regulatórios, existem

sequências de DNA responsáveis pelo *imprinting* genômico, chamadas de regiões controladoras de *imprinting* (ICRs - do inglês, *Imprinting Control Regions*). Os dinucleotídeos CpG das ICRs são submetidos a padrões de metilação diferenciados dependendo da gametogênese ser feminina ou masculina (revisto por Wood and Oakey, 2006).

Apesar da marca epigenética para o *imprinting* genômico ser definida na gametogênese, é muito comum que genes se comportem de maneira “imprintada” em um tecido, e em outro apresentem expressão bialélica (Schulz *at al.*, 2006). Um exemplo intrigante é o do gene humano *GRB10* (*Growth Factor Receptor-bound protein 10*). Este gene, localizado no cromossomo 7 possui um padrão complexo de expressão que combina variantes de *splicing* e expressão diferencial dependendo do tecido (Monk *at al.*, 2009). Os autores confirmaram expressão do alelo paterno exclusiva no cérebro, mas também demonstraram pela primeira vez *imprinting* materno para o mesmo gene nas vilosidades trofoblásticas da placenta.

Não são raras as evidências na literatura de que placenta e cérebro estejam relacionados evolutivamente no que diz respeito à conservação de regulação de expressão gênica por *imprinting* genômico, à qual foram atribuídos valores adaptativos (revisto em Keverne, 2012; Renfree *at al.*, 2013). Esta hipótese é apoiada pelo fato de que a maioria dos genes submetidos a *imprinting* identificados são assim regulados na placenta e no cérebro, e estão geralmente associados ao comportamento materno, lactação, transferência materno-fetal de nutrientes e desenvolvimento da placenta (revisto em Renfree *at al.*, 2013).

A identificação de genes regulados por *imprinting* no cromossomo X é um desafio, pois este passa por um processo de inativação no qual o cromossomo inteiro é transcricionalmente silenciado (inativação do cromossomo X, ou ICX). Isso dificulta a identificação de um gene ligado ao X que tenha sua expressão regulada por *imprinting* e não pela ICX.

No embrião de camundongo, em estágios pré-implantacionais até a fase de mórula e nos tecidos extra-embrionários, a ICX é “imprintada” e o cromossomo de origem paterna é preferencialmente silenciado. Entretanto nesses animais alguns genes ligados ao X foram descritos como regulados por *imprinting*. É o caso da família de genes *Xlr*: em 2005, o gene *Xlr3b* foi descrito como regulado por *imprinting* em cérebro de camundongo, sendo expresso a partir do alelo materno (Davies *et al.*, 2005). No mesmo volume do periódico, foi publicado um artigo de outro grupo que verificou que, além de *Xlr3b*, os parálogos também ligados ao X *Xlr4b* e *Xlr4c* são igualmente expressos exclusivamente a partir do alelo de origem materna em células do cérebro (Raefski and O’Neill, 2005). Já na placenta, Davies e colaboradores (2005) não observaram evidências de *imprinting* e os resultados expostos por Raefski e O’Neill (2005) não foram conclusivos.

Por haver inativação preferencial do cromossomo X de origem paterna em camundongos nos estágios iniciais do desenvolvimento e em seus tecidos extra-embrionários, os genes *Xist* e *Tsix*, intimamente envolvidos no início do processo de ICX, também são considerados como regulados por *imprinting*, sendo *Xist* expresso a partir do alelo paterno e *Tsix* do alelo materno (GF *et al.*, 1994; Lee, 2000; Sado *et al.*, 2001; Zuccotti and Monk, 1995)

Os trabalhos de Kobayashi e colaboradores (2006, 2010, 2013) trouxeram grandes contribuições à lista de genes “imprintados” no cromossomo X de camundongos. Nesses três trabalhos o grupo identificou sete genes com expressão de origem paterna em embriões de camundongo em estágios pré-implantacionais. São eles, *Rhox5* (Kobayashi *et al.*, 2006), *Fthl17* (Kobayashi *et al.*, 2010), e mais cinco genes mapeados na região que compreende o centro de inativação do cromossomo X (*Jpx*, *Ftx*, *Zcchc13* e os microRNAs *miR-374-5p* e *miR-421-3p*) (Kobayashi *et al.*, 2013). Os três genes e os dois microRNAs identificados estão envolvidos com o processo de ICX “imprintada” (Kobayashi *et al.*, 2013; Okamoto *et al.*, 2005); tendo sido *Jpx* e *Ftx* relacionados à regulação positiva da expressão

de *Xist* tanto na ICX aleatória (Chureau *at al.*, 2011; Tian *at al.*, 2010) quanto na “imprintada” (Kobayashi *at al.*, 2013).

Em humanos, há muito se especula sobre a existência de genes do cromossomo X que possam ser regulados por *imprinting*. Entretanto, até o momento, as evidências são apenas indiretas e estão baseadas em traços fenotípicos de pessoas portadoras da síndrome de Turner (monossômicas para o cromossomo X ou com grandes deleções de parte deste cromossomo) (Bishop *at al.*, 2000; Kesler *at al.*, 2004; Skuse *at al.*, 1997). Em um primeiro trabalho publicado em 1997, Skuse e colaboradores analisaram competências sociais de mulheres portadoras da síndrome de Turner relacionando o resultado dos testes com a origem parental do único cromossomo X. Eles observaram que as habilidades sócio-cognitivas ficam prejudicadas naquelas que possuem apenas o X de origem materna em comparação com as que possuem unicamente o X de origem paterna. Dada essa diferença, os autores especulam a existência de um ou mais genes regulados por *imprinting* no cromossomo X, inativados no X materno. Além disso, ao verificarem que mulheres portadoras de deleções parciais envolvendo o braço curto do cromossomo X paterno também apresentam o fenótipo daquelas com monossomia do X (Skuse *at al.*, 1997), os autores propuseram a localização de um *locus* imprintado em Xp.

Em outro trabalho com a participação do mesmo grupo (Good *at al.*, 2003), foram observadas alterações no volume de massa cinzenta na amígdala e do córtex orbitofrontal de mulheres com síndrome de Turner quando comparadas a mulheres normais, sugerindo mais uma vez que deve haver um *locus* regulado por *imprinting* no cromossomo X. Mais especificamente, os autores identificaram três genes no braço curto deste cromossomo relacionados ao desenvolvimento estrutural e funcional dessas áreas do sistema nervoso, *MAOA* (Xp11.3, montagem genômica de referência GRCh38), *MAOB* (Xp11.23, GRCh38) e *USP9X* (Xp11.4, GRCh38), e propuseram que esses sejam genes candidatos a sofrerem *imprinting* no cromossomo X humano.

Em minha dissertação de mestrado, foi aplicada uma nova abordagem para a análise de expressão alelo-específica de genes presentes no cromossomo X, utilizada para estudar o padrão de ICX em placenta a termo (Moreira de Mello, 2010; Moreira de Mello *et al.*, 2010). No presente trabalho, essa mesma estratégia foi empregada para investigar o padrão de expressão alelo-específico desses genes ligados ao X candidatos a serem regulados por *imprinting*.

2. OBJETIVOS ESPECÍFICOS

- Verificar o padrão de expressão alelo-específico dos genes *MAOA*, *MAOB* e *USP9X* por sequenciamento de regiões transcritas contendo polimorfismos em heterozigose em amostras de placenta e cérebro humanos;
- Identificar a origem parental do alelo expresso nas amostras de placenta;
- Verificar se os genes *MAOA*, *MAOB* e *USP9X* são regulados por *imprinting* genômico. Se confirmado, esses serão os primeiros genes “imprintados” identificados no cromossomo X humano.

3.1. Laboratórios

Esta parte do trabalho foi desenvolvida em colaboração com o grupo da profa. Dra. Wendy P. Robinson no *Department of Medical Genetics University of British Columbia Child & Family Research Institute* em Vancouver, Canadá, durante estágio de um mês.

3.2. Seleção dos polimorfismos e síntese dos iniciadores

Variantes como os polimorfismos de base única (SNPs) são utilizados nesse projeto como ferramenta de identificação de expressão alelo-específica. A comparação dos resultados de expressão com os da genotipagem do DNA parental permite ainda identificar a origem do alelo expresso.

Os polimorfismos utilizados nesse projeto foram selecionados por estarem presentes em região codificadora dos genes *MAOA*, *MAOB*, *USP9X* e *XIST*, localizados no cromossomo X (UCSC Genome Browser (<https://genome.ucsc.edu/>) e NCBI (<https://www.ncbi.nlm.nih.gov/snp/>), ambos com o filtro para dbSNP 138. Foram aceitos aqueles SNPs presentes em todos os possíveis transcritos alternativos descritos de cada gene aqui estudado, em região não traduzida 5' ou 3', ou, quando em região traduzida, levando a troca sinônima de aminoácido.

Na tabela 7 estão listados, em ordem citogenética, os 13 SNPs escolhidos de acordo com os critérios estabelecidos, bem como a sigla do nome dos genes que os contêm, a identidade do SNP e sua variante alélica e posição no gene.

Tabela 7 Genes e polimorfismos selecionados				
Localização no cromossomo	Gene	SNP	Alelos	Posição no gene
Xp11.4	<i>USP9X</i>	rs10463	A G	3' UTR
Xp11.3	<i>MAOA</i>	rs1800464	A C	5' UTR
		rs6323	G T	exon 9
		rs1137070	C T	exon 15
		rs3027407	A G	3' UTR
Xp11.23	<i>MAOB</i>	rs17462	A G	3' UTR
		rs3027439	A G	3' UTR
		rs2072745	A T	3' UTR
Xq13.2	<i>XIST</i>	rs1894271	A C G T	exon 1
		rs1620574	C T	exon 6
		rs1794213	G T	exon 6
		rs16992443	G T	exon 6
		rs16992436	C T	exon 6

NCBI: GRCh37/Hg19; dbSNP BUILD 138 (<http://www.ncbi.nlm.nih.gov/snp>)

Dentre os genes selecionados é importante ressaltar a escolha de um em particular, o *XIST*. Como descrito, a placenta humana é um mosaico em relação ao padrão de ICX apresentando regiões com inativação preferencial do X paterno e outras do X materno (Moreira de Mello *at al.*, 2010). Sendo o gene *XIST* expresso unicamente a partir do X inativo (Brown *at al.*, 1991), e por seu perfil de expressão alelo-específico refletir o perfil geral de inativação do cromossomo X em cada célula, sua expressão foi utilizada para se inferir o padrão de ICX de cada amostra.

Os pares de iniciadores (*primers*) para a amplificação por PCR e sequenciamento tipo Sanger ou pirosequenciamento para a maioria dos SNPs (Anexos7) foram desenhados por nosso grupo (Moreira de Mello *at al.*, 2010; Stabellini *at al.*, 2009) tese Raquel, Stabellini, 2008. A sequência dos *primers* para a amplificação da região em torno do SNP rs1620574 no gene *XIST* foi adquirida do trabalho de Carrel e Willard (2005).

O programa Primer3 (Koressaar and Remm, 2007; Untergasser *at al.*, 2012) foi utilizado para se desenhar os *primers* para amplificação por PCR e sequenciamento do tipo Sanger. Ao se desenhar *primers* para a detecção de expressão alelo-específica é fundamental evitar polimorfismos na região de hibridação com a sequência alvo, e também garantir que pelo menos um deles esteja a no mínimo 100 bases de distância do polimorfismo para um sequenciamento apropriado.

Os *primers* específicos para pirosequenciamento foram desenhados evitando-se polimorfismos na região de hibridação com a sequência alvo, com o auxílio do Pyrosequencing™ assay design software (Biotage).

3.3. Obtenção das amostras

3.3.1. Tecidos

– Placenta

Foram incluídas neste estudo apenas placentas resultantes de partos normais de meninas saudáveis nascidas a termo. As amostras de placenta a termo (pl.) foram coletadas na maternidade pública Amparo Maternal, localizada à Rua Loefgreen, nº 1901, Vila Clementino, São Paulo, SP, sempre com o consentimento informado da mãe e com a autorização do Comitê de Ética em Pesquisa do Instituto de Biociências (Protocolo nº 039/2006). Também foram utilizadas amostras de placenta a termo femininas do grupo da profa. Dra. Wendy P. Robinson, coletadas e estocadas de maneira semelhante à descrita a seguir, sempre com dois ou três segmentos de áreas não adjacentes.

Segmentos de não mais que 0,5 cm de espessura foram extraídos da porção fetal, lavados em solução tampão fosfato-salina (PBS) pH 7,4 (para remoção do sangue), e imediatamente submersos em 3 mL de solução estabilizadora de RNA total (RNA^{later}™ QIAGEN); a refrigeração para estoque a -70 °C seguiu as especificações do fabricante. Quando possível, foram separados três ou quatro segmentos coletados também da porção fetal mas de cotilédones diferentes e acondicionados em tubos individuais.

– Cérebro

Com colaboração da profa. Dra. Iscia Teresinha Lopes Cendes e inestimável ajuda da pesquisadora Dra. Danyella Dogini, ambas do laboratório de genética molecular da FCM-UNICAMP, foram doadas três amostras de lobo temporal de mulheres diagnosticadas com epilepsia e três casos controles, com as idades dos indivíduos à ocasião da autópsia

variando de 44 a 66 anos. No momento da coleta, as amostras foram congeladas diretamente em nitrogênio líquido e posteriormente estocadas à -80 °C.

3.3.2. Material parental

A fim de se obter o DNA genômico parental, amostras de células da mucosa oral da mãe, e sempre que possível de ambos os pais, foram coletadas raspando-se por alguns instantes a parede interna da bochecha com o auxílio de uma escova cervical estéril. O material genômico das células aderidas às cerdas foi estabilizado imergindo imediatamente a ponta da escova em tubos plásticos estéreis com tampa de rosca contendo 500 µL de solução hidróxido de sódio 50 mM, para posterior extração do DNA.

No laboratório da Profa. Dra. Wendy P. Robinson é rotina recolher sangue materno à ocasião do parto. Por esta razão, o DNA genômico destas amostras foi isolado a partir de sangue periférico total.

3.4. Obtenção dos ácidos nucleicos

3.4.1. DNA PARENTAL

O DNA genômico dos pais foi extraído de acordo com protocolo descrito por Richards e colaboradores (1993) a partir das amostras de mucosa oral, e estocado a -20 °C. De forma resumida, os tubos com a solução de hidróxido de sódio contendo a ponta da escova cervical com as células aderidas às cerdas foram agitados e mantidos em banho-maria a 96 °C por 5 minutos. Posteriormente, as escovas foram removidas e ao conteúdo do tubo foram adicionados 80 µL de solução Tris-HCl 1,0 M com pH 8,0. As amostras foram então submetidas à centrifugação de 17.900 rcf por 6 minutos (Richards *et al.*, 1993). O sobrenadante, contendo o DNA dos pais, pôde então ser removido para um novo tubo estéril e estocado sob refrigeração de -20 °C.

Do laboratório da Dra. Wendy Robinson tive também acesso ao banco de DNA materno previamente obtido a partir de células sanguíneas estocado a -80 °C.

3.4.2. Tecido Extra-embrionário e cérebro

– DNA e RNA

Para se ter acesso ao DNA genômico e RNA total das amostras de placenta e cérebro, segmentos de aproximadamente 100 mg (4 a 8 mm³ de tecido) foram triturados em almofariz de porcelana adicionando-se nitrogênio líquido. Ao pó gerado foi adicionado 1,5 mL do reagente comercial TRIzol® (Invitrogen) com 20 µL de proteinase K (20 mg/mL). O material em solução foi deixado por 30 minutos em banho-maria a 55 °C, e a purificação dos ácidos nucleicos seguiu de acordo com as especificações do fabricante. O RNA obtido foi estocado a -70 °C e o DNA a 4 °C.

Para se evitar qualquer possibilidade de contaminação com material genômico residual, todas as amostras de RNA foram tratadas com a enzima Turbo DNA-free (Ambion®), utilizando o protocolo rigoroso sugerido pelo fabricante.

– Síntese de cDNA

A partir de 1 a 2 µg do RNA total previamente tratado com DNase, foi sintetizado o cDNA utilizando-se iniciadores aleatórios (*random primers*) e a transcriptase reversa Superscript III (Invitrogen™) de acordo com o sugerido pelo fabricante da enzima. Para se verificar eventual resistência de contaminação com DNA genômico mesmo após tratamento com DNase, o protocolo também foi seguido com uma alíquota da amostra onde apenas a transcriptase reversa não era adicionada (RT-). As amostras de cDNA (RT+), bem como as RT-, foram mantidas sob refrigeração a -20 °C.

3.5. Genotipagem do DNA genômico e verificação do alelo expresso

Para a genotipagem das amostras de placenta e cérebro, a região do DNA genômico compreendendo cada polimorfismo foi amplificada por PCR e posteriormente sequenciada. Das amostras informativas foi então sequenciada a região correspondente a partir do cDNA para a verificação do padrão de expressão.

3.5.1. Reação de sequenciamento direto tipo Sanger

De 50 a 100 ng de DNA ou cDNA foram amplificados usando 1,0 U de Taq DNA Polimerase (Invitrogen™), 1X de tampão (Invitrogen™), 200 µM de dNTPs (Invitrogen™), 150 nM de cada iniciador, em volume final de 25 µL.

Os produtos amplificados a partir de DNA foram separados por eletroforese a 100 V por cerca de 20 minutos em gel de agarose 1% 1X TAE (5,5 x 6,0 cm), em solução 1X TAE, corados com 0,5 µg/mL de brometo de etídio (Invitrogen™) e observados sob luz ultravioleta.

O material amplificado a partir de cDNA (RT+ e RT-) foi submetido à eletroforese a 100 V por 1 h em gel de poliacrilamida 10% 1X TBE (10 x 10 cm), em solução 1X TBE. Após a eletroforese, procedeu-se a coloração por impregnação por nitrato de prata (protocolo levemente modificado de Santos e colaboradores 1993: mergulha-se o gel em solução fixadora (10% de etanol; 0,5% de ácido acético) por 10 minutos, segue-se a imersão do mesmo em solução corante (nitrato de prata 2 g/L) por 10 minutos, breve enxágue e imersão em solução reveladora (hidróxido de sódio 30 g/L e 0,002% de formaldeído 37%) até a visualização das bandas. Por fim, o gel era mergulhado novamente em solução fixadora por 10 minutos para a documentação fotográfica.

Os produtos de PCR foram purificados com as enzimas ExoI (5 U) e Sap (1 U) por uma hora a 37 °C e posterior interrupção da atividade enzimática a 85 °C por quinze minutos.

A reação de sequenciamento direto foi realizada com o *primer* mais adequado, F ou R, de acordo com a distância em relação à posição do SNP (≥ 100 bases) utilizando-se o Kit BigDie® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems®) e 3 μ L do produto de PCR purificado. As reações de sequenciamento foram purificadas por coluna de sephadex (Sephadex G-50 DNA Grade F, GE Healthcare).

A eletroforese capilar das amostras foi realizada em sequenciadores automáticos Applied Biosystems 3730DNA ou 3130XL do GaTE Lab no Departamento de Botânica do Instituto de Biociências da Universidade de São Paulo (Docente Responsável: Profa. Dra. Marie-Anne Van Sluys; Técnica Responsável: Dra. Tatiana C. Corrêa).

Através do sequenciamento de regiões que possuem SNP em heterozigose é possível traçar um paralelo entre o padrão observado em eletroferogramas de um gene com expressão monoalélica preferencial e aquele esperado para ICX aleatória. Das amostras geneticamente informativas (*i.e.* heterozigotas para *XIST* e um dos genes candidatos, *MAOA*, *MAOB* ou *USP9X*) sequenciou-se também o respectivo cDNA. No caso de uma amostra informativa apresentar os dois alelos expressos para *XIST* e apenas um único alelo expresso para um dos outros genes em questão inferir-se-ia expressão monoalélica preferencial, e ainda, da comparação com o genótipo dos pais, nos experimentos com placenta, seria possível identificar a origem parental do alelo expresso.

3.5.2. Reação por pirosequenciamento

O pirosequenciamento é uma técnica sensível que possibilita detecção de polimorfismos, além de quantificar a porcentagem dos alelos presentes na amostra. Desta maneira, esta metodologia torna a análise de expressão alelo-específica bastante precisa e prática.

As reações de PCR com DNA para a genotipagem e cDNA para a verificação do alelo expresso dos genes candidatos foram realizadas com 100 ng de DNA genômico ou 1,0 μ L

de cDNA em volume final de 15 µL contendo 0,2 mM de dNTP, uma unidade de DNA polimerase Hot StarTaq (Qiagen) e 26 µM de cada *primer*. O sucesso da reação de PCR foi verificado em eletroforese em gel de agarose como descrito no subitem anterior (2.5.1).

Cinco microlitros do produto final da PCR foram purificados e utilizados na reação de pirosequenciamento de acordo com as especificações do fabricante (Qiagen) utilizando-se PyroMark™ Q96MD.

– Sensibilidade do teste

Os dados gerados por pirosequenciamento são diretamente quantitativos, o que torna a determinação da proporção dos alelos expressos imediata. Com o intuito de se verificar a sensibilidade da metodologia aplicada na detecção dos casos de expressão diferencial, foi construída uma curva padrão. O DNA obtido de duas amostras homozigotas para os diferentes alelos do polimorfismo rs137070 (presente no gene *MAOA*) foi misturado em concentrações conhecidas a fim de simular diferentes taxas de desvio de inativação, com proporções variando entre 100:0, 80:20, 60:40, 50:50, 40:60, 20:80, e 0:100 da presença de cada alelo. As sete diferentes concentrações foram então amplificadas e sequenciadas nas mesmas condições de todos os outros experimentos de pirosequenciamento descritas neste subitem.

4.1. Análise de expressão alelo-específica em placenta humana a termo

4.1.1. Genotipagem

Ao todo, nosso banco de amostras possui 36 espécimes, sendo que de 10 destes foi possível coletar mais de um segmento (entre três e quatro vilos). Infelizmente, desses últimos não foi possível colher material para a obtenção do DNA genômico parental. Do laboratório da profa. Dra. Wendy P. Robinson foram disponibilizados 33 conjuntos de amostras de placenta (com quatro vilos coletados) e DNA materno.

Foram realizados com sucesso 595 ensaios de genotipagem de DNA de amostras de placenta humana por sequenciamento direto do tipo Sanger e 312 por pirosequenciamento. Destes, 69 resultaram em polimorfismos em heterozigose nas amostras de placenta distribuídos entre os genes *USP9X*, *MAOA*, *MAOB* e *XIST*. Na tabela 8 estão representados apenas os conjuntos placenta/pais informativos, ou seja: SNP em heterozigose para um dos genes candidatos (*MAOA*, *MAOB* e *USP9X*) na placenta, mas em homozigose em um dos pais, o que torna possível a identificação da origem parental do alelo expresso. Sempre que homozigoto o genótipo dos pais é indicado com as bases nitrogenadas correspondentes nos alelos identificados.

Tabela 8 | Amostras informativas após genotipagem de material genômico extra-embriônico e parental

Tabela 3. Polimorfismos informados nos tipos genotípicos de material genético extra embrionário e parental.																
	Gene	USP9X				MAOA				MAOB			XIST			
Amostra	SNP	rs10463	rs1800464	rs6323	rs3027407	rs1137070	rs17462	rs3027439	rs2072745	rs1894271	rs16992443	rs1794213	rs1620574	rs16992436		
Mat.02		-	AA	HTR	-	-	AA	-	-	-	-	x	HTR	-		
pl.02		-	HTR	HTR	-	-	HTR	-	-	-	-	HTR	HTR	-		
Mat.08		-	-	TT	-	-	-	-	-	-	-	x	-	-		
pl.08		-	-	HTR	-	-	-	-	-	-	-	HTR	-	-		
Mat.10		-	AA	HTR	-	-	AA	-	-	-	-	x	-	-		
pl.10		-	HTR	HTR	-	-	HTR	-	-	-	-	HTR	-	-		
Mat.11		-	HTR	HTR	-	-	HTR	-	-	-	-	-	TT	-		
Pat.11		-	A	G	-	-	T	-	-	-	-	-	C	-		
pl.11		-	HTR	HTR	-	-	HTR	-	-	-	-	-	HTR	-		
Mat.17		-	AA	TT	x	x	-	-	-	AA	-	-	HTR	-		
pl.17		-	HTR	HTR	HTR	HTR	-	-	-	HTR	-	-	HTR	-		
Mat.18		-	AA	-	-	-	-	HTR	HTR	HTR	-	-	HTR	-		
Pat.18		-	C	-	-	-	-	T	T	A	-	-	C	-		
pl.18		-	HTR	-	-	-	-	HTR	HTR	HTR	-	-	HTR	-		
Mat.27		-	-	TT	GG	CC	-	-	-	-	x	x	-	x		
pl.27		-	-	HTR	HTR	HTR	-	-	-	-	HTR	HTR	-	HTR		
pl.31*		HTR	HTR	-	-	-	-	-	-	HTR	-	-	-	-		
pl.32*		-	-	HTR	HTR	HTR	-	-	-	HTR	-	-	HTR	-		
pl.37*		HTR	HTR	-	-	HTR	-	-	-	-	-	-	HTR	-		
pl.40*		-	-	-	HTR	HTR	-	-	-	-	-	HTR	HTR	-		
pl.43*		-	-	-	-	-	HTR	-	-	-	HTR	-	-	HTR		
Mat.264		-	AA	CC	CC	-	-	-	-	HTR	-	-	HTR	-		
pl.264*		-	HTR	HTR	HTR	-	-	-	-	HTR	-	-	HTR	-		
Mat.273		-	HTR	HTR	HTR	-	-	-	-	HTR	-	-	CC	-		
pl.273		-	HTR	HTR	HTR	-	-	-	-	HTR	-	-	HTR	-		
Mat.285		-	AA	CC	CC	-	-	-	-	CC	-	-	CC	-		
pl.285*		-	HTR	HTR	HTR	-	-	-	-	HTR	-	-	HTR	-		

"HTR": polimorfismo em heterozigose; "A", "T", "G" ou "C": genótipo parental; "x": não foi possível genotipar; "-": não informativo; células vazias: não genotipado

"pl.": genótipo da amostra; "Mat.": genótipo materno; "Pat.": genótipo paterno; "*": amostras com mais de um segmento coletado
casos informativos evidenciados em verde

Outra condição para se classificar um caso como informativo é que a amostra de placenta apresente algum SNP em heterozigose no gene *XIST*, para se determinar o padrão de ICX da amostra. Como já foi mencionado, a ICX em placenta humana a termo é aleatória, mas o órgão se apresenta como um mosaico com grandes áreas de inativação preferencial do X paterno ou do X materno, e áreas de intersecção que apresentam o clássico padrão aleatório de ICX (Moreira de Mello *et al.*, 2010).

Na tabela 8 os casos informativos estão evidenciados em verde. Nela, pode-se observar que 14 conjuntos de amostras são informativos para o gene *MAOA*, cinco para *MAOB* e apenas dois para *USP9X*. Infelizmente as amostras numeradas de pl.31 a pl.43 não possuem respectivo DNA parental, entretanto destas foram coletados mais de um segmento de cotilédones distintos. Assim, se um gene for regulado por *imprinting*, sua expressão deverá ser coerente entre todas as áreas mostrando sempre a expressão do mesmo alelo.

4.1.2. Análise de expressão alelo-específica

- Sequenciamento tipo Sanger

As amostras de placenta identificadas como pl.02 a pl.43 (Tabela 8) tiveram o seu perfil de expressão alelo-específico determinado por sequenciamento tradicional tipo Sanger dos respectivos cDNAs. As amostras pl.02, pl.08, pl.10, pl.11 (Tabela 9) e pl.37 (Tabela 10) apresentaram claro desvio total de ICX de acordo com a expressão de *XIST*, sendo que apenas para a pl.11 (Tabela 9) foi possível identificar o alelo paterno como o expresso, determinando-o como inativo. A expressão dos genes *USP9X* (Tabela 10), *MAOB* (Tabela 9) e *MAOA* (Tabelas 9 e 10) para esses mesmos casos estão de acordo com o padrão de ICX, ou seja expressão monoalélica exclusiva proveniente do único X ativo: pl.02; pl.10 e pl.11 expressando os genes *MAOA* e *MAOB* unicamente a partir do X de origem materna (Tabela 9). Os quatro vilos coletados da pl.37 expressam unicamente o mesmo alelo (Tabela 10) reforçando o desvio total de ICX observado pela expressão de *XIST*. Como já explicado, uma das condições para se considerar um caso como informativo é a observação de ICX aleatória através da expressão de *XIST* e consequentemente os casos pl.02, pl.08, pl.10, pl.11 e pl.37 não apresentam resultado relevante para este estudo.

Tabela 9| Alelo expresso determinado por sequenciamento direto tipo Sanger das amostras com um fragmento coletado

Amostra	Gene SNP	<i>XIST</i>			<i>MAOA</i>		<i>MAOB</i>	
		rs1894271	rs1794213	rs1620574	rs1800464	rs6323	rs17462	rs3027439
Mat.02			x	HTR	AA			
pl.02	-		T	T	A	x	x	-
Mat.08			x		-	x	-	-
pl.08	-		G					
Mat.10			x		AA	HTR	AA	
pl.10	-		T		A	G	A	-
Mat.11				TT	HTR	HTR	HTR	
Pat.11	-		-	C	A	G	T	-
pl.11				C	C	T	G	
Mat.17	AA		-	x	AA	x	-	-
pl.17	A/g				A/c			
Mat.18	HTR			HTR	AA			HTR
Pat.18	A		-	C	C	-	-	T
pl.18	G/A			C/T	C/A			T/C

"HTR": polimorfismo em heterozigose; "x": não foi possível o resultado; "-": não informativo;

"pl.": amostra de placenta; "Mat.": amostra materna; "Pat.": amostra paterna

Alelos resultados da genotipagem parental em letra maiúscula

Expressão materna N/N = expressão equivalente os alelos

Expressão paterna N/n = expressão com alguma diferença entre os alelos

Origem do alelo expresso indefinida

Os casos das placentas 17 e 18 mostram que os dois alelos do gene *XIST* estão expressos (Tabela 9), padrão esperado para áreas onde as regiões de ICX materna e paterna se encontram, sendo que pl.17 apresentou um leve desvio de ICX a favor do alelo materno (representado como A/g, onde a fonte maiúscula indica o alelo preferencialmente expresso). O Padrão de expressão do gene *MAOA* para pl.17 e pl.18 e *MAOB* para pl.18 estão de acordo com o observado para *XIST*, expressão bialélica para pl.18 e com leve desvio a favor do alelo materno para pl.17 (representado como A/c), se estes genes fossem regulados por *imprintitng* e não pela ICX o padrão esperado seria o de expressão monoalélica exclusivo.

Assim como na tabela 9, a tabela 10 tem os resultados de expressão bialélica destacados em verde, e apesar da ausência do material genético parental, foi possível observar desvio total de ICX para um ou outro alelo nos casos das amostras pl.31 e pl.40, com respectivamente três e quatro vilos coletados.

Tabela 10| Alelo expresso determinado por sequenciamento direto tipo Sanger das amostras com mais de um segmento coletado

Amostra	Gene SNP	<i>XIST</i>			<i>USP9X</i>	<i>MAOA</i>		
		rs1894271	rs1794213	rs1620574	rs10463	rs1800464	rs3027407	rs1137070
pl.31(A)		G/A	-	-	A/G	C/A	-	-
pl.31(B)		A/g	-	-	G	A	-	-
pl.31(C)		G/a	-	-	A	C	-	-
pl.37(A)		-	-	T	G	A	-	C
pl.37(B)		-	-	T	G	A	-	C
pl.37(C)		-	-	T	G	A	-	C
pl.37(D)		-	-	T	G	A	-	C
pl.40(A)		-	G	C/t	-	-	A	T
pl.40(B)		-	T	T	-	-	G	C
pl.40(C)		-	G/t	C/T	-	-	A	T
pl.40(D)		-	G/t	C/T	-	-	A	T

"pl.": amostra de placenta; "-": não informativo

as letras entre parênteses identificam os diferentes fragmentos de uma mesma amostra

Alelo expresso de mesma origem parental

Alelo expresso de mesma origem parental

N/N = expressão equivalente os alelos

N/n = expressão com alguma diferença entre os alelos

A placenta pl.31 vilo A apresenta ICX aleatória de acordo com a expressão de *XIST* e coerente expressão dos genes *USP9X* e *MAOA*. Já para os segmentos B e C, o padrão de expressão de *XIST* indica desvio de inativação para alelos de origem parental oposta, o alelo mais expresso apresenta uma adenina (A) na posição do polimorfismo (rs1894271) no vilo pl.31B e guanina (G) no vilo pl.31C. Entretanto quando o padrão de expressão dos genes *USP9X* e *MAOA* foram analisados para estes mesmos vilos, pl.31B e pl.31C, o que se observou foi expressão monoalélica, coerente com desvio total de ICX ainda que para alelos opostos em concordância com o desvio observado para *XIST*.

Resultado comparável ao da pl.31 está apresentado no caso da pl.40, na tabela 10 e figura 17, onde é possível observar que diferentes padrões de ICX puderam ser detectados entre os quatro vilos (A, B, C e D) de acordo com a expressão de *XIST*. Já o padrão de expressão de *MAOA* indica ICX completamente desviada, sendo que os vilos A, C e D expressam o alelo de mesma origem parental, e o vilo B apresenta expressão do outro alelo.

Figura 17 | Eletroferogramas representativos do caso pl.40. A identificação de cada vilo encontra-se a esquerda; a identificação do gene e do polimorfismo sequenciado a cima; asteriscos evidenciam a posição do polimorfismo. Os eletroferogramas são resultados do sequenciamento tipo Sanger realizado a partir de cDNA de cada amostra.

Um exame cuidadoso da figura 17 revela a dificuldade que pode levar a imprecisão ao se classificar visualmente casos de desvio de ICX ou expressão diferencial entre os alelos de um mesmo gene, tendo como base o sequenciamento do tipo Sanger de polimorfismos de base única. No caso da pl.40 (Figura 17), os polimorfismos no gene *XIST* indicam um certo grau de desvio de ICX (vilo A rs1620574; vilos C e D rs1794213) enquanto que outros sugerem ICX aleatória (vilos C e D rs1620574). Já a expressão do gene *MAOA* é exclusivamente monoalélica para os quatro segmentos. Essa inconsistência pode ser devida ao fato de que variações na fluorescência, inerente aos quatro diferentes corantes utilizados na reação de sequenciamento tipo Sanger, resultem em diferenças artificiais entre as alturas dos picos, dificultando a interpretação do resultado biológico. Esse tipo de

problema foi extensamente discutido em minha dissertação de mestrado (Moreira de Mello, 2010), onde a questão da análise visual foi contornada utilizando-se o programa PeakPicker (Ge *et al.*, 2005). Este programa estima a altura dos picos na posição de um dado polimorfismo de diversos eletroferogramas ao mesmo tempo, normalizando os valores de acordo com picos de referência indicados em posições não polimórficas em torno da posição do SNP. Infelizmente esta solução não pôde ser aqui aplicada, pois a análise estatística para se validar os resultados do PeakPicker requer um número amostral maior do que o aqui apresentado (Moreira de Mello, 2010).

– Pirosequenciamento

Ao contrário do sequenciamento por Sanger, os dados gerados por pirosequenciamento são diretamente quantitativos, o que torna a determinação da porcentagem dos alelos expressos imediata. Com o objetivo de se testar a sensibilidade da técnica de pirosequenciamento em detectar diversos graus de desvio de ICX, foi construída uma curva padrão (Figura 18). Sua criação se deu com concentrações conhecidas de diferentes amostras de placenta homozigotas para os dois possíveis alelos da variante rs0137070 do gene *MAOA*. O gráfico na figura 18 mostra o sucesso do teste ($R^2=0,9977$), e a comparação entre os valores esperados e obtidos demonstra que a técnica é capaz de detectar com precisão proporções entre os alelos variando de 100:0, 80:20, 60:40, 50:50, 40:60, 20:80, e 0:100.

Figura 18| Curva padrão construída com os valores obtidos por pirosequenciamento. À esquerda, gráfico gerado com os valores esperados e observados para o alelo T (timina) por técnica de pirosequenciamento da variante rs1137070 do gene *MAOA*. A direita, quadro com os valores esperados de acordo com a proporção entre os alelos e valores observados. $R^2 = 0.9977$.

As regiões alvo dos genes *XIST* (rs162574) e *MAOA* (rs1137070 e rs6323) das amostras pl.264 e pl.285 foram sequenciadas por pirosequenciamento a partir do cDNA. A figura 19 apresenta de forma gráfica os valores associados à razão de cada alelo expresso, onde o gráfico de barras representa a proporção entre a expressão dos alelos materno (em rosa) e paterno (em azul). Acima de cada gráfico se apresenta um quadro com os valores em porcentagem obtidos dos ensaios de pirosequenciamento referentes a cada alelo expresso para cada vilo e sua origem parental.

pl.264

pl.285

Figura 19| Origem parental do alelo expresso do gene *MAOA* em placenta a termo. Resultados de expressão alelo-específica obtidos por pirosequenciamento de regiões polimórficas nos genes *XIST* (rs1620574) e *MAOA* (rs1137070 e rs6323) das amostras de

placenta pl.264 e pl.285. O gráfico de barras indica a porcentagem da altura dos picos no pirograma referente a cada alelo. Acima, pl.264 com as proporções entre os alelos materno e paterno para os vilos 1, 2 e 3; abaixo, pl.285 com as proporções para os vilos 1, 3 e 4. Os valores em porcentagem da presença de cada alelo são fornecidos pelo pirosequenciamento realizado a partir de cDNA de cada amostra.

Como nos casos analisados por sequenciamento tradicional, as amostras pl.264 e pl.285 também apresentaram diferentes níveis de expressão de *XIST* e concordantes níveis de expressão de *MAOA* (Figura 19). No caso da amostra pl.264, o vilo 1 apresentou cerca de 70% de expressão de *XIST* proveniente do X de origem paterna e o gene *MAOA* estava expresso aproximadamente na mesma proporção a partir do cromossomo X de origem materna. Já no vilo 3 da mesma amostra o cromossomo X predominantemente inativo é o de origem materna (cerca de 80% da expressão de *XIST*), sendo *MAOA* expresso majoritariamente a partir do alelo paterno. Infelizmente não foi possível detectar a expressão de *XIST* para o vilo 2 da amostra pl.264. Entretanto, a expressão de *MAOA* foi detectada a partir dos dois alelos com um certo grau de desvio (72:28).

O vilo 1 do caso pl.285 também apresenta um desvio de ICX a favor do X materno (cerca de 70% de expressão de *XIST* detectada do alelo materno). Coerentemente, o gene *MAOA* se mostrou expresso preferencialmente a partir do alelo paterno. Já os vilos 3 e 4 da mesma amostra apresentam o clássico padrão de ICX aleatória, com cerca de 50% de cada alelo de *XIST* expresso. Resultado semelhante foi observado para a expressão de *MAOA* (Figura 19).

Em conjunto, os resultados expostos neste item revelam que *MAOA*, *MAOB* e *USP9X* não tem expressão monoalélica dependente de origem parental ao menos em placenta humana a termo e seu padrão de expressão segue aquele da ICX.

4.2. Análise de expressão alelo-específica em cérebro humano:

4.2.1. Genotipagem:

Das seis amostras de cérebro doadas, três se apresentaram informativas para *MAOA* (C1, C3 e C4) e outras duas para *MAOB* (C2 e C5); entretanto, nenhuma foi informativa para o gene *USP9X*. Apesar de apenas uma amostra ser informativa para variantes em *XIST* (C5), decidiu-se seguir com as análises de expressão alelo-específica dos genes candidatos em todas aquelas que apresentassem algum polimorfismo em heterozigose nestes genes.

4.2.2. Expressão alelo-específica:

Foi possível obter RNA e sintetizar cDNA de boa qualidade para todas as cinco amostras de cérebro. Este material foi então submetido a sequenciamento tradicional tipo Sanger e o resultado da análise de expressão pode ser observado na tabela 11 e figura 20.

Tabela 11 Alelo expresso determinado por sequenciamento direto tipo Sanger das amostras de cérebro humano						
Amostra	Gene SNP	<i>MAOA</i>		<i>MAOB</i>	<i>XIST</i>	
		rs1800464	rs6323	rs17462	rs1894271	rs1794213
C1*		C/A	-	-	-	-
C2*		-	-	G/a	-	-
C3*		-	T/g	-	-	-
C4**		C/A	-	-	-	-
C5**		-	-	A/G	-	T/G

* casos de epilepsia no lobo frontal; ** casos controle

N/N = expressão equivalente os alelos

N/n = expressão com alguma diferença entre os alelos

Figura 20| Eletroferogramas representativos do sequenciamento de cDNA das amostras de cérebro. O símbolo de cada gene encontra-se acima do conjunto de eletroferogramas. A identificação de cada amostra situa-se à esquerda, com a identidade do SNP entre parênteses; asteriscos evidenciam a posição do polimorfismo. Os eletroferogramas resultam do sequenciamento realizado a partir de cDNA de cada amostra.

Apesar de não existirem dados suficientes para se definir o padrão de ICX pela expressão de *XIST* em todas as amostras (C1, C2, C3 e C4 não são heterozigotas para nenhum polimorfismo neste gene, tabela 11), fica claro que os genes *MAOA* e *MAOB* não apresentam expressão regulada por *imprinting* em cérebro adulto humano, pois em todas elas, foi possível detectar expressão proveniente dos dois alelos (Tabela 11). Com os resultados aqui apresentados também não é possível afirmar haver nenhuma evidência de diferença de expressão entre amostras de doadoras diagnosticadas com casos de epilepsia ou de casos controle.

O gene *USP9X* (*ubiquitin specific peptidase 9, X-linked*) codifica uma enzima de

remoção de resíduos de ubiquitina substrato-específica que antagoniza a via ubiquitina-proteassoma; sua proteína apresenta 2.570 aminoácidos e seu mRNA pode ser detectado em quase todos os tecidos e estágios do desenvolvimento (<http://www.ncbi.nlm.nih.gov/UniGene>). Homan e colaboradores (2014) destacam que sua sequência é altamente conservada e intolerante a variantes, sendo que mutações de ponto em *USP9X* foram relacionadas a retardo mental ligado ao cromossomo X (MIM 300072). Neste estudo funcional envolvendo os ortólogos humanos e camundongos, os autores revelam a importância de *USP9X* no desenvolvimento do cérebro por estar envolvida em vias que contém proteínas do citoesqueleto, sendo seu pleno funcionamento indispensável para a adequada migração de células neuronais e crescimento dos axônios (Homan *et al.*, 2014). A dificuldade no presente trabalho de se encontrar variantes em heterozigose nas regiões traduzidas de *USP9X* possivelmente se deve ao fato deste gene ser altamente conservado e polimorfismos em sua sequência podem resultar em fenótipos graves.

Os genes das monoamina oxidases A e B (*MAOA* e *MAOB*) codificam enzimas que se posicionam na membrana externa das mitocôndrias e que participam das vias de degradação de neurotransmissores da classe das monoaminas. Apesar da alta similaridade entre a sequência de aminoácidos, 73% de identidade entre MAO-A e MAO-B (<http://www.ncbi.nlm.nih.gov/protein>), MAO-A tem afinidade por substratos como serotonina, norepinefrina e epinefrina, enquanto que MAO-B tem maior afinidade por feniletilamina, sendo dopamina e tiramina substratos comuns às duas isoformas (Wang *et al.*, 2013).

Por participarem das vias de degradação de importantes neurotransmissores e por sua relevância clínica e farmacológica, a investigação sobre esses genes, seus padrões de expressão, regulação, suas funções, interações, vias de ação etc, têm sido principalmente estudadas tendo como foco o sistema nervoso. De fato, a atividade de ambas puderam ser relacionadas com distúrbios de ansiedade, distúrbios de

humor, depressão, hiperatividade e mesmo doenças como mal de Parkinson e Alzheimer (revisto em Bortolato *at al.*, 2008).

Bonnin e colaboradores (2011) mostraram que no desenvolvimento embrionário, a serotonina é fundamental nos processos de neurogênese e crescimento dos axônios; o triptofano, seu precursor, é exclusivamente transferido pela mãe e convertido em serotonina na porção fetal da placenta em camundongos entre os estágios E10.5 e E15.5. Essa característica não é exclusiva destes animais, foi demonstrado em humanos que a porção fetal da placenta com 11 semanas de gestação também apresenta alta atividade de síntese de serotonina (Bonnin *at al.*, 2011). Assim, em humanos e camundongos a atividade metabólica da placenta está diretamente relacionada com o desenvolvimento do cérebro do feto.

Durante a gestação a serotonina age também como vasoconstritor na placenta, e níveis aumentados de serotonina circulante na gestante, provocado pela reduzida atividade de MAO-A na placenta foram observados em casos de pré-eclâmpsia. Esta é uma doença multi-sistêmica caracterizada por hipertensão, proteinúria e restrição de crescimento intrauterino podendo resultar em diversas complicações e mesmo levar o feto a óbito. Foi detectada elevada atividade de MAO-A principalmente no sincício-trofoblasto, sendo a pré-eclâmpsia também associada a um estresse oxidativo crônico nesta porção da placenta (Sivasubramaniam *at al.*, 2002).

Em concordância ao que foi apresentado por Moreira de Mello e colaboradores (2010) pode-se observar aqui que a placenta humana a termo se apresenta como um mosaico em relação à inativação do cromossomo X, e que vilos não adjacentes podem apresentar padrões distintos de ICX.

Apesar de, no total, este trabalho apresentar dados de apenas seis amostras de placenta informativas, a interpretação dos resultados de sequenciamento tipo Sanger e pirosequenciamento torna evidente que *MAOA*, *MAOB* e *USP9X* não são regulados por *imprinting* na placenta humana a termo, pois os níveis alelo-específicos de sua expressão seguem o padrão de ICX de cada vilo, (Tabelas 9 e 10 e figuras 17 e 19). As análises em cérebro tampouco indicam *imprinting* para *MAOA* e *MAOB*. Assim, podemos afirmar que estes genes não possuem expressão dependente de origem parental nas duas principais fontes de identificação de genes “imprintados”, cérebro e placenta.

Como descrito, genes presentes no cromossomo X inativo podem ainda assim escapar à inativação apresentando expressão bialélica (Carrel and Willard, 2005), fato que pode ser facilmente confundindo com ICX aleatória quando se determina o padrão de ICX pela expressão de um pequeno número de genes (Moreira de Mello *et al.*, 2010). Sabe-se também que os genes que escapam à ICX não são necessariamente os mesmos entre indivíduos e até mesmo entre tecidos de uma mesma pessoa (Berletch *et al.*, 2011; Cotton *et al.*, 2013). Ao propor que *MAOA*, *MAOB* e *USP9X* fossem genes candidatos a serem regulados por *imprinting* Skuse e colaboradores (1997) e Good e colaboradores (2003) ressaltaram que estes genes escapam à ICX como publicado por Distech e colaboradores (1999) e por Carrel e Willard (2005). Em 2009 Stabellini e colaboradores mostraram que *MAOA* é submetido à inativação em fibroblastos humanos e os resultados aqui expostos

também indicam que na placenta humana a termo os genes *MAOA*, *MAOB* e *USP9X* não escapam à ICX pois foi possível identificar expressão monoalélica em 11 das 13 amostras heterozigotas para *MAOA*, duas das três para *MAOB* e seis das sete para *USP9X* (Tabelas 9 e 10 e figura 17). Nas amostras de cérebros humanos, entretanto, não é possível afirmar se os genes escapam ou não à ICX, pois para apenas uma amostra foi possível determinar ICX aleatória de acordo com a expressão de *XIST* e todas apresentaram expressão de ambos os alelo para os genes *MAOA* e *MAOB* (Figura 20 e tabela11).

Conclusões gerais

Desde a proposição de Mary Lyon, em 1961, de que mamíferos apresentam um mecanismo de compensação de dose dos cromossomos sexuais entre fêmeas e machos, a inativação do cromossomo X (ICX) tem se tornado um campo bastante empolgante da ciência básica e aplicada. Muito do conhecimento que se tem hoje advém de estudos em organismos modelo como os camundongos. Entretanto, ao longo dos últimos cinquenta anos foram observadas diferenças expressivas entre humanos e camundongos: diferenças entre genes regulatórios envolvidos no início processo de ICX (como *XIST/Xist*, *Tsix* e *XACT*); entre a escolha do X inativado (“imprintada” em tecidos extra-embrionários de camundongo e aleatória em humanos); entre o poder que as células tronco embrionárias murinas tem em recapitular o processo de ICX *in vitro* enquanto que as humanas não, são apenas alguns exemplos.

Ao se avaliar o perfil de atividade alelo-específico de genes do cromossomo X de embriões humanos em diferentes estágios pré-implantacionais o presente estudo trouxe contribuições significativas à compreensão do início do processo de ICX e escolha do X inativo:

- Foi identificado aqui que o gene *XIST*, responsável pelo início do processo ICX, somente é expresso no estágio de oito células, ao mesmo tempo em que se observa a ativação do genoma próprio do embrião;

- Este mesmo gene pode ter seus transcritos detectados em células femininas ou masculinas, entretanto sua expressão é estabilizada nas células femininas no estágio de blastocisto onde ocorre regulação positiva do gene;

- O silenciamento transcricional dos genes do cromossomo X tem início no estágio de blastocisto e a escolha do X inativo se dá de forma aleatória.

Entretanto, a propagação global do sinal de silenciamento gênico deve ocorrer em estágio pós-implantacional;

–Adicionalmente foi demonstrada pela primeira vez que em células embrionárias a proporção de genes autossômicos com expressão monoalélica é muito maior do que se esperava coerente com o observado recentemente para células adultas.

Em relação à existência de genes regulados por *imprinting* no cromossomo X humano, resultados fenotípicos de mulheres com síndrome de Turner parecem apontar para tal fenômeno. Foram testados no presente trabalho três genes do cromossomo X candidatos a serem assim regulados (*MAOA*, *MAOB* e *USP9X*) entretanto, não se observou qualquer evidência de que estes fossem genes “imprintados” em amostras de cérebro ou de placenta humanos. A falta de evidência de que estes sejam genes regulados por *imprinting* não invalida as observações feitas em mulheres com síndrome de Turner que indicam a presença de genes “imprintados” no X. São cada vez mais abundantes dados de sequenciamento de última geração de exoma e transcriptoma humanos depositados em bancos públicos, utilizados nas buscas por genes expressos de acordo com a origem parental. As técnicas de bioinformática aqui aplicadas se mostraram eficientes em estudar genes do cromossomo X a partir de dados de RNA-Seq, em geral negligenciados nesses tipos de estudo dada a existência da ICX.

Resumo

Eventos epigenéticos como o *imprinting* genômico e a inativação do cromossomo X (ICX), já foram amplamente estudados em camundongos. Nesses animais muitos dos processos epigenéticos que levam à ICX já estão profundamente esclarecidos. Em humanos entretanto, o conhecimento sobre a ICX é mais limitado, em particular os eventos iniciais do processo durante o desenvolvimento embrionário.

O desenvolvimento e aprimoramento de ensaios que envolvem o sequenciamento em larga escala do transcriptoma (RNA-Seq) de células únicas iniciam uma nova era nos estudos sobre a ICX. São crescentes os dados de RNA-Seq depositados em bancos de dados públicos e em 2013 os trabalhos de Xue e colaboradores e de Yan e colaboradores tornaram disponíveis os resultados de RNA-Seq de células individuais isoladas de embriões humanos a partir do estágio de duas células até a fase de blastocisto. Através de técnicas de bioinformática avaliamos o nível de expressão do gene *XIST*, intimamente envolvido no processo de ICX, nos diferentes estágios do desenvolvimento. Alinhamos também as leituras geradas por RNA-Seq contra o genoma humano de referência no intuito de se identificar variantes em regiões transcritas e assim verificar a origem do alelo expresso. Com isso, pudemos observar que o gene *XIST* tem sua expressão iniciada em embriões humanos no estágio de oito células, e que o silenciamento transcricional dos genes do cromossomo X já se iniciou no estágio de blastocisto de forma aleatória mas ainda não se disseminou, *i.e.* a ICX não está completa.

Devido ao fenômeno de ICX, a caracterização de genes “imprintados” neste cromossomo é desafiadora. Ainda assim em camundongos foram relatados alguns genes do X que são assim regulados. Mulheres portadoras da síndrome de Turner (45,X) apresentam diferenças fenotípicas dependentes da origem parental do

cromossomo X herdado, sugerindo a existência de genes “imprintados” no X humano. Em particular os genes *MAOA*, *MAOB* e *USP9X* foram indicados como candidatos a serem regulados por *imprinting*. Através do sequenciamento de regiões transcritas contendo SNPs em heterozigose foram avaliados o padrão de expressão alelo-específico dos três genes indicados. Nenhum sinal de regulação por *imprinting* pôde ser detectado nem em placenta nem em cérebro humano, pois a procedência dos alelos expressos era independente da origem parental. Isso não significa que a variabilidade fenotípica em mulheres com Turner não possa ser explicada por *imprinting* em genes do X. Experimentos de RNA-Seq em diversos tecidos humanos ou a partir de células únicas são uma abordagem conveniente para se elucidar este fenômeno.

Abstract

Epigenetic phenomena as genomic imprinting and X chromosome inactivation (XCI) have been widely studied in mice. While most of the processes and steps involved in XCI in mice are well studied, in humans our knowledge is still very limited, specially during early embryo development.

Advances in single-cell whole transcriptome high throughput sequencing techniques (RNA-Seq) bring a new era to the XCI field. Single-cell RNA-Seq results of from 2-cell to the blastocyst stage of human embryos were published by Xue et cols and Yan et cols in 2013.

Using bioinformatics techniques we searched for the *XIST* gene expression level (a gene closely involved in XCI) throughout the human pre-implantation embryo development. We aligned reads generated by RNA-Seq assays to the human reference genome looking for variants in gene transcriptional regions and to identify the origin of the expressed allele. Our results show that *XIST* expression starts from the 8-cell stage and is stabilized and upregulated at the female blastocyst stage. We also show that the transcriptional silence of X-linked genes started at the blastocyst stage and is independent of parental origin but this does not apply for all genes. We concluded that the completion of the transcriptional silence step is probably established during post-implantation stage.

The search for X-linked imprinted genes is challenging due to the XCI phenomenon. Nevertheless, X-imprinted genes were reported in mice. In humans, no X-imprinted genes were found so far, but phenotypic differences reported in Turner's syndrome (45,X) women was related to the parental origin of the X chromosome inherited. This suggests the existence of X-linked imprinted genes, in particular *MAOA*, *MAOB* and *USP9X* seemed good candidates. By sequencing transcript regions

containing heterozygous SNPs in these genes we could access their expression pattern. Our results show no sign of imprinting regulation of *MAOA*, *MAOB* or *USP9X*, neither in human brain nor in human term placenta. This does not rule out the possibility that the phenotypic differences observed in Turner's syndrome women could be the consequence of other unknown X-linked imprinted genes. RNA-Seq of different human female tissues is a powerful approach to finally find the genes involved in such phenotypes.

Referências bibliográficas

- Amos-Landgraf, J.M., Cottle, A., Plenge, R.M., Friez, M., Schwartz, C.E., Longshore, J., and Willard, H.F. (2006). X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. *Am. J. Hum. Genet.* 79, 493–499.
- Anderson, CL e Brown, CJ Polymorphic X-chromosome inactivation of the human TIMP1 gene. *Am.J.Hum.Genet.*, 1999. 65:699-708.
- Avner, P. and Heard E. (2001). X-chromosome inactivation: counting, choice and initiation. *Nature Reviews. Genetics* 2, 59–67.
- Bacher, C.P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., and Heard, E. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat. Cell Biol.* 8, 293–299.
- Ballestar, E., Esteller, M., and Richardson, B.C. (2006). The epigenetic face of systemic lupus erythematosus. *J. Immunol.* 176, 7143–7147.
- Berletch, J.B., Yang, F., Xu, J., Carrel, L., and Disteche, C.M. (2011). Genes that escape from X inactivation. *Hum. Genet.* 130, 237–245.
- Bermejo-Alvarez, P., Ramos-Ibeas, P., and Gutierrez-Adan, A. (2012). Solving the “X” in embryos and stem cells. *Stem Cells Dev.* 21, 1215–1224.
- Bishop, D., Canning, E., Elgar, K., Morris, E., Jacobs, P., and Skuse, D. (2000). Distinctive patterns of memory function in subgroups of females with Turner syndrome: evidence for imprinted loci on the X-chromosome affecting neurodevelopment. *Neuropsychologia* 38, 712721.
- Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*.
- Bonnin, A., Goeden, N., Chen, K., Wilson, M.L., King, J., Shih, J.C., Blakely, R.D., Deneris, E.S., and Levitt, P. (2011). A transient placental source of serotonin for the fetal forebrain. *Nature* 472, 347–350.
- Borel, C., Ferreira, P.G., Santoni, F., Delaneau, O., Fort, A., Popadin, K.Y., Garieri, M., Falconnet, E., Ribaux, P., Guipponi, M., *et al.* (2015). Biased allelic expression in human primary fibroblast single cells. *Am. J. Hum. Genet.* 96, 70–80.
- Bortolato, M., Chen, K., and Shih, J. (2008). Monoamine oxidase inactivation: from pathophysiology to therapeutics. *Advanced Drug Delivery Reviews* 60, 1527–1533.
- Boumil, R.M., and Lee, J.T. (2001). Forty years of decoding the silence in X-chromosome inactivation. *Hum. Mol. Genet.* 10, 2225–2232.
- Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 71, 515–526.

- Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349, 38–44.
- Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71, 527–542.
- Bruck, T., and Benvenisty, N. (2011). Meta-analysis of the heterogeneity of X chromosome inactivation in human pluripotent stem cells. *Stem Cell Res* 6, 187–193.
- Butler, M.G. (2009). Genomic imprinting disorders in humans: a mini-review. *J. Assist. Reprod. Genet.* 26, 477–486.
- Calvanese, V., Lara, E., Kahn, A., and Fraga, M.F. (2009). The role of epigenetics in aging and age-related diseases. *Ageing Res. Rev.* 8, 268–276.
- Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 434, 400–404.
- Chureau, C., Chantalat, S., Romito, A., Galvani, A., Duret, L., Avner, P., and Rougeulle, C. (2011). Ftx is a non-coding RNA which affects Xist expression and chromatin structure within the X-inactivation center region. *Hum. Mol. Genet.* 20, 705–718.
- Clemson, C., McNeil, J., Willard, H., and Lawrence, J. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *The Journal of Cell Biology*.
- Cotton, A.M., Ge, B., Light, N., Adoue, V., Pastinen, T., and Brown, C.J. (2013). Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol.* 14, R122.
- Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J. Cell Biol.* 153, 773–784.
- Daniels, R., Zuccotti, M., Kinis, T., Serhal, P., and Monk, M. (1997). XIST expression in human oocytes and preimplantation embryos. *Am. J. Hum. Genet.* 61, 33–39.
- Davies, W., Isles, A., Smith, R., Karunadasa, D., Burrmann, D., Humby, T., Ojarikre, O., Biggin, C., Skuse, D., Burgoyne, P., *et al.* (2005). Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. *Nature Genetics* 37, 625–629.
- Degner, J., Marioni, J., Pai, A., Pickrell, J., Nkadori, E., Gilad, Y., and Pritchard, J. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics (Oxford, England)* 25, 3207–3212.
- Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343, 193–196.
- Disteche, C.M. (1999). Escapees on the X chromosome. *Proc. Natl. Acad. Sci. U.S.A.* 96, 14180–14182.

- Dupont, C., and Gribnau, J. (2013). Different flavors of X-chromosome inactivation in mammals. *Curr. Opin. Cell Biol.* 25, 314–321.
- Eckersley-Maslin, M., and Spector, D. (2014). Random monoallelic expression: regulating gene expression one allele at a time. *Trends in Genetics* 30, 237244.
- Ge, B., Gurd, S., Gaudin, T., Dore, C., Lepage, P., Harmsen, E., Hudson, T.J., and Pastinen, T. (2005). Survey of allelic expression using EST mining. *Genome Res.* 15, 1584–1591.
- Kay, G.F., Barton, S.C., Surani, M.A., and Rastan, S. (1994). Imprinting and X chromosome counting mechanisms determine Xist expression in early mouse development. *Cell* 77, 639–650.
- Good, C.D., Lawrence, K., Thomas, N.S., Price, C.J., Ashburner, J., Friston, K.J., Frackowiak, R.S., Orelund, L., and Skuse, D.H. (2003). Dosage-sensitive X-linked locus influences the development of amygdala and orbitofrontal cortex, and fear recognition in humans. *Brain* 126, 2431–2446.
- Gupta, V., Parisi, M., Sturgill, D., Nuttall, R., Doctolero, M., Dudko, O., Malley, J., PS, E., and Oliver, B. (2006). Global analysis of X-chromosome dosage compensation. *Journal of Biology* 5, 3.
- Homan, C.C., Kumar, R., Nguyen, L.S., Haan, E., Raymond, F.L., Abidi, F., Raynaud, M., Schwartz, C.E., Wood, S.A., Gecz, J., *et al.* (2014). Mutations in USP9X are associated with X-linked intellectual disability and disrupt neuronal cell migration and growth. *Am. J. Hum. Genet.* 94, 470–478.
- Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. *Science* 293, 1074–1080.
- Johnston, C., Lovell, F., Leongamornlert, D., Stranger, B., Dermitzakis, E., and Ross, M. (2008). Large-scale population study of human cell lines indicates that dosage compensation is virtually complete. *PLoS Genetics* 4, e9.
- Kesler, S., Garrett, A., Bender, B., Yankowitz, J., Zeng, S., and Reiss, A. (2004). Amygdala and hippocampal volumes in Turner syndrome: a high-resolution MRI study of X-monosomy. *Neuropsychologia* 42, 19711978.
- Keverne, E.B. (2012). Significance of epigenetics for understanding brain development, brain evolution and behaviour. *Neuroscience*.
- Kobayashi, S., Fujihara, Y., Mise, N., Kaseda, K., Abe, K., Ishino, F., and Okabe, M. (2010). The X-linked imprinted gene family *Fthl17* shows predominantly female expression following the two-cell stage in mouse embryos. *Nucleic Acids Res.* 38, 3672–3681.
- Kobayashi, S., Isotani, A., Mise, N., Yamamoto, M., Fujihara, Y., Kaseda, K., Nakanishi, T., Ikawa, M., Hamada, H., Abe, K., *et al.* (2006). Comparison of Gene Expression in Male and Female Mouse Blastocysts Revealed Imprinting of the X-Linked Gene, *Rhox5/Pem*, at Preimplantation Stages. *Current Biology*.
- Kobayashi, S., Totoki, Y., Soma, M., Matsumoto, K., Fujihara, Y., Toyoda, A., Sakaki, Y., Okabe, M., and Ishino, F. (2013). Identification of an imprinted gene cluster in the X-inactivation center. *PLoS ONE* 8, e71222.

- Koboldt, D., Chen, K., Wylie, T., Larson, D., McLellan, M., Mardis, E., Weinstock, G., Wilson, R., and Ding, L. (2009). VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics (Oxford, England)* 25, 2283–2285.
- Koboldt, D., Zhang, Q., Larson, D., Shen, D., McLellan, M., Lin, L., Miller, C., Mardis, E., Ding, L., and Wilson, R. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 22, 568–576.
- Kocabas, A.M., Crosby, J., Ross, P.J., Otu, H.H., Beyhan, Z., Can, H., Tam, W.-L.L., Rosa, G.J., Halgren, R.G., Lim, B., *et al.* (2006). The transcriptome of human oocytes. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14027–14032.
- Koressaar, T., and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Lee, J. (2000). Disruption of Imprinted X Inactivation by Parent-of-Origin Effects at Tsix. *Cell* 103.
- Lee, J. (2011). Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nature Reviews. Molecular Cell Biology* 12, 815–826.
- Lee, J.T., and Bartolomei, M.S. (2013). X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. *Cell* 152, 1308–1323.
- Lee, J.T., and Lu, N. (1999). Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* 99, 47–57.
- Lee, J.T., Davidow, L.S., and Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. *Nature Genetics* 21, 400–404.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Lopez-Gines, C., Gil-Benso, R., Ferrer-Luna, R., Benito, R., Serna, E., Gonzalez-Darder, J., Quilis, V., Monleon, D., Celda, B., and Cerdá-Nicolas, M. (2010). New pattern of EGFR amplification in glioblastoma and the relationship of gene copy number with gene expression profile. *Mod. Pathol.* 23, 856–865.
- Lyon, M.F. (1961). Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* 190, 372–373.
- Mak, W., Nesterova, T.B., de, N.M., Appanah, R., Yamanaka, S *et al* Reactivation of the paternal X chromosome in early mouse embryos. *Science*, 2004. 303:666-669.
- Metsalu, T., Viltrop, T., Tiirats, A., Rajashekar, B., Reimann, E., Köks, S., Rull, K., Milani, L., Acharya, G., Basnet, P., *et al.* (2014). Using RNA sequencing for identifying gene imprinting and random monoallelic expression in human placenta. *Epigenetics* 9, 1397–1409.

- Lyon, M.F. (1962). Sex chromatin and gene action in the mammalian X-chromosome. *American Journal of Human Genetics* 14, 135–148.
- Mietton, F., Sengupta, A., Molla, A., Picchi, G., Barral, S., Heliot, L., Grange, T., Wutz, A., and Dimitrov, S. (2009). Weak but uniform enrichment of the histone variant macroH2A1 along the inactive X chromosome. *Molecular and Cellular Biology* 29, 150–156.
- Migeon, B.R. (2003). Is Tsix repression of Xist specific to mouse? *Nat. Genet.* 33, 337; author reply 337–8.
- Migeon, B.R., Chowdhury, A.K., Dunston, J.A., and McIntosh, I. (2001). 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, 951–960.
- Monk, D., Arnaud, P., Frost, J., Hills, F., Stanier, P., Feil, R., and Moore, G. (2009). Reciprocal imprinting of human GRB10 in placental trophoblast and brain: evolutionary conservation of reversed allelic expression. *Human Molecular Genetics* 18, 3066–3074.
- Moreira de Mello, J. C. Estudo do padrão de inativação do cromossomo X em tecido extra-embriônico humano. 2010. 110f. Dissertação (Mestrado em Ciências) - Instituto de Biociências, Universidade de São Paulo, São Paulo.
- Moreira de Mello, J.C., de Araújo, E.S.S., Stabellini, R., Fraga, A.M., de Souza, J.E., Sumita, D.R., Camargo, A.A., and Pereira, L.V. (2010). Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. *PLoS ONE* 5, e10947.
- Moreira de Mello, J.C., de Araújo, E.S.S., Stabellini, R., Fraga, A.M., de Souza, J.E., Sumita, Nguyen, D.K., and Disteche, C.M. (2006). Dosage compensation of the active X chromosome in mammals. *Nat. Genet.* 38, 47–53.
- Ohno S (1967) Sex chromosomes and sex linked genes. Springer, Berlin Heidelberg New York
- Ohno, S., Kaplan, W.D., and Kinoshita, R. (1959). Formation of the sex chromatin by a single X-chromosome in liver cells of *Rattus norvegicus*. *Exp. Cell Res.* 18, 415–418.
- Okamoto, I., Arnaud, D., Le Baccon, P., Otte, A.P., Disteche, C.M., Avner, P., and Heard, E. (2005). Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature* 438, 369–373.
- Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D., and Heard, E. (2004). Epigenetic dynamics of imprinted X inactivation during early mouse development. *Science* 303, 644–649.
- Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., Wolf, J.-P.P., Renard, J.-P.P., Duranthon, V., *et al.* (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature* 472, 370–374.
- Patrat, C., Okamoto, I., Diabangouaya, P., Vialon, V., Baccon, P., Chow, J., and Heard, E. (2009). Dynamic changes in paternal X-chromosome activity during imprinted X-chromosome inactivation in mice. *Proceedings of the National Academy of Sciences of the United States of America* 106, 5198–5203.

- Payer, B., and Lee, J. (2008). X chromosome dosage compensation: how mammals keep the balance. *Annual Review of Genetics* 42, 733–772.
- Payer, B., Lee, J.T., and Namekawa, S.H. (2011). X-inactivation and X-reactivation: epigenetic hallmarks of mammalian reproduction and pluripotent stem cells. *Hum. Genet.* 130, 265–280.
- Pessia, E., Engelstädter, J., and Marais, G.A. (2014). The evolution of X chromosome inactivation in mammals: the demise of Ohno's hypothesis? *Cell. Mol. Life Sci.* 71, 1383–1394.
- Pinter, S., Sadreyev, R., Yildirim, E., Jeon, Y., Ohsumi, T., Borowsky, M., and Lee, J. (2012). Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Res.*
- Postovit, L.-M.M., Costa, F.F., Bischof, J.M., Seftor, E.A., Wen, B., Seftor, R.E., Feinberg, A.P., Soares, M.B., and Hendrix, M.J. (2007). The commonality of plasticity underlying multipotent tumor cells and embryonic stem cells. *J. Cell. Biochem.* 101, 908–917.
- R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Raefski, A.S., and O'Neill, M.J. (2005). Identification of a cluster of X-linked imprinted genes in mice. *Nat. Genet.* 37, 620–624.
- Rastan, S., and Cattanaach, B.M. (1983). Interaction between the Xce locus and imprinting of the paternal X chromosome in mouse yolk-sac endoderm. *Nature* 303, 635–637.
- Ray, P.F., Winston, R.M., and Handyside, A.H. (1997). XIST expression from the maternal X chromosome in human male preimplantation embryos at the blastocyst stage. *Hum. Mol. Genet.* 6, 1323–1327.
- Genet. 6, 1323–1327.
- Renfree, M.B., Suzuki, S., and Kaneko-Ishino, T. (2013). The origin and evolution of genomic imprinting and viviparity in mammals. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 368, 20120151.
- Richards, B., Skoletsky, J., Shuber, A.P., Balfour, R., Stern, R.C., Dorkin, H.L., Parad, R.B., Witt, D., and Klinger, K.W. (1993). Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. *Hum. Mol. Genet.* 2, 159–163.
- Richards, E. (2006). Inherited epigenetic variation--revisiting soft inheritance. *Nature Reviews. Genetics* 7, 395–401.
- Sado, T., Wang, Z., Sasaki, H., and Li, E. (2001). Regulation of imprinted X-chromosome inactivation in mice by Tsix. *Development* 128, 1275–1286.
- Santos, F.R., Pena, S.D., and Epplen, J.T. (1993). Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. *Hum. Genet.* 90, 655–656.

- Schultz, R.M. (2002). The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. *Human Reproduction Update* 8, 323–331.
- Schulz, R., Menheniott, T.R., Woodfine, K., Wood, A.J., Choi, J.D., and Oakey, R.J. (2006). Chromosome-wide identification of novel imprinted genes using microarrays and uniparental disomies. *Nucleic Acids Res.* 34, e88.
- Sharman, G. (1971). Late DNA replication in the paternally derived X chromosome of female kangaroos. *Nature* 230, 231–232.
- Silva, S.S., Rowntree, R.K., Mekhoubad, S., and Lee, J.T. (2008). X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4820–4825.
- Sivasubramaniam, S.D., Finch, C.C., Billett, M.A., Baker, P.N., and Billett, E.E. (2002). Monoamine oxidase expression and activity in human placentae from pre-eclamptic and normotensive pregnancies. *Placenta* 23, 163–171.
- Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., *et al.* (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423, 825–837.
- Skuse, D.H., James, R.S., Bishop, D.V., Coppin, B., Dalton, P., Aamodt-Leeper, G., Bacarese-Hamilton, M., Creswell, C., McGurk, R., and Jacobs, P.A. (1997). Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function. *Nature* 387, 705–708.
- Stabellini, R. Análise funcional dos genes XIST e DNMT1 na manutenção do processo de inativação do cromossomo X humano através do silenciamento gênico por RNAi. 2008. 143f. Tese (Doutorado em Ciências) - Instituto de Biociências, Universidade de São Paulo, São Paulo.
- Stabellini, R., Vasques, L.R., de Mello, J.C., Hernandez, L.M., and Pereira, L.V. (2009). MAOA and GYG2 are submitted to X chromosome inactivation in human fibroblasts. *Epigenetics* 4, 388–393.
- Stavropoulos, N., Lu, N., and Lee, J. (2001). A functional role for Tsix transcription in blocking Xist RNA accumulation but not in X-chromosome choice. *Proceedings of the National Academy of Sciences*.
- Surani, M.A., Hayashi, K., and Hajkova, P. (2007). Genetic and epigenetic regulators of pluripotency. *Cell* 128, 747–762.
- Takagi, N e Sasaki, M Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature*, 1975. 256:640-642.
- Talebizadeh, Z., Simon, S.D., and Butler, M.G. (2006). X chromosome gene expression in human tissues: male and female comparisons. *Genomics* 88, 675–681.
- Tang, W.Y., and Ho, S.M. (2007). Epigenetic reprogramming and imprinting in origins of disease. *Rev Endocr Metab Disord* 8, 173–182.

- Tian, D., Sun, S., and Lee, J. (2010). The Long Noncoding RNA, Jpx, Is a Molecular Switch for X Chromosome Inactivation. *Cell* 143.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 40, e115.
- Vallot, C., Huret, C., Leseque, Y., Resch, A., Oudrhiri, N., Bennaceur-Griscelli, A., Duret, L., and Rougeulle, C. (2013). XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells. *Nat. Genet.*
- van den Berg, I.M., Laven, J.S., Stevens, M., Jonkers, I., Galjaard, R.-J.J., Gribnau, J., and van Doorninck, J.H. (2009). X chromosome inactivation is initiated in human preimplantation embryos. *Am. J. Hum. Genet.* 84, 771–779.
- Vassena, R., Boué, S., González-Roca, E., Aran, B., Auer, H., Veiga, A., and Izpisua Belmonte, J.C. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. *Development* 138, 3699–3709.
- Veitia, R., Veyrunes, F., Bottani, S., and Birchler, J. (2015). X chromosome inactivation and active X upregulation in therian mammals: facts, questions and hypotheses. *J Mol Cell Biol.*
- Wang, X., Sun, Q., McGrath, S.D., Mardis, E.R., Soloway, P.D., and Clark, A.G. (2008). Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS ONE* 3, e3839.
- Wang, X., Soloway, P.D., and Clark, A.G. (2011). A survey for novel imprinted genes in the mouse placenta by mRNA-seq. *Genetics* 189, 109–122.
- Wang, X., Miller, D.C., Harman, R., Antczak, D.F., and Clark, A.G. (2013). Paternally expressed genes predominate in the placenta. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10705–10710.
- Wang, C.C., Billett, E., Borchert, A., Kuhn, H., and Ufer, C. (2013). Monoamine oxidases in development. *Cell. Mol. Life Sci.* 70, 599–630.
- Wang, X., Miller, D.C., Clark, A.G., and Antczak, D.F. (2012). Random X inactivation in the mule and horse placenta. *Genome Res.* 22, 1855–1863.
- Wei, Y., Su, J., Liu, H., Lv, J., Wang, F., Yan, H., Wen, Y., Liu, H., Wu, Q., and Zhang, Y. (2014). Metalprint: an information repository of mammalian imprinted genes. *Development* 141, 2516–2523.
- West, J.D., Frels, W.I., Chapman, V.M., and Papaioannou, V.E. Preferential expression of the maternally derived X chromosome in the mouse yolk sac. *Cell*, 1977. 12:873-882.
- Wood, A.J., and Oakey, R.J. (2006). Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genet.* 2, e147.
- Xu, N., Tsai, C.-L.L., and Lee, J.T. (2006). Transient homologous chromosome pairing marks the onset of X inactivation. *Science* 311, 1149–1152.
- Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.-Y.Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., *et al.* (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature*.

- Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., *et al.* (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.*
- Zuccotti, M., and Monk, M. (1995). Methylation of the mouse Xist gene in sperm and eggs correlates with imprinted Xist expression and paternal X-inactivation. *Nat. Genet.* 9, 316–320.

Endereços na internet

- Dados do sequenciamento:
<http://www.ncbi.nlm.nih.gov/geo/>
número de acesso: GSE44183 (Xue *at al.*, 2013); GSE36552 (Yan *at al.*, 2013)
- UCSC Genome Browser:
<https://genome.ucsc.edu/>
- Banco de SNPs NCBI:
<https://www.ncbi.nlm.nih.gov/snp/>
- UNIGENE:
<http://www.ncbi.nlm.nih.gov/UniGene>
- Montagem do genoma humano:
<http://hgdownload.cse.ucsc.edu/goldenpath/hg19/chromosomes/>
- Relação dos transcritos humanos:
<http://hgdownload.cse.ucsc.edu/goldenpath/hg19/bigZips/>
- Banco de SNPs:
ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/BED/
- Anotação dos genes humanos:
ftp://ftp.sanger.ac.uk/pub/gencode/Gencode_human/release_19
<http://www.ensembl.org/biomart/martview>.
- Programas e ferramentas:
<http://nebc.nerc.ac.uk/>
<http://www.ncbi.nlm.nih.gov/>

Anexos

Anexo 1 | Reads alinhadas em cada fase do processo e SNPs anotados por amostra da série GSE36552

Estágio	Amostra	célula	ID	reads			alinhamento com Hg19			alinhamento com refmRNA			resultado do alinhamento			SNPs		
				totais	com boa qualidade		alinhadas	%	total	%	alinhadas		total	%	total	%	total	total
					com	total					%	total						
Oócito		#1	SRR445718	32943665	94.16	31018141	58.09	18017783	30.19	55.79	5224419	23242202	75.00	279451	75.00	279451	279451	279451
		#2	SRR445719	30690825	94.24	28923322	59.31	17155098	29.27	55.64	4709832	21864930	76.00	288772	76.00	288772	288772	288772
		#3	SRR445720	25552876	94.09	24041774	58.59	14087275	29.94	55.17	3971673	18058948	75.00	262788	75.00	262788	262788	262788
		*		29729122	94.16	27994412.33	58.66	16420052	29.8	55.53	4635308	21055360	75.00	277003.66	75.00	277003.66	277003.66	277003.66
Zigoto		#1	SRR445721	36528347	95.22	34781848	43.51	15133236	41.55	22.98	3321017	18454253	53.00	276186	53.00	276186	276186	276186
		#2	SRR445722	32865315	95.15	31270021	44.31	13856344	40.39	21.46	2710609	16566953	53.00	241946	53.00	241946	241946	241946
		#3	SRR445723	41057289	93.62	38438544	46.19	17754116	34.95	29.85	4009912	21764028	57.00	313592	57.00	313592	313592	313592
		*		36816983.67	94.66	34830137.67	44.67	15581232	38.96	24.76	3347179.333	18928411.33	54.00	277241.33	54.00	277241.33	277241.33	277241.33
2-células	emb.#1	#1	SRR445724	50896528	93.39	47533782	42.22	20070188	39.15	28.09	5227321	25297509	53.00	294759	53.00	294759	294759	294759
		#2	SRR445725	51954904	93.45	48550067	44.31	21510869	37.98	31.01	5719033	27229902	56.00	309490	56.00	309490	309490	309490
	emb.#2	#1	SRR445726	49180773	93.83	46146024	46.40	21411806	35.55	28.67	4703057	26114863	57.00	285419	57.00	285419	285419	285419
		#2	SRR445727	52518698	93.44	49075626	45.80	22474673	35.30	30.45	5275210	27749883	57.00	306226	57.00	306226	306226	306226
4-células	emb.#3	#1	SRR445728	48727188	93.18	45405208	47.09	21380837	35.14	29.81	4755834	26136671	58.00	278658	58.00	278658	278658	278658
		#2	SRR445729	59155370	92.91	54960016	46.63	25628951	35.68	29.52	5789709	31418660	57.00	314204	57.00	314204	314204	314204
	emb.#1	*		52072243.5	93.36	48611787.17	45.4	22079554	36.46	29.59	5245027.333	27324581.33	56.00	296126.75	56.00	296126.75	296126.75	296126.75
		#1	SRR490961	49127668	94.59	46471492	56.10	26068582	29.23	53.87	7317756	33386338	72.00	498698	72.00	498698	498698	498698
		#2	SRR490962	45583212	95.5	43532982	54.91	23902088	30.20	53.00	6967542	30869630	71.00	418186	71.00	418186	418186	418186
		#4	SRR490964	31651487	95.25	30146725	52.89	15944097	31.90	48.41	4655457	20599554	68.00	310167	68.00	310167	310167	310167
	emb.#2	#1	SRR490965	58021187	92	53378355	56.20	29999050	22.01	47.56	5587246	35586296	67.00	425973	67.00	425973	425973	425973
		#2	SRR490966	53970537	91.9	49596442	56.46	28002177	22.65	47.60	5346584	33348761	67.00	397096	67.00	397096	397096	397096
		#3	SRR490967	48487342	92.28	44742642	55.47	24817902	23.18	45.31	4698999	29516901	66.00	340612	66.00	340612	340612	340612
		#4	SRR490968	62850846	91.73	57655058	56.53	32592879	22.12	49.32	6289421	38882300	67.00	442549	67.00	442549	442549	442549
	emb.#3	#1	SRR490969	53928812	92.21	49730380	55.53	27615301	24.92	47.54	5889959	33505260	67.00	398606	67.00	398606	398606	398606
		#2	SRR490970	52735750	92.33	48689862	56.43	27477895	25.65	53.08	6629595	34107490	70.00	444320	70.00	444320	444320	444320
		#3	SRR490971	38896352	93.25	36271442	55.72	20209707	25.51	50.16	4641755	24851462	69.00	351919	69.00	351919	351919	351919
		#4	SRR490972	58235449	93.14	54238253	55.32	30004366	25.31	49.10	6739001	36743367	68.00	440482	68.00	440482	440482	440482
8-células	emb.#1	*		50317149.27	93.1	46768512.09	55.59	26057640.36	25.69	49.54	5887574.091	31945214.45	68.00	406237.09	68.00	406237.09	406237.09	406237.09
		#1	SRR490973	31911442	91.51	29203533	46.17	13484722	36.79	27.40	2943633	16428355	56.00	223739	56.00	223739	223739	223739
		#2	SRR490974	30282390	89.62	27137822	44.09	11966044	40.52	23.74	2610583	14576627	54.00	211185	54.00	211185	211185	211185
		#3	SRR490975	33565816	86.99	29199190	33.07	9655330	53.31	12.24	1905202	11560532	40.00	167444	40.00	167444	167444	167444
		#4	SRR490976	33261944	91.54	30447017	52.54	15996922	28.38	38.20	3300584	19297506	63.00	266944	63.00	266944	266944	266944
		#1	SRR490977	27602538	89.67	24752197	47.29	11706378	33.73	31.09	2595747	14302125	58.00	203893	58.00	203893	203893	203893
		#2	SRR490978	32482770	91.02	29567293	45.18	13357718	35.93	28.45	3022492	16380210	55.00	216618	55.00	216618	216618	216618
		#3	SRR490979	28271452	90.04	25456162	43.39	11045739	39.36	21.69	2172737	13218476	52.00	176324	52.00	176324	176324	176324
	emb.#2	#4	SRR490980	31915232	91.93	29339298	50.18	14721241	32.07	34.91	3284531	18005772	61.00	245523	61.00	245523	245523	245523
		#5	SRR490981	34098900	95.67	32622235	51.53	16810645	31.20	30.36	3089746	19900391	61.00	337895	61.00	337895	337895	337895
		#6	SRR490982	32376268	95.77	31008084	53.06	16452509	29.25	32.15	2915301	19367810	62.00	308345	62.00	308345	308345	308345
		#7	SRR490983	19792800	94.96	18794296	51.85	9744450	29.36	32.29	1782144	11526594	61.00	177210	61.00	177210	177210	177210
		#8	SRR490984	22364554	95.09	21266940	53.25	11324580	28.24	34.30	2060083	13384663	63.00	215896	63.00	215896	215896	215896

Anexo 1| Reads alinhadas em cada fase do processo e SNPs anotados por amostra da série GSE36552 (continuação)

Estágio	Amostra	célula	ID	reads			alinhamento com Hg19			alinhamento com refmRNA			resultado do alinhamento			SNPs	
				totais	com boa qualidade		%	alinhadas		%	alinhadas		%	total	%	total	total
					%	total		%	total		%	total					
8-células	emb.#3	#1	SRR490985	47990468	94.18	45197216	55.98	25299437	23.19	42.99	4504902	29804339	66.00	377077			
		#2	SRR490986	42243898	94.22	39801334	56.30	22407791	23.53	46.65	4370110	26777901	67.00	343892			
		#3	SRR490987	36089791	94.92	34257643	55.96	19171282	23.35	49.31	3944611	23115893	67.00	328291			
		#4	SRR490988	48887034	94.6	46247755	55.71	25763895	23.13	50.50	5400883	31164778	67.00	372862			
		#5	SRR490989	42851497	94.64	40556694	56.41	22876250	23.51	51.67	4926020	27802270	69.00	361261			
		#6	SRR490990	47500857	94.32	44801736	55.75	24974747	24.15	52.01	5628241	30602988	68.00	340872			
		#7	SRR490991	33545010	95.32	31973638	54.74	17503377	24.40	48.59	3790617	21293994	67.00	300212			
		#8	SRR490992	43263115	95.32	41240467	56.03	23105749	23.74	50.71	4964134	28069883	68.00	359895			
Mórula	emb.#1	*		35014888.8	93.06	32643527.5	50.92	16868440.3	30.357	36.96	3460615.05	20329055.35	62.00	276768.9			
		#1	SRR490993	44585235	88.72	39554218	54.90	21714004	18.37	38.56	2801675	24515679	62.00	279288			
		#2	SRR490994	28690806	89.12	25568351	50.39	12884166	21.19	33.06	1790777	14674943	57.00	166716			
		#3	SRR490995	27850586	89.47	24918333	54.20	13505723	16.07	17.09	684300	14190023	57.00	139753			
		#4	SRR490996	39065873	88.74	34668523	50.36	17457575	23.49	33.38	2718849	20176424	58.00	238703			
		#5	SRR490997	33894431	89.92	30477190	50.48	15384962	20.23	33.51	2066029	17450991	57.00	184629			
		#6	SRR490998	29489887	90.49	26683957	52.06	13892951	19.84	41.94	2220948	16113899	60.00	209838			
		#7	SRR490999	24327382	90.57	22032348	48.98	10791952	20.97	29.55	1365248	12157200	55.00	127901			
		#8	SRR491000	29435822	91.68	26985633	57.38	15483971	15.19	19.18	785871	16269842	60.00	192905			
		#1	SRR491001	39037531	92.34	36048260	50.03	18033712	27.58	37.73	3751463	21785175	60.00	291299			
		#3	SRR491003	25135705	92.89	23347550	51.82	12098113	26.67	39.55	2462738	14560851	62.00	207068			
		#4	SRR491004	41041777	91.42	37519596	51.50	19322797	24.98	36.85	3453597	22776394	61.00	256834			
Blastocisto	emb.#2	#5	SRR491005	33698770	91.89	30966498	51.83	16050311	25.47	40.77	3215697	19266008	62.00	274954			
		#6	SRR491006	36456083	91.53	33369040	46.47	15508136	30.94	29.34	3028524	18536660	56.00	207776			
		#7	SRR491007	26736090	94.17	25177291	46.26	11645886	31.23	25.20	1981833	13627719	54.00	161599			
		#8	SRR491008	36989021	94.06	34790242	44.31	15414736	34.83	27.64	3349491	18764227	54.00	252775			
		*		33095666.6	91.13	30140468.67	50.73	15279266.33	23.8	32.22	2378469.333	17657735.67	59.00	212802.53			
		#1	SRR491009	44798733	91.32	40909596	54.30	22214087	17.55	52.62	3776950	25991037	64.00	219294			
		#2	SRR491010	21392946	92.06	19693470	54.15	10663993	19.28	49.16	1866798	12530791	64.00	178598			
		#3	SRR491011	33286654	90.7	30192573	50.09	15123151	23.29	33.65	2366433	17489584	58.00	240550			
		#4	SRR491012	36693969	92.13	33804647	54.80	18524121	17.00	48.97	2814644	21338765	63.00	223608			
		#5	SRR491013	33877559	91.76	31086579	54.90	17067537	18.32	52.09	2965755	20033292	64.00	263173			
		#6	SRR491014	24177697	91.98	22239761	55.22	12280554	16.65	46.69	1728427	14008981	63.00	183423			
		#7	SRR491015	32500760	92.19	29960830	55.39	16595952	17.08	52.38	2680751	19276703	64.00	235529			
		#8	SRR491016	28250519	90.09	25451233	50.59	12876445	19.69	37.39	1873757	14750202	58.00	170572			
		#9	SRR491017	53804158	90.54	48716076	55.90	27232460	17.11	56.12	4676528	31908988	65.00	302524			
		#10	SRR491018	26834187	90.83	24373248	55.02	13409961	18.02	53.27	2339907	15749868	65.00	222151			
		#11	SRR491019	18049650	92.33	16665325	54.70	9116055	19.20	55.20	1765781	10881836	65.00	163008			
		#12	SRR491020	22083970	91.61	20230173	47.70	9650036	21.05	28.58	1217142	10867178	54.00	126832			

Anexo 1 Reads alinhadas em cada fase do processo e SNPs anotados por amostra da série GSE36552 (continuação)															
Estágio	Amostra	célula	ID	reads		alinhamento com Hg19			alinhamento com refmRNA			resultado do		SNPs	
				com boa qualidade		alinhadas		%	não alinhadas		%	alinhamento		alinhamento	
				%	total	%	total		%	total		total	%	total	total
Blastocisto	emb. #2	#1	SRR491021	89.14	26387606	50.58	13346110	20.34	25.03	1343508	14689618	56.00	131424	131424	
		#2	SRR491022	90.39	25787464	56.82	14651252	17.20	48.18	2136924	16788176	65.00	233346	233346	
		#3	SRR491023	90.48	20894453	55.16	11525954	15.28	47.35	1511823	13037777	62.00	158176	158176	
		#4	SRR491024	90.61	23005299	57.70	13274390	21.27	63.58	3111380	16385770	71.00	273504	273504	
		#5	SRR491025	91.24	23360606	54.69	12776876	19.95	60.53	2820716	15597592	67.00	197040	197040	
		#6	SRR491026	91.12	25317482	55.44	14036222	23.05	65.69	3832962	17869184	71.00	270775	270775	
		#7	SRR491027	91.21	16881915	54.45	9192719	23.44	63.88	2527626	11720345	69.00	215146	215146	
		#8	SRR491028	91.63	20695930	56.14	11618950	14.96	46.11	1427438	13046388	63.00	156697	156697	
		#9	SRR491029	91.01	30951748	54.59	16895912	18.23	54.54	3078357	19974269	65.00	156628	156628	
		#10	SRR491030	91.56	31834841	55.77	17755704	22.18	61.31	4327990	22083694	69.00	279885	279885	
emb. #3		#1	SRR491031	88.45	43351089	53.79	23319656	16.26	34.73	2447961	25767617	59.00	168426	168426	
		#2	SRR491032	88.62	35518452	52.34	18589842	20.81	49.85	3684726	22274568	63.00	239141	239141	
		#3	SRR491033	89.35	26771477	54.11	14486747	24.77	61.49	4078395	18565142	69.00	287004	287004	
		#4	SRR491034	88.39	32680961	50.74	16581663	20.08	53.81	3531298	20112961	62.00	204693	204693	
		#5	SRR491035	88.75	29290126	51.18	14991589	22.13	53.91	3494955	18486544	63.00	235108	235108	
		#6	SRR491036	88.16	33164388	50.76	16833177	20.5	49.14	3341734	20174911	61.00	159539	159539	
		#7	SRR491037	90.28	35562166	52.43	18644783	22.9	60.33	4912938	23557721	66.00	272291	272291	
		#8	SRR491038	90.01	42764948	52.67	22522442	20.02	49.82	4265441	26787883	63.00	243000	243000	
		*		90.59	28918148.73	43.41	15526611.33	15.47	40.51664333	2864968.167	18391579.5	64.00	213702.83	213702.83	
		#2	SRR445731	92.8	34602265	56.25	19464314	19.18	54.45	3613308	23077622	67.00	351628	351628	
hESCp0		#4	SRR491039	93.16	25447265	55.75	14186932	19.89	56.53	2862066	17048998	67.00	290729	290729	
		#5	SRR491040	93.8	25023801	56.7	14188085	17.62	51.17	2256359	16444444	66.00	264255	264255	
		#6	SRR491041	93.82	22899634	57.54	13176793	18.39	52.39	2206185	15382978	67.00	262946	262946	
		#7	SRR491042	93.71	18318683	55.32	10133054	19	49.27	1714806	11847860	65.00	195654	195654	
		#8	SRR491043	93.27	27724444	55.97	15517520	18.99	48.55	2555610	18073130	65.00	270465	270465	
		#10	SRR491045	93.37	18987415	55.52	10541181	18.85	50.26	1799212	12340393	65.00	201064	201064	
		*		93.42	24714786.71	56.15	13886839.86	18.84	51.80285714	2429649.429	16316489.29	66.00	262391.57	262391.57	

Anexo 2 | Reads alinhadas em cada fase do processo e SNPs anotados por amostra da série GSE44183

Estágio	Amostra	célula	ID	reads			alinhamento com Hg19			nhemamento com refmR			resultado do			SNPs	
				com boa qualidade			alinhadas			alinhadas			alinhamento			anotados	
				totais	%	total	%	total	%	%	total	%	total	%	total	total	
Oócito	DNA Paterno		SRR689250	47632194.00	93.88	44716981.00	48.52	21694619	50.63				21694619.00	48.52	1811877		
		#1	SRR893046	18817693.00	93.42	17578996.00	25.54	4489296	65.62	44.06	5082459.00	44.06	9571755.00	54.45	87677		
		#3	SRR893047	21485629.00	93.93	20181867.00	29.19	5890889	60.56	45.07	5508194.00	45.07	11399083.00	56.48	131081		
		#4	SRR893048	15910213.00	93.07	14806844.00	23.81	3525538	68.05	36.68	3695442.00	36.68	7220980.00	48.77	82270		
Pronúcleo		*		18737845.00	93.47	17522569.00	26.18	4635241	64.74	41.93	4762031.67	41.93	9397272.67	53.63	100342.6667		
		#2	SRR893049	25367535.00	93.16	23631533.00	26.36	15141418	64.07	48.74	7380473.00	48.74	22521891.00	95.30	233508		
		#3	SRR893050	21635482.00	94.53	20451866.00	25.85	5286576	66.16	45.06	6096565.00	45.06	11383141.00	55.66	238435		
		#4	SRR893051	24632986.00	93.23	22965542.00	26.91	6180182	64.77	56.22	83633220.00	56.22	14543402.00	63.33	272842		
Zigoto		*		23878667.67	93.64	22349647.00	26.37	8869392	65	50	7280086.00	50	16149478.00	72.26	248261.6667		
		#1	SRR893052	24018461.00	93.03	22345206.00	27.33	6108041	65.66	43.3	6352725.00	43.3	12460766.00	55.76	244122		
		#2	SRR893053	20674358.00	93.09	19245822.00	25.23	4856221	65.97	35.83	4548796.00	35.83	9405017.00	48.87	137775		
		*		22346409.50	93.06	20795514.00	26.28	5482131	65.815	39.565	5450760.50	39.565	10932891.50	52.57	190948.5		
2-células		#1	SRR893054	22356935.00	93.72	20952114.00	25.99	5446335	64.71	52.72	7147674.00	52.72	12594009.00	60.11	281630		
		#5	SRR893055	18795422.00	95.19	17891303.00	26.13	4674156	63.81	39.12	4466469.00	39.12	9140625.00	51.09	151673		
		#6	SRR893056	19629441.00	94.08	18466549.00	21.75	4015723	68.72	29.49	3742854.00	29.49	7758577.00	42.01	146096		
		*		20260599.33	94.33	19103322.00	24.62	4712071.333	65.74	40.44	5118999.00	40.44	9831070.33	51.46	193133		
4-células		#1	SRR893057	19608479.00	93.06	18247788.00	25.19	4596140	62.58	49.98	5707362.00	49.98	10303502.00	56.46	277403		
		#2	SRR893058	20905573.00	94.35	19724076.00	24.71	4873848	62.45	25.52	3144058.00	25.52	8017906.00	40.65	151285		
		#3	SRR893059	20202289.00	93.14	18816259.00	21.67	4076631	64.41	16.83	2039252.00	16.83	6115883.00	32.50	69175		
		#4	SRR893060	24264089.00	94.53	22937143.00	23.53	5397843	63.74	28.48	4164489.00	28.48	9562332.00	41.69	187810		
8-células		*		21245107.50	93.77	19931316.50	23.77	4736115.5	63.295	30.2	3763790.25	30.2	8499905.75	42.65	171418.25		
	emb.#a	#3a	SRR893061	19513222.00	94.96	18529144.00	25.21	4670291	63.51	46.71	5496997.00	46.71	10167288.00	54.87	224394		
		#6a	SRR893062	17930367.00	95.03	17039994.00	25.35	4319532	62.65	43.85	4681780.00	43.85	9001312.00	52.82	205455		
		#7a	SRR893063	18704103.00	94.86	17743071.00	26.14	4637314	62.73	51.62	5745900.00	51.62	10383214.00	58.52	236999		
	emb.#b	#1b	SRR893064	21136195.00	80.59	17032954.00	19.78	3369373	64.49	27.74	3046718.00	27.74	6416091.00	37.67	89443		
		#2b	SRR893065	21721484.00	81.59	17722581.00	18.28	3239016	69.23	28.47	3492704.00	28.47	6731720.00	37.98	115065		
		#3b	SRR893066	21657472.00	85.18	18447278.00	22.09	4075903	61.42	38.98	4416369.00	38.98	8492272.00	46.04	141230		
	emb.#c	#4b	SRR893067	20776447.00	83.55	17358371.00	21.15	3671459	63.9	34.44	3820384.00	34.44	7491843.00	43.16	141943		
		#5b	SRR893068	16444543.00	85.76	14102381.00	20.96	2955969	64.18	30.96	2801875.00	30.96	5757844.00	40.83	97781		
		#3c	SRR893069	13836735.00	80.44	11130569.00	16.78	1867927	69.05	20.66	1587529.00	20.66	3455456.00	31.04	52398		
	emb. íntegro	#5c	SRR893070	14593516.00	77.11	11252958.00	12.12	1363554	77.75	11.19	979204.00	11.19	2342758.00	20.82	39599		
		*	SRR893071	20522657.00	85.87	12925726.00	19.02	2457892	67.9	21.17	1857994.00	21.17	4315886.00	33.39	62358		
Mórula		*		18306067.36	85.9	15753184.27	20.62	3329839.091	66.07	32.34	3447950.36	32.34	6777789.45	43.02	127878.6364		
	emb. íntegro	#1	SRR893072	15724202.00	94.72	14893376.00	24.83	3697505	58.35	50.77	4412279.00	50.77	8109784.00	54.45	153779		
		#4	SRR893073	16352201.00	94.92	15521270.00	24.5	3802231	58.82	38.45	3509824.00	38.45	7312055.00	47.11	113624		
		#5	SRR893074	22020236.00	94.3	20764732.00	23.43	4864470	61.73	43.35	5555976.00	43.35	10420446.00	50.18	131942		
		*		180322213.00	94.64	17059792.67	24.25	4121402	59.63	44.19	4492693.00	44.19	8614095.00	50.49	133115		

Anexo 3 Número de variantes anotadas por amostra/por cromossomo, GSE36552																															
Estágio	Amostra	célula	ID	chr1	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	chr11	chr12	chr13	chr14	chr15	chr16	chr17	chr18	chr19	chr20	chr21	chr22	chrX	chrY	chrA	Total		
Oócito		#1	SRR445718	26502	22978	17867	13124	14854	13666	13993	13157	11940	12631	13516	14342	7876	10820	10295	11679	12791	5360	11487	7188	3461	4833	5089	2	274360	279451		
		#2	SRR445719	27901	23843	18683	13541	15344	14120	14240	13623	12603	12944	14040	15053	8458	11098	10676	11738	13103	5540	11547	6989	3478	4877	5331	2	283439	288772		
		#3	SRR445720	25169	21842	16865	12054	13625	12764	13213	11932	11599	11693	12985	13819	7770	10233	9943	10824	11803	4956	10594	6633	3174	4472	4734	2	258052	262788		
	*		26524	22888	17805	12906	14608	13517	13815	12904	12047	12423	14243	13514	14405	8035	10717	10305	11414	12596	5285	11209	6937	3371	4727	5051	2	271950	277004		
Zigoto		#1	SRR445721	26205	23117	17851	12167	15061	12929	13582	13225	11036	12691	13412	14418	7926	11414	10490	11550	12621	5447	10840	7135	3046	4859	5162	2	271022	276186		
		#2	SRR445722	23014	20114	15116	10426	13198	11301	12476	11299	9986	10792	12517	12331	6079	9799	9106	10274	11576	4680	10255	6637	2495	4321	4512	2	237432	241946		
		#3	SRR445723	29980	25766	20000	13542	14763	15605	14121	13557	13957	15917	16319	7832	12700	11833	13574	15328	5934	13101	8491	3544	5835	5663	2	307927	313592			
	*		26400	22999	17656	12045	14829	12998	13888	12882	11526	12480	13829	14356	7279	11304	10476	11799	13175	5354	11399	7421	3028	5005	5112	2	272127	277241			
2-células	emb.#1	#1	SRR445724	27922	23840	18566	12427	15634	13955	15097	13756	11876	13175	15212	15340	7988	11762	11140	12807	14075	5466	12882	7959	3156	5313	5409	2	289348	294755	289348	294755
		#2	SRR445725	29493	25454	19598	13166	16220	14861	15681	14313	12656	13891	15460	15955	8394	12468	11776	13208	15111	5849	13309	8140	3300	5529	5656	2	303832	309490	303832	309490
	emb.#2	#1	SRR445726	26839	23716	17931	11979	14996	12825	14397	13250	11476	12767	14752	14792	7971	11820	11066	12368	13726	5406	12027	7871	3021	5049	5372	2	280045	285419	280045	285419
		#2	SRR445727	29102	24620	19121	13119	15814	13815	15310	14122	12480	13699	15657	15634	8102	12367	11912	13598	15219	5725	13427	8567	3282	5854	5678	2	300546	306226	300546	306226
4-células	emb.#3	#1	SRR445728	26272	23024	17813	11958	14883	12775	14128	12267	11517	12476	13737	14615	7727	10815	10420	12286	13395	5290	12202	7804	3012	5094	5146	2	273510	278658	273510	278658
		#2	SRR445729	29740	25317	20144	13240	16363	14213	15956	13914	13540	14020	15567	16452	8486	12061	12070	13743	15384	6037	14362	8737	3353	5891	5612	2	308590	314204	308590	314204
	*		28228	24329	18862	12648	15652	13741	15095	13604	12258	13338	12508	15064	15465	8111	11882	11397	13002	14485	5629	13035	8180	3187	5455	5479	2	292645	298126	292645	298126
	emb.#1	#1	SRR490961	46076	40608	32643	21433	23971	21924	24160	22023	22208	23012	25835	25466	11413	18312	19437	22999	25189	9554	24257	13352	5336	9867	9621	2	489075	498698	489075	498698
emb.#2		#2	SRR490962	37631	31821	25761	17332	20554	18962	21768	18909	18663	18342	22784	20681	9688	15940	16808	19227	21161	7941	21177	11583	5024	8662	7766	1	410419	418186	410419	418186
		#4	SRR490964	28000	23742	19163	12852	15074	13730	15893	14691	13414	14836	16838	15508	7295	11977	11684	14560	15039	5536	15612	8858	3628	6432	5805	0	304362	310167	304362	310167
		#1	SRR490965	40419	32301	26554	17069	20130	19655	21414	19108	18243	19154	23858	21262	9224	15125	15511	20377	21200	7727	22544	12187	4275	9327	9307	2	416664	425979	416664	425979
		#2	SRR490966	36026	31128	25167	17381	19414	17682	20146	17886	17634	18293	21337	20644	8670	14718	14972	18217	19358	7666	19925	11372	3871	8152	7435	2	389659	397096	389659	397096
emb.#3		#3	SRR490967	31394	27190	20775	13455	16333	14969	17426	14746	14702	14846	18156	17798	7251	11934	12513	17100	17881	6378	18248	10348	3294	7204	6669	2	333941	340612	333941	340612
		#4	SRR490968	40252	34946	27948	17766	21770	19755	22858	19078	19299	20241	23767	23230	9923	15455	17500	20766	22544	8653	22011	12686	4403	9176	8520	2	434027	442549	434027	442549
		#1	SRR490969	38116	30923	24751	15473	18750	17871	19473	16998	17473	17007	21229	20436	8373	14543	15101	19863	21167	7767	21649	11921	3823	8615	7282	2	391322	398606	391322	398606
		#2	SRR490970	41177	35206	27841	18594	21147	19388	21770	18687	20240	19885	23158	22509	9355	15805	17361	22493	23339	8649	22394	13154	4721	9281	8164	2	436154	444320	436154	444320
8-células		#3	SRR490971	33024	27123	21682	13752	16919	15375	17722	14944	15393	15078	18441	18159	7517	12901	13084	17141	18726	7044	19115	10870	3199	8114	6594	2	345323	351919	345323	351919
		#4	SRR490972	41222	33773	26965	17568	20945	19619	21663	18946	19848	18829	23003	22767	9267	15866	15959	21507	23523	8738	23932	13436	4586	10044	8474	2	432066	440482	432066	440482
	*		37576	31706	25386	16607	19546	18085	20390	17820	17920	18138	21673	20769	8907	14780	15448	19477	20830	7787	20988	11797	4196	8625	7785	2	398450	406237	398450	406237	
	emb.#1	#1	SRR490973	24577	17369	11959	8455	10785	9793	9373	7944	9280	10954	11291	12667	4290	9103	7053	9514	12817	3384	13975	7346	1944	4384	5481	1	218257	223739	218257	223739
emb.#1		#2	SRR490974	22695	15910	12782	7679	9433	8915	9074	7461	8790	9553	10410	11101	4255	7764	6399	12311	12125	3411	13381	6729	1658	4393	4955	1	206229	211185	206229	211185
		#3	SRR490975	19901	13652	9460	7075	7239	7435	6758	6250	6346	7638	8293	8739	4000	5823	5617	6365	9665	2542	10666	4728	1651	3262	4338	1	163105	167444	163105	167444
		#4	SRR490976	29529	19899	16145	10007	12809	11710	11500	10338	11079	12478	13490	15946	6471	9917	8658	10628	14070	4497	15832	7672	2532	5421	6314	2	260628	266944	260628	266944
		#1	SRR490977	22689	14936	11489	7350	9343	8486	8302	7999	8805	9310	9940	11509	4165	7481	6148	9148	12034	2858	13577	6624	1904	4836	4960	0	198933	203893	198933	203893
emb.#2		#2	SRR490978	24507	15595	12826	8077	10226	9859	9423	7762	8957	10040	10986	12079	4261	7961	6964	7706	12581	3402	14487	7015	1782	4787	5330	5	211283	216618	211283	216618
		#3	SRR490979	19494	13473	10077	6655	8145	7613	7856	6571	7404	8041	8688	10128	3497	6396	5183	6609	9934	2637	12004	5800	1435	4020	4664	0	171660	176324	171660	176324
		#4	SRR490980	27587	19200	15082	9358	11561	11055	11136	9646	10366	11117	12284	13568	6040	9518	7936	8326	12326	3995	15641	7542	1910	4576	5753	0	239770	245523	239770	245523
		#5	SRR490981	32923	27802	20572	15441	16356	16717	16239	14294	13021	15399	18499	17879	9483	12680	12680	14738	14935	5842	14906	9960	3323	6550	7653	3	330239	337895	330239	337895
emb.#2		#6	SRR490982	29623	25283	18930	13970	16823	16161	14058	12743	13363	15172	15336	16103	8679	11486	12287	12832	13243	5966	12022	8316	3206	5217	7526	0	300819	308345	300819	308345
		#7	SRR490983	18678	14370	9927	7639	8898	9475	7962	7430	7600	8729	8593	16503	5104	7302	6216	6953	7698	3076	9088	4757	1705	3489	3618	2	173590	177210	173590	177210
		#8	SRR490984	20526																											

Anexo 3 Número de variantes anotadas por amostra/por cromossomo, GSE3552 (continuação)																														
Estágio	Amostra	célula	ID	chr1	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	chr11	chr12	chr13	chr14	chr15	chr16	chr17	chr18	chr19	chr20	chr21	chr22	chrX	chrY	chrA	Total	
8-células	emb.#3	#1	SRR490985	36885	27561	22642	15441	17232	12759	18404	14899	16797	17637	18642	19617	8121	14333	12515	19376	12196	7014	21301	11647	4801	7845	5777	35	371265	377077	
		#2	SRR490986	34700	26080	21990	13380	15653	15680	16743	12703	14361	16046	16522	18152	7479	13546	11004	17946	17919	6775	19709	9867	4510	7326	5724	77	338091	343892	
		#3	SRR490987	32546	23987	19651	11788	20171	15028	15095	12388	14260	14297	16815	17365	6445	12807	9847	17380	18646	5690	18687	9987	3474	7225	4657	55	323579	328291	
		#4	SRR490988	39187	27975	22001	15133	17307	17511	16945	15216	14855	17008	18481	19377	8247	13881	12210	19616	19458	6502	21356	11629	4473	8270	6130	94	366638	372862	
		#5	SRR490989	36717	26667	20052	13856	21170	15802	15833	13640	15134	16385	18056	17975	8098	13748	11465	19107	19352	6233	21208	11185	4521	7970	5220	47	355994	361261	
		#6	SRR490990	36080	24916	20231	13054	16304	15414	15852	12398	13929	15533	16700	17202	6687	12634	11316	19057	18750	5838	21372	10500	4349	7550	5169	37	335666	340872	
		#7	SRR490991	31510	22355	17881	11591	13340	13927	14050	11203	12527	13921	14941	15696	6196	10824	9564	16127	16696	4717	18125	9588	3685	6926	4786	36	295390	300212	
		#8	SRR490992	37944	25697	21131	13681	18708	15843	15485	13215	15962	16457	18296	19707	7145	14162	11045	18959	19315	6504	21457	10908	4718	7873	5617	66	354212	359895	
Mórula	*			28915	20964	16330	10976	13624	12777	12524	10677	11602	12759	13925	14799	6232	10492	9085	13145	14644	4743	15934	8398	2985	5781	5437	23	271309	276769	
	#1	SRR490993	36525	16249	19865	12696	15475	9987	15019	8271	9280	4577	11118	11730	7526	8907	6658	13674	17372	5438	21367	9772	3148	9020	5561	53	273674	279288		
	#2	SRR490994	18794	12736	9743	6223	7643	7458	6456	6053	6861	7907	8053	9760	3753	6447	5248	6697	9750	2572	10359	4912	1449	4679	3127	36	163553	166716		
	#3	SRR490995	4422	12905	8500	8071	10301	7997	8808	6190	6123	10353	7935	7618	5443	5522	6828	1377	8174	5568	2221	1036	1375	2234	716	36	139001	139753		
	#4	SRR490996	30529	13920	16445	10646	12868	8147	11538	6699	7134	11353	8467	9548	6767	6947	5902	11791	15504	4331	17640	8093	2360	7645	4387	42	234274	238703		
	#5	SRR490997	20964	14711	10454	6871	7980	8470	6983	6802	7569	8718	9204	10814	3783	7403	5552	7869	10198	2735	11794	5645	1426	5347	3297	40	181292	184629		
	#6	SRR490998	24114	15844	11938	7636	9214	9112	8527	7873	8613	9880	10052	12112	4724	7750	6667	8863	11137	3347	14344	6567	1953	6040	3476	55	206307	209838		
	#7	SRR490999	13999	10013	7119	4942	5738	5586	4803	4972	5540	5973	5962	7768	2611	5087	4059	5103	7661	1703	8262	4125	950	3748	2130	47	125724	127901		
	#8	SRR491000	5283	18917	10533	13090	14344	12123	10012	11188	8983	13268	9111	10955	8533	7891	9489	1804	11493	6250	2190	1066	3207	2103	1044	28	191833	192905		
	#1	SRR491001	30819	21488	16662	9804	13334	12005	12858	12455	12740	12630	14963	15800	5008	10345	8916	13677	16871	4380	20331	9531	2902	7093	6642	45	284612	291299		
	#3	SRR491003	23246	15209	11358	6981	9586	8149	8696	7976	9099	9220	10688	12766	3542	7544	6731	10600	12512	2712	15011	7335	2212	5395	451	49	206568	207068		
	#4	SRR491004	28034	18452	15129	9184	12128	11209	11327	10240	11355	11707	12582	15858	4593	9906	8613	11878	14899	4244	17707	9115	2406	5915	306	47	256481	256834		
#5	SRR491005	28187	19750	15524	9350	12537	11485	11515	16123	12322	11418	13834	15458	4991	9248	7978	11843	15987	4523	19136	8927	2195	6368	6205	50	268699	274954			
Blastocisto	emb. #1	#6	SRR491006	24269	14997	11695	6654	8900	9734	8871	8000	8571	9126	10725	12523	3222	7521	6494	10058	14050	2690	15166	6930	1855	5386	302	37	207437	207776	
		#7	SRR491007	18784	11857	9179	5758	7112	7594	6674	6085	6972	7417	8008	9655	2802	5860	5013	7235	10196	2410	11944	5495	1512	3857	144	36	161419	161599	
		#8	SRR491008	27385	18432	14902	8894	11853	11529	10350	9390	10907	11370	12395	14805	4532	8727	7487	11106	14669	3935	17991	8506	2025	5939	5593	53	247129	252775	
		*			22357	15699	12603	8453	10601	9372	9496	8554	8805	9661	10206	11811	4789	7674	6776	8905	12698	3789	13698	6470	2065	5385	2892	44	209867	212803
		#1	SRR491009	24892	16841	12523	9383	10095	10404	10572	9296	9908	9779	10784	12075	4555	7309	7215	8271	11951	4084	11909	5893	2392	5072	4018	73	215203	219294	
		#2	SRR491010	20938	13815	10224	7223	8877	7883	7923	6859	7998	8413	8725	9834	4135	6135	5615	7454	10154	2810	9388	5541	1876	3695	3050	33	175515	178598	
		#3	SRR491011	26581	19475	14052	11199	12879	12457	11382	9748	10581	12119	11108	13236	6232	9007	8060	8842	12066	4293	9865	6636	2658	4222	3776	76	236698	240550	
		#4	SRR491012	25175	17933	13600	8857	10704	11313	9748	8489	9666	10279	11036	11830	5568	7533	6934	8634	12243	3686	12044	6395	2466	5235	4144	96	219368	223608	
#5	SRR491013	28731	20913	15777	11692	13162	12449	12011	11108	11412	11726	12892	13551	5825	8645	8346	11522	14058	4920	12993	7631	3181	5879	4659	90	258424	263173			
emb. #1		#6	SRR491014	20506	13981	10290	7404	8842	8685	8178	7799	8088	8775	9443	9343	4175	6204	6250	8388	9880	2974	9375	5609	2272	2443	53	247129	252775		
		#7	SRR491015	27576	18218	13419	9620	12227	11079	9889	9767	10169	10976	12049	12891	5144	8420	7416	9376	13147	3891	11862	6820	2482	5320	3727	44	231758	235529	
		#8	SRR491016	19515	13271	9934	7165	8473	8001	7339	6944	6957	8145	8424	9586	3889	5968	5208	6956	9309	2928	8721	4768	2339	3968	2704	60	167808	170572	
		#9	SRR491017	33528	24648	17769	12496	14882	14242	13139	12390	13364	14171	14862	16964	6994	10192	9235	11716	16231	5608	15712	8434	3725	6590	5566	66	296892	302524	
		#10	SRR491018	25326	18182	12503	8657	10911	11158	9625	10242	9893	10355	11164	12263	4865	7201	6754	9306	11523	3440	11623	6405	2237	4677	3752	89	218310	222151	
		#11	SRR491019	19436	13446	9299	6892	8368	8153	7151	6694	6954	7437	7953	8676	3484	5303	4935	6253	8907	2665	7828	4823	1874	3484	2906	87	160015	163008	
		#12	SRR491020	15083	10151	7242	5420	6465	6289	4765	5349	5254	5989	6429	6863	2694	4227	3774	5267	6691	1991	6585	3895	2998	1993	23	124816	126832		

Anexo 3 Número de variantes anotadas por amostra/por cromossomo, GSE36552 (continuação)																															
Estágio	Amostra	célula	ID	chr1	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	chr11	chr12	chr13	chr14	chr15	chr16	chr17	chr18	chr19	chr20	chr21	chr22	chrX	chrY	chrA	Total		
Blastocisto	emb. #2	#1	SRR491021	15352	11404	7415	5326	6888	6146	5309	6033	5903	6091	6584	7176	2563	4597	3888	4974	6982	1897	6650	3709	1416	2922	2133	66			129225	131424
		#2	SRR491022	25577	19905	14185	10301	11687	11408	11088	8908	10518	10592	11343	11977	4826	8527	6680	9641	12335	3497	12185	6721	2825	4739	3800	81			229465	233346
		#3	SRR491023	18250	13236	9237	6155	7523	8050	6539	5638	6498	7076	7954	9435	3225	5571	4271	5927	8244	3084	9134	4964	1858	3800	2480	27			155669	158176
		#4	SRR491024	30664	22853	16472	11163	14331	13452	12433	10960	11126	13936	13159	14314	6012	10056	8660	10434	14209	5265	12377	8464	3531	5087	4440	106			268958	273504
		#5	SRR491025	21972	16331	11202	7930	10900	8195	8132	7611	8636	9518	10593	10793	4580	7435	5974	8443	11004	3242	9584	5422	1828	4193	3438	84			193518	197040
		#6	SRR491026	29115	21166	16498	10715	13666	11854	11921	10623	11225	12683	14109	14339	7076	8564	8050	11474	16065	4477	14408	8118	3391	6085	5065	88			265622	270775
		#7	SRR491027	25169	17619	12734	8375	10980	10913	9602	8462	9152	8863	10316	11461	4795	7598	6397	8801	11764	3824	11544	6182	2368	4618	3519	90			211537	215146
		#8	SRR491028	17490	11833	8755	7010	7811	8053	6326	5952	6476	7726	8313	8662	3009	6020	4763	6349	8202	2496	8656	4872	1780	3193	2900	50			153747	156697
		#9	SRR491029	18800	12214	8782	6223	7768	7512	6867	6294	5314	6419	8562	8781	3024	5696	4570	6860	9014	2493	8902	3971	1635	4151	2715	61			153852	156628
		#10	SRR491030	31303	22870	15775	12208	14770	13260	11337	11712	11277	12449	13413	15619	7738	9082	8401	11032	14206	5504	15112	7861	3817	5843	5204	92			274589	279885
hESp0	emb. #3	#1	SRR491031	18317	14137	8770	5542	7069	7329	7556	5684	7252	7026	7377	9499	2679	4584	4984	8481	10328	2866	12602	5045	2127	5073	4058	41			164327	168426
		#2	SRR491032	25693	17836	13275	8957	11638	10814	9833	7636	11420	10906	12837	12335	4648	7734	7427	11606	13748	3651	16026	7040	2678	5801	5557	45			233539	239141
		#3	SRR491033	30914	22105	17175	10731	14246	15016	12862	11258	12558	13432	14196	15133	6835	10428	9295	12528	14549	4936	15387	8285	3218	6423	5452	42			281510	287004
		#4	SRR491034	21867	16254	12550	8078	9886	8706	9066	7841	8600	9335	10263	10759	3651	5667	6147	9207	11827	2877	13286	6870	2441	4585	4884	46			199763	204693
		#5	SRR491035	24599	16702	12857	8303	9691	9994	10688	8304	10946	9786	12352	12634	4192	7368	6877	13009	14951	3264	16371	7753	3002	6435	4985	45			230078	235108
		#6	SRR491036	17406	11967	8361	5976	7029	7078	7571	6185	6974	6754	8689	8296	2988	4413	4926	7517	9392	2423	10727	4592	1855	4847	3532	41			155966	159539
		#7	SRR491037	29404	21303	14654	9801	12728	12351	12072	10785	12052	11663	14446	13097	4825	9285	7990	13236	16409	4755	16392	8509	3383	7052	6054	45			266192	272291
		#8	SRR491038	19104	18331	14096	9017	11171	10829	11176	10285	11197	10565	13973	13245	4734	7697	7161	12420	14269	4153	16568	7579	2332	7142	5913	43			237044	243000
		*		23609	16965	12314	8594	10522	10102	9403	8495	9246	9766	10778	11489	4632	7216	6540	9131	11789	3600	11794	6294	2479	4919	3964	62			209677	213703
		#2	SRR445731	35643	27515	20750	15268	17377	18532	17632	14066	15866	14996	16584	18765	9210	11618	10827	15142	17896	5401	20275	9348	3417	7453	7940	107			343581	351628
hESp0	emb. #3	#4	SRR491039	30393	21931	17583	11322	15048	13388	13321	11185	12647	12676	14942	16010	6086	9324	8733	13390	16095	4576	17751	8735	2968	6438	6071	116			284542	290729
		#5	SRR491040	28510	22097	15317	10270	13563	13085	11688	10916	9948	11976	14014	15300	5890	9017	8234	10996	13310	4184	14893	7021	2866	5639	5413	108			258734	264255
		#6	SRR491041	28247	20285	15838	9766	13280	13435	12141	9391	11209	12086	12931	14456	6019	9340	8010	11948	13382	4029	15430	7978	2888	5669	5080	108			257758	262946
		#7	SRR491042	20388	15868	11845	7533	9665	10062	8559	7782	8222	8563	9760	10787	4199	6385	5303	8910	10696	3019	11362	6093	2216	4165	4189	83			191382	195654
		#8	SRR491043	29502	22398	15159	10906	13167	13568	12765	10861	11583	12632	13137	14490	5851	9118	8106	11670	14594	4262	14735	7463	2619	5739	6033	107			264325	270465
		#10	SRR491045	21328	15669	11914	7500	10549	9946	9272	7866	8623	9152	10084	11380	4371	6719	5930	8860	11044	3097	11103	6284	1883	4219	4190	81			196793	201064
		*		27716	20823	15487	10366	13236	13145	12197	10295	11157	11726	13065	14455	5947	8789	7878	11559	13860	4081	15078	7560	2694	5617	5559	101			256731	262392

Anexo 4 Número de variantes anotadas por amostra/por cromossomo_GSE44183																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
Estágio	Amostra	célula	ID	chr1	chr2	chr3	chr4	chr5	chr6	chr7	chr8	chr9	chr10	chr11	chr12	chr13	chr14	chr15	chr16	chr17	chr18	chr19	chr20	chr21	chr22	chrX	chrY	chrA	Total																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Oócito	DNA Paterno	#1	SRR689250	187976	131623	108479	77307	80347	86743	87729	68398	81819	75643	112640	95178	36586	56085	59264	79509	101146	27751	110834	51065	20642	40915	33936	262	1777679	1811877																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#3	SRR893046	7839	7855	5286	4327	4726	5033	5037	4570	3671	4353	4607	4565	2166	3202	3135	2925	3330	1556	3387	1905	1230	1339	1633	0	86044	87677																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#4	SRR893047	12866	11048	8745	6277	7412	6742	7159	6805	5666	6293	6401	6670	3560	4538	4571	4464	5025	2591	5117	2829	1555	2035	2712	0	128369	131081																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		*	SRR893048	7969	7046	5134	4112	4632	4705	4388	3424	3822	4264	4134	2048	2970	2767	2613	3206	1520	3166	1783	1029	1219	1685	0	80585	82270																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
Pronúcleo		#2	SRR893049	9558	8650	6388	4905	5590	5470	5634	5254	4254	4823	5091	5123	2591	3334	3854	1889	3890	2172	1271	1531	2010	0	98333	100343																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																
		#3	SRR893050	22383	18183	14560	9696	10615	10977	11887	10931	10298	10218	12579	11587	5143	8303	8383	11025	11901	3908	12480	6897	2270	4905	4373	6	229129	233508																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#4	SRR893051	23141	20007	15335	10878	12627	11573	12613	11158	10258	10627	12259	11988	6213	8861	8616	9390	10498	4656	10178	6162	2388	4146	4854	9	233572	238435																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		*	SRR893052	25936	21767	16522	11640	13585	12678	13902	12639	12192	12077	14551	13642	6239	9659	10151	12504	13154	4786	13714	7749	2907	5521	5319	8	267515	272842																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Zigoto		#1	SRR893053	23820	19986	15472	10738	12276	11743	12801	11576	10916	10574	13130	12406	5865	8941	9050	10973	11851	4450	12124	6936	2522	4857	4849	8	243405	248262																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#2	SRR893054	23424	20009	15683	10580	12559	11961	12801	11238	10880	10796	12513	12380	6429	8645	8696	10238	11831	4799	11992	6226	2291	4568	4546	8	239568	244122																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#5	SRR893055	13215	11207	8409	6213	7496	7110	6954	6676	5776	6426	7368	6943	3657	5038	5048	5409	5998	2530	6189	3533	1559	2276	2742	3	135030	137775																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		*	SRR893056	18320	15608	12046	8397	10028	9536	9672	8957	8328	8611	9941	9662	5043	6862	6872	7824	8615	3665	9091	4880	1925	3422	3644	6	187299	190949																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
2-células		#1	SRR893057	26661	23855	18441	12245	14395	13565	14518	12073	12057	12979	14401	14330	6782	11609	10732	11609	13119	5397	13274	7578	3006	4987	5457	6	276174	281630																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#3	SRR893058	13756	12861	10185	6599	8603	6891	8080	6613	6570	7147	7772	7052	3585	5477	5712	5961	6835	2759	6844	3762	1665	2862	2688	6	148591	151285																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#5	SRR893059	6252	5780	4560	3126	4027	3728	3594	3392	3263	3213	3944	3365	1651	2576	2732	2450	3055	1234	3142	1794	884	1318	3	67854	69175																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															
		*	SRR893060	17447	15588	12252	7791	9935	9394	9273	8533	8169	9091	9777	9619	3991	6779	6870	7815	8610	3338	9070	5186	2184	3700	3392	6	184412	187810																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
4-células		#1	SRR893057	26035	22920	17729	11452	13986	13007	14016	11282	11769	12203	14674	13519	6240	10484	10349	13051	14476	5118	12947	8010	3113	5563	5457	3	271943	277403																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#3a	SRR893061	24171	17132	13409	8703	11116	10478	10220	8653	9538	9690	11394	12731	4878	8648	6650	9755	11882	3599	14007	6592	2176	4727	4194	51	220149	224394																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#6a	SRR893062	22344	15042	12380	8082	10068	9141	9177	8338	8578	8665	10243	11266	4815	8065	6944	8685	11134	3318	13752	5329	1733	4693	3635	28	201792	205455																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#7a	SRR893063	26079	17308	13427	8690	11172	10714	10435	8618	10151	11020	11772	12661	5422	9616	7526	10671	12940	3799	15716	6993	2540	5605	4066	58	232875	236999																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
8-células	emb#b	#2b	SRR893064	9463	6168	4612	3327	3824	4279	4259	3396	3886	3517	5034	4410	1839	3015	3145	3712	4983	1442	7489	2530	861	2183	2052	17	87374	89443																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#3b	SRR893065	13075	8793	6377	4356	5456	5149	5010	4524	4676	5389	6257	6610	2116	4618	3785	4549	6197	1221	8147	3437	958	2425	1914	26	113125	115065																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#4b	SRR893066	15182	9538	7676	4938	6180	6984	6097	5113	6137	6057	8041	7268	3202	4725	4673	6489	7897	1781	11582	4675	1176	3403	2371	45	138814	141230																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#5b	SRR893067	15256	10452	7614	4953	6830	6086	6574	5344	6101	5915	8489	6711	2925	5353	4147	6947	7494	2086	10772	4104	1500	3352	2903	35	139005	141943																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
emb. #c	emb. #c	#3c	SRR893068	11318	6412	5350	3029	4490	4184	4491	3281	4104	3761	5635	5193	1645	3600	3593	4742	5760	1064	8002	2911	780	2558	1859	19	95903	97781																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#5c	SRR893069	5925	3699	2978	1851	2212	2027	2134	1753	2241	2238	2298	2965	941	2158	1372	2400	3353	529	4491	1615	277	1380	850	11	51537	52398																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		*	SRR893070	4041	2890	2215	1208	1740	1667	1567	1206	1899	1491	2188	2345	578	1486	1276	1822	2390	375	3621	1091	226	1123	1143	11	38445	39599																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		*	SRR893071	7584	4490	3377	2356	2345	3166	2703	2313	2894	2779	3680	3243	1516	2351	2038	2594	3426	701	3869	1766	722	1372	1060	13	61285	62358																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
Mórula	emb. integro	#1	SRR893071	14040	9266	7220	4681	5948	5807	5697	4776	5473	5502	6885	6855	2716	4876	4104	5670	7041	1810	9223	3731	1177	2984	2368	29	125482	127879																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#4	SRR893072	16258	11988	8483	5461	6877	6645	6405	6026	6776	6440	8375	7776	2901	5421	4603	6995	9506	2072	11004	5199	1475	4362	2701	30	151048	153779																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		#5	SRR893073	13101	10052	7028	5108	6293	5506	4859	4617	4704	5724	5413	5748	3024	4383	3424	3869	5283	1924	5618	3300	1112	1335	2162	37	111425	113624																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
		*	SRR893074	14497	11208	7983	6121	7339	6455	5564	5329	5413	6522	6421	7220	3485	5186	3863	4109	6089	2172	6699	3800	1413	2405	2606	43	129293	131942																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														

Anexo 5 | Desvios de alinhamento em favor do alelo de referência, GSE36552

<i>reads alinhadas</i>							
Estágio	Amostra	célula	ID	referência	alternativo	total	referência (%)
Oócito		#1	SRR445718	279514	260755	540269	51.73607962
		#2	SRR445719	263247	244753	508000	51.82027559
		#3	SRR445720	220775	205529	426304	51.78816056
Zigoto		#1	SRR445721	192167	173218	365385	52.59301832
		#2	SRR445722	150226	135059	285285	52.65821897
		#3	SRR445723	181888	167317	349205	52.08631033
2-células	emb.#1	#1	SRR445724	273867	266553	540420	50.6766959
		#2	SRR445725	285017	278791	563808	50.55213832
	emb.#2	#1	SRR445726	248743	237009	485752	51.20781798
		#2	SRR445727	278090	268739	546829	50.85502049
	emb.#3	#1	SRR445728	279154	247546	526700	53.00056958
		#2	SRR445729	334018	299055	633073	52.76137191
4-células	emb.#1	#1	SRR490961	304526	283127	587653	51.82071733
		#2	SRR490962	257839	240798	498637	51.70875807
		#4	SRR490964	169617	160040	329657	51.4525704
	emb.#2	#1	SRR490965	230571	206266	436837	52.7819301
		#2	SRR490966	243417	215164	458581	53.08048087
		#3	SRR490967	204515	186262	390777	52.33547522
	emb.#3	#4	SRR490968	282940	250734	533674	53.01738515
		#1	SRR490969	273754	246813	520567	52.58765923
		#2	SRR490970	287410	259069	546479	52.59305481
		#3	SRR490971	213092	188324	401416	53.08507882
		#4	SRR490972	311918	277052	588970	52.95991307
		#1	SRR490973	100449	94086	194535	51.63543835
8-células	emb.#1	#2	SRR490974	90193	84815	175008	51.53650119
		#3	SRR490975	64571	61059	125630	51.39775531
		#4	SRR490976	117628	111593	229221	51.31641516
		#1	SRR490977	88984	81257	170241	52.26942981
	emb.#2	#2	SRR490978	104693	99983	204676	51.150599
		#3	SRR490979	78257	72391	150648	51.94692263
		#4	SRR490980	120754	115557	236311	51.09961026
		#5	SRR490981	141604	121518	263122	53.81686062
		#6	SRR490982	133607	121320	254927	52.40990558
		#7	SRR490983	76341	69296	145637	52.41868481
		#8	SRR490984	90129	83518	173647	51.90357449
		#1	SRR490985	187111	172520	359631	52.0286071
	emb.#3	#2	SRR490986	158055	152434	310489	50.90518505
		#3	SRR490987	158680	143386	302066	52.53156595
		#4	SRR490988	240333	221479	461812	52.04130685
		#5	SRR490989	199245	190556	389801	51.11454306
		#6	SRR490990	201934	177767	379701	53.18237245
		#7	SRR490991	158580	148733	307313	51.6021125
		#8	SRR490992	207866	185961	393827	52.78104345
Mórula	emb. #1	#1	SRR490993	86355	78717	165072	52.31353591
		#2	SRR490994	62125	58407	120532	51.54232901
		#3	SRR490995	14069	10555	24624	57.13531514
		#4	SRR490996	94131	86455	180586	52.12530318
		#5	SRR490997	79133	75377	154510	51.21545531
		#6	SRR490998	75425	71591	147016	51.30393971
		#7	SRR490999	47126	43704	90830	51.88373885
		#8	SRR491000	22970	16766	39736	57.80652305

Anexo 5 | Desvios de alinhamento em favor do alelo de referência, GSE36552 (continuação)

<i>reads alinhadas</i>							
Estágio	Amostra	célula	ID	referência	alternativo	total	referência (%)
Mórula	emb. #2	#1	SRR491001	126357	119969	246326	51.29665565
		#3	SRR491003	88704	83355	172059	51.55440866
		#4	SRR491004	137408	127848	265256	51.80203275
		#5	SRR491005	103035	92959	195994	52.57048685
		#6	SRR491006	122623	113940	236563	51.83524051
		#7	SRR491007	87496	82396	169892	51.50095355
		#8	SRR491008	134013	120748	254761	52.60342046
Blastocisto	emb. #1	#1	SRR491009	103963	90745	194708	53.39431354
		#2	SRR491010	66950	62975	129925	51.52972869
		#3	SRR491011	84050	76877	160927	52.22865026
		#4	SRR491012	93418	81536	174954	53.39574974
		#5	SRR491013	108066	97635	205701	52.53547625
		#6	SRR491014	54932	52518	107450	51.12331317
		#7	SRR491015	94178	84444	178622	52.72474835
		#8	SRR491016	71240	66743	137983	51.62954857
		#9	SRR491017	144987	129324	274311	52.85497118
		#10	SRR491018	77911	69477	147388	52.86115559
		#11	SRR491019	61339	56414	117753	52.09124184
		#12	SRR491020	46109	41398	87507	52.69178466
	emb. #2	#1	SRR491021	53579	49233	102812	52.11356651
		#2	SRR491022	83827	69839	153666	54.55142972
		#3	SRR491023	48168	43465	91633	52.56621523
		#4	SRR491024	108954	95496	204450	53.29126926
		#5	SRR491025	91993	77901	169894	54.14729184
		#6	SRR491026	136906	119570	256476	53.37965346
		#7	SRR491027	78634	74189	152823	51.4542968
		#8	SRR491028	46041	40022	86063	53.49685695
		#9	SRR491029	90076	72966	163042	55.24711424
		#10	SRR491030	158857	139954	298811	53.16303617
	emb. #3	#1	SRR491031	53493	53261	106754	50.10866103
		#2	SRR491032	123665	114152	237817	52.00006728
		#3	SRR491033	138786	129206	267992	51.78736679
		#4	SRR491034	93688	82928	176616	53.04615663
		#5	SRR491035	97008	89514	186522	52.00887831
		#6	SRR491036	97234	87375	184609	52.67023818
		#7	SRR491037	158061	146476	304537	51.90206773
		#8	SRR491038	140636	132980	273616	51.39904099
hESCp0		#2	SRR445731	156220	144148	300368	52.00953497
		#4	SRR491039	111758	104929	216687	51.5757752
		#5	SRR491040	106225	98477	204702	51.89250716
		#6	SRR491041	96488	88509	184997	52.15652146
		#7	SRR491042	72668	66026	138694	52.39447993
		#8	SRR491043	129956	120451	250407	51.8979102
		#10	SRR491045	73867	68541	142408	51.86997921

Anexo 6| Verificação da presença de células aneuploides nos embriões em diversos estágios. Cada linha no gráfico representa uma célula individual. A partir dos dados de expressão (rpkm) foi calculada a mediana do nível de expressão dos genes de cada cromossomo separadamente e os valores plotados foram obtidos do cálculo da razão de cada um deles com a mediana da totalidade dos genes autossômicos.

Anexo 6| Verificação da presença de células aneuploides nos embriões em diversos estágios. Cada linha no gráfico representa uma célula individual. A partir dos dados de expressão (rpkm) foi calculada a mediana do nível de expressão dos genes de cada cromossomo separadamente e os valores plotados foram obtidos do cálculo da razão de cada um deles com a mediana da totalidade dos genes autossômicos.

Anexo 7 Sequência dos primers utilizados					
Primers para genotipagem por pirosequenciamento					
Gene	SNP	Primer F	Primer R	Primer de sequenciamento	Amplicon pb
MAOA	rs1800464	TGGAATCCCATTCGATATTG*	CTCCTTCCCATGTTATCTATTG	CCCATGTTATCTATTGTCC	66
	rs6323	GACTGCCAAGATTCACTTCAGA*	ATACATCATGCACCTTAATGACAGC	GCACTTAATGACAGCTCC	97
	rs1137070	TAAATGGTCTCGGGAAGTG	CTTTGATTCAGGTTCTTGACCC*	GGGAAGGTGACCGAG	59
	rs3027407	CAAAGCATGCCAATTTCTCTAA*	GGCCGACATTACTCTTAAGATAGA	GCCCTAATTTCTTTCTGT	208
	rs1894271	CACCCCATAGACCCCTTTATT*	CGATTACGCACATAAAATGTCCA	TTGTGGTCACTTACAA	90
XIST	rs1620574	GCACAAACATTTATTATCATTTATGAACC*	GCCCCACAGAAAAGTAATCACCAAT	GTATAGAACTGTAGGCTT	127
Primers para verificação de expressão alelo-específica por pirosequenciamento					
Gene	SNP	Primer F	Primer R	Primer de sequenciamento	Amplicon pb
MAOA	rs6323	GACTGCCAAGATTCACTTCAGA*	TGATCATGCAGCCACAGTAAT	TTAATGACAGCTCCCA	148 (DNA 993)
	rs1137070	ACAAAGTGGA GCGGCTACAT	CTTTGATTCAGGTTCTTGACCC*	GAAGGTGACCGAGAAA	123 (DNA 326)
NCBI dbSNP BUILD 138 (http://www.ncbi.nlm.nih.gov/snp)					
* primer biotilado					
Primers para sequenciamento tipo Sanger					
Gene	SNP	Primer F	Primer R	Amplicon pb	
MAOA	rs1800464	TCTGACACAGTAGGGTCTGG	TTTACCTCTTCCCATGTT	237	
	rs6323	GCAGAGAGAAACCAGTTAATTCAGCG	TCATTATGTGTGGCCAAAGGA	275	
	rs3027407	GTAACATTCTATTTCAACTTCAG	GGGGAAAAACAAGAGGTGGG	362	
	rs1137070	AAATGGTCTCGGGAAGGTGA	ACCGTGGCAGGAGCTTGAT		
	rs17462	TGGCTTAGCGTTCTGTTTCA	TGTGGGCTTCAGTTTGCTC	502	
MAOB	rs2072745				
USP9X	rs3027439	GACCATTTAGTCTCACCTTCTG	CGGGACACCATCGGTAACATA	198	
	rs10463	TGGTTCCATCAAGACTGTGC	ACCTTGGCATTACACATTAGC	222	
XIST	rs1620574	TCATTCTATCTGTATAGAACTGTAGGATT	AGCTATATCTGCTGAATGATCATTGATTAC	139	
	rs1894271	CAGCATGGTTGGTGACACCTAAG	TGGAATGAGCAGTGTGCGAT	241	
	rs16992436	GGCATTATTCTGACTTGGAGATG	CCTGCTTATCGTAGTGGCCAG	431	
	rs1794213	GATGGTTATGGAAAGGGTGACTG	CAAAATGTTAACTGCATGATTGCC	311	
	rs16992443	GTGACACAAGGCCAACGACC	GAATCAGGCTTAAAGATAAACAGGAG	369	
NCBI dbSNP BUILD 138 (http://www.ncbi.nlm.nih.gov/snp)					