Vrije Universiteit Brussel

FACULTEIT GENEESKUNDE EN FARMACIE
Research Group Reproduction and Genetics

Pluripotent stem cells as research models: the examples of trinucleotide repeat instability and X-chromosome inactivation

Doctoral dissertation submitted in the fulfilment of the requirements for the degree of Doctor of Philosophy in Medical Sciences

Anna Seriola Petit

Brussels, September 2015

Supervisor VUB
Professor Claudia Spits

Supervisor UAB:
Professor Josep Santalo

Members of the jury

Prof. Dr. Claudia Spits, supervisor
Research Group Reproduction and Genetics, Faculty of Medicine and Pharmacy,
Vrije Universiteit Brussel
Brussels (Belgium)

Prof. Dr. Josep Santalo, supervisor
Departament de Biologia Cel·lular, de Fisiologia i d'Immunologia
Universitat Autonoma de Barcelona
Barcelona (Spain)

Prof. Dr. Peter In't Veld, chairman
Diabetes Research Center
Vrije Universiteit Brussel
Brussels (Belgium)

Dr. Pau Sancho-Bru
Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS)
Barcelona (Spain)

Prof. Dr. Catherine Verfaille
Department of Development and Regeneration, Stem Cell Institute Leuven, Cluster Stem Cell Biology and Embryology, KU Leuven Medical School,
Leuven (Belgium)

Prof. Dr. Björn Heindryckx
Ghent Fertility and Stem cell Team (G-FaST), Department for Reproductive Medicine
Ghent University Hospital
Ghent (Belgium)

Prof. Dr. Frank Kooy
Cognitieve Genetica
Departement Biomedische Wetenschappen
Universiteit Antwerpen
Antwerpen (Belgium)

Prof. Dr. Susana Chuva de Sousa Lopes
Department of Anatomy and Embryology
Leiden University Medical Center
The Netherlands

Prof. Dr. Fanny Vidal
Departament de Biologia Cel·lular, de Fisiologia i d'Immunologia
Universitat Autonoma de Barcelona
Barcelona (Spain)

Table of contents

List of abbreviations

AMD	Age-related macular degeneration
ALS	Amyotrophic lateral sclerosis
ARX	Aristaless related homeobox gene
bFGF	Basic fibroblast groth factor (FGF2)
BAC	Bacterial artificial chromosome
BMP4	Bone morphogenetic protein 4
DM1	Dystrophia myotonica Type 1
DMPK	Dystrophia myotonica protein kinase
DMEM	Dulbecco's Modified Eagle Medium
DNA	Desoxyribonucleic acid
EB	Embryoid body
ECM	Extracellular Matrix
ERK	Extracellular signal-regulated kinases
ESC	Embryonic stem cells
EpiSC	Epiblast stem cells
FGFR3	fibroblast growth factor receptor 3
FISH	Fluorescent in situ hybridisation
FXS	Fragile X syndrome
FRAXA	Fragile X syndrome
FRDA	Friedreich Ataxia
GSK3β	Glycogen synthase kinase 3-beta
HD	Huntington's disease
hESC	Human embryonic stem cells
hiPSC	Human induced pluripotent stem cells
hPSC	Human pluripotent stem cells
HLA	Human leukocyte antigen
HFF	Human foreskin fibroblasts
HTT	Huntingtin gene
ICM	Inner cell mass
IVF	*In vitro* fertilization
JNK	c-Jun N-terminal kinase
KSR	Knockout serum replacement
LIF	Leukemia inhibitory factor
MEK	Mitogen-activated protein kinase
MEF	Mouse embryonic fibroblasts
mEpiSC	Mouse epiblast stem cells
mESC	Mouse embryonic stem cells

MFS	Marfan syndrome
MMR	Mismatch repair
MPS	Massive parallel sequencing
mRNA	Messenger ribonucleic acid
NEAA	Non-essential aminoacids
NHSM	Naïve human stem cell medium
NSC	Neural stem cells
NT-ESC	Nuclear transferred-embryonic stem cells
OPL	Osteoprogenitor like cells
OSKM	Oct3/4, Sox2, Klf4 and c-Myc
PBS	Phosphate buffered salin
PCR	Polymerase chain reaction
PGD	Preimplantation genetic diagnosis
PICMI	Post inner cell mass intermediate
PKC	Protein kinase C
PRC2	Polycomb repressive complex 2
p.f.	Post fertilization
PSC	Pluripotent stem cells
p38	Mitogen-activated protein kinase
RNA	Ribonucleic acid
ROCK	Rho-associated protein kinase
RPE	Retinal pigmented epithelium
RT-PCR	Real-time PCR
SCA1	Spinocerebellar ataxia 1
SCID	Severe combined immunodeficiency
SCNT	Somatic cell nuclear transfer
siRNA	Small interference ribonucleic acid
SNP	Single nucleotide polymorphism
SR	KnockOut Serum Replacement
SSC	Somatic stem cells
SSEA	Stage-specific embryonic antigens
STR	Short tandem repeats
TE	Trophectoderm
TGFβ	Transforming growth factor β
TRA	Tumor-rejection antigens
XCI	X chromosome inactivation
XIST	X inactive specific transcript
XaXa	Two X chromosome actives
XaXi	One X chromosome inactivated

Summary

Disease modelling is an essential tool for the understanding of human disease. Currently, much of the information we have on human diseases is based on animal models. However, animal models differ molecularly and phenotypically from humans, and are not always suitable to reproduce with fidelity human diseases. In the past decades, human pluripotent stem cells (hPSC) have emerged as an interesting option in the field of cellular modelling, this development recently having taken up much momentum. In this work, we aimed at characterizing hPSC as models for the study of Myotonic dystrophy type 1 (DM1) and Huntington's disease (HD) trinucleotide repeat (TNR) instability and to investigate the status of the X-chromosome inactivation with an eye on using these cells as models for early human development.

In the first part of our work, we observed a significant TNR instability for the DM1 locus in hESC, and that differentiation resulted in a stabilization of the repeat. This stabilization was concommitant with a downregulation of the mismatch repair (MMR). Our results were later replicated in hiPSC by other researchers, showing their reproducibility and suggesting they may be extrapolated to other hPSC lines worldwide. Regarding the HD repeat, we found it was very stable in all conditions studied, both in undifferentiated hESC and cells differentiated into osteogenic progenitor-like cells, teratoma cells and neural progenitors. This is in line with other studies showing that hESC show very limited TNR in the HD locus. On the other hand, some groups have now reported some instability of this locus in cells differentiated into the neuronal lineage. The instability seen in neuronal lineage in later studies, not in our study, is probably explained by the use of hPSC derived neurons more similar to the cells showing *in vivo* instability than the ones we were able to generate at the time of the study.

In the second part of the thesis we studied the X-chromosome inactivation in 23 female hPSC lines. We found that hPSC rapidly progress from a *XIST*-dependent XCI state to a culture-adapted, *XIST*-independent XCI state with loss of repressive histone modifications and erosion of methylation. We also report a remarkably high incidence of non-random XCI patterns, and that this skewing of the methylation patterns is independent from the transition to the *XIST*-independent XCI state, the origin of the X chromosome or chromosomal aberrations. These results suggest that XCI skewing is possibly driven by the activation or repression of a specific allele on the X chromosome, conferring a growth or survival advantage to the cells.

Overall, hPSC appear to be a good *in vitro* model for the study of both DM1 and HD

TNR instability, as the repeat follows *in vitro* the same patterns as found *in vivo*, including its dependency of the MMR machinery, particularly in the case of DM1. However, our results on the study of the X chromosome inactivation (XCI) state suggest caution when using hPSC as early human developmental research models. The eroded state of XCI found in many of the hPSC lines, and the frequency of skewed XCI patterns suggests that these cells are not a good proxy to early embryonic cells, at least what XCI is concerned. Conversely, they may still provide an interesting model to study gene function and mechanisms implicated.

Samenvatting

Ziektemodellen zijn een essentiëel instrument voor het verwerven van inzicht in ziekten in de mens. Het merendeel van de momenteel beschikbare informatie over deze ziekten is gebaseerd op diermodellen. Deze diermodellen verschillen echter zowel moleculair als fenotypisch van de mens en zijn daarom niet altijd geschikt om humane ziekten waarheidsgetrouw te reproduceren. In de laatste decennia hebben humane pluripotente stamcellen (hPSC) hun opmars gemaakt als een interessante optie in het gebied van cellulaire modelling; een ontwikkeling die recent veel momentum kreeg. Onze doelstelling in dit werk was om hPSC te karakteriseren als onderzoeksmodel voor de instabiliteit van de trinucleotide repeat (TNR) in Myotone dystrofie type 1 (DM1) en de ziekte van Huntington (HD) en om de status van X chromosoom inactivatie (XCI) te onderzoeken met een oog op het gebruik van deze cellen als model voor vroege humane ontwikkeling.

In het eerste deel van ons werk vonden we een significante instabiliteit van de TNR in de DM1 locus tijdens hESC cultuur; terwijl differentiatie resulteerde in de stabilisatie van de repeat. Deze stabilisatie vond gelijktijdig plaats met een verminderde transcriptie van de mismatch repair machinery (MMR). Onze resultaten werden later door andere onderzoekers bevestigd in hiPSC, wat hun reproduceerbaarheid aantoont en suggereert dat ze kunnen geëxtrapoleerd worden naar andere hPSC lijnen wereldwijd. De HD repeat daarentegen werd zeer stabiel bevonden in alle onderzochte condities, en zowel in ongedifferentiëerde hESC als in osteoprogenitor-gedifferentiëerde cellen, teratomas en neurale voorlopercellen. Deze resultaten zijn in overeenstemming met andere studies die een gelimiteerde TNR instabiliteit van hESC in de HD locus aantonen. Anderzijds rapporteerden recent enkele groepen dat een zekere instabiliteit van deze locus bestaat in cellen gedifferentieerd naar de neuronale lijn. Deze instabiliteit, die aangetoond werd in latere studies, niet in ons werk, kan waarschijnlijk verklaard worden door het gebruik van hPSC-afgeleide neuronen, meer gelijkend op de cellen die *in vivo* instabiliteit vertonen dan diegene die wij konden afleiden tijdens onze studie.

In het tweede deel van deze thesis onderzochten we X chromosoom inactivatie in 23 vrouwelijke hPSC lijnen. We vonden dat hPSC een snelle progressie vertonen van een *XIST*-afhankelijke naar een cultuur-geadapteerde, *XIST*-onafhankelijke XCI status met verlies van repressieve histone modificaties en erosie van DNA methylatie. We rapporteerden ook een opmerkelijk hoge incidentie van niet-random, asymmetrische XCI patronen en vonden dat deze scheeftrekking van de patronen niet gebonden is

aan de transitie naar een *XIST*-onafhankelijke XCI status, aan de oorsprong van het X chromosoom of aan een chromosomale afwijking. Deze resultaten suggereren dat de asymmetrie in het XCI patroon mogelijk gedreven wordt door de activatie of repressie van een specifiek allel op het X chromosoom, dat een groei- of overlevingsvoordeel toekent aan de cellen.

Globaal gezien lijken hPSC goede *in vitro* modellen voor de studie van zowel DM1 en HD TNR instabiliteit, aangezien de repeat hetzelfde patroon volgt als *in vivo* gevonden wordt, inclusief de afhankelijkheid van de MMR machinerie, in het bijzonder in het geval van DM1. Onze resultaten uit de studie naar XCI suggereren echter dat men behoedzaam moet zijn bij het gebruik van hPSC als onderzoeksmodel voor humane ontwikkeling. De geërodeerde XCI status die in vele hPSC lijnen aangetroffen werd en de frekwentie van de asymmetrische XCI patronen suggereren dat deze cellen geen goede proxy tot vroege embryonale cellen zijn, ten minste wat XCI betreft. Daarentegen kunnen ze nog steeds een interessant model bieden voor de studie naar de functie van de genen en mechanismen die hierbij betrokken zijn.

Resum

Els models de recerca són una eina bàsica per la comprensió de les malalties humanes. Actualment, la majoria de la informació de la que disposem de malalties humanes es basa en models animals. Tot i això, els models animals difereixen molecular i fenotípicament dels humans, i no sempre reprodueixen fidelment la malaltia humana. En les últimes dècades, les cèl·lules mare humanes s'han establert com una opció molt interessant en el camp de la modelització cel·lular. En aquest treball hem volgut caracteritzar les cèl·lules mare embrionàries com a models per a l'estudi de la inestabilitat de la repetició de trinucleotids a la distròfia miotònica tipus 1 (DM1) i la malaltia de Huntington (HD). Així mateix, hem volgut estudiar la inactivació del cromosoma X amb la intenció de fer servir línies cel·lulars com a models per l'estudi del desenvolupament embrionari humà.

A la primera part d'aquest treball, hem observant una inestabilitat de repeticions de trinucleotids significativa al locus de la malaltia DM1 de les cèl·lules mare estudiades. La diferenciació d'aquestes cèl·lules estabilitza el número de repeticions. L'estabilització de les repeticions es concomitant amb la repressió de l'expressió de gens involucrats en els mecanismes de reparació de ADN. Posteriorment a la publicació del nostre article, altres grups han reproduït els nostres resultats, però en aquest cas utilitzant cèl·lules mare induïdes. Els estudis recolzen la reproductibilitat dels nostres resultats, suggerint que poden ser extrapolats a altres línies de cèl·lules mare arreu del mon. Referent a la mutació de HD, vam trobar que era estable en totes les condicions estudiades, en cèl·lules indiferenciades, diferenciades a progenitors d'os, teratomes i progenitors neurals. Aquests resultats estan en concordança amb els resultats obtinguts per altres grups que descriuen un baix nombre de repeticions al locus de HD. Per altra banda, varis grups han descrit la presencia de inestabilitat de les repeticions en cèl·lules diferenciades a la línia neural. La discrepància entre els nostres resultats i aquests últims podria ser deguda a la obtenció de cèl·lules neurals menys madures en el moment del nostre estudi.

A la segona part d'aquesta tesi hem estudiat la inactivació del cromosoma X en 23 línies femenines de cèl·lules mare pluripotents. Vàrem observar una ràpida progressió de les cèl·lules de dependència de XIST en la inactivació del cromosoma X cap a un estat d'adaptació al cultiu que es caracteritza per un estadi de inactivació independent de l'expressió de XIST i amb una erosió de la metilació. També descrivim un patró d'inactivació esbiaixat en la majoria de les línies estudiades, contrari al patró aleatori observat en cèl·lules femenines adultes. A més a més, aquest patró és independent de XIST, de l'origen del cromosoma X i d'aberracions cromosòmiques. Aquests

resultats suggereixen que el patró esbiaixat observant esta dirigit provablement per l'activació o repressió d'al·lels específics que es troben en el cromosoma X i que li confereixen a la cèl·lula un avantatge respecte a les altres cèl·lules.

En conclusió, les cèl·lules mare pluripotents semblen ser un bon model in vitro per a l'estudi d'ambdues malalties, DM1 i HD, ja que presenten el mateix patró d'inestabilitat de la repetició del trinucleotid que s'observa in vivo. Cal remarcar també la dependència. D'altra banda, els resultats de l'estudi de la inactivació del cromosoma X suggereixen precaució amb l'ús de cèl·lules mare pluripotents com a models de desenvolupament embrionari humà. L'estat erosionat de la inactivació del cromosoma X, i la freqüència amb que el patró de metilació es esbiaixat demostren que les cèl·lules mare no son equivalents a cèl·lules embrionàries, almenys pel que fa a la inactivació del cromosoma X. D'altra banda, encara poden suposar un model interessant per estudiar funció gènica i els mecanismes associats a la inactivació del cromosoma X.

1 Aims and structure of this thesis

The general aim of this thesis was to evaluate the suitability of hESC as research models. We focused on two aspects. First, we aimed at investigating if hESC carrying Huntington's disease or Myotonic dystrophy type 1 mimic a number of the key molecular aspects of the trinucleotide instability that drive these diseases. Secondly, we aimed characterizing the patterns of XCI in hPSC with an eye in its impact on the use of these cells as models for X-linked disease, early human development, or as models for the research into molecular mechanisms driving the X-chromosome inactivation process.

This PhD dissertation has been structured in 9 chapters. Chapter 2 provides a general introduction to human pluripotent stem cells. Chapter 3 reviews the literature relevant to our first objective, taking into account the timing of that first part of the study. Chapter 4 presents the aims of the first work, followed, in chapter 5, by the results for the first part, as published in Seriola et al., Human Molecular Genetics, 2011, 20:176-85. Chapter 6 is an introduction specifically oriented to provide the background to the rationale and aims of the second study of this thesis. The aims of this second part are listed in chapter 7, and the study itself in chapter 8. In the general discussion (chapter 9), we put all our work in perspective to the most recent publications, and provide some future perspectives.

2 Human pluripotent stem cells

2.1 POTENCY AND DIFFERENTIATION

2.1.1 Early embryonic development

In mammals, two highly committed cells, the sperm cell and the oocyte, merge to give rise to a single uncommitted cell named the zygote. The zygote carries the paternal and maternal genomes, maternal mitochondria and a large stock of maternally provided messenger RNA. During development, the differential expression of genes will not only be responsible for the specific traits of the organism, but also regulate the commitment of cells to a certain type.

After fertilization, the unicellular zygote develops into a multicellular embryo by mitotic cell divisions. During the first days of development, the newly formed cells are called blastomeres. At day 2 after fertilization, the embryo consists of 4 cells and is at the 8-cell stage by day 3. During the first cell division, the embryo fully relies on the mRNA and proteins stored in the cytoplasm of the MII oocyte. The first waves of activation of the embryonic genome take place at 2-cell stage, both in mouse and human embryos (Flach et al. 1982; Wang et al. 2004). The embryonic genome activation is a sequential process that involves the activation of the transcription of key genes for cell division and correct development of the embryo and will progress until the complete activation of the whole embryonic genome (Rashid et al., 2010; Vassena et al., 2011). During the embryonic genome activation, there is a progressive degradation of the maternal mRNA and an initiation of the pluripotency program.

The next key process is compaction, during which blastomeres establish adherent junctions, resulting in a morula-stage embryo. During the next round of cleavages the blastomeres become polarized and generate two groups of cells: apolar inside cells and polar outside cells. By day 3.5 in mouse and day 5 in human, the embryo reaches the blastocyst stage. At this stage, the embryo undergoes the first identifiable lineage differentiation into two populations of cells: the inner blastomeres differentiate into inner cell mass (ICM) and the outer cells into trophectoderm (TE). The ICM is composed by a group of 20 to 30 cells that give rise to the embryo and its associated allantois, amnion and yolk sac. The TE cells contribute to the embryonic portion of the placenta and the chorion, however, do not contribute to embryonic structures. Right before implantation, the ICM cells diverge into two lineages, the epiblast and primitive

endoderm or hypoblast. The epiblast is a cluster of 10 to 20 cells that will separate into the embryonic epiblast layer that give rise to the whole embryo and the amniotic endoderm. The hipoblast cells form the extraembryonic endoderm, which forms the yolk sac.

2.1.2 Potency

The ability of a cell to give rise to different types of differentiated cells is known as potency. Totipotency is the ability of a cell to differentiate into any cell type of an organism, including embryonic (cells from the three germ layers: endoderm, mesoderm and ectoderm) and extra-embryonic tissues (placenta, corion and umbilical cord) and the capacity to generate a viable offspring (Edwards and Beard, 1997; McCracken et al., 2014). During the development of the mammalian embryo, the totipotency is gradually lost from totipotent to pluripotent, multipotent and unipotent (De Paepe et al., 2014; Hochedlinger and Plath, 2009; Eguizabal et al., 2013; Marchetto et al., 2010) (Figure 1).

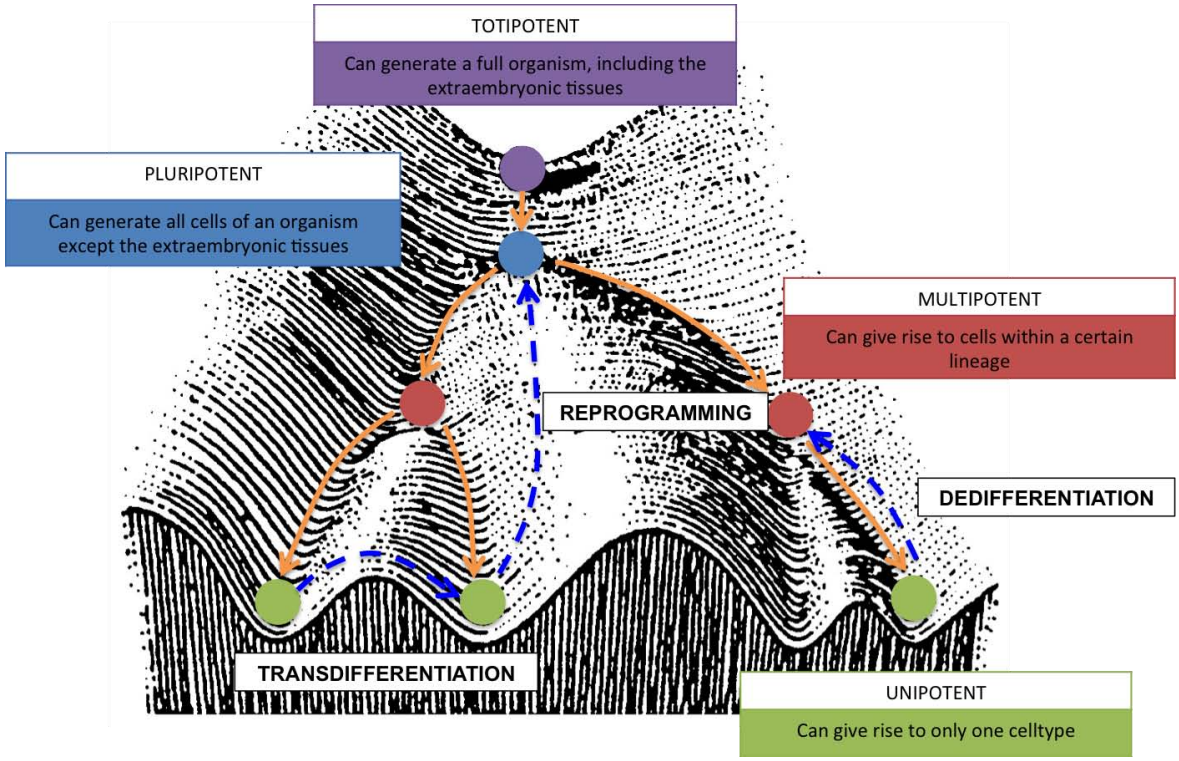


Figure 1. Developmental potential and processes of cells at different stages of development. The marbles rolling down the landscape into different valleys (cell fates) represent cell populations at different developmental stages. Reprogramming, transdifferentiation and dedifferentiation processes are indicated in blue dashed arrows. Colour marbles correspond to the differentiation states (purple, totipotent; blue, pluripotent; red, multipotent; green, unipotent). This figure is an adaptation of C. H. Waddington's epigenetic landscape model (1957).

During embryogenesis, in mouse embryos, zygotes and blastomeres from early pre-implantation embryos are considered totipotent. While in mouse, totipotency is easily shown by transferring embryos derived from single blastomeres into pseudopregnant mice, in humans, totipotency is difficult to prove due to both ethical and legal restrictions. However, the birth of a child from a blastocyst obtained *in vitro* from the culture of a single viable blastomere of a 4-cell stage embryo (Ben-Nun and Benvenisty, 2006; Veiga et al., 1987), prove the totipotency of human blastomeres at this developmental stage (Van de Velde et al., 2008; Yamashita et al., 2014). Furthermore, cells from both the ICM and TE from day-5 human blastocysts are not fully committed to their respective lineages. Isolated TE cells placed in ICM central position could change its developmental fate and become ICM cells (Bellin et al., 2012; De Paepe et al., 2013). However, at day 6 of embryo development, both cell types have lost their plasticity to change lineage direction (Cauffman et al., 2009).

Regarding stem cells, pluripotent stem cells have the ability to become any cell type of an organism, however they cannot give rise to trophoblast cells. An example of pluripotent stem cells are mouse embryonic stem cells derived from post-implantation mouse embryos, which are not able to generate extra-embryonic tissues. Human embryonic stem cells, on the other hand, are derived from pre-implantation embryos and have the ability to differentiate into any cell type from the three germ layers, germ cells, extra-embryonic tissues as well as trophoblast, making them totipotent. However, since it cannot be proven *in vivo*, they strictly should be considered to be pluripotent.

Adult stem cells or somatic stem cells (SSC) are overall multipotent or unipotent cells found in probably all adult tissues. Multipotent SSC are generally committed to differentiate to different cells of a certain lineage, although it has been shown that, in some cases and under well defined conditions, they are capable of switching between lineages (Eguizabal et al., 2013). Examples of multipotent SSC are haematopoetic stem cells and mesenchymal stem cells. While multipotent cells still have the capacity to give rise to different cell types, unipotent cells are not yet fully differentiated cells and can just differentiate into one cell type. An example of unipotent cells is spermatogonial stem cells.

SSC are mainly involved in self-renewal and regeneration of damaged tissues, and are frequently used in regenerative medicine. As from the late 50's the first trials of haematopoietic stem cells transplants took place. By early 70's bone marrow

transplantation took off with the first successful haematopoietic stem cells transplant, this currently being the most common medical use of stem cells. Other applications of SSC include stem cell-based skin grafts for third degree burns and mesenchymal stem cells have been used in clinical trials to treat osteogenesis imperfecta, myocardial infarction and multiple sclerosis (www.clinicaltrials.org). Another example is limbal stem cells, which are used to restore vision after corneal damage (Rama et al., 2010).

At the end of the spectrum of potency are the completely differentiated cells, such as, cardiomyocytes and skeletal muscle cells. Although these cells cannot give rise to other cell types, they may undergo a process of dedifferentiation *in vivo* or *in vitro* (see next section on cell commitment).

2.1.3 Cell commitment

A cell is considered to be committed when it loses its differentiation potential and develops a specific function in a tissue. Even though cell commitment was believed to be unchangeable, it is now known that under certain circumstances cells can undergo dedifferentiation. In this process, the cells lose commitment and go from a completely differentiated state to a more potent one, becoming more similar to its embryonic state. Dedifferentiation can happen both *in vivo* and *in vitro* and appears to be involved in the regeneration process of tissues (Eguizabal et al., 2013). Cell dedifferentiation in mammals has only been observed, *in vivo*, in mouse in Schwann cells. No cell dedifferentiation has been described *in vivo* in human cells. However, human cell dedifferentiation *in vitro* studies have been recently published involving human tyroid follicular cells (Suzuki et al., 2011), keratinocytes (Sun et al., 2011), adult human islet (Hanley et al., 2011) and fat cells (Shen et al., 2011), among others. Furthermore, committed cells can also switch between lineages in a process known as lineage conversion or transdifferentiation (Jopling et al., 2011; Eguizabal et al., 2013). Transdifferentiation takes place in situations when the regeneration of a lost part takes place in animals, for example the regeneration of the striatal muscle in the jellyfish (Schmid et al., 1988). In human, no evidence exist of transdifferentiation events *in vivo*, however, *in vitro*, human mesenchymal stem cells showed improved cardiomyogenic transdifferentiation efficiency when treated with angiotensin receptor blockers (Numasawa et al., 2011). Another example of human transdifferentiation *in vitro* is transdifferentiation of human dermal fibroblasts into multilineage blood derivatives lineage using ectopic expression of *OCT4* (Szabo et al., 2010).

A striking example of dedifferentiation is nuclear reprogramming, in which a cell is

dedifferentiated up to the toti/pluripotent state (Hochedlinger and Jaenisch, 2006). Cell nuclear reprogramming involves deep epigenetic changes, and can be achieved through somatic nuclear transfer (SCNT) (Tachibana et al., 2013), cell fusion (Cowan et al., 2005) and direct induction of reprogramming through the ectopic expression of certain embryonic transcription factors (Takahashi et al., 2007). This last point will be further elaborated in section 2.2.2 on human induced pluripotent stem cells.

2.2 HUMAN PLURIPOTENT STEM CELLS

Human pluripotent stem cells (hPSC) are uncommitted cells with the ability of unlimited self-renewal while retaining their capacity to differentiate into any cell type of a human body. At present, there are several approaches to generate hPSCs: from embryos at different stages (human embryonic stem cells (hESC)), from parthenogenetic embryos and from embryos obtained by somatic cell nuclear transfer (SCNT-ESC) and from somatic cells by reprogramming (human induced pluripotent stem cells (hiPSC)).

2.2.1 Human embryonic stem cells

The first ESC line was obtained from the inner cell mass of mouse blastocyst embryo (Evans and Kaufman, 1981). One decade later, the first hESCs line was established from the ICM from pre-implantation embryos (Thomson et al., 1998).

A. SOURCES AND TYPES OF EMBRYOS FOR THE DERIVATION OF hESC

Next to the most commonly used approach of deriving hESC from the ICM from a blastocyst-stage embryo, hESC can also be derived from single blastomeres from embryos at different developmental stages: from single blastomeres of 4-cell stage embryos (Flach et al., 1982; Geens et al., 2009; Vassena et al., 2011), from 8-cell stage embryos (Klimanskaya et al., 2006; Klimanskaya et al., 2007) and from early cleavage-stage arrested embryos (Lerou et al., 2008; Zhang et al., 2006) The rationale of these approaches is to avoid the destruction of the embryo and, at the same time, obtain a hESC cell line, or to increase the probability of obtaining an ESC line from a specific embryo. Also, ESCs have been obtained from mouse, primate and human activated non-fertilized oocytes. Parthenogenetic ESC lines have the advantage that they could reduce the genetic variability of hESC lines, making it easier to obtain an immune matching for regenerative medicine. Most recently, hESC have been obtained from embryos generated by SCNT of adult somatic cells into occytes (Tachibana et al., 2013).

B. DERIVATION OF hESC

Human ESCs are usually obtained from supernumerary embryos from IVF cycles. Donated embryos are kept in culture until blastocyst stage and the whole embryo or only the ICM is plated in hESC culture conditions. In order to improve the efficiency of hESC derivation, different derivation techniques, such as isolation of the ICM by immunosurgery, mechanical dissection or laser-mediated separation and elimination of TE cells have been used, as well as improved culture media Recent studies have described a transient structure, the post inner cell mass intermediate or PICMI, which appears right before the appearance of the first hESC colony, during the first 7 days after embryo plating. The PICMI is a transient epiblast-like structure that results from the progression of the ICM into an ESC state, with a distinct state of cell signalling. The PICMI is both necessary and sufficient for the derivation of an hESC line (O'Leary et al., 2012).

C. NAÏVE AND PRIMED STEM CELLS

Culture requirements and developmental potential for mouse ESC (mESC) and hESC are different. While mESCs are dependent on LIF and BMP4 growth factors to maintain its pluripotency and grow in culture as thick and tight colonies; hESC are bFGF and activin/TGFβ dependent and grow as flat 2D colonies (Figure 2).

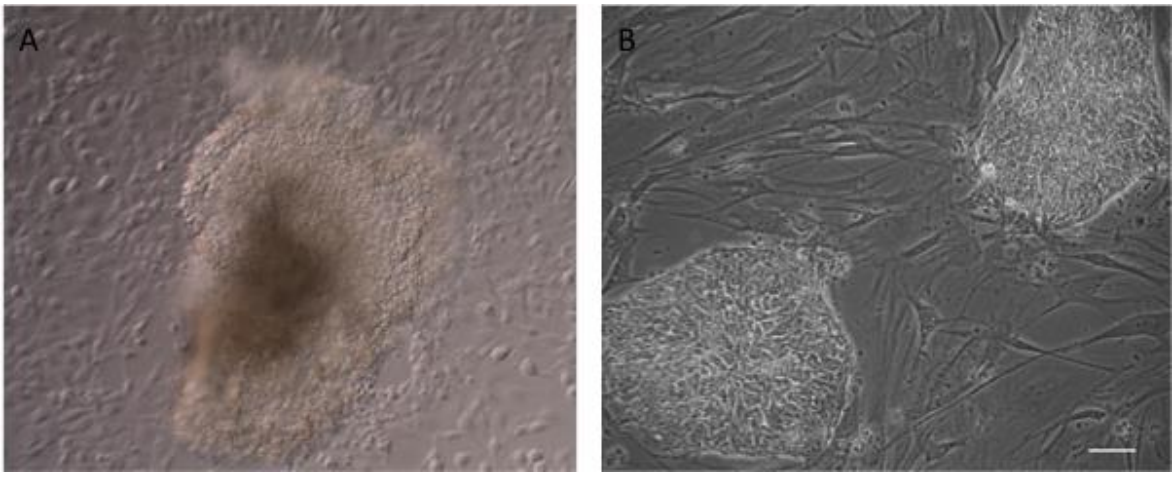


Figure 2. Morphological differences between mouse ESC and human ESC. A: mESC grown on mouse embryonic fibroblast (MEF) feeders. B: hESC grown on human foreskin fibroblasts (HFF).

At first, it was believed that the differences were interspecies, however, the derivation of mouse pluripotent stem cells from epiblast cells from post-implantation embryos, the mouse epiblast stem cells (mEpiSC), led to the concept of two distinct pluripotent states of ESC, namely "naïve" and "primed" (Brons et al., 2007; Tesar et al., 2007). Therefore, mouse pluripotent stem cells exist in two states, the naïve mESC and the

primed state represented by the post-implantation epiblast derived mEpiSC.

Naïve mESC are derived from the epiblast cells of the peri-implantational blastocyst and express pluripotency factors Oct4, Sox2, Nanog, Klf2 and Klf4 and are very clonogenic. Regarding epigenetic status, female mESC show two active X chromosomes (XaXa) (Chambers et al., 2003; Edwards and Beard, 1997; Eguizabal et al., 2013; Hochedlinger and Plath, 2009; Nichols and Smith, 2009; van den Berg et al., 2009; Veiga et al., 1987). Conversely, primed mEpiSC are derived from epiblast cells from post-implantation mouse embryos, express core pluripotency factors Oct4, Sox2 and Nanog, but differ from mESC in the expression of other early transcripts as Klf4. EpiSC cells do not survive efficiently when grown as single cells, its morphology is more 2D and epithelial and are dependent on bFGF and activin (mEpiSC are derived under hESC culture conditions) (Cauffman et al., 2009; Van de Velde et al., 2008). EpiSC present already one X chromosome inactivated (XaXi) as seen in further steps of the development. In conclusion, mEpiSC show morphology, molecular and epigenetic characteristics more similar to hESC than to mESC (Nichols and Smith, 2009). An overview of the differential traits of naïve and primed stem cells is shown in Table 1.

Table 1. Naïve and primed characteristic features.

Property	**Naïve**	**Primed**
Embryonic tissue of origin	Early epiblast	Egg cylinder or embryonic disk
Blastocyst chimeras	Yes	No
Teratomas	Yes	Yes
Pluripotency markers	Oct4, Nanog, Sox2, Klf2, Klf4	Oct4, Sox2, Nanog
Response to LiF	Self-renewal	Differentiation
Response to FGF/ERK	Differentiation	Self-renewal
Response to 2i	Self-renewal	Differentiation
Clonogenicity	High	Low
X chromosome status	XaXa	XaXi

Much work has been devoted to the generation of human naïve ESC. They have been derived using different approaches, such as the use of a cocktail of inhibitors of the

GSK3β and MEK/ERK signalling pathways, commonly known as 2i culture, associated with self-renewal and pluripotency maintenance (Ying et al., 2008); derivation under hypoxic conditions (Lengner et al., 2010), and the combination of overexpression of transgenic factors and specific culture media (Eguizabal et al., 2013). Even though, the naïve state in hESC is difficult to maintain for long-term cultures. Great advanced was made with the definition of a culture conditions that allow the generation of naïve stem cells from primed pluripotent stem cells, stem cells generated from somatic cells or directly from blastocysts. The culture medium used was termed NHSM and contained LIF, TGFβ1, bFGF2, ERK1/2 inhibitor GSK3β inhibitor, JNK inhibitor, p38 inhibitor, ROCK inhibitor and PKC inhibitor. The obtained cells present molecular and functional characteristics similar to those of mouse ESC (Gafni et al., 2013). Recently, Ware et al. 2014 reported the generation of non transgenic naïve hESC line and its unlimited culture in 2i media (Ware et al., 2014). Figure 3 shows a schematic overview of approaches to obtain human and mouse, primed and naïve stem cells.

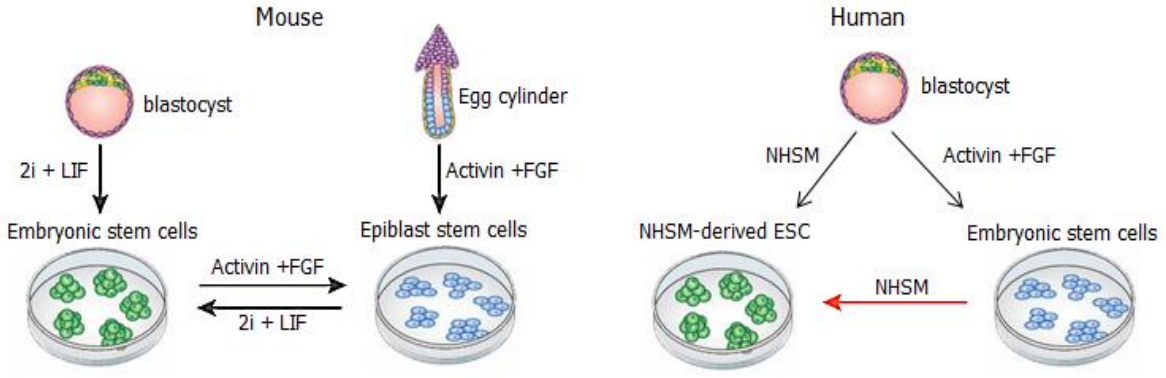


Figure 3. Graphical overview of mouse and human naïve and primed PSC. In mouse, naïve and primed PSC are derived from the ICM of pre-implantation embryos and epithelial epiblast stage embryos, respectively. In human, naïve PSC were obtained by ectopic over-expression of transcription factors POU5F1, KLF4, KLF2 or by culturing conventional hESC in NHSM. Upon certain changes in culture conditions, these PSC can be reversibly converted from one state to the other. Image adapted from (Mascetti and Pedersen, 2014).

2.2.2 Human induced pluripotent stem cells

In 2006, Takahashi and Yamanaka described the reprogramming of an adult somatic cell type, mouse fibroblasts, into a pluripotent state without the use of SCNT. The reprogrammed cells received the name of induced pluripotent stem cells (iPSC) (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). The reprogramming was carried out by the ectopic overexpression of 4 genes; Oct4 (also known as POU5F1),

Sox2, Klf4 and c-myc, involved in pluripotency, self-renewal and cell cycle-control. The ectopic expression of these pluripotency transcription factors activates a cascade of transcription that will lead to the reactivation and sustained expression of the endogenous pluripotency regulating factors (Takahashi et al., 2007).

At first, the technique was inefficient and the miPSC showed lower expression of endogenous pluripotent genes than the mESC and were considered as "partially reprogrammed". Furthermore, the viral transgenes were not silenced; some of the pluripotent genes were still active and miPSC cells gave rise to teratomas when injected into severe combined immunodeficiency (SCID) mice but not adult chimeras (Okita et al., 2007). Some modifications were included in the reprogramming protocol in order to increase miPSC generation efficiency and, in 2007, the first human iPSC (hiPSC) lines were generated (Park et al., 2007; Takahashi et al., 2007; Yu et al., 2007).

hiPSC are morphologically and molecularly very similar to hESCs (González et al., 2011). Despite their resemblance, unlike hESC, hiPSC do not require the use of embryos for their derivation, and can be patient-specific. This minimizes the issue of immune rejection in regenerative medicine, as well as providing the opportunity to generate hPSC lines to study patient specific diseases (Hanley et al., 2011).

A. GENERATION METHODS AND CELL TYPES

The first combination of transcription factors used for reprogramming, Oct4, Sox2, Klf4 and c-myc, is often referred to as the "Yamanaka cocktail" and it is usually abbreviated as OSKM. Since then, there has been an exponential evolution in delivery methods, combinations of factors and alternatives to transgenic overexpression. Several methods have been used for the delivery of the factors needed to reprogram the somatic cells. They can be classified in: integrative or non-integrative viral systems, non-viral systems and DNA-free systems.

Constitutive integrative viral systems use retroviral vectors and lentiviral vectors. Retroviral vectors integrate in the genome of dividing cells, while lentiviral vectors integrate into both dividing and non-dividing cells. In this system, exogenous transgenes become usually silenced after the reprogramming (Hotta and Ellis, 2008). Reprogrammed cells will then endogenously express the core pluripotency factors. The efficiencies obtained with integrative techniques are 0.1% in miPSC generation and 0.01% in human. The integrative system has the risk of random integration of the transgene into an oncogene or tumour suppressor genes and alter its expression.

Non-integrative systems solve the risk of random and/or permanent integration of the transgene. There are 4 categories of non-integrative systems: integration defective viral systems, episomal, RNA and DNA-free systems. Examples of integration defective viral systems are the adenoviral virus and the Sendai virus. DNA-free systems include modified proteins (Zhou et al., 2009) and cell extracts from genetically modified cell lines (Kim et al., 2009). But DNA-free systems are very inefficient techniques and require multiple transfections (Stadtfeld et al., 2008).

Another point of attention has been the regulation of the expression of the transgene. One of these systems is the doxycycline-inducible. This is a drug dependent system by which, doxycycline is included in the culture media during the reprogramming process and later on is removed. Without the administration of doxycycline to the media, the partially reprogrammed cells will not survive. This will work as a selection system for cells that do not endogenously express the transcription factors transferred. Also, strategies have been devised to remove the transgenes from host cell. For this, excisable systems have been designed. Two examples of excisable systems are transposon elements such as the piggyback, which can be removed by transient expression of transposase, and systems with a loxP sequence that can be excised from the genome by the addition of transient Cre-recombinase (Kaji et al., 2009; Soldner et al., 2009).

Since first generation of hiPSC from fibroblasts, hiPSC have been generated from different cell types as keratinocytes (Raya et al., 2010; Aasen et al., 2008; Maherali et al., 2008), neural cells, adipose cells and cord blood (Giorgetti et al., 2010) amongst others, reviewed in (Gonzalez et al., 2011). The primary tissue from which hiPSC are generated has an effect on the requirements for reprogramming factors: fibroblast are reprogrammed without the need of *C-MYC* overexpression (Nakagawa et al., 2007) and hepatocytes can be reprogrammed with or without *C-MYC* (Aoi et al 2008), cord blood cells can be reprogrammed with *OCT4* and *SOX2* (Giorgetti et al., 2009), and neural stem cells (NSC) by just one factor, *OCT4* (Kim et al., 2009). Why one cell type requires more or less factors to get reprogrammed it is still unclear; however the probable cause is due to the epigenetic characteristics of each cell.

2.2.3 Human pluripotent stem cell culture and characterization

A. hPSC CULTURE

hPSC cells can be grown efficiently *in vitro* in a system that includes a supporting

matrix, a specific growth medium including essential growth factors and physiological conditions (typically 37ºC, 5% CO_2 and atmospheric oxygen tension). For long-term culture, hPSC cells need to be either mechanically or enzymatically dissociated (eg. Dispase, Trypsine or non-enzymatic EDTA solution) and plated on new culture plates for expansion, this is known as *passage*. The plates are covered with either feeder layers or matrices that provide adherent support and secrete growth factors and cytokines that further enable hPSC proliferation. Mitotically inactivated mouse embryonic fibroblasts (MEF) or human foreskin fibroblasts (HFF) are typically used as feeder cells.

Since feeder dependent cultures are not well defined and can be a source of animal contaminants to the hPSC, as feeder-free cell culture support other extracellular components are used, such as Matrigel, a semichemicallly defined xenogeneic substrate and extracellular matrix (ECM) proteins as fibronectin, vitronectin, laminin-521 and E-cadherine, amongst others (Villa-Diaz et al., 2012)(Chen et al., 2014).

Depending on the support used, the cells grow as colonies or compact aggregates of hundreds of cells with defined borders at the periphery with high nucleus to cytoplasm ratio and prominent nuclei when grown on MEFs and HFF. However, when grown on laminin, hPSC grow as confluent monolayers. When differentiation appears in the culture, the cells need to be passaged. Each passage can vary between 5 to 7 days depending on the cell line.

In addition to the cell support, hPSC require specific growth media. The conventional growth media used to grow hPSC on feeder cells is based on KO-DMEM as basal component that is supplemented with non-essential aminoacids (NEAA), KO Serum replacement (KOSR, defined serum-free formulation), 2-β-mercaptoethanol, Glutamine and bFGF. Due to the variability of the media and risk of transmission of animal contaminants to the stem cell culture, growth media has evolved to chemically defined and xenogenic-free culture media, being the most commonly used mTeSR™ (StemCell Technologies) and E8™ (LifeTechnologies).

B. hPSC CHARACTERIZATION

It is generally accepted that for a cell line to be considered pluripotent it has to be characterized and to comply with specific morphological and molecular attributes. The main pluripotency tests carried out include: morphological tests, assessment of alkaline phosphatase expression and core pluripotency markers expression, silencing of the transgenes used for reprogramming when viral vectors have been used, as well as karyotype and genome stability analysis (Martí et al., 2013).

Additionally to the routinely morphological analysis, molecular phenotype and differentiation potential are the main characteristic that all hPSC share. All pluripotent stem cells express cell surface markers and transcription factors including OCT4 (or also known as POU5F1), SOX2, NANOG, stage-specific embryonic antigen-4 (SSEA-4), SSEA-3 and do not express differentiation markers as SSEA-1.

One of the most commonly used functional tests for the evaluation of the differentiation potential is the formation of embryoid bodies (EBs) and posterior evaluation of their capacity to contribute to the 3 germ layers (ectoderm, mesoderm and endoderm) both *in vitro* and *in vivo*. For the *in vitro* assay, hPSC cells are placed in non adhere culture plates to assess their differentiation potential by direct or sporadic differentiation into embryoid bodies (EBs). The *in vivo* assay is performed by the evaluation of hPSC capacity to form teratomas when injected (~10.000 cells) into SCID mice. SCID mice do not produce T nor B cells so do not have immunosuppression when human cells are transplanted.

Genetic stability is also an important test to take into account since it has been made evident that, during long term *in vitro* culture, hPSC suffer cell culture adaptation and can acquire chromosomal abnormalities, which eventually take over the culture based on selective advantage (Avery et al., 2013; Lund et al., 2012; Nguyen et al., 2013b). The cultures can be routinely screened using, amongst others, G-banding, FISH or DNA microarrays.

C. STEM CELL REGISTRIES

Since the derivation of the first hPSC lines, the stem cell field has found its way into laboratories worldwide and these cells have become the basis of a very active research domain with more than 1.5 million publications in 2012. To help researchers identifying potential lines to use in their research, different stem cell registries have been set up worldwide. In Europe, for instance, there is the European Human Pluripotent Stem Cell Registry (hPSCreg) (Kurtz et al. 2014). This is an initiative funded by the European Commission. The hPSCreg supports transparency and traceability of ethical and regulatory compliance of the hPSC generated. The registry collects information on the cell line characteristics such as donor phenotype, derivation process, genotype and pluripotency profile. Currently, hPSCreg has registered 672 hESC and 70 hiPSC lines that comply with the informed consent and follow the corresponding regulatory compliance to be eligible for their use in EU-funded projects. The cell lines included come from more than 84 laboratories of 24 countries (http://www.hescreg.eu).

2.3 STEM CELL APPLICATIONS

Given their unique characteristics of unlimited self-renewal and differentiation capacity, pluripotent stem cells hold great promise for regenerative medicine, cell replacement therapies and as research models for the study of human disease.

2.3.1 Regenerative medicine and clinical trials

Regenerative medicine aims at replacing or regenerating human cells, tissues or organs to restore its normal function. For this, access to a sufficient quantity of specific cell types is required. Since the derivation of the first cell lines, hPSC have been considered prime candidates to providing these cells.

Until now, hPSC have been successfully differentiated into many cell types of a human body, as for example, cardiomyocytes (Shiba et al., 2012), insulin-producing β cells (Kroon et al., 2008) and neurons (Kriks et al., 2011). Tissues and organs with a more complex organization are more difficult to generate, although some promising progress has been made by the *in vitro* formation of human mini-brains (Lancaster et al., 2013; Messer et al., 1999; Monckton et al., 2001; Savouret, 2003) and renal tissues with a kidney-like organization (Owen et al., 2005; Takasato et al., 2014).

Neural cell types were the first lineages to be obtained through directed differentiation of hPSC, for instance, dopamine neurons - main cell type to be replaced in Parkinson disease, striatal neurons – main cell type affected in Huntington's disease and glial cells - the cells used in the first hESC-based therapy to reach clinical trials, with the aim of treating spinal cord injuries.

Another example of successful directed differentiation is found in hPSC cell–based therapies for cardiac repair. Cardiomyocites have been successfully obtained through monolayer-based differentiation with BMP4 and Activin A as main supplements and by embryoid bodies. Both strategies have provided successful pre-clinical results. One million hESC derived cardiomyocites engrafted in myocardial ischemia primate model presented extensive remuscularization of the injured heart and functional integration of the cell graft (Chong et al., 2014). The group also concluded the possibility of clinical scale cell production as well as good viability of the cells after cryopreservation.

Pluripotent stem cells clinical translation is moving slowly but onwards. Prove of it is

the increasing number of clinical trials using hPSC around the globe. Table 2 shows an overview of the on-going clinical trials worldwide.

Table 2. Ongoing clinical trials of human embryonic stem cells-derived cell transplantation.

Trial number[1]	Start	Sponsor	Title	Condition	Remarks
NCT01469832	2011	Advanced Cell Tecnhology	Safety and Tolerability of Sub-retinal Transplantation of hESC Derived Retinal Pigmented Epithelial Cells in Patients With Stargardt's Macular Dystrophy	Stargardt's Macular Dystrophy; Fundus Flavimaculatu; Juvenile Macular Dystrophy	Phase I/II, Open-Label, Multi-Center, Prospective Study
NCT01344993	2011	Advanced Cell Technology	Safety and Tolerability of Sub-retinal Transplantation of hESC Derived RPE Cells in Patients With Advanced Dry Age Related Macular Degeneration	Dry Age Related Macular Degeneration	Phase I/II, Open-Label, Multi-Center, Prospective Study
NCT01345006	2011	Advanced Cell Technology	Sub-retinal Transplantation of hESC Derived RPE Cells in Patients With Stargardt's Macular Dystrophy	Stargardt's Macular Dystrophy	Phase I/II, Open-Label, Multi-Center, Prospective Study
NCT01674829	2012	CHA Bio & Diostech	Determine the Safety and Tolerability of Sub-retinal Transplantation of hESC Derived Retinal Pigmented Epithelial Cells in Patients With Advanced Dry Age-related Macular Degeneration	Dry Age Related Macular Degeneration	Phase I/IIa, Open-Label, Single-Center, Prospective Study
NCT01691261	2012	Pfizer	Implantation Of hESC Derived Retinal Pigment Epithelium In Subjects With Acute Wet Age Related Macular Degeneration And Recent Rapid Vision Decline	Age Related Macular Degeneration	Phase I, Open-label, Safety And Feasibility
NCT02057900	2013	Assistance Publique - Hôpitaux de Paris	Transplantation of hESC-derived Progenitors in Severe Heart Failure (ESCORT)	Ischemic heart disease	Phase I, Open Label, feasibility and safety study

NCT02122159	2014	University of California, Los Angeles	Research With Retinal Cells Derived From Stem Cells for Myopic Macular Degeneration	Myopic Macular Degeneration	Phase I/II, Open-Label, Prospective Study to Determine the Safety and Tolerability of Sub-retinal Transplantation
NCT02445612	2015	Ocata Therapeutics	Long Term Follow Up of Sub-retinal Transplantation of hESC Derived RPE Cells in Stargardt Macular Dystrophy Patients	Stargardt's Macular Dystrophy	Phase I/II, Open-Label, Multi-Center, Prospective Study to Determine the Safety and Tolerability of Sub-retinal Transplantation
NCT02463344	2015	Ocata Therapeutics	Long Term Follow Up of Sub-retinal Transplantation of hESC Derived RPE Cells in Patients With AMD	Age-related Macular Degeneration	Phase I/II, Open-Label, Multi-Center, Prospective Study to Determine the Safety and Tolerability of Sub-retinal Transplantation
NCT02286089	2014	Cell Cure Neurosciences Ltd.	Safety and Efficacy Study of OpRegen for Treatment of Advanced Dry-Form Age-Related Macular Degeneration	Age-related Macular Degeneration	Phase I/IIa Dose Escalation Safety and Efficacy Study

[1] www.ClinicalTrial.gov

The first clinical trial using hESC-derived cells involved oligodendrocyte progenitor cells to treat spinal cord injury. It was initiated in 2010 and halted one year later for lack of funding, and only recently restarted (ClinicalTrials.gov study: NCT01217008).

HPSC-based treatment of retinal degenerative diseases such as acute age-related macular degeneration (AMD), retinitis pigmentosa (RP) and Stargardt's macular dystrophy are now leading in the charts of hPSC-based clinical trials. These diseases are characterized by a loss of retinal-pigmented epithelium (RPE) cells, which leads to the death of the photoreceptors causing loss of vision. The first successfully completed phase I clinical trial with hESC-derived cells concluded on September 2014 in England with the transplantation of hESC derived RPE sheets into 18 patients in total, nine affected with age related macular degeneration and nine with Stargardt's macular dystrophy. The clinical trial did not only show the feasibility and safety of the procedure but revealed an improvement in visual acuity of at least 15 letters in eight of 18 patients (doubling of the visual angle) during the first year after surgery, with no adverse effects seen in any of them (Schwartz et al., 2015). The studies are

registered at ClinicalTrials.gov with numbers NCT01345006 for Stargardt's macular dystrophy and NCT01344993 for age-related macular degeneration.

Recently, Philippe Menasche, head of the Heart Failure Surgery Unit of the Hôpital Européen Georges Pompidou and director of INSERM (France's National Institute of Health and Medical Research), informed about the first worldwide clinical trial that will transplant cardiac progenitor cells derived from hESC to six patients. These patients will receive purified hESC derived cardiac progenitors expressing CD15+ Isl-1+ markers, which will be included in a biocompatible fibrin gel patch and transplanted into the infarcted area of the heart. The cells will be transplanted when the patients undergo scheduled coronary artery bypass surgery or mitral valve procedures. Results are expected by June 2016 (ClinicalTrials.gov study: NCT02057900).

Concerning hiPSC, the world's first clinical trial using derivates of these cells was announced in September 2014, by the Laboratory for Retinal Regeneration, RIKEN (Japan). This clinical trial will evaluate the safety and feasibility of using autologous hiPSC to treat blind age-related macular degeneration (AMD) patients.

2.3.2 Research models

Another interesting use of hPSC is as research models (Figure 4). hPSC have been proposed as models to study, amongst others, human genetic disorders, early human development, linneage commitment and pluripotency (Braam et al., 2009; Brennand et al., 2011; Foiry et al., 2006; Messer et al., 1999; Rubin, 2008; Savouret et al., 2004; Wheeler, 1999).

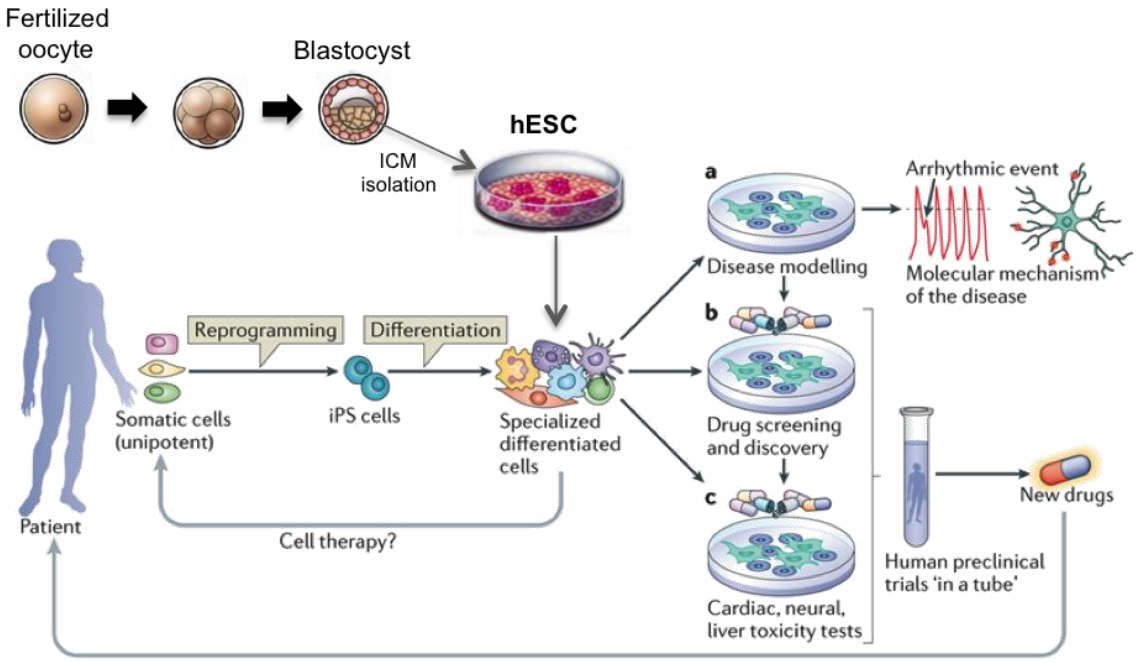


Figure 4. Applications for human pluripotent stem cells. hPSCs can be used for disease modelling **a)** or for drug development or toxicity tests **(b,c)**. The result of these applications can be directly coupled back to the patient. This image is adapted from (Bellin et al., 2012).

Thanks to the capacity of hPSC to differentiate into different cell types, hPSC can be used to modelling diseases as neurodegenerative disorders, heart diseases, vision degenerative diseases, diabetes, amyothropic lateral sclerosis (ALS) amongst others (Sterneckert et al., 2014). In the case of hESC, the work is often focused on known monogenic diseases or diseases with a suspected genetic component.

HPSC will not only fit as better models compared to animal models but also help to reduce the number of animals used in these studies. For many monogenic diseases, the research animal models do not resemble or show only an approximate resemblance to the human disease, being difficult to create a model with the right behavioral and psychological conditions.

To obtain hPSC lines with a specific genetic background, two approaches are used. First, hESC lines can be derived from embryos carrying a monogenic disease. The VUB stem cell bank, for example, currently holds more than 25 hESC cell lines carrying monogenic diseases, including Huntington's disease, Myotonic dystrophy, Marfan syndrome, Spinocerebral ataxia, Fragile X syndrome, Osteogenesis imperfecta and Cystric fibrosis. The second possibility is to generate hiPSC from cells of patients with the disease of interest. This approach have been used, for instance, to study Spinal muscular atrophy (Ebert et al., 2008), Schizophrenia (Brennand et al., 2011),

inherited liver diseases (Rashid et al., 2010), and gastric diseases (McCracken et al., 2014). The use of hPSC specifically in the context of the study of dynamic mutations will be discussed in point 3.3 of this thesis.

Because of the fact that hESC are derivatives of the human inner cell mass, they have been considered for the study of early human embryonic development. Currently, this research is almost exclusively limited to the pre-implantation phase, due to obvious ethical and practical concerns, and even then is severely limited by the scarcity of the material. Given the substantial differences during embryogenesis between different animal species (for example: the gestation period in human is about 14 times longer than in mouse), extrapolation from animal models to the human is limitted. hESC can provide interesting information on early embryogenesis, the roles of different transcription factors in pluripotent or more committed stem cells, and the relationship of epigenetics and cellular identity in a controlled environment (De Paepe et al., 2014).

Finally, they can also be used in cytotoxicity testing and drug screening, and are becoming increasingly used in pharmaceutical companies in this context (Marchetto et al., 2010)(Ben-Nun and Benvenisty, 2006). For instance, Yamashita and colleagues discovered that Statin could be used to rescue skeletal dysplasia caused by gain-of-function mutation of fibroblast growth factor receptor 3 (FGFR3) using patient-specific hiPSC lines (Yamashita et al., 2014). In a similar line, hPSC can provide a stepping stone towards personalized medicine. Because hiPSC can be used to test drug toxicity and efficacy, patient-specific hiPSC could be used to establish the suitability of a treatment for individual patient, and provide tailored medication. However, given the actual technologies and the cost of the generation of patient-specific hiPSC, it is still prohibitive to afford a personalized treatment approach.

2.4 CHALLENGES AND BOTTLENECKS IN STEM CELL RESEARCH AND CLINICAL APPLICATION

Although high expectations surround the application of hPSC, there are limitations that are of critical importance for its use both for research and in the clinic.

For the use of hPSC in the clinic they need to be generated and cultured under GMP conditions. as well as grown with xeno-free conditions, both media and matrices. The amount of cells needed for a single patient is also a limiting factor. In order to produce enough hPSC cells, several approaches have been put in place. Cell culture on laminin-521 allows the expansion of hPSC as single cells instead of colonies,

making possible to greatly expand the number of cells generated in fewer passages and in lesser time. Another approach that will allow easy expansion of hPSC are the bioreactors. Bioreactors are devices with controlled conditions and automatic change of media and the option of different volume capacities.

A major challenge that remains a harness is the broad differentiation potential of hPSC and the generation of fully specialized and functional cells, suitable for clinical use. Standardized, replicable and feasible differentiation protocols are needed in order to overcome this problem. Most differentiation protocols are not reproducible and the resulting cell populations are heterogeneous. Robust tests for cellular processing, optimization of chemically defined culture media and the integration of those differentiation protocols to GMP guidelines are in order.

As previously mentioned, it has been described that during *in vitro* culture, stem cells acquire genetic and epigenetic mutations (Eguizabal et al., 2013; Jopling et al., 2011; Maitra et al., 2005; Nguyen et al., 2013a; Numasawa et al., 2011; Shen et al., 2011; Spits et al., 2008). Several groups have reported that undifferentiated hPSC carry abnormal methylation patterns that will pass once the cells are differentiated (Cowan et al., 2005). The abnormal methylation pattern has been observed at early passages after derivation (Allegrucci et al. 2007). Some of this chromosomal abnormalities acquired while in culture, can provide selective advantage, which could be translated into abnormal growth when transplanted *in vivo* (Hanna et al., 2009). Moreover, these acquired alterations are often related to the silencing of a number of genes.

Once transplanted, the main concerns are the survival of the transplanted cells, their functional integration in the new environment and the engraftment in the specific location needed.

Finally, prior to using hPSC as research models for a specific disease or condition, and drawing final conclusions from the results obtained with these cells, they need to be fully characterized. A deep understanding is required of the similarities and differences between hPSC or hPSC-derived cells and the cells affected by the disease. It is necessary to fully investigate if the molecular pathways leading to the disease *in vivo* are recapitulated in the hPSC or hPSC-derived cells. Much of this work is on-going in many laboratories worldwide, and contributing to this increasing body of knowledge is the focus of this thesis.

3 Modelling genetic conditions using human pluripotent stem cells: the example of dynamic mutations

3.1 DYNAMIC MUTATIONS AND HUMAN DISEASE

Repetitive DNA sequences are long tracks of tandem repeats that are located in coding and non-coding regions of the genome. The repetitive sequences are heterogeneous in size and include trinucleotide repeats (TNR), tetranucleotides, pentanucleotides, microsatellites (1-4 bp), minisatellites (6-64 bp) and megasatellites (several kb).

Most individuals have short and stable repeats that remain identical across generations. The change in the number of repeat units in a determined locus is known as repeat instability. The basic unit of a disease-causing TNR typically is a (CNG)n, where N represents any nucleotide and n the number of repeats. The number of repeat units is variable amongst individuals. Longer tracks are more prone to undergo instability, and different TNR show different thresholds beyond which they become unstable.

TNR instability is associated with disease, each disease being caused by a different mutation at a different locus. The location of the TNR is also different for every disease. Based on the location of the expansion, they are categorized in coding or non-coding repeats. Overall, the wild type allele is stable and in order to become a disease-causing mutation, and to become unstable, the number of repeats needs to overcome a specific length threshold, known as pre-mutation. Depending on the location of the repeat (coding or non-coding region), this threshold is different.

When TNR instability happens in a coding region of the gene, the threshold length for the repeat to become unstable is approximately 29-35 repeats (Kovtun and McMurray, 2008; Mirkin, 2007). TNR expansions found in coding regions can give rise to polyglutamine (CAG) tracts and occasionally polyalanine (CGG) and polyaspartic tracts, which result in an altered protein conformation. The mutant protein may for instance aggregate or establish aberrant interactions with other proteins, which lead to a dysfunction that induces toxicity and cell death. On the other hand, when the variation of the number of repeat units takes place in a non-coding region of the gene, the pre-mutation allele has a threshold of 55-200 units (Kovtun and McMurray,

2008; Mirkin, 2007). Non-coding regions tolerate substantially higher number of repeats than coding regions. The expansion of TNR in a non-coding region can result for example in aberrant RNA function. The altered RNA can interact with proteins, leading to neuronal and systemic malfunction. An overview of diseases caused by TNR instability can be found in McMurray 2010 review (McMurray, 2010). In Table 3 we included the diseases mentioned in this thesis; this dissertation focuses on Huntington's disease (HD) and myotonic dystrophy type 1 (DM1), on which it will further elaborate in the next sections.

Table 3. Examples of trinucleotide repeat diseases.

Disease	Gene	Location	Repeat	Normal repeat range[1]	Pre-mutation repeat range[2]	Disease repeat range[3]
Spinocerebellar ataxia 1 (SCA1)	*ATXN1*	Chr 6 (exon 8)	CAG	6 - 39	40	41 - 82
Huntington's Disease (HD)	*HTT*	Chr 4 (exon 1)	CAG	9 - 29	29 - 37	37 - 121
Fragile X syndrome (FRAXA)	*FMR1*	Chr X (5'UTR)	CGG	6 – 52	55-200	230 - 2000
Myotonic dystrophy type 1 (DM1)	*DMPK1*	Chr 19 (3'UTR)	CTG	5 – 37	37 - 50	90 - 6500
Friedreich Ataxia (FRDA)	*FXN*	Chr 9 (intron 1)	GAA	6 – 32	31 - 100	>200

[1]Average repeat size range in unaffected population.

[2]Repeat size threshold range beyond which the disease is manifested

[3]Repeat size range in the population affected by the disease

Another hallmark of TNR diseases is the transmission of an increasing TNR length from one generation to the next. The consequence of this progressive increase is that over the generations, the phenotype becomes more severe and the disease manifests earlier in time in the progeny. This is because the length of the repeat positively correlates with the severity of the symptoms of the disease and inversely correlates with the age of onset. This phenomenon is known as genetic anticipation.

Germline instability can occur during the replication of the DNA in proliferating cells, during meiosis or during the repair of dormant haploid or arrested germ cells. It is assumed that if expansions occur after meiosis is completed it implies the action of

DNA damage repair or transcription, while mutations occurring before meiosis is finished derive from DNA recombination, repair or transcription. Depending on each specific TNR, the transmission of the TNR to the next generation presents a characteristic paternal or a maternal expansion or contraction bias, specific for spermatogenesis and oogenesis respectively, and the number of repeats can be different between individuals of the same family. This bias in type of transmission correlates to the repeat type. For example the intergenerational instability of the CAG repeats expansion is through paternal transmission, while CTG repeat expansion happens preferentially through maternal transmission (Pearson, 2003).

Finally, TNR in coding regions increase at a rate of less than 10 repeat units from one generation to the next, whilst mutations in non-coding regions can increase from 10 to 10.000 repeat units (Kovtun and McMurray, 2008; Mirkin, 2007).
Next to germ-line instability, TNR can have variable numbers of repeats amongst tissues of the same individual, and this instability is different for each locus and tissue-, developmental stage- and cell-specific (Castel et al., 2010; Pearson et al., 2005). This tissue- and locus-specific instability is most likely linked to the activity of repair systems in different cell types and development stages, replication programmes, epigenetics marks, chromatin packaging and transcription levels of specific loci.

This germ-line and somatic instability is the reason mutations caused by TNR instability are termed dynamic mutations: they keep changing through generations and within individuals during their lifetime.

3.1.1 Huntington's disease

Huntington's Disease is a monogenic disease with progressive neurodegenerative symptoms caused by a CAG (coding for a glutamine) expansion in the exon 1 of the Huntingtin gene (*HTT*). This gene is located in the short arm of the chromosome 4 and codifies for the huntingtin protein. The prevalence of the disease is of 1:10.000 with a mean age of onset of 35-55 years; however, 10% of the total HD cases are juvenile. HD is currently incurable, and there is no effective treatment.

From a genetic point of view, repeats of 6 until 29 CAG units are considered normal and the individuals will not present symptoms of the disease. When the number of repeats reaches 29 to 37 units the symptoms rarely appear but the patients show an increased susceptibility towards further expansion of the mutation in the following germline transmissions. This intermediate state of the expansion is known as pre-mutation. Repeat sizes of >37 repeats are considered as full mutations and will result

in the disease.

The expansion of the pre-mutation occurs mainly in pre-meiotic germ cells during male gametogenesis. Consequently, the majority of the affected individuals inherited the expansion from their father; while few cases have been described to inherit the expansion of the pre-mutation from the mother with no more than a 20 repeats expansion (Nahhas et al., 2005). Studies done on human testicular germ cells isolated by laser capture micro-dissection from HD patients show that CAG repeat expansions are already present before the first meiotic division starts (Yoon et al., 2003). Expansion of the CAG repeat is evident in the early primordial germ cells segregation *in utero* and through the individuals life-long post-pubertal spermatogonial stem cell divisions (Pearson, 2003).
HD patients present somatic TNR instability in adult life but not in foetal stages (Benitez et al., 1995). This instability is restricted to the striatum (largest repeat tracks) and cerebellum (shortest repeat tracks) of the brain. The number of repeats in the same patient present a mosaic pattern in the different tissues presenting the instability (Takano et al., 1996).

At the cellular level, the disease is caused by the fact that the mutated *HTT* protein (m*HTT*) acquires an increasingly large PolyQ tract as the number of CAG repeats increases. The m*HTT* generates fibrillar aggregates, which in turn accumulate in the cytoplasm and the nucleus of neurons. The brain damage induced by HD includes progressive death of projection neurons at the striatum region of the brain, following of motor, cognitive and psychiatric degeneration and eventual death (Harper et al., 1999). The main symptoms are involuntary movements, weight loss, abnormal gait and speech and swallowing difficulties. HD also presents psychiatric manifestations such as personality changes, depression, aggression and early onset dementia.

3.1.2 Myotonic distrophy type 1 disease

Myotonic dystrophy type 1, DM1, is a monogenic neurodegenerative disease with a worldwide prevalence of 1:8000. The clinical symptoms span a continuum from mild to severe and have been categorized into three somewhat overlapping phenotypes: mild, classic, and congenital. In mild DM1, while life span is normal, the patients have cataracts and mild myotonia. In classic DM1, life span may be shortened, and the patients show muscle weakness and wasting, myotonia, cataract, and often show cardiac conduction abnormalities. Congenital DM1 is characterized by hypotonia and severe generalized weakness at birth, often with respiratory insufficiency and early

death. In congenital DM1, intellectual disability is common. The symptoms of DM1 were first described by Hans Steinert in 1909, but was not until 1992 when several groups described the molecular cause as an expansion of a CTG in the 3' non-coding promotor region of the dystrophia myotonica protein kinase (*DMPK*) gene on chromosome 19q13.2-q13.3 (Themeles et al., 2009, Wang et al., 2009, Cole et al., 2009, Ye et al., 2009).

DM1 is a multi-systemic disorder caused by an aberrant RNA with long CUG tracts that deregulates several transcripts involved in a wide range of physiological processes. The expanded RNAs will bind to RNA-splicing proteins, such as CUG triplet repeat RNA-binding protein 1 (CUGBP1) and ETR3-like factors (CELFs proteins), and musclebind-like (MBNL proteins), altering its function. The *DMPK* RNA carrying an expansion forms nuclear foci that upregulate MBNL proteins, which are antagonists of CUGBP1, resulting in an upregulation of CUGBP1. It also results in the aberrant protein expression of, amongst others, insulin receptor, chloride channel CLCN1, cardiac troponin T and MTMR1, resulting in insulin resistance, myotonia, cardiac abnormalities and muscle weakness, respectively (reviewed in Gatchel et al. 2005).

The DM1 CTG repeats number varies in the general population from 5 up to 35 repeats. Individuals with 38 to 49 CTG repeats do not present symptoms during their lifetime but are considered carriers of the pre-mutation. At this point, the TNR can increase from one generation to the next by up to thousand or more repeats per generation (Kovtun and McMurray, 2008; Mirkin, 2007). Patients with mild DM1 have 50 to 150 repeats, patients with classic DM1, 100 to 1000 repeats, and those with congenital DM1 can have more than 2000 repeats.

In DM1, the length of the CTG expansion is often correlated to the type of germ-line transmission. CTG repeats of <100 repeats are mainly the result of a paternal transmission, 200-600 repeats equally originate from both paternal and maternal transmissions and very large expansions (>1000 repeats) are mostly exclusively transmitted through the maternal lineage (Pearson, 2003).

The differences in expansion rate of the repeat between paternal and maternal transmissions could be due to the long time the oocytes remain arrested. Unlike male germ cells, female germ cells start meiosis during *in utero* development and after spend years arrested in dictyotene (the last phase of meiosis I's prophase). When women reach sexual maturity, meiosis restarts with every menstrual cycle and it is finally finished after fertilization. Large CTG expansions have been documented early in primary oocytes, in the maternal uterus, during oogenesis and before meiosis I is

completed. This suggests that these expansions are DNA repair-dependent (Dean et al., 2006; TEMMERMAN et al., 2004). In male germ cells with full mutations, the repeats tend to contract in diving spermatogonia, and the longer the repeat; the more probable they contract (Usher et al., 2009, Chen et al., 2009).
In somatic cells, DM1 patients present CTG instability already at the 13-16th week of development and for certain tissues this instability continues during the entire life of the individual (Niclis et al., 2009; Pardo et al., 2004). In the adult, in the case of DM1, patients present instability of the CTG repeat in muscle and lymphocytes of the blood (Castel et al., 2010; Delmas et al., 2009), as well as in certain regions of the brain.

3.2 MECHANISMS INVOLVED IN TNR INSTABILITY

Twenty years ago, the unstable expansion of the TNR was described as the cause of several neuromuscular and neurodegenerative diseases. However, the molecular mechanisms causing TNR instability are not yet fully understood. Nevertheless, it is well known that the sequence symmetry inherent to the repetitive sequences of (CNG)n can form unusual secondary DNA structures known as looped or slipped-stranded intermediates (Castel et al., 2010; Gacy et al., 1995; McMurray, 1999). The most generally accepted mechanism considers DNA slipped sequences as the origin of TNR instability.

Slipped-stranded intermediates are mainly formed in single stranded DNA during processes involving DNA denaturalization and renaturalization, such as DNA replication, recombination and/or repair (Castel et al., 2010; Pearson and Sinden, 1996). Transcription or epigenetic modifications can also lead to the formation of slipped DNA structures (Cleary and Pearson, 2003; Pearson et al., 2005; Pook, 2012). Slipped-DNA structures are not only composed by standard base pair matches (A:T; C:G) but also by mismatches (Gacy et al., 1995). The degree of instability of the sequence is determined by the composition of the repetitive sequence and the energy contribution of the mismatched base pairs. The stability decreases following the order: CGG>CTG>CAG=CCG (Moore et al., 1999).

Studies of TNR instability, in animal models, have shown that hairpin or slipped-stranded repetitive sequences can be either included or excluded in the DNA sequence as TNR expansions, leading to DNA expansions or contractions of the original locus. Both expansions and contractions take place during DNA replication, DNA repair and/or DNA recombination mechanisms (Dean et al., 2006; Mirkin, 2007). Since the main focus of our research was on DNA repair, in the next sections we cover more

extensively the mechanisms of DNA repair than DNA replication and recombination.

3.2.1 DNA replication

Slipped-DNAs are known to stall various DNA polymerases (Gacy et al., 1998; Telenius et al., 1994; Usdin and Woodford, 1995) during the progression of the replication fork. Occasionally, this stalling of the replication fork results in a misalignment of repetitive DNA which causes expansions or contractions depending on whether the slippage happens either in the template or the daughter DNA strand, respectively (Kang et al., 1995; Shelbourne et al., 2007) (Figure 5a).

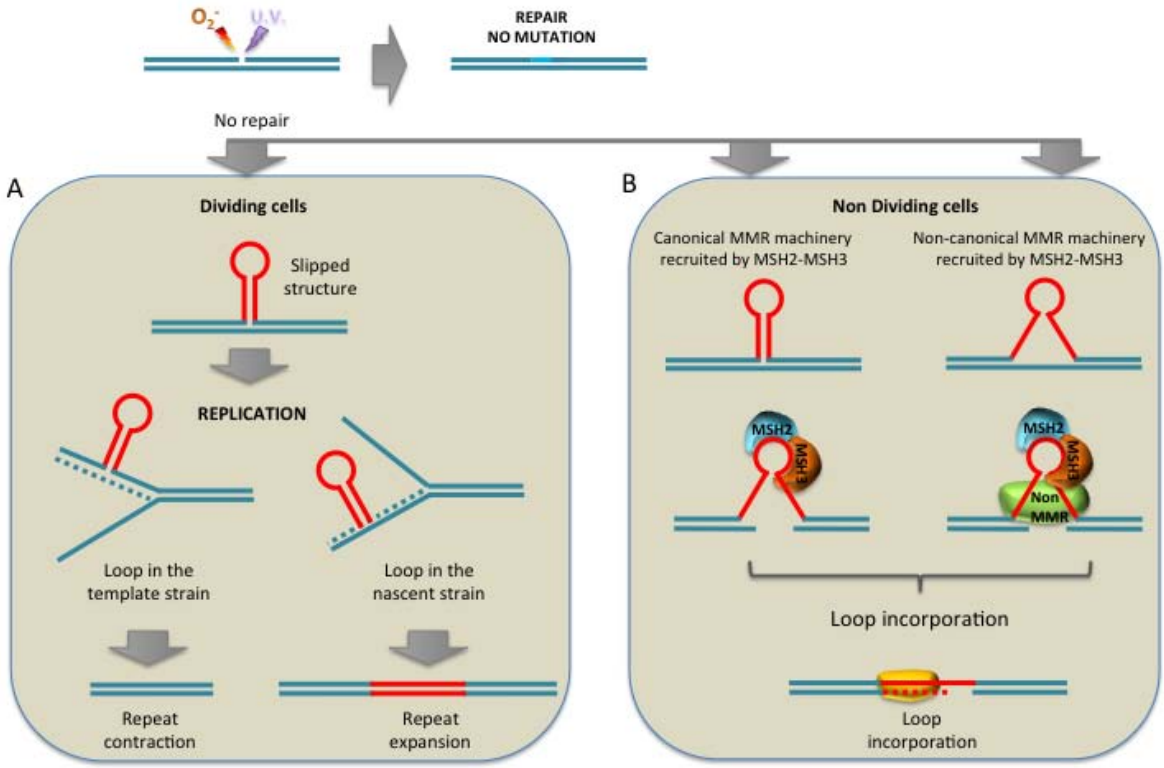


Figure 5. Mechanisms of repeat instability. A: In dividing cells, during replication, if the repeat forms a loop in the template or nascent strand the repeat will be contracted or expanded respectively. **B:** In non diving cells, breaks in the strand caused by different factors

Some proteins from the replication complex have been identified as mediators of TNR instability. The current hypothesis suggests that proteins that stabilize replication forks protect against repeat expansions, as for example mediator of replication checkpoint 1 (Mrc1), topoisomerase 1-associated factor 1 (Tof-1), chromosome segregation in meiosis protein 1 (Csm1), Rec proteins and Sr2. Conversely, proteins involved in template switching and replication fork restart promote TNR expansion

(endonuclease FEN1, the DNA ligase 1 and the proliferating cell nuclear antigen (PCNA)) (Razidlo et al., 2008; Castel et al., 2010).

3.2.2 DNA repair

TNR expansions are observed both in proliferating tissues and post-mitotic tissues, such as brain or skeletal muscle (Fortune et al., 1993; 2000; Kennedy, 2003; Manley et al., 1998; Thornton et al., 1994). The fact that tissues with none or limited capacity for cell division can show TNR instability indicates that non replicative processes, such as DNA repair and maintenance, can also contribute to the acquisition of TNR instability.

One of the mechanisms that carry out DNA repair is the Mismatch repair (MMR) system. DNA MMR recognizes and repairs erroneous DNA pairings, insertions and deletions or DNA damaged caused by several physical and chemical agents such as UV and oxidative stress. Because post-mitotic tissues are exposed to the lifetime accumulation of oxidative stress that induces nicks or gaps in the DNA, DNA repair mechanisms are particularly active in these tissues. For instance, oxidative stress damage is detected and repaired by the MMR machinery, specifically by base excision repair 8-oxoguanine glycosylase enzyme (OGG1), which has also been linked to somatic instability in HD mouse models (Benitez et al., 1995; Slean et al., 2008).

CTG DNA slipped-outs are repaired by different mechanisms depending on its length. Loops not exceeding from 1-20 nucleotides are efficiently repaired in a MMR dependent manner while slip-outs of more than $(CTG)_{20}$ escape repair, resulting in expansions. By contrast, slipped structures up to 216 nucleotides are repaired through other paths other than MMR (Panigrahi et al., 2010).

MMR also participates in nucleotide excision repair (NER), Base excision repair (BER), Double strand break (DSB), Transcriptional coupled repair (TCR), meiotic recombination, single strand annealing, apoptosis and insertion, deletions and loops Indels repair (Iyer et al., 2006; Kunkel and Erie, 2005; Modrich, 2006).

3.2.3 MUT proteins

The early work on MMR in *Escherichia coli* and *Saccharomyces cerevisae* (Kunkel et al., 2005) led to the identification of several protein complexes, MutS, MutL, MutH and UvrD, that, when mutated or inactivated, resulted in a hyper-mutator phenotype. They were called "Mut" for that reason, and they turned out to play a central role in detecting and repairing base pair mismatches.

In human cells, five MutS homologues (MSH) have been identified (MSH2 to 6), from which *MSH2*, *MSH3* and *MSH6* participate in MMR in the form of heterodimers (Table 4). The repair of base-base mismatches and small insertion/deletion loops (IDLs) of one or two extrahelical nucleotides is initiated by the mismatch-binding factor composed by MSH2 and MSH6 (MutSα). The repair of larger IDLS is initiated by MSH2 and MSH3 heterodimers (MutSβ).

Human cells express four MutL homologues, MLH1, MLH3, PMS1 and PMS2, which function as three different heterodimers (Table Mut proteins). MLH1 can form heterodimers with PMS1 (MutLα complex), PMS2 (MutLβ complex) or MLH3 (MutLγ complex), which interacts with MSH2- containing complexes bound to mismatches.

Table 4. Mut and human homolog, MSH protein complexes involved in MMR.

Protein Complex	Human Mut homolog	Function
MutSα	MSH2:MSH6	Recognition of single base pair mismatch, single and double base insertions
MutSβ	MSH2:MSH3	Recognition of 2-13 base pair loops, single base pair mismatches and small extrahelical loops (IDLs)
MutLα	MLH1:PMS2	Form heterodimers that will bind to the MSH2:MSH3 bound to the mismatch. PMS2 has endonuclease activity, which will generate nicks into the DNA strand.
MutLβ	MLH1:PMS1	Form heterodimers that will bind to the MSH2:MSH3 bound to the mismatch.
MutLγ	MLH1:MLH3	Involved in meiotic recombination

3.2.4 Mismatch repair and trinucleotide repeat instability

DNA repair is initiated by the MutS complexes, which identify the mismatch and bind to the DNA chain. Once bound to the mismatch, MutS will recruit MutL forming a ternary structure that in the presence of RFC and PCNA will activate the endonuclease activity of PMS2. PMS2 endonuclease activity will introduce single strand breaks close to the mismatch allowing new areas for the endonuclease EXO1 to degrade the new generated un-methylated strand containing the mismatch. MutLα (MLH1-PMS2) recruits the DNA polymerase III to the mismatch site by directly interacting with the clamp loader subunits of the DNA polymerase. MLH1 can also form heterodimers with MLH3, known as MUTLγ complex which is involved in cell division during meiosis (Jiricny, 2006). The role of the MMR in TNR instability is illustrated in Figure 4b, in

which the two main proteins MSH2 and MSH3 are shown.

The involvement of these proteins in TNR instability has been proven in numerous experiments. For instance, MSH2-MSH3 heterodimer repairs single-base insertions and small loop-outs appeared during DNA replication. However, it can work as a mutational mechanism instead of as a repair mechanism when encountering CTG/CAG repeats. It has been revealed that human hMSH2–hMSH3 (the hMutSβ complex) plays a role in CTG/CAG instability since short DNA slip-outs of 1 – 3 CTG units are repaired in a hMutSβ dependent manner. At higher hMutSβ protein concentrations, the correct repair of slipped-CTG repeats is escaped and likely error-prone repair leads to expansions (Panigrahi et al., 2010), similar to the MSH3 concentration sensitivity observed in DM1 mice (Foiry et al., 2006)

Both MutS and MutL complexes have been identified as promoters of TNR expansion in transgenic DM1 and HD mice. Mutational inactivation of MSH2 or MSH3 decreases the frequency of repeat expansions in mice. Homozygous knockouts of MSH2-MSH3 suppress >50% of the TNR instability while shifting the pattern towards contractions (G Jansen, 1994; Martorell, 1997; Monckton et al., 2001; Wohrle et al., 1995; Owen et al., 2005).

In this sense, in Hdh^{Q111} HD mouse model, MSH2-MSH6 complex protects against paternal intergenerational contractions, while striatal instability is MSH2-MSH3 dependent (De Temmerman et al., 2008; Dragileva et al., 2009). $MSH2^{-/-}$ transgenic mice or mice with a mutation in the ATPase activity of MSH2 show a tendency towards contractions and completely abolish expansion of the CAG TNR both in germ and somatic cells (Monckton et al., 2001; Niclis et al., 2009). Similarly, $MSH2^{-/-}$ and $MSH3^{-/-}$ HD or DM1 mouse models show stabilization or contraction of the expanded CAG/CTG repeat (Dragileva et al., 2009; Wheeler, 1999). Importantly, MSH3 is a limiting factor for somatic instability, its deficiency improves the contraction of expanded CTG in DM1 mice (Foiry et al., 2006). Differently than MSH2, MSH3 deficiencies cause weak mutator phenotype and no cancer predisposition (Slean et al., 2008; Thomson et al., 1998). Altogether reinforce the idea of a decrease in TNR instability when there is a deficiency of either MSH2 or MSH3 proteins.

The MutLα complex also plays an important role in MMR. Probe of that is that mice cells lacking either MLH1 or PMS2 exhibit mutator phenotype and microsatellite instability that is comparable to cells lacking MSH2 (Marra and Jiricny, 2005). The rate of somatic expansions in PMS2-/- mice present is ~50% lower than the wildtype mouse (Gomes-Pereira et al., 2004).

In human cellular models, results obtained from knocking down, by siRNAs, MLH1 and PMS2 proteins in a fibrosarcoma cell line revealed a requirement of these proteins for the instability of integrated CTG tracts (Lin and Wilson, 2009). Several groups have been testing the involvement of MSH2 and MSH3, MSH6 in TNR instability by transgenic knock down in cells carrying a CTG.CAG repeat. Overall the results shows that knock out of MSH2 and MSH3 result in a significant decrease in TNR instability (Halabi et al., 2012). Conversely knock down of MSH6 appears to have strongest effect on cells carrying large expansions, increasing the number of repeats (Nakatani et al., 2015). Other work on Friedreich Ataxia hiPSC has shown results apparently contradicting the above. The knock down of MSH2 and MSH6 in hiPSC resulted in the stabilization of the repeat. This differences maybe due to the use of cells with different repeat length and different cell types (Du et al., 2013). All in all, the link between a decreased TNR instability and a natural down-regulation of MMR proteins in an endogenous disease locus was never shown at the time of the start of our work.

3.2.5 DNA recombination

In addition to repair and replication, there is evidence suggesting that DNA recombination plays a role in TNR instability. For instance, the massive germline TNR expansion seen at the DM1 locus may involve recombination processes during the intra-chromosomal annealing between sisters chromatids. However, data derived from experiments using mammals as models organisms show contradictory conclusions on the implication of the recombination process in the instability of the TNR (Pearson et al., 2005). More research is required in order to determine the impact of implication of recombination processes in TNR instability.

3.3 RESEARCH INTO TNR INSTABILITY: MODELS AND LIMITATIONS

The majority of the current knowledge on the mechanisms behind TNR instability derive from the study of bacteria, yeast, flies, worms and mice models (reviewed in the table S3 of (Castel et al., 2010)). These systems have provided very valuable information for the understanding of the TNR biology. However, some of the molecular mechanisms and developmental processes in human greatly differ from the ones in the laboratory models.

Several animal models have been used for the study of the TNR instability, e.g. C. elegans, D. melanogaster, rodents and large animals. However, from all the animal

models, rodents are the most commonly used. Since rodents do not present natural TNR instability, different approaches have been taken in order to generate suitable models for the study. Nevertheless, no single strain has been obtained that can fully reproduce the symptoms of the human disease. Each rodent model recapitulates a unique set of phenotypes and molecular features of the TNR diseases, including the severity and the time of appearance of the symptoms, and the manner of germline transmission. Also the short lifespan of the rodent models is a limitation for the study of TNR diseases because age-dependent factors do not have time to manifest in rodents. Regarding the germline transmission of TNR, differences exist between mice and human. During spermatogenesis, for instance, the germline in humans undergoes a greater number of mitotic divisions than in mice. The number of divisions undergone by spermatogonial stem cells during the lifetime of a human is also estimated to be of one order of magnitude larger than in mice, potentially increasing the number of mutations acquired in humans (Pearson et al. 2003).

Before the description of the genetic basis of TNR diseases, rodent models for HD and DM1 diseases were generated by neurotoxin striatal mediated lesions. For example, glutamate agonists, who induce an acute lesion in the brain, were used for the generation of HD models. Glutamate agonist produces selective loss of GABAergic neurons, which are the most affected neurons in HD disease. In addition to the neurotoxins, mitochondrial toxins like malonate and 3-nitropropionic acid (3-NPA) were also used for the generation of the HD models. However, the sharp nature of the lesions and the localized damage did not reproduce the slow progression of the symptoms that takes decades to appear in patients, neither the damage in the peripheral nervous system.

With the discovery of the genetic loci responsible for the different TNR diseases, scientists started creating genetically manipulated animal models. For instance, knock-in transgenic animal models with fragments or with the full length of the *HTT* and *DMPK* genes have been generated. One hindrance to this technology is the large size of the *HTT* (170kb) and *DMPK* (>45kb) loci, making it very difficult to generate an error-free cDNA and to clone it into a plasmid, which can harbour inserts of about 10 to 15kb. To overcome the size issue, *HTT* and *DMPK* genes were cloned in genomic plasmids as yeast artificial chromosome (YAC) and bacterial artificial chromosomes (BAC). YAC and BACs have been used for the generation of the transgenic knock-in animal models, which include the full length of the genes of interest.

3.3.1 Huntington's disease models

Currently, there are more than 20 different HD rodent models have been generated. The more widespread strains are the R6/1 and R6/2, which include a truncated N-terminal fragment of the *HTT* gene; the $HdhQ^{111}$ model that is a full length knock-in *HTT* that includes the exon 1 *HTT* plus a variable number of CAG repeats and YAC128 that is a transgenic mouse model including the full human *mHTT* transgene.

R6/1 and R6/2 mice present somatic and germline instability (Gonitel et al., 2008). This model exhibits a rapid onset of the motor, cognitive and behavioural symptoms that are more severe than the human diseases.

The $HdhQ^{111}$ knock-in mice recapitulate the intergenerational expansion bias by paternal transmission and specific instability in the striatum, both characteristic features found in HD patients (Wheeler, 1999). Studies performed in this model shown that Msh3 protein is a limiting factor in the generation of CAG striatal instability. Since the instability is CAG length dependent allows the possibility to determine the effect of the length in the pathogenesis progress. To further study this CAG length dependence, the impact of Msh2 loss was tested (Wheeler et al., 2003) as well as the loss of Msh3, Msh6 and Xpc repair proteins (Dragileva et al., 2009).

The transgenic YAC128 mice present progressive motor, psychiatric and cognitive disturbances as well as striatal and cortical atrophy. The CAG repeat is however somatically stable, contrary to the situation in the human. In addition, the YAC and BAC systems have low copy number integration resulting in low expression levels of the transgene, in contrast with the endogenous gene expression seen in patients. Nevertheless, they do not present a consistent neurodegeneration and behavioural pattern during the lifespan (2-3 years).

Regarding the germline repeat instability, in HD mouse models it is first detected in haploid spermatid cells whereas in humans is detected earlier, during mitotic divisions of spermatogonial diploid cells (Reichelt et al., 2009, Wang et al., 2009).

Finally, there are also three transgenic models using large animals. Transgenic Rhesus macaque and Tibetian miniature pig have been generated using fragments of the *mHTT* gene. Transgenic sheep were developed using the full-human *HTT* mutation. All three models present severe phenotype with early onset of the symptoms. Monkeys with a 29 CAG-repeat tract (a not-pathological genotype in the humans) did not survive for longer than 6 months, but both the sheep and the miniature pig were able to transmit the expansion to the progeny. Therefore, none of the large animal models

can be considered fully useful for the study of human HD.

3.3.2 Myotonic dystrophy type 1 models

There are no animal models that show the extensive and large expansions seen in human DM1 tissues. However, several groups have attempted to create mouse models that reproduce the large intergenerational repeat length changes, the high mutation rate observed in germ line as well as the characteristic symptoms of the DM1 disease. The most commonly used models to study DM1 are the DM300 transgenic mice, which carry the full human *DMPK* gene plus more than 300 CTG repeats, and the knock-in mouse model $(CTG)_{84}$ with CTG repeats inserted in the 3' region of the humanized *DMPK* gene. DM300 presents the repeat instability characteristic of DM1 disease and brain and muscle abnormalities, including myotonia. The $(CTG)_{84}$ model shows a shift from expansion towards contractions in the somatic tissues and also across generations in an $Msh2^{-/-}$ background (Savouret et al., 2003). In the complete absence of Msh3, it shows a complete blocking of the CTG instability in somatic tissues (VandenBroek et al., 2002).

3.3.3 Stem cells for the study of Huntington's disease and Myotonic dystrophy type 1

To overcome the limitations of animal and cellular models, hPSC have been proposed as a better alternative for the study of specific aspects of the disease. The principle is that hPSC can be differentiated into the relevant cells for HD and DM1 disease, such as striatal projection neurons (DARPP32+/GABA+), oligodendrocytes and astrocytes, providing cells with an endogenous human mutation. For these two diseases, the possibility of having an in principle infinite source of these specific neural cells is particularly appealing, given the difficulties to obtain these cells and to keep them in culture.

To produce hPSC carrying TNR expansions, either hESC can be derived from PGD spare embryos affected of HD and DM1, or hiPSC can be generated from patients. This last approach has the advantage that the patient characteristics are known, and this may be of interest to correlate this to the results of the molecular research.
The hPSC laboratory of the Vrije Universiteit Brussel currently holds over 25 hESC lines carrying monogenic diseases (3 of which carry the DM1 mutation and 1 cell line the HD). Regarding hiPSC, numerous cell lines have been obtained from patients with HD and DM1. More recently, the improvement of the genome editing approaches such

as the method using CRISPR/TALEN has enable researchers to edit, correct or insert mutations in hiPSC cell lines generated from patients. This allows for the generation of lines with identical genetic background, except for the mutation.

In order to determine whether hPSC are suitable models for the study of dymanic mutations, much preliminary work needs to be carried out in terms of characterizing the behaviour of these loci in this specific cell type. At the time of the start of this thesis, only two papers had been published carrying out this type of work for HD and DM1, and one paper showed that the repeat causing fragile X syndrome had a very particular behavior in a male hESC line carrying a large CGG expansion (Eiges et al., 2001). Regarding DM1, the study of Temmerman et al. by our group, assessed the instability of the DM1 locus in one hESC line (Temmerman et al., 2004), and Niclis et al studied two HD-derived hESC cell lines and their differentiated progeny (Niclis et al., 2009). Since the publication of our work, other groups have published similar studies; these will we discussed in the general discussion, and put in perspective to our work and the works published before the start of our study.

4 Aims of the study (I)

In this work, we proposed hPSC as possible models for the study of mechanisms of TNR instability in HD and DM1.

We aimed at characterizing the behaviour of the TNR in the DM1 and HD derived hESC lines in undifferentiated pluripotent cells and cells differentiated into osteoprogenitors and neural progenitor cells, and its link to the levels MMR machinery proteins.

Concretely, we aimed at:

1. Characterising the levels of instability of the CAG and CTG repeat during *in vitro* culture of hESC carrying expansions in the HD and DM1 locus.
2. Investigating the effect of differentiation into osteogenic progenitors-like cells and neural progenitors on the stability of the TNR in both diseases.
3. Investigate the levels of gene-expression of the MMR genes in both differentiated and undifferentiated cells, and its correlation to the behaviour of the TNR.

5 Huntington's and myotonic dystrophy hESCs: down-regulated trinucleotide repeat instability and mismatch repair machinery expression upon differentiation

Seriola A*, Spits C*, Simard J.P., Hilven P., Haentjens P., Pearson C.E., Sermon K.

Published in Human Molecular Genetics. 2011, 20:176-85. * Joint first authorship.

5.1 INTRODUCTION

Various diseases are caused by the genetic expansion of trinucleotide repeats (TNR), including Huntington's disease (HD [MIM 143100]) and myotonic dystrophy type I (DM1 [MIM 160900]). HD is caused by an expansion of a CAG repeat tract in the coding region of the Huntingtin gene (HTT) to more than 35-40 repeats, but rarely above 60 repeats (Castel et al., 2010). The expansion gives rise to a toxic-gain-of-function protein with an expanded polyglutamine tract, which subsequently leads to neural cell death. The severity of disease symptoms, age-of-onset and progression are related to the size of the CAG expansion. DM1 is caused by a CTG repeat expansion to at least 90 repeats in the 3' untranslated region of the DMPK gene, which produces a toxic-gain-of-function CUG RNA. DM1 expansions are frequently expanded to thousands of units (Castel et al., 2010). For DM1, as for HD, the severity of disease symptoms, age-of-onset and progression are related to the size of the CTG expansion. Data from patients and model systems support a contribution of somatic instability at these disease loci to age-of-onset, disease severity and progression (Castel et al., 2010).

Although HD and DM1 are caused by dynamic mutations of complementary repeats CAG/CTG and thus share common characteristics, as noted above, there are also striking differences. The degree of instability of the repeat does not only depend on its size, but also on its genomic context and on the tissue studied (Cleary and Pearson, 2003; Cleary et al., 2010; Pearson et al., 2005).

Intergenerational instability occurs during gametogenesis when the repeat is over a certain threshold size and can change (often increasing in size), leading to the phenomenon of genetic anticipation (i.e. the aggravation of the symptoms and earlier

age of onset from one generation to the next) (Pearson et al., 2005; Pearson, 2003; Yoon et al., 2003; Dean et al., 2006).

In somatic tissues, the HD CAG repeat is stable, except for the brain, where the largest expansions are observed in the striatum (Telenius et al., 1994; Shelbourne et al., 2007). In contrast, the DM1 repeat presents high levels of somatic instability, with major length differences between different tissues (Thornton et al., 1994). In Huntington's fetuses no repeat instability can be detected between tissues (Benitez et al., 1995; De Temmerman et al., 2008) whereas in DM1 fetuses, considerable levels of instability have been detected (G Jansen, 1994; Lerou et al., 2008; Martorell, 1997; Tachibana et al., 2013; Wohrle et al., 1995; Zhang et al., 2006). While inter-tissue repeat length differences are limited in early fetuses, some instability has been detected. However, the precise timing of the onset of somatic instability is unclear (G Jansen, 1994; Martorell, 1997; Mateizel et al., 2008; Wohrle et al., 1995). Repeat instability has (Chang et al., 2000; De Temmerman et al., 2008; Panigrahi et al., 2010) and one differentiated HD hESC line (Cannavo et al., 2005; Charbonneau et al., 2014; Niclis et al., 2009).

Various processes have been implicated in non-human model systems to contribute to CTG/CAG instability. DNA replication, repair as well as transcription are suggested to drive repeat instability (Brons et al., 2007; Cleary et al., 2010; Goula et al., 2009; López Castel et al., 2009; Mascetti and Pedersen, 2014; Tesar et al., 2007). Proteins in the DNA mismatch repair pathway (MMR) are involved in TNR instability (Panigrahi et al., 2010; Slean et al., 2008). The MMR machinery is composed of various interacting proteins, including MSH2, MSH3, MSH6, MLH1, PMS1, PMS2, and MLH3. The MSH2, MSH3 and PMS2 proteins have been shown to be involved in TNR expansion in transgenic DM1 and HD mice. MSH2 forms a heterodimer with MSH3, to direct the repair of insertion-deletion loops. MSH2 also forms a heterodimer with MSH6, to repair base substitutions and small insertion/deletion loops up to 12 nucleotides. MLH1 can form a heterodimer with either PMS1 or PMS2 or MLH3, which interacts with MSH2-containing complexes bound to mismatches. The involvement of these proteins in TNR instability has been proven in numerous experiments. For instance, Msh2-/- and Msh3-/- HD or DM1 mouse models show stabilization or contractions of the expanded CTG/CAG repeat (Cannavo et al., 2005 006ep; Dragileva et al., 2009); PMS2 is known to be an enhancer of somatic mosaicism of trinucleotide CAG•CTG repeat (Gomes-Pereira et al., 2004). Importantly, the concentration of Msh3 in DM1 mice determines the level of CTG expansions (Foiry et al., 2006). Curiously, in the absence of either Msh2 or Msh3, large expanded CTG repeats show 90%

contractions, as opposed to the 90% expansion in the presence of these proteins (Savouret et al., 2004). A mechanistic role of the human hMSH2-hMSH3 (the hMutSβ complex) in CTG/CAG instability has been revealed: short DNA slip-outs of 1-3 CTG units are repaired in a manner depending upon hMutSβ. At higher MutSβ protein concentrations the proper repair of slipped CTG repeats is hampered and likely error-prone leading to expansions (Panigrahi et al., 2010), similar to the MSH3 concentration sensitivity in DM1 mice (Foiry et al., 2006).

Human cell models where MMR proteins were knocked-down by siRNAs revealed a requirement of these proteins for the instability of integrated CTG tracts (Lin and Wilson, 2009). However, there has been no demonstration of a natural down-regulation of MMR proteins with instability of an endogenous disease locus.

Human Embryonic Stem cells (hESC) are pluripotent cells derived from surplus or discarded embryos after in vitro fertilization treatments. The cells possess the ability to differentiate into the different cell types present in the adult individual and are thus considered pluripotent (Gerecht-Nir and Itskovitz-Eldor, 2004). Next to their obvious potential in regenerative medicine, hESC have also been proposed as a valuable tool to study early development, lineage commitment, tissue growth and maturation during embryonic development. They can also be used for the identification of biomarkers of disease and could be models to test the toxicity and efficacy of new drugs and chemicals (Brivanlou et al., 2003; Keller and SNODGRASS, 1999). This is of particular interest in disorders such as DM1 and HD, for which the existing mouse models do not fully represent the human pathogenesis or the culture of the relevant cell types is difficult.

The aims of this study were to characterize the behavior of the TNR in the DM1- and HD-derived hESC lines in undifferentiated and differentiated cells (osteogenic progenitor-like cells, neural progenitor cells and teratoma cells). We also studied the expression of various MMR proteins and correlated this with the repeat instability.

5.2 RESULTS

5.2.1 HD and DM1 repeats in the hESC lines

The analysis of the CAG repeat in HTT, in passages 3, 12, 20, 30, 31, 41, 51, 58, 65, 75, 86, 97, 103, 106 of VUB05_HD, showed a stable genotype of 23 and 44 repeats (supplementary figures 1a and 1b).

On the other hand, the study of the size of the CTG expansion in DMPK in VUB19_DM1 and VUB24_DM1, showed an unstable genotype, presenting both expansions and

contractions (Figure 5). VUB19_DM1 carried an expansion of approximately 250 repeats at passage 1, which enlarged during cell culture, to give rise to a population of hESC carrying mostly an expansion of 370 repeats by passage 37, and of 465 repeats at passage 54. VUB24_DM1 started with a much larger expansion of 1800 repeats, first measured at passage 5. Over time in culture, predominantly contractions of this repeat were detected. Wild type alleles remained unchanged through the whole time period studied.

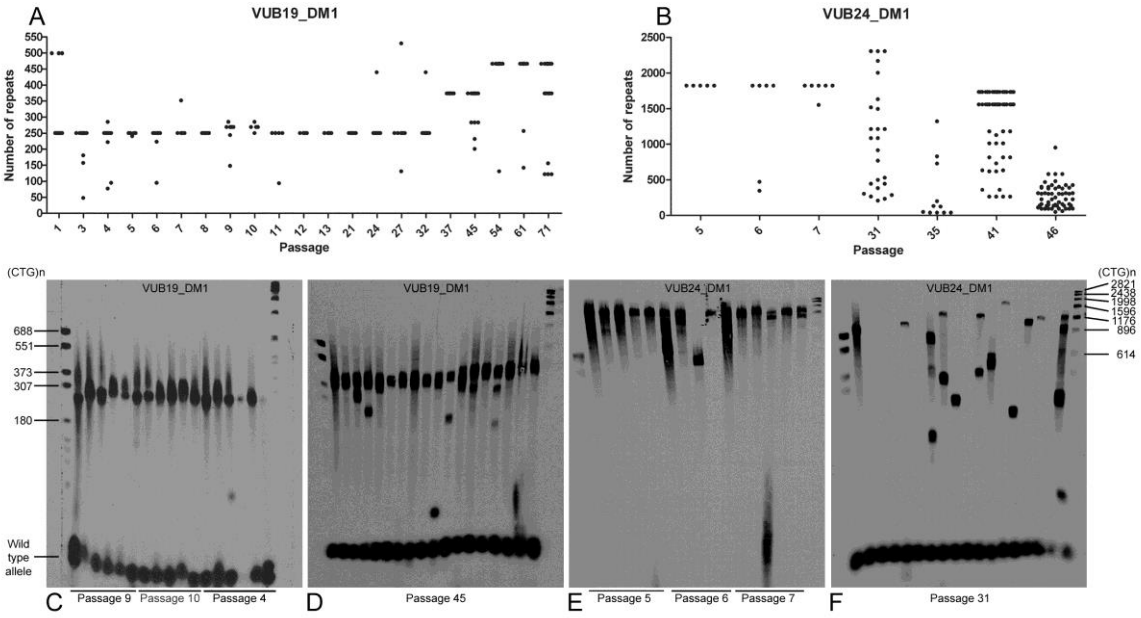


Figure 5. Evolution of the size of the CTG repeat in the hESC lines VUB19_DM1 and VUB24_DM1. A: Allele representation for different passages of VUB19_DM1. Each dot represents an allele, as detected by small-pool PCR. The repeat shows a statistically significant trend to an increase in size ($P<0.001$). B: Allele representation for different passages of VUB24_DM1. The repeat shows a statistically significant trend to a decrease in size ($P<0.001$). Panels C to F: Small-pool PCR and Southern blot results for VUB19_DM1 and VUB24_DM1. The different lanes show replicates of small-pool PCR, and each blot contains the molecular weight markers VI and VII at the first and last lanes, respectively. The marker has been converted to its equivalent in (CTG)n to simplify the interpretation. C: VUB19_DM1 passage 9 (lanes 2-7), passage 10 (lanes 8-12) and passage 4 (lanes 13-18). D: VUB19_DM1 passage 45. E: VUB24_DM1 passage 5 (lanes 2-6), passage 6 (lanes 7-11) and passage 7 (lanes 12-16). The wild type allele is not visible because the gel was run long to resolve the high molecular weigh bands. F: VUB24_DM1, passage 31.

The results for VUB03_DM1 were previously published and are here summarized for the clarity of the other results in this work (De Temmerman et al., 2008). At the initial passages, VUB03_DM1 carried a (CTG)470. Small-pool PCR of later passages, (up to passage 120), showed an increasing number of alleles, with predominantly expansions going up to 2100 repeats (upper detection limit of the method).

Quantification of the relationship between repeat size and passage number using

multiple regression analysis indicated that this relationship was statistically significant ($P<0.001$), with a standardized regression coefficient of 0.91 for VUB19_DM1 (Figure 5A) and of -0.499 for VUB24_DM1 (Figure 5B). The sign of the standardized regression coefficient describes the direction of the relationship between repeat size (Y, dependent variable) and passage number (X, independent variable), i.e., repeat numbers in the higher passages are significantly higher in VUB19_DM1 and lower in VUB24_DM1.

5.2.2 HD and DM1 repeats in differentiated cells

VUB03_DM1, VUB05_HD and VUB19_DM1 were differentiated into osteogenic progenitor-like cells (OPL, see Methods). This particular differentiation protocol was chosen because it yields a very homogeneous population of differentiated cells and is easily reproducible. DNA and RNA samples were collected at most passages, until the senescence of the OPL cells at passage 11.

VUB05_HD did not show any changes in repeat size during the OPL differentiation (Figure 6A, B and C). On the other hand, VUB03_DM1 and VUB19_DM1 showed stabilization of the repeat.

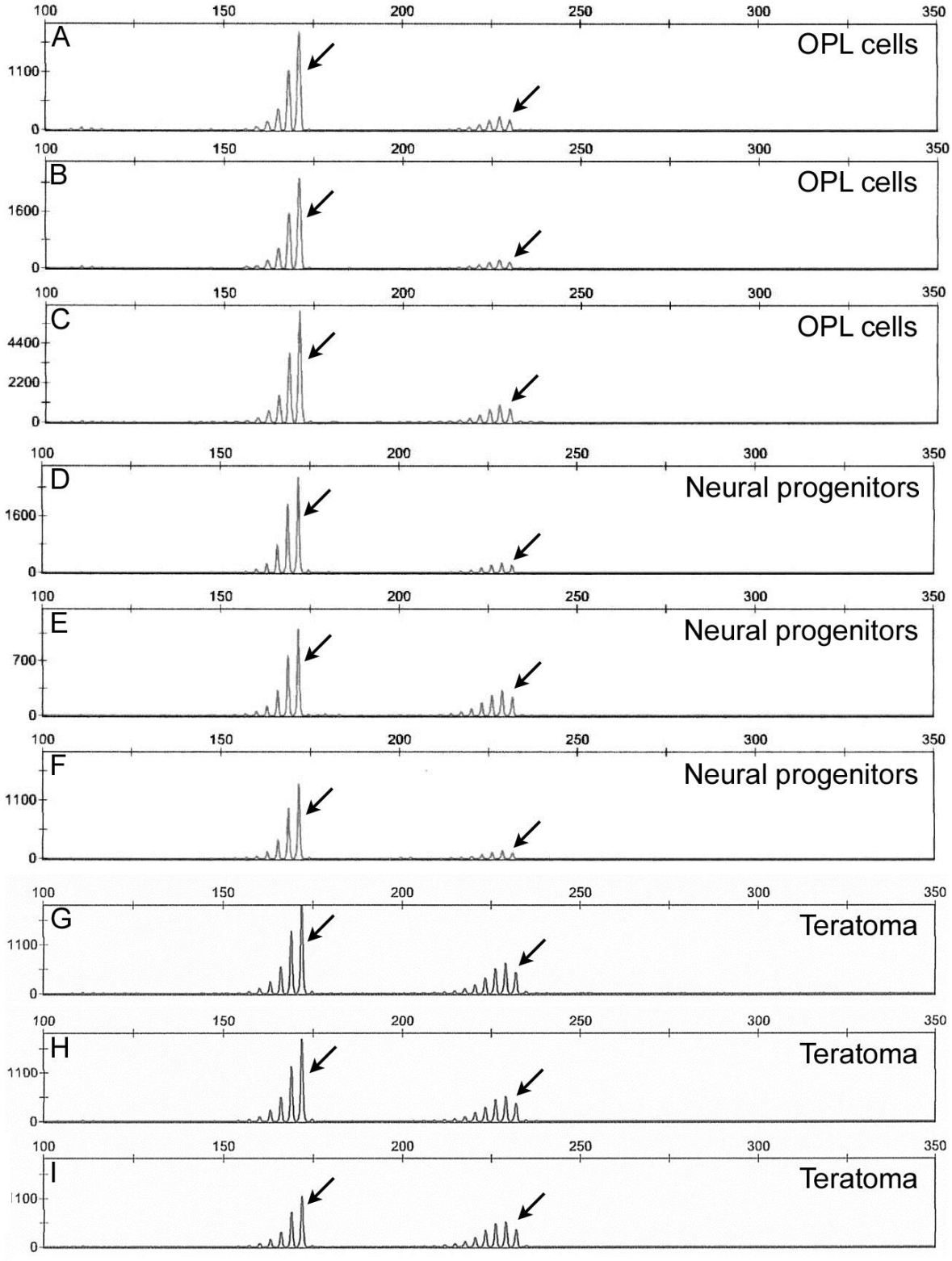


Figure 6. Representative results of the small-pool PCR analysis of the HD repeat in VUB05_HD. The horizontal scale indicates fragment size in basepairs, the vertical shows fluorescent signal intensity. The arrows indicate the CAG alleles. A, B and C: results for the OPL cells obtained from VUB05_HD. D, E and F: VUB05_HD derived neural cells. G, H and I: DNA from a VUB05_HD derived teratoma.

Figure 7 shows the results for the OPL derived cells of VUB03_DM1 and VUB19_DM1. The undifferentiated hESC were considered as passage 0, in which a large variety of alleles could be seen. Within two passages, most of the alleles were lost due to the clonal and focal growth of the culture during these first two passages. Morphologically, the culture becomes homogeneous at passage 3 (Mateizel et al., 2008), which coincides with the genetic homogenization. Through the subsequent passages, no significant repeat instability was detected.

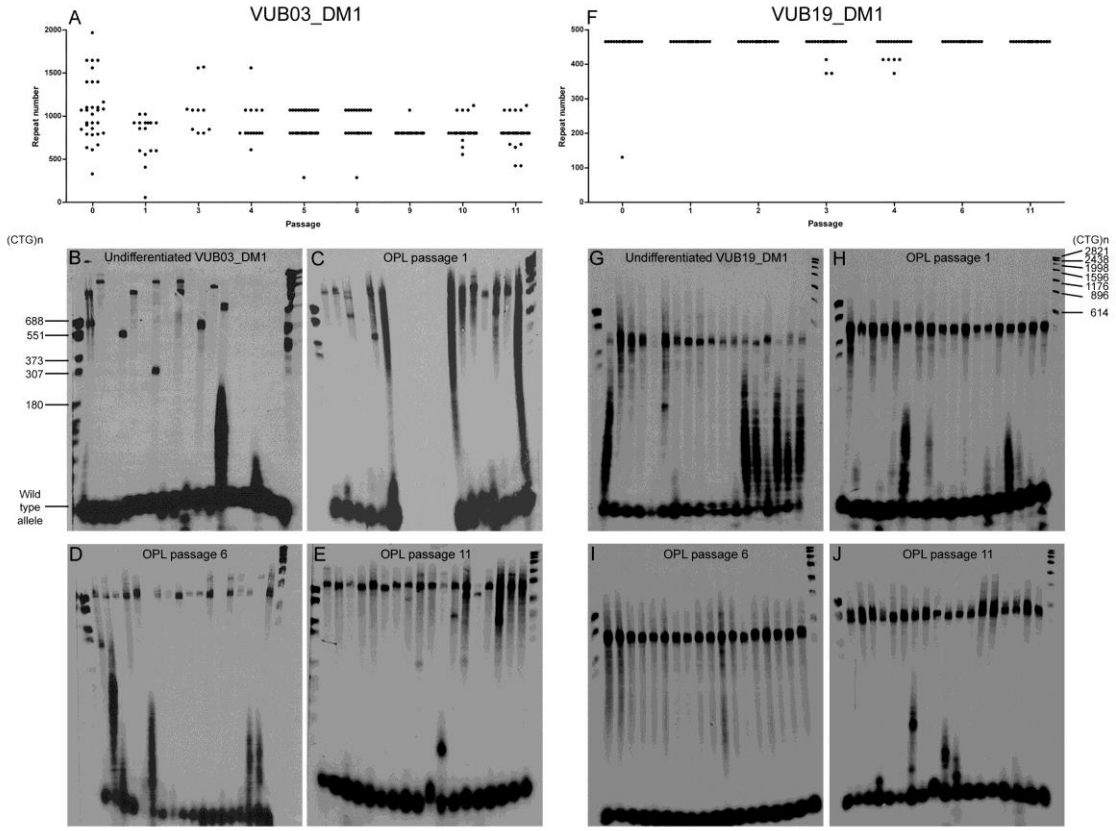


Figure 7. CTG repeat stabilization during OPL cell differentiation of the hESC lines VUB03_DM1 (panels A-E) and VUB19_DM1 (panels F-J). A: CTG allele representation for different passages of the differentiation of VUB03_DM1. Each dot represents an allele, as detected by small-pool PCR. A variety of alleles can be seen at passage 0, which are lost upon differentiation, leaving a population with predominantly two alleles. B: Small-pool PCR and Southern blot results for passage 0 (undifferentiated VUB03_DM1 hESC passage 88), C: OPL differentiated cells passage 1, D: passage 6 and E: passage 11. The different lanes show replicates of small-pool PCR, and each blot contains the molecular weight markers VI and VII at the first and last lanes, respectively. The marker has been converted to its equivalent in (CTG)n to simplify the interpretation. A large variety of alleles can be seen for passages 0 and 1, whereas in passages 6 and 11 there are two predominant alleles. F: Allele representation of the differentiation of VUB19_DM1. G: Small-pool PCR and Southern blot results for passage 0 (undifferentiated VUB19_DM1 hESC passage 54), H: passage 1, I: passage 6 and J: passage 11. The starting hESC population (passage 0) was homogeneous, and no new alleles appear during the differentiation process.

Since we hypothesized that the HD repeat could be unstable in neural progenitors derived from hESC, we tested by small-pool PCR the DNA obtained from a teratoma formed by VUB05_HD (which should contain a mix of cell types, including neural progenitors) in a SCID mouse and DNA from different passages of neural progenitor cells, obtained by directed in vitro differentiation of VUB05_HD. No change in repeat size was detected in any of the samples analyzed (Figure 6, D to I).

5.2.3 Gene expression of MMR

The semi-quantitative study of gene expression of the MMR genes *hMSH2*, *hMSH3*, *hMSH6*, *hMLH1*, *hMLH3*, *hPMS1* and *hPMS2* showed a statistically significant down-regulation of each of their transcripts upon differentiation into OPL cells (results for VUB03_DM1 in figure 8A, results for VUB05_HD were analogous and can be found in the supplementary data). For VUB03_DM1 MMR expression was significantly different between undifferentiated and differentiated cells ($P < 0.001$, one-way ANOVA). Subsequent pair-wise comparisons indicated that expression in undifferentiated cells was significantly higher than the differentiated stages p1 to p11 ($P < 0.001$, Bonferoni test). No significant differences were observed among the various stages post-differentiation. Similar findings were observed for VUB05_HD ($P<0.001$)

These transcript levels suggest that MMR protein levels are also reduced. However, the protein levels of several MMR factors, particularly the relative levels of *hMSH2*, *hMSH3* and *hMSH6*, does not always reflect the level of the transcript, but is dependent upon protein-protein interactions (Chang et al., 2000; Panigrahi et al., 2010). To this end we assessed the levels of MMR proteins.

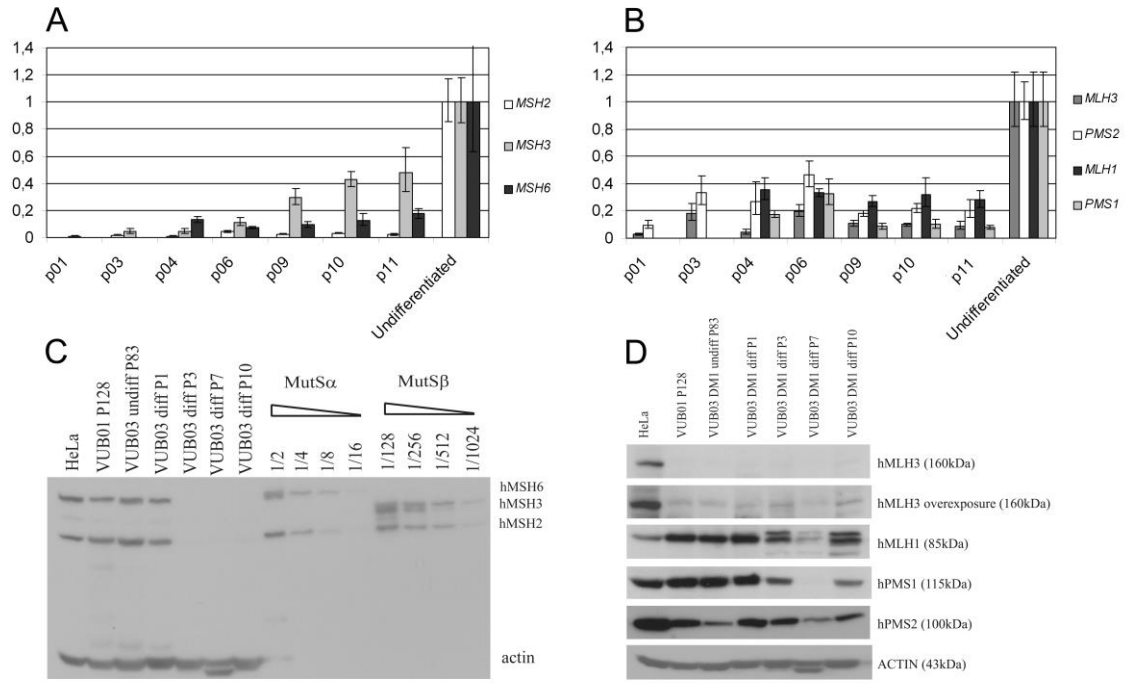


Figure 8. MMR expression in the VUB03_DM1 hESC line in the undifferentiated and differentiated state. A: relative quantification of the gene transcript expression of *MSH2*, *MSH3* and *MSH6*. B: relative quantification of the gene transcript expression of *MLH1*, *MLH3*, *PMS1* and *PMS2*. In panels A and B, the error bars show intra-experimental error. The expression levels of all genes are statistically significantly lower in all passages during OPL differentiation as compared to the undifferentiated cells ($P < 0.001$). For VUB03_DM1, the one-way ANOVA indicated that there was a statistically significant difference between the samples from differentiated and undifferentiated cells, with an $F_{7,24}$ value of 19.63, and a corresponding P value of less than 0.001. Subsequent pair-wise comparisons using the Bonferroni test indicated that the mean relative quantification value of the undifferentiated cells was significantly higher than the differentiated categories p1 to p11. No significant differences were observed among the mean relative quantification values of the differentiated categories. Similar findings were observed for VUB05_HD after one-way ANOVA ($F_{10,33}$ value of 10.25, and $P<0.001$ and subsequent Bonferroni test procedures. C: Western-blot results for hMSH2, hMSH3 and hMSH6. A dilution series of hMutSα and hMutSβ purified proteins are included to equalize antibody signal of their constituent protein subunits. The reproducibility is evident in replicates of the Westerns, and darker exposures (Supplementary Figure 3). D: Western-blot results for hMLH3, hMLH1, hPMS1, hPMS2. At post-differentiation passage P10, the levels of protein do not actually go up, but a larger amount was loaded onto the blot. This can be seen by comparison to the intensity of the actin band.

5.2.4 Mismatch repair protein expression

The undifferentiated control (VUB01) and DM1 (VUB03_DM1) hESC lines show similar levels of hMSH2, hMSH3 and hMSH6; comparable to the well-characterized MMR-proficient HeLa cell line (Figure 8C), these expression changes are highly reproducible (Supplementary Figure 3). Protein levels did not vary with passage number of the undifferentiated hESCs (data not shown). Upon differentiation of the hESCs the protein expression remained relatively unchanged at the first passage following differentiation (VUB03_DM1 P1). However, in later passages of the differentiated VUB03_DM1 (P3, P7 and P10) hMSH2, hMSH3, and hMSH6 proteins are not detectable (Figure 8C; Supplementary Figure 3). This result is coincident with gene expression profiles obtained. The delay in protein loss, relative to transcript down-regulation, is likely due to the time for protein turnover. This analysis reveals that the levels of these MMR proteins do not vary with passage number of the undifferentiated hESCs but dramatically drops upon differentiation.

The levels of hPMS1, hPMS2, hMLH1 and hMLH3 were also assessed by Western blotting (Figure 8B). The levels of hMLH3 were low in the hESCs, as previously noted for most cell lines and tissues (Cannavo et al., 2005; Charbonneau et al., 2014), with higher levels in HeLa extracts (Cannavo et al., 2005). The levels of hPMS1, hMLH1 and hMLH3 did not vary in the control and DM1 undifferentiated hESCs and the earliest differentiated passage, P1. A subsequent reduction in hPMS1 and hMLH1 levels was observed in later differentiated passages (P3, P7 and P10). Curiously, hMLH1 showed a triplet banding pattern in the post-differentiation passages, possibly due to degradation or post-translational modification. The level of hPMS2 protein was somewhat higher in the control hESC lines compared to the DM1 hESC, and this level was mildly increased in the earliest and diminished in later post-differentiation passages. Together these results suggest that the levels of many of the MMR genes are down-regulated at the transcriptional level and proteins are degraded to varying degrees.

5.3 DISCUSSION

In DM1 hESCs the expanded DM1 CTG repeat is highly unstable from very early passages onward, and is apparent both between different passages and between cells within the same colony. When comparing the instability results of VUB03_DM1 (De Temmerman et al., 2008), with those presented herein for VUB19_DM1 and VUB24_DM1, it is apparent that, although all lines show some levels of instability from

the beginning, at later passages VUB03_DM1 and VUB24_DM1 present a much broader range of alleles than VUB19_DM1, with very large size changes. On the other hand, VUB19_DM1 shows a progressive increase in size for the most common allele in the cell culture, without such large and sudden changes. Furthermore, while VUB03_DM1 and VUB19_DM1 show a trend towards an increase in repeat size, VUB24_DM1 seems to present rather a tendency towards contractions. Possible factors influencing these differences include the differences in the size of the original expansion, the different 'parental origins' of the lines (the affected donor of the embryos from which these lines were derived were female in the case of VUB03_DM1 and VUB24_DM1, and male in the case of VUB19_DM1), or the genetic backgrounds of each line. Regarding the tendency for DM1 CTG contractions, rare large transmitted contractions have been observed in some DM1 families particularly when large alleles are paternally transmitted (T Ashizawa, 1994), which contrasts with the very large expansions that often occur by maternal DM1 transmission.

Since hESC are derived from preimplantation embryos, which have been shown to have a stable CTG repeat (Dean et al., 2006; TEMMERMAN et al., 2004), one would expect the TNR to be stable in hESC. The results presented here suggest that hESC do not always behave like the inner cell mass cells they are derived from. One explanation for these counter-intuitive findings is that the *in vitro* culture could be influencing the stability of an already expanded repeat. Repeat instability has been observed in other *in vitro* cultures of diverse cell types such as fetal tissues (Wohrle et al., 1995; Yang et al., 2003) and lymphoblastoid cell lines (Ashizawa et al., 1996; Khajavi et al., 2001). In our hESC lines we observed the same two types of mutations as previously described (Ashizawa et al., 1996; Goula et al., 2009; López Castel et al., 2009; Yang et al., 2003), namely frequent small changes of the repeat size and more rare large changes, although the frequency of these changes differed from line to line.

Another explanation could be that this unstable behavior is not culture related, but typical for stem cells. For instance, it has been recently shown that the CGG repeat causing fragile X syndrome has a very particular behavior in a male hESC line carrying a large CGG expansion (Eiges et al., 2007; Shelbourne et al., 2007). The repeat is unstable and unmethylated in the undifferentiated state, and stabilizes and becomes methylated upon differentiation; consistent with experimental systems revealing a role of CpG methylation upon CCG instability (Kennedy, 2003; Nichol Edamura et al., 2005). Moreover, in our work, the stem cells display a similar pattern of stabilization of the expansion upon differentiation. The analysis of the CTG repeat in the OPL cells derived from VUB03_DM1 and VUB19_DM1 showed that the repeat undergoes

stabilization, as compared to the undifferentiated source hESCs. After a few passages, only a few allele sizes remain observable in the culture, and no new allele lengths tend to appear. The stabilization of repeat instability with the arrest of proliferation coincident with differentiation may indicate a link to cell growth (Cleary et al., 2010) or variations of *trans*-factors.

Since several publications point towards the involvement of the MMR proteins Msh2, Msh3 and Pms2 in the instability of large CTG/CAG tracts in transgenic mice (Dragileva et al., 2009; Foiry et al., 2006; Okita et al., 2007) we tested the evolution of the relative gene expression of these genes in the different hESC passages. We found that the cultured hESCs expressed many of the MMR proteins, as previously reported by large-scale expression microarray gene expression in hESC (Manley et al., 1999). We show that upon differentiation of control DM1- and HD-derived hESCs the cells quickly lose expression of MMR genes. Interestingly, this reduced expression is concomitant with the stabilization of the DM1 repeat. At the protein level, the reduction was most dramatic for the hMS2, hMSH3, and hMSH6 proteins, the former two which are known to dramatically contribute to CTG/CAG instability in transgenic mice. The down-regulation of MMR protein expression and the coincident loss of high levels of CTG instability present in the DM1 undifferentiated hESCs is consistent with the current model that indicates a requirement of MMR proteins for the instability of large CTG/CAG tracts (Dragileva et al., 2009; Foiry et al., 2006; Gomes-Pereira et al., 2004; Savouret, 2003; Tomé et al., 2009; van den Broek et al., 2002). Direct evidence supports altered TNR instability can be induced by variations in the levels of DNA repair proteins (Goula et al., 2009; López Castel et al., 2009). Our observation is particularly relevant as it is the first time a correlation between altered repeat instability is seen in cells undergoing a pseudo-natural biologically-induced reduction in MMR expression.

In contrast to the results for the DM1-derived hESC lines, the HD repeat of VUB05_HD proved very stable in all studied conditions. This finding is similar to the *in vivo* behavior of the HD repeat, where only germ cells and brain tissue present limited levels of instability (Shelbourne et al., 2007; Kennedy, 2003). Higher levels of somatic instability arise in various tissues of DM1 patients, but not in HD patients. Two explanations for this difference have been put forward: the HD CAG/CTG expansion is considerably shorter than the DM1 CTG/CAG expansion making it less prone to mutation. Secondly, the DM1 locus may harbor a stronger *cis*-element that drives its instability compared to the HD locus (Cleary et al., 2010). Data favoring this is the much higher levels of instability observed in HD transgenic mice that is likely due to

the larger CAG expansions in the transgenes relative to those present in HD patients (180-225 repeats compared to <80 repeats) (Wheeler, 1999). Various attempts have failed to observe high levels of CAG instability in patient-derived HD cells (Cannella et al., 2009) and only limited instability was observed in cells derived from transgenic HD mice with >155 repeats (Messer et al., 1999). The absence of instability of the (CAG)44 HD tract in our HD-hESCs may be due to either the short expansion length, insufficient culture time, and/or the absence of a strong *cis*-element. Recently, Niclis and collaborators (Niclis et al., 2009) reported the same CAG stability in undifferentiated hESC cells carrying HD expansions of (CAG)37 or (CAG)51. On the other hand, they detected low-level instability in neurospheres derived from these hESC lines. In our case, it is probable that none of the cells we used as differentiated models were close enough to the *in vivo* cell type to mimic the somatic instability of this particular repeat. Another possibility is that the cells were not kept in culture long enough to mimic the age-related gains in length in HD neural tissue.

In summary, in this study we have shown that the expanded CTG/CAG tract is highly unstable in DM1- but not HD-derived hESCs. Differences exist between different hESC lines in the general behavior of the CTG expansion, some lines showing a preference for expansions whereas others show predominantly contractions. This trait may be variable, and it cannot be excluded that other hESC lines carrying these mutations behave differently. In our hands, upon differentiation, the DM1 repeat looses instability, and this process occurs at the same time as a loss of expression of the MMR machinery. This is the first demonstration of the correlation between a biologically natural down-regulation of the MMR proteins and changes in the stability of an endogenous disease locus.

5.4 MATERIALS AND METHODS

5.4.1 Human embryonic stem cell culture and differentiation

For this study, we used the hESC lines VUB03_DM1, VUB19_DM1 and VUB24_DM1 carrying DM1, and VUB05_HD carrying HD, all of them derived in our laboratory. The hESC lines are registered at the European Human Embryonic Stem Cell Registry (www.hescreg.eu) and are available upon request. The hESC were derived from human pre-implantation embryos that were donated for research after *in vitro* fertilization and/or pre-implantation genetic diagnosis treatments. All donor couples gave informed consent and the local ethical committee as well as the Belgian Federal

Commission for Embryo Research approved the different studies involving the derivation and study of the hESC. Derivation, culture, cryopreservation and characterization of their stemness properties were performed as described by Mateizel and collaborators (Mateizel, 2005). Details on the culture conditions can be found in the supplementary experimental procedures.

Teratomas were obtained by injection of undifferentiated hESC into the rear leg muscle of SCID-beige mice (Mateizel, 2005). During teratoma formation, the hESC differentiate into tissues of the three germ cell layers: endoderm, ectoderm and mesoderm. The presence of these three layers was investigated in the teratomas to prove the pluripotency of the hESC lines used in this work. DNA was extracted from the teratoma formed by VUB05_HD to study the behavior of the HD repeat.

The hESC lines VUB03_DM1, VUB05_HD and VUB19_DM1 were differentiated into osteogenic progenitor-like cells (OPL)(Mateizel et al., 2008). Briefly, the hESCs were collected using collagenase type IV (Invitrogen, Carlsbad, CA, USA), and the cell clumps re-suspended in differentiation medium and plated on 0.1% gelatin-coated plastic dishes (passage 1). The differentiation medium contained Knockout Dulbecco's modified Eagle's medium, 2 mmol/L L-glutamine, 1% non-essential amino acids, 0.1 mmol/L β-mercaptoethanol and 20 % heat-inactivated fetal calf serum (Gibco, Invitrogen). The differentiation medium was changed every 2 days, and after 24 days in culture, the cells were passaged and maintained in the same conditions. When possible, during passaging, a sample was collected for DNA and for RNA extraction.

DNA and RNA samples from neural progenitors derived from VUB05_HD were kindly provided to us by the group of M. Peschanski, I-stem, Evry, France (Aubry et al., 2008).

5.4.2 Analysis of TNR instability

Details on all PCR and detection procedures can be found in the supplementary experimental procedures. The amplification of the CTG repeat in *DMPK* was performed as previously described (De Temmerman et al., 2008). Passages 1 and 3 of the hESC line VUB19_DM1 were studied using fragments of colonies. From passage 4 of VUB19_DM1 onwards, and for all samples of VUB24_DM1, as well as for all samples from differentiated cells, analysis was done by small-pool PCR on 200 pg of DNA (between 4 and 27 replicates). The number of CAG repeats in the *HTT* gene was studied in all DNA samples by small-pool PCR (at least 16 replicates) using the primer set and PCR conditions described by Sermon and collaborators (Sermon et al., 1998).

Fragment size of the wild type allele of the CTG in *DMPK* and of both alleles of the CAG repeat in *HTT* was established using an ABI PRISM 3100 - Avant Genetic analyser (Applied Biosystems, Life technologies, Carlsbad, CA, USA). The expanded CTG repeats in *DMPK* were detected by Southern blot using a $(CAG)_5$ probe.

5.4.3 Gene transcript expression analysis by Real-time-PCR

Real-time-PCR (RT-PCR) was used to quantify the mRNA levels of *hMSH2*, *hMSH3*, *hMSH6*, *hPMS1*, *hPMS2*, *hMLH3* and *hMLH1* in undifferentiated and differentiated cultures from VUB03_DM1 and VUB05_HD. The RT-PCR reactions were performed on RNA obtained from undifferentiated hESCs and OPLs. Cells were collected using collagenase type IV or trypsin-EDTA (Gibco, Invitrogen) and RNA was extracted using the RNeasy micro or RNeasy mini kits from Qiagen, including a DNAse treatment and according to the instructions of the manufacturer (Qiagen, Venlo, The Netherlands). cDNA was synthesized using the first-strand cDNA synthesis kit, following the protocol provided by the manufacturer (GE healthcare,).

RT-PCR reactions were performed in a final volume of 25 µl containing 1x TaqMan Universal PCR Master Mix (Life technologies), 900 nM of the forward and reverse primers (Eurogentec, Seraing, Belgium) and 250 nM of the probe (Applied Biosystems, Life technologies), or a Taqman assay (Applied Biosystems, Life technologies). Sequences and assay numbers can be found in the supplementary information. RT-PCR was performed on the Sequence Detector System ABI-Prism 7500 (Applied Biosystems, Life technologies) under the following conditions: 2 min at 50°C and 10 min at 95°C, followed by 40 cycles of 15 s of denaturation and 1 min of annealing/extension at 60°C. For each test, two endogenous controls were used (*GAPDH* and *UBC*) and mouse cDNA samples and no-template control were taken along.

Data analysis was done using the 7500 System SDS Software (Applied Biosystems, Life technologies). Relative quantification of gene expression was achieved by normalization against the endogenous control *GAPDH* or *UBC* using the ΔΔCt method. Fold changes were calculated as $2^{-\Delta\Delta Ct}$.

5.4.4 Protein expression analysis by Western blotting

Western blot for hMSH2, hMSH3 and hMSH6 was performed as follows. Thirty µg of cell lysate, and multiple dilutions of hMutSα and hMutSβ were separated at 170 volts on an 8% SDS-PAGE. Proteins were transferred to a PVDF membrane at 300mAmps.

The membranes were blocked and then simultaneously probed overnight at 4°C in primary antibodies anti-MSH3 (BD Transduction Laboratories, San Jose, CA, USA, cat.#611390), anti-MSH2 (Calbiochem, Merck KGaA, Darmstadt, Germany, cat.# CA8000-852), anti-MSH6 (BD Transduction Laboratories cat.# 610919) and anti-actin (BD Transduction Laboratories cat.#612656). Immunoblots were washed briefly and incubated in HRP-conjugated sheep anti-mouse secondary antibody. Chemiluminescent signals were generated using GE Healthcare ECLTM Western Blotting Detection Reagent (GE Healthcare, Uppsala, Sweden). Images were captured on Denville HyBlot film (Denville Scientific Inc., NJ, USA) with multiple exposures.

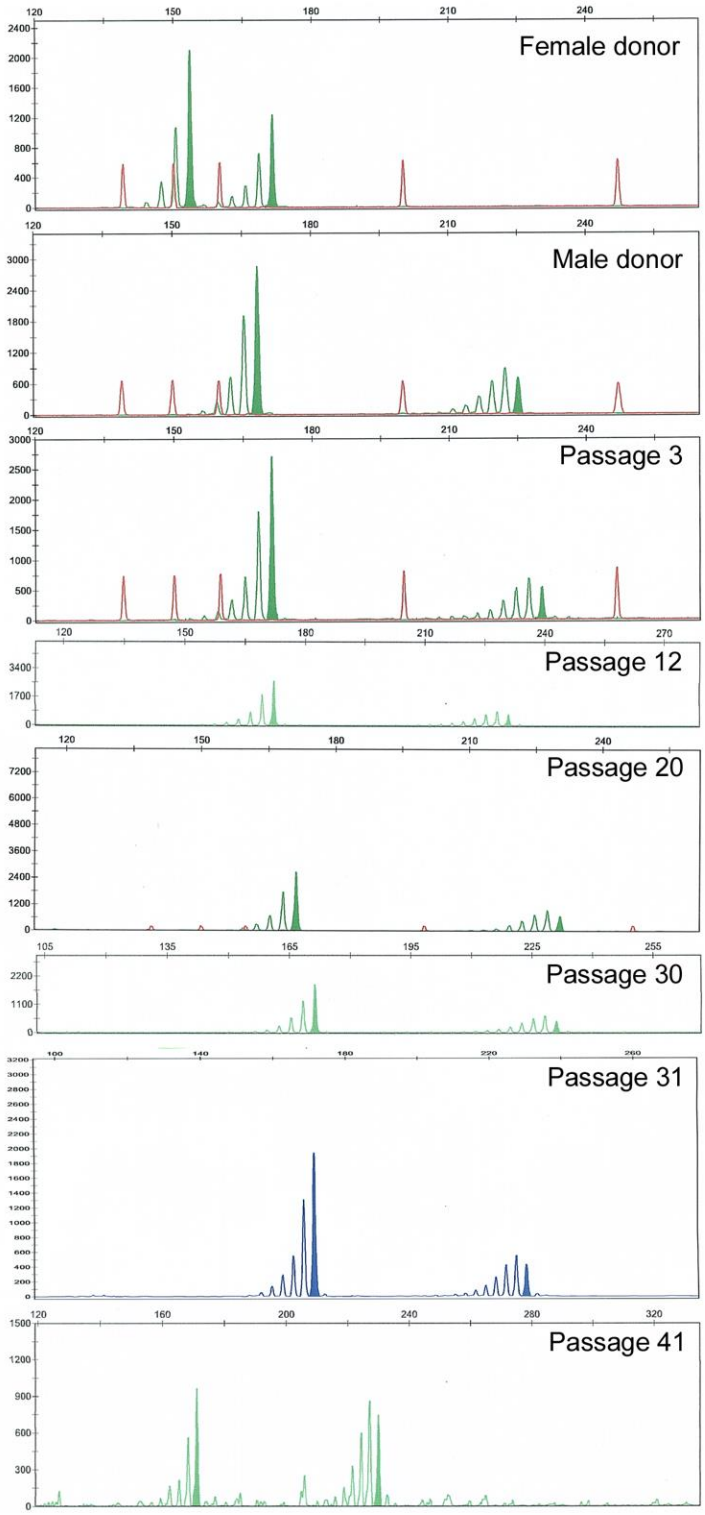
For hPMS1, hPMS2, hMLH1 and hMLH3, 30 µg of each cell lysate was separated and transferred as above, membranes were probed independently with anti-hPMS1 (Santa Cruz, Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA, cat.# SC-615), anti-hPMS2 (BD Transduction Laboratories cat.# 556415), anti-hMLH1 (BD Transduction Laboratories cat.#554073) and anti-hMLH3 (Santa Cruz cat.# SC-25313). Membranes were washed briefly and incubated in HRP-conjugated secondary antibodies. Chemiluminescent signals were generated using GE Healthcare ECLTM Western Blotting Detection Reagent. Images were captured on Denville HyBlot film with multiple exposures

5.4.5 Statistical analysis

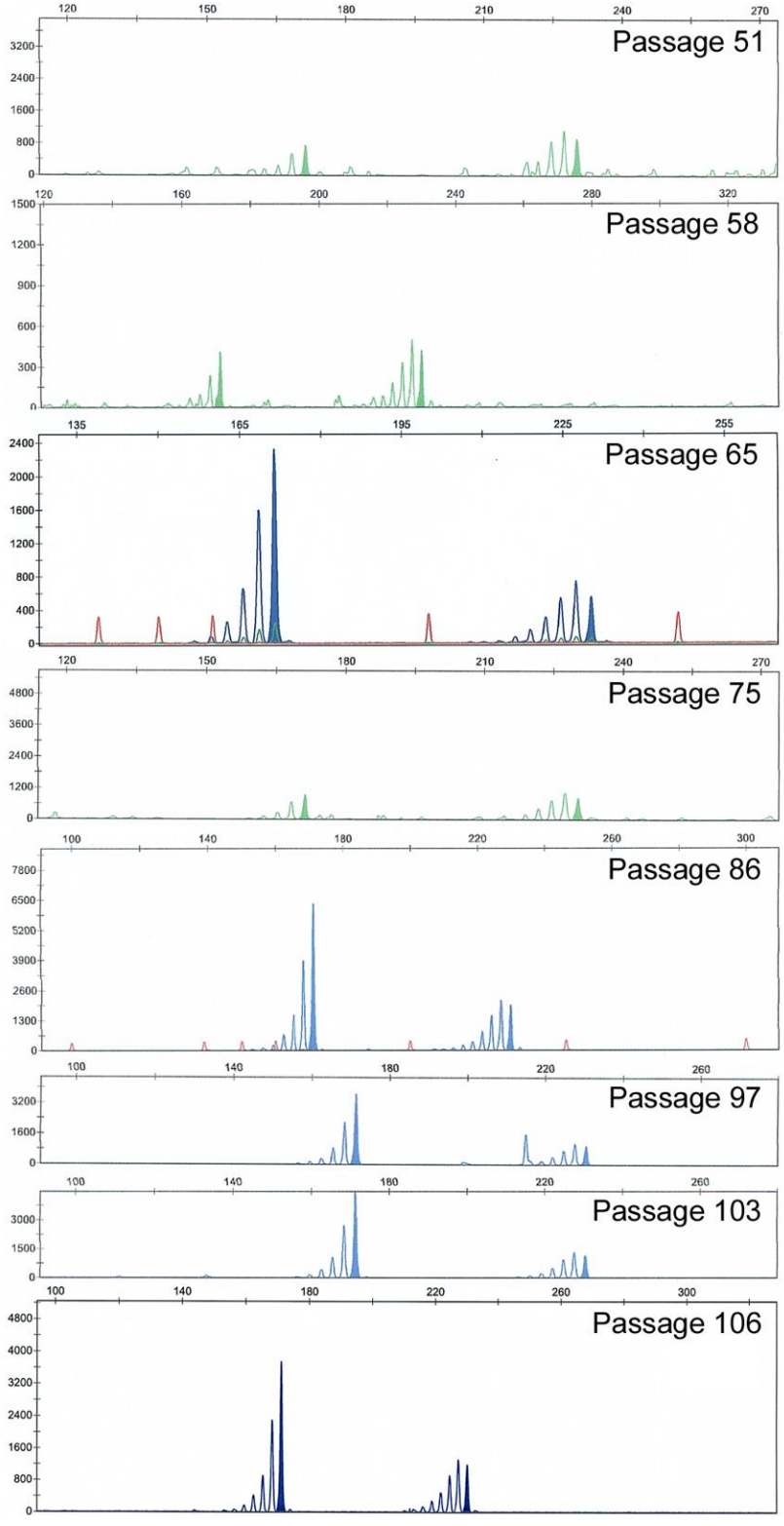
A multiple regression analysis was conducted to explore the potential relationship between passage number (X, independent variable) and repeat size in the *DMPK* gene (Y, dependent variable). The magnitude of the strength of this relationship was quantified by the standardized regression coefficient (labeled Beta). Tests were two-tailed, and a P-value of less than 0.05 was considered to indicate statistical significance.

In the case of the gene expression study, one-way analysis of variance (ANOVA) was used to test the null hypothesis that the mean relative quantification values are the same for all categories of samples, i.e., the differentiated samples labeled p1 to p11 and the undifferentiated sample. The Bonferroni test procedure was used for subsequent multiple comparisons. One-way ANOVA and Bonferroni testing were conducted, in turn, for VUB05_HD and VUB03_DM1.

5.5 SUPPLEMENTARY DATA



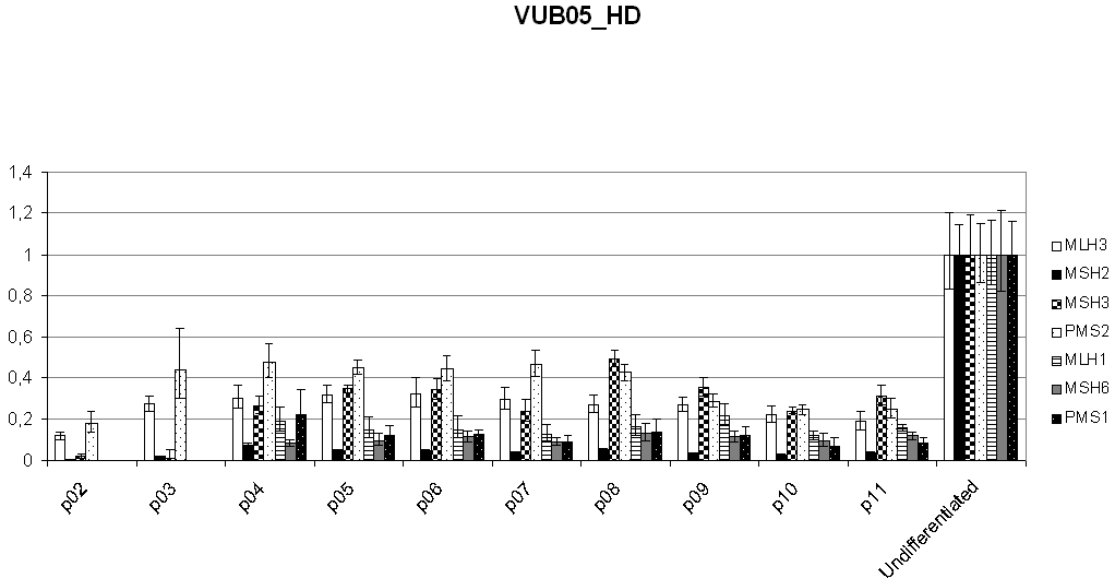
Supplementary Figure 1a



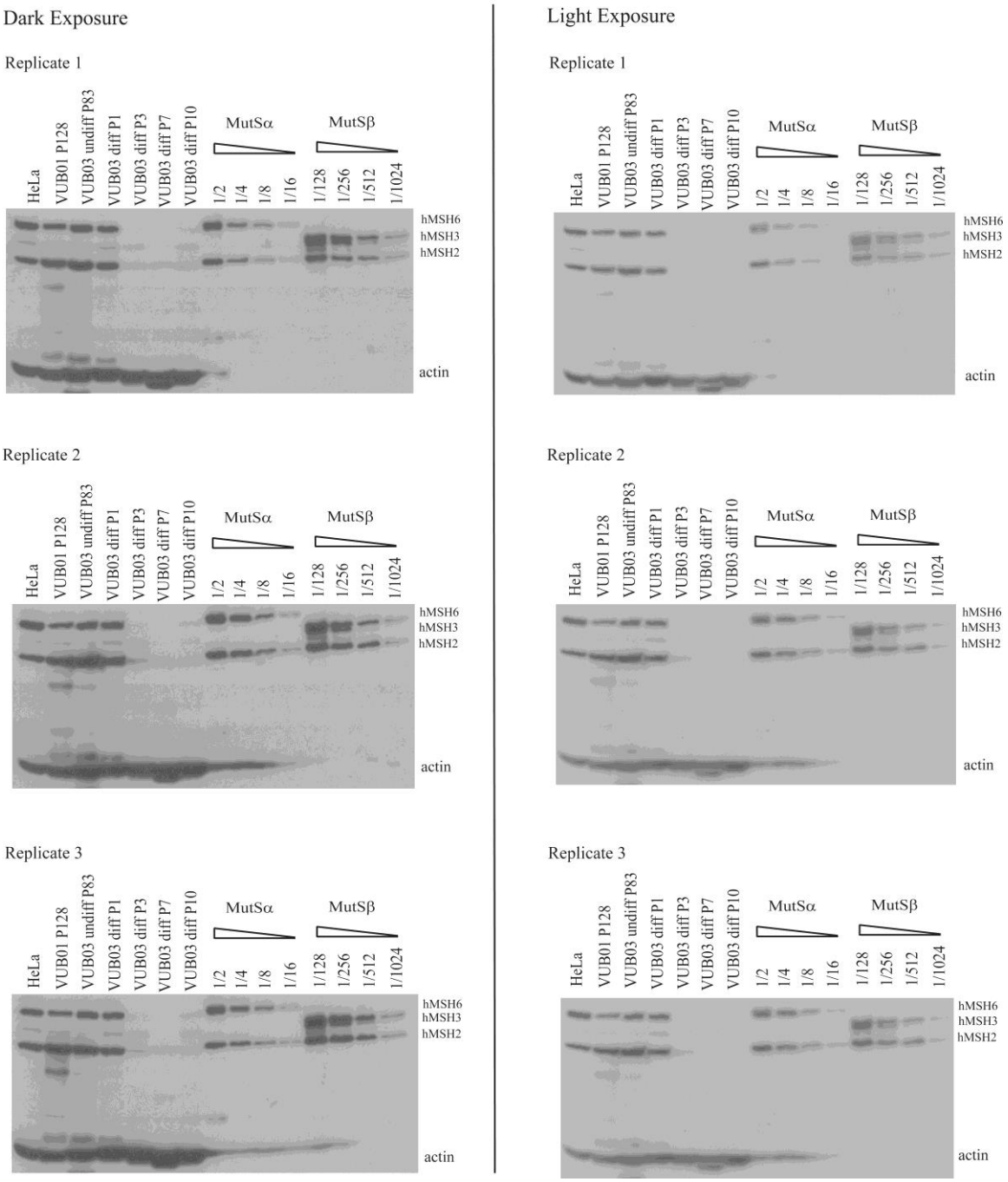
Supplementary Figure 1b

Supplementary figures 1a and 1b. HD repeat during long-term culture of the hESC line VUB05_HD. The analysis of the CAG repeat in *HTT*, in passages 3, 12, 20, 30, 31, 41, 51, 58, 65, 75, 86, 97, 103, 106 of VUB05_HD, showed a stable genotype of 23 and 44 repeats. Supplemental figures 1a and 1b show the results of the PCR for the HD repeat for all these passages, in addition to the genotypes of the donors of the embryo used to derive the line. Every panel of the figure is an out print of an automated

sequencer. Since these analysis were performed over a timespan of 5 years, not all out prints are in the same scale. Nevertheless, the alleles shown are always of 23 and 44 repeats.



Supplementary figure 2. Gene expression of the MMR machinery genes by relative quantification. Results of the relative quantification of the gene expression of MSH2, MSH3, MSH6, MLH1, MLH3, PMS2 and PMS1 during the OPL cell differentiation of VUB05_HD. The error bars show intra-experimental error. The results for all passages as compared to the undifferentiated cells are statistically significantly different ($P < 0.001$).



Supplementary figure 3. Technical replicates of the MMR protein expression study by Western-blot for hMSH2, hMSH3 and hMSH6. The analysis of the MMR protein expression is highly reproducible. A dilution series of hMutSα (hMSH2-hMSH6) and hMutSβ (hMSH2-hMSH3) purified proteins are included to equalize antibody signal of their constituent protein subunits. We show three replicate Western blots of extracts from control, undifferentiated and differentiated cells. On the left are the dark exposures and on the right are lighter exposures of the same blots. At post-differentiation passage P10, the levels of protein do not actually go up, but a larger amount was loaded onto the blot. This can be seen by comparison to the intensity of the actin band.

5.6 SUPPLEMENTARY EXPERIMENTAL PROCEDURES

5.6.1 Human embryonic stem cell culture and differentiation

The VUB hESC lines were derived from mechanically isolated inner cell masses of human blastocysts. All lines were derived and cultured on 0.1% gelatin-coated culture dishes (Sigma-Aldrich, Schnelldorf, Germany) containing mitomycin C inactivated CF1 mouse embryonic fibroblast feeders. The cells were kept 37°C and 10 % CO2 in medium consisting of 80 % Knockout Dulbecco's modified Eagle's medium (KO DMEM; Invitrogen, Carlsbad, CA) supplemented with 20 % KO-serum replacement (KO SR, Invitrogen), 2 mmol/l L-glutamine (Invitrogen), 1 % non-essential amino acids (Invitrogen), 0.1 mmol/l β-mercaptoethanol (Sigma-Aldrich) and 4 ng/ml human recombinant basic fibroblast growth factor (Invitrogen). Cells were mechanically passaged every 5 to 7 days and medium was changed daily. All hESC lines were characterized as such by standard tests (Mateizel, 2005).
VUB05_HD and VUB03_DM1 were kept in culture for over 100 passages and VUB19_DM1 and VUB24_DM1 for over 50 without any apparent senescence of the cells. Karyotypical stability of the lines was checked by array-based comparative genomic hybridization (Spits et al., 2008). VUB03_DM1 presented a normal karyotype at passage 30, and a duplication of 20q11.21 at passage 98. VUB05_HD presented a stable fully normal karyotype up to passage 106, when the experiments were concluded. VUB19_DM1 presented a normal karyotype at passages 5 and 32; and so did VUB24_DM1 at passages 10 and 31.

5.6.2 Analysis of TNR instability

DNA was phenol/chloroform extracted from hESC colonies detached using collagenase type IV. Briefly, the cells were lysed using 0.2 mg/ml proteinase K (Merck, Brussels, Belgium) and 1% SDS, and incubation at 45°C for 2 hours. Genomic DNA was extracted twice by phenol and twice by 24:1 chlorophorm:2-propanol. The DNA was ethanol-precipitated and resuspended in sterile milliQ water.

The CTG expansion in DMPK was amplified using a specific long PCR protocol. The PCR was performed in a total volume of 50 µl, containing 0.3 µM primers DM101 (6-FAM 5' CTT CCC AGG CCT GCA GTT TGC CCA TC 3') and DM102 (5' GAA CGG GGC TCG AAG GGT CCT TGT 3'), 500 µM of each dNTP (DNA Polymerisation Mix, GE Healthcare, Roosendaal, The Netherlands), 1x Expand Long Template (ELT) PCR buffer

2 provided by the manufacturer (Roche Diagnostics, Vilvoorde, Belgium), 2.5 % dimethylsulfoxic oxide (DMSO, Sigma Aldrich), 2 % Tween 20 (Sigma Aldrich) and 1.4 U DNA Polymerase from the ELT kit (Roche Diagnostics). The PCR programme consisted of a first denaturation step at 95° C for 5 min, followed by 10 cycles at 95° C for 30 s, 65° C for 30 s and 68° C for 45 s, followed by 30 cycles at 95° C for 30 s, 65° C for 30 s and 68° C for 45 s, with an increase of 20 s per cycle for the elongation step at 68° C.

The PCR products were separated by a denaturing agarose gel electrophoresis on a 1.5% agarose gel and transferred to a positively charged nylon membrane. For the hybridization, DIG Wash and Block buffer set (Roche Diagnostics) was used. After hybridisation with a DIG-labelled oligoprimer (CAG)5-probe, the repeat sizes could be determined in comparison to the molecular size markers VI and VII (Roche Diagnostics). Fragment size of the wild type allele was established using an ABI PRISM 3100 - Avant Genetic analyser (Applied Biosystems, Life technologies, Carlsbad, CA, USA).

The CAG repeat in the *HTT* gene was amplified by PCR in a final volume of 25 µl containing: 0,2 µM primers (HU4: 6-FAM-5'-ATG GCG ACC CTG GAA AAG CTG ATG A-3'; HU3: 5'-GGC GGC TGA GGA AGC TGA GGA-3'), 200 µM of each dNTP (DNA Polymerisation Mix, GE Healthcare), 5% (Sigma Aldrich), 1X of reaction buffer II (Applied Biosystems, Life technologies), 20 mM $MgCl_2$ (Applied Biosystems, Life technologies), and 1.25 IU Taq polymerase (Applied Biosystems, Life technologies). The PCR programme consisted of a first denaturation step at 96° C for 5 min, followed by 10 cycles at 96° C for 30 s, 70° C for 30 s and 72° C for 30 s, followed by 28 cycles at 94° C for 30 s, 70° C for 30 s and 72° C for 30 s. Fragment analysis was done using an ABI PRISM 3100 - Avant Genetic analyser (Applied Biosystems, Life technologies).

5.6.3 Gene expression analysis by Real-time PCR

Primer sets and assay numbers

GAPDH forward	ATG-GAA-ATC-CCA-TCA-CCA-TCT-T
GAPDH reverse	CGC-CCC-ACT-TGA-TTT-TGG
GAPDH probe	CAG-GAG-CGA-GAT-CC
UBC forward	CGC-AGC-CGG-GAT-TTG
UBC reverse	TCA-AGT-GAC-GAT-CAC-AGC-GA
UBC probe:	TCG-CAG-TTC-TTG-TTT-GTG
MSH2	Hs00179887_m1
MSH3	Hs00989003_m1
PMS2	Hs00416527_m1
MLH3	Hs00271778_m1
MSH6	Hs00264721_m1
PMS1	Hs00922270_g1
MLH1	Hs00179866_m1

6 X-chromosome inactivation

6.1 MOLECULAR BASIS OF DOSAGE COMPENSATION BY X CHROMOSOME INACTIVATION

One of the earliest epigenetic modifications that take place during mammalian development is the gene dosage compensation of the X chromosome. Female embryonic cells contain two X chromosomes and need to achieve the same expression dosage of the X-linked genes than males, whose cells contain a single copy of the X chromosome paired to a Y chromosome (Lyon, 1962). To this purpose, the majority of the genes of one X chromosomes are silenced in a complex genetic process called X-chromosome inactivation (XCI). The XCI happens during pre-, peri- and post-implantation stages of female mammalian embryos and it is organized by a number of regulatory genes located at the X inactivation center of the X chromosome. Most of the information we have on the XCI process has been obtained from the mouse. The inactivation process consists of a series of sequential events, namely: counting and choice, initiation, spreading, establishment and maintenance of X chromosome inactivation during the lifetime of the individual.

The first step the cells need to do in order to determine the amount of X chromosomes is to **count**. The cells are able to establish the ratio of the number of X chromosomes in respect to the number of autosomes. The choice of X chromosome to be inactivated depends on a pairing between the two homologous X inactivation centers of the two X chromosomes that happens very briefly during the initial steps of the process. The chromosomes bind to each other through the X pairing region of the X inactivation center. Only when two X inactivation center are present, a molecular crosstalk happens between them in *trans* that triggers the XCI process. This mechanism is known as **competence or sensing** (Heard et al., 1999). In the presence of a single X inactivation center, a cell cannot detect the presence of two X chromosome and XCI will not be initiated.

The next step of the XCI is the **choice** of the X chromosome to be inactivated. The selected X chromosome will upregulate the expression of the X-inactive specific transcript (*XIST*) gene in *cis.* The expression of *XIST* is a crucial step to **initiate** XCI both in mouse and human. *XIST* encodes for a long non-coding RNA transcript of 17kb that will coat the X chromosome, creating the so called "*XIST* RNA cloud" which will **spread** and cover the entire X chromosome from which *XIST* is expressed (Brown et al., 1991). In parallel, *XIST* induces a progressive chromatin reorganization forming a

silent nuclear compartment isolated from the transcription machinery of the cell and devoid of polymerase II (Chaumeil et al., 2006). *XIST* also recruits the Polycomb Repressive Complex 2 (PRC2) into the silent nuclear compartment. PRC2 will deposit trimethylated histone 3 on lysine 27 (H3k27me3), a chromatin repressive histone mark, which will in turn recruit the Polycomb Repressive Complex 1 (PRC1). PRC1 will ubiquitinate the activating Macrohistone H2A that will contribute to chromatin compaction. The coordinated action of all these factors will **establish** the silencing of the genes in the future inactive X chromosome (Xi). However, a small portion of genes (15%, Carrel and Willard, 2005) will not be silenced and will remain active outside the silent nuclear compartment in somatic cells. These genes are mainly localized, but not exclusively, in PAR regions, which are areas involved in the chromosome pairing during meiosis (Berletch et al., 2010).

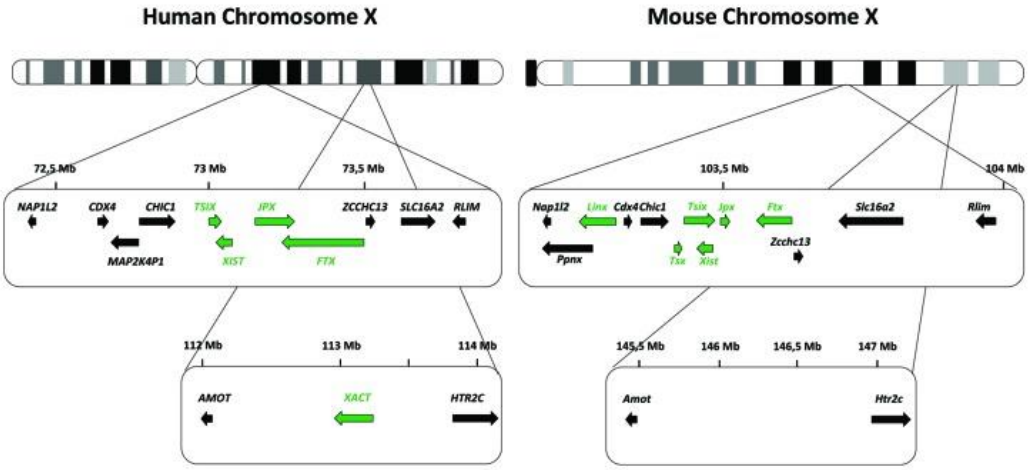


Figure 9. Regulation of *XIST* and Xist expression. This figure shows a comparison between the human and mouse X-chromosomes regulation maps. The upper part of the figure corresponds to the X-inactivation center region, and the lower part includes the *XACT* region (specific for human), according to the RefSeq annotation. Protein-coding genes are coloured in black and genes producing long non-coding RNAs are coloured in green. Image copied from Vallot et al. (Vallot et al., 2013)

The regulation of XIST expression is different in mouse and human (Vallot et al. 2013)(Figure 9). In mouse, the regulation of *Xist* expression is carried out by several loci that acts at the DNA, RNA or protein level. Different X-linked loci function as activators or repressor elements that contribute to the XCI by regulating *Xist* expression. One of these is the Ring finger protein 12 (*Rnf12*), which is a protein that activates the expression of *Xist* in a threshold dependent manner. The *Rnf12* concentration threshold required to induce *Xist* expression requires the gene expression from two active X chromosomes (XaXa). *Rnf12* regulates its own expression by a negative feedback loop, which prevents overexpression of *Rnf12* and

in turn prevents over-activation of *Xist* that could lead to the silencing of the second X chromosome.

Many regulatory elements downstream of *Xist* have been identified in mouse including *Jpx, Tsix, DxPas34, Xite* and *Tsx*, which are repressors of *Xist* expression. *Jpx* codes for a long non-coding RNA that specifically inactivates *Xist* expression in female cells. *Tsix* plays a central role in the mono-allelic expression of *Xist* during the random choice of the XCI initiation. A deletion of the *Tsix* promoter or enhancer produces skewing towards inactivation of the mutant copy. *Tsix* represses *Xist* by coding an RNA that is complementary (antisense) and binds to *Xist*. Therefore, *Tsix* is critical in rodents for the regulation of *Xist* expression and in turn, for XCI. However, murine *Tsix* is very different from human *TSIX* in patterns of expression and extent of transcription across the sense strand. The lack of regulatory elements in the human *TSIX* gene supports the hypothesis of *TSIX* not being a regulator of *XIST* in humans. The regulation of XCI in human is still not well understood.

Few studies have addressed the mechanisms of *XIST* regulation, yielding no conclusive results on the function of human factors such as *TSIX, JPX, FTX* and *RNF12*. Recently, the *XACT* long non-coding RNA has been reported to coat the active X chromosome in human cells. *XACT* expression is restricted to human pluripotent cells and early differentiating cells and it is a feature of eroded XCI. When *XIST* is absent, and the two X chromosomes are active in human cells, a cloud of *XACT* RNA covers both X chromosomes. Therefore, *XACT* is believed to participate in human-specific mechanisms for the regulation of the X chromosome inactivation (O'Leary et al., 2012; Vallot and Rougeulle, 2014).

The *XIST* RNA coating cloud is not enough for the silencing of the X-linked genes. It has been described that *XIST* is necessary to initiate the XCI but not to maintain it throughout the lifetime of the individual (Brown and Willard, 1994), and other genes and processes are involved in XCI. In order to **maintain** this gene silencing during the lifetime of the female, repressive histone MacroH2A1 is incorporated as well as H3K27me3 (Eggan et al., 2000; Wutz, 2011), H2AK119ub, histone H4 lysine 20 methylation (H4K20me1) and hyper-methylation of H3 lysine K9 (H3K9me) together with DNA methylation of CpG islands of the genes subjects to X inactivation (Mermoud et al., 1999).

In summary, once the XCI process is completed, the inactivated X chromosome is distinguishable from the active one by differential transcription of the *XIST* gene, distinct DNA methylation patterns and presence of epigenetic chromatin modifications

such as H3K9me and H3K27me3 foci, H4K20me1 and macroH2A1 (Brockdorff, 2002; Monkhorst et al., 2008).

In mouse, which has been the main model of study of XCI process, the XCI happens following two distinct patterns: the imprinted XCI pattern and the random XCI pattern. Imprinted XCI is characterized by a directed inactivation of one of the two X chromosomes in a parental specific way, whereas random inactivation has no preference for the parental origin of the chromosome to be inactivated.

6.1.1 Imprinted XCI

The imprinted X inactivation has been described in marsupials and in the extra-embryonary tissues of some eutherian mammals and, temporarily, in mouse pre-implantation embryos.

In two-cell stage mouse embryos, the paternal X chromosome (Xp) is preferentially inactivated, from which *Xist* is expressed. *Tsix* in turn, is expressed from the maternal X-chromosome (Xm) keeping the chromosome active. Imprinted Xp inactivation is maintained in the trophoblast and extra-embryonary tissues. Conversely, there is a reactivation of the paternal X chromosome in the ICM of the blastocysts which then present two active X chromosome (XaXa) (Lee and Bartolomei, 2013; Mak et al., 2004; Starmer and Magnuson, 2008). This process is characterized by the progressive loss of the X inactive heterocromatin marks.

The XaXa state of the ICM is short and transient and as the cells from the ICM start to differentiate, the process of random XCI will resume. The primordial germ cells (PGCs) also reactivate the imprinted Xp chromosome in a similar process to the ones undergone by the ICM cells. The X chromosome reactivation in female germ line takes place during the migration of the PGCs due to the inactivation of *Xist* expression, which correlates with the acquisition of pluripotency.

In vitro, X chromosome reactivation has been accomplished by the reprogramming of somatic cells into mouse iPSC (Bacher et al., 2006; Maherali et al., 2007) and by SCNT (Eggan et al., 2000).

6.1.2 Random XCI

The random XCI of either the paternal or the maternal X chromosome takes place in all cells from the epiblast of the ICM. When the epiblast cells start to differentiate, they progressively lose the expression of the pluripotency genes, which induces the mono-allelic expression of *Xist* in a random pattern.

Mammalian females are, therefore, a mosaic with a ≈50:50 ratio of cells that will inactivate the Xm or the Xp. Once the XCI is established, it will be maintained in the progeny of the cell for the lifetime of the individual.

How cells inactivate the X chromosome in a random manner is still unknown. The **stochastic model** proposes that the autosomes produce a *trans*-acting factor that induces *Tsix* expression and subsequent repression of *Xist* (Monkhorst et al., 2008). A second X-linked factor would bind and upregulate *Xist*. The autosomic and X-linked factor would compete for *Xist* and as a result of this competition, stochastically one or the other chromosome would be inactivate in the majority of the cells, and to a lesser extent, some cells would inactivate both or none chromosome.

A more sophisticated model is the **feedback model**, which combines aspects of older models. The feedback model hypothesizes the coordinated action of three major components: a signalling feedback loop, a mechanism for choice, and X-linked loci that directly affect XCI initiation (Starmer and Magnuson, 2008). Other mechanisms have been proposed, but experimental evidence is still needed to validate the actual mechanism that causes a random inactivation of the X chromosome (Augui et al., 2007; Bacher et al., 2006; Masui et al., 2011; Xu et al., 2006).

6.2 TIMING AND MODES OF X-CHROMOSOME INACTIVATION DURING EMBRYONIC DEVELOPMENT

Most of the research carried out on this topic has been done in mouse. In humans, insight into the early mechanisms of XCI has been obtained through a limited number of analyses of human embryos and human embryonic stem cells (hESC), based *in vitro* systems. Figure 10 represents a scheme of the XCI state in mouse and in human.

A. XCI IN THE MOUSE

At 26-29 hours post-fertilization (p.f.) mouse embryos undergo the embryonic genome activation. The *Xist* gene is transcriptionally upregulated at this point and as early as 2-cell stage, the first signs of XCI appear (1.5 day p.f.) The first detectable sign of XCI is the appearance of a small area with *Xist* RNA that surrounds the X chromosome that will be inactivated (Xi) which, in the mouse, will be the imprinted Xp. At around the 4- to 8-cell stage, the *Xist* RNA spreads creating the full "Xist RNA cloud" which will coat Xp, leading to its silencing (Mak et al., 2004; Okamoto et al., 2004).

Later on, at the blastocyst stage, the epiblast cells from the inner cell mass (ICM) start the reactivation of the inactivated Xp chromosome. The cells from the ICM show high levels of pluripotency markers such as *Oct4, Sox2* or *Nanog* (i.e. eighty to eighty-eight per cent of the cells are Nanog positive) (Augui et al., 2011). Those cells will be temporaly XaXa and the *Xist* coating will be absent. Pluripotency markers such as *Oct4*, *Nanog*, *Sox2*, *Prdm14* and *Rex1*, repress the expression of *Xist* through different mechanisms. Among other mechanisms *Oct4* upregulates *Tsix* and *Xite* which are repressors of *Xist* (Donohoe et al., 2009). *Nanog* and *Oct4* can also bind to the intron 1 of the *Xist* gene repressing its expression and contributing to maintain the X chromosome in an activated state. *Sox2* cooperates in the binding of *Tsix/Xite* site that will also maintain the repression of *Xist* expression(Navarro et al., 2011). At this stage, the paternally imprinted Xi is only still found in the trophectoderm (TE) and the pre-epiblast (PrE) cells. The definitive XCI is closely linked to the loss of pluripotency of the cells from the ICM and the start of cell differentiation (Augui et al., 2011). Upon differentiation, those cells of the ICM will again inactivate one X chromosome but this time in a random manner (Schulz et al., 2014).

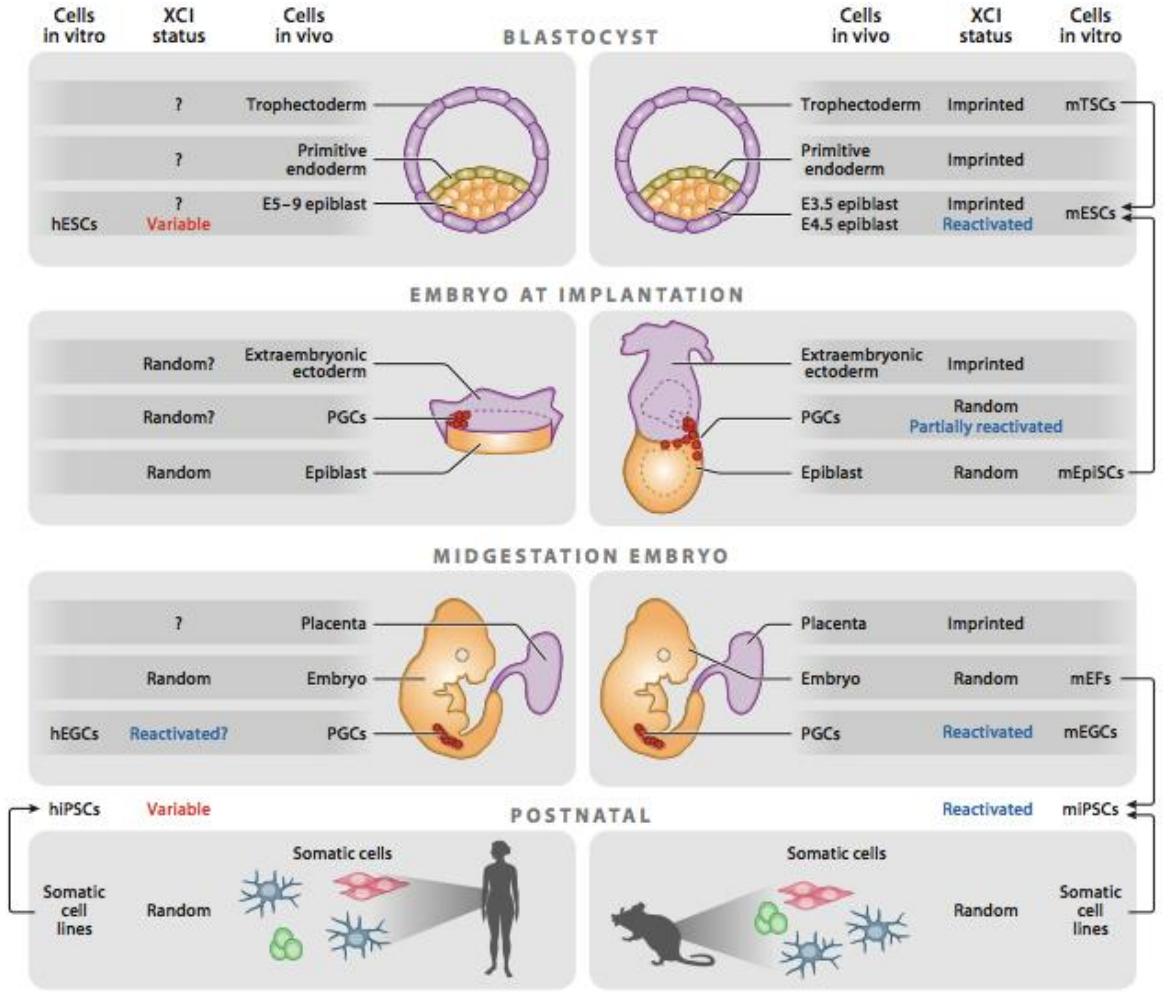


Figure 10. Comparison of mouse and human XCI state in female embryos and PSC. Figure adapted from (Lessing et al., 2013; van den Berg et al., 2009). Black arrows indicate *in vitro* reprogramming. "E" followed by a number indicates days of embryonic development; EF, embryonic fibroblast; EpiSC, epiblast stem cell; ESC, embryonic stem cell; iPSC, induced pluripotent stem cell; TSC, trophoblast stem cell.

B. XCI IN THE HUMAN

Human embryos undergo the embryonic genome activation wave as early as 2-cell stage (1.5 day p.f.) (Teklenburg et al., 2012; Vassena et al., 2011). Single *XIST* foci start to form in each blastomere at the 6- to 8-cell stage. *XIST* is still present in the ICM at the blastocyst stage (Vallot and Rougeulle, 2014; van den Berg et al., 2009). Despite that, these results suggest that human embryos initiate at the 6 to 8 cell stage, they have not completed the XCI process at the blastocyst stage, resulting in a mosaic of XaXa/XaXi cells. Furthermore, whether the ICM cells in human embryos undergo imprinted or random Xi remains to be elucidated (Maherali et al., 2007; van den Berg et al., 2009). However, after implantation the XCI is completed in the majority of the cells (Teklenburg et al., 2012).

6.3 X-CHROMOSOME INACTIVATION IN PLURIPOTENT STEM CELLS

Mouse Embryonic Stem Cells (mESC) have two active X chromosomes (XaXa) in their pluripotent undifferentiated state, because they are derived from the ICM cells which have reactivated the imprinted Xp. Mouse stem cells derived from post-implantation epiblastic cells (mEpiSC) have already inactivated one X chromosome (XaXi) and actively express *Xist*. Similarly to mESC, mouse induced pluripotent stem cells (miPSC) will also show Xi reactivation as one of the last steps of the reprogramming process (Hall et al., 2008; Maherali et al., 2007; Shen et al., 2008; Silva et al., 2008).

Differently, hESC have been shown to be a more complex XCI situation than their murine counterparts. Since hESC are derived from the ICM cells, as are mESC, it was expected that they also were XaXa. Nevertheless, the vast majority of hESC lines present already one inactivated X chromosome (XaXi) and XCI can be detected already at early passages after derivation. To explain these differences it is assumed that, since human embryos present *XIST* expression, transcriptional silencing and repressive chromatin modifications in the ICM cells at the blastocyst stage (van den Berg et al., 2009), the XCI has already taken place, and therefore hESC present the same marks as the cells from which they are derived. In a recent study investigating the transition of the ICM cells towards the first hESC colony, the authors showed the existence of an intermediate state, named Post-ICM intermediate (PICMI). They found that the PICMI was always consistently presenting foci of H3K27me3 (indicative of XCI) in all *OCT4* positive cells. On the other hand, two out of the four hESC cell lines derived from these XaXi PICMIs were negative for H3K27me3, showing a XaXa state. This suggested that hESC can undergo an X-chromosome reactivation during the early stages of derivation (O'Leary et al., 2012).

6.3.1 hESC classification according to their XCI status

Currently, there is general consensus in classifying hESC lines in three classes according to their XCI status and their epigenetic characteristics (Silva et al., 2008) (Figure 11).

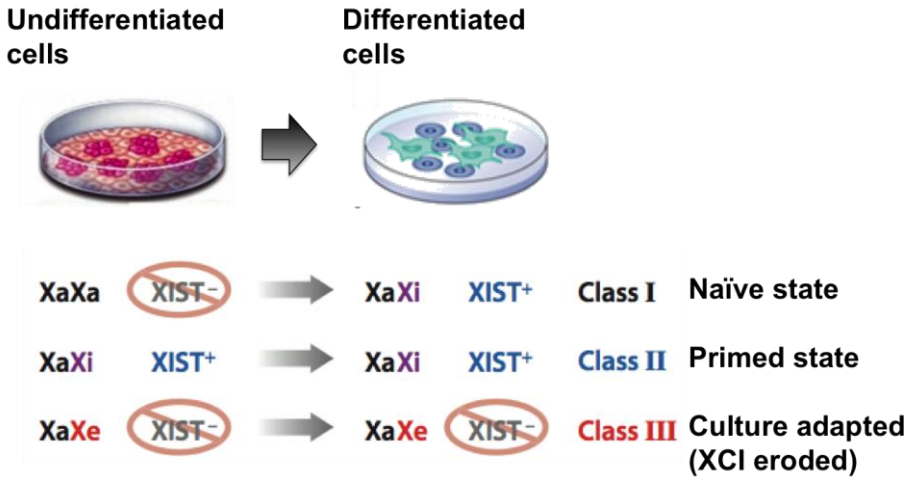


Figure 11. hPSC classification based on XCI status. Xa, X chromosome active; Xi, X chromosome inactive; Xe, X chromosome eroded.

Class I: include those cells that present two active X chromosomes at the pluripotent state (XaXa) and inactivate one of the X chromosomes upon differentiation (XaXi). At the pluripotent state, cells express low or undetectable *XIST* RNA and do not express other XCI markers such as H3K27me3, H4K20me1 or macroH2A1. Class I characterizes the ground state of pluripotency, and stem cells in this state are usually referred to as naïve pluripotent stem cells.

Class II cells show XIST expression in the pluripotent and in the differentiated state, as well as other XCI marks. Primed pluripotent stem cells are included in this class (XaXi).

Cells belonging to **Class III** have lost XIST coating and do not show any other XCI marks. Class III cells do not reactivate *XIST* expression upon differentiation (Shen et al., 2008; Silva et al., 2008). Cells from class III present an "erosion" of the Xi termed XaXe. The erosion of the XCI is characterized by loss of *XIST* and H3K27me3 expression and a loss of repression of the X-linked genes from the inactive X chromosome (Mekhoubad et al., 2012). The class III status is typically acquired during the adaptation of cells to the *in vitro* culture.

It has been suggested that the classification into three classes of the hESC represents the epigenetic progression undergone by cells during adaptation to the culture. In line with this, it appears that the derivation and cell culture conditions can have a significant impact in the XCI pattern of the hESC. For instance, subcultures that exhibit excessive cell death during expansion and which display abnormal nucleus morphology tends to lose their XCI marks and progress from class II towards class III. (Shen et al., 2008)

It has also been shown that an oxygen tension close to physiological conditions (5%

O_2) during the derivation of hESC helps maintaining pluripotency, suppresses spontaneous differentiation and prevents precocious XCI, keeping the cells in the naive state (class I). However, when cells are cultured under atmospheric conditions (20% $O_{2,}$), they irreversibly initiate XCI, undergo extensive modifications in their gene expression profile, and initiate induction of differentiation (Lengner et al., 2010). Taken together, it is fair to conclude that the XaXa state of hESC is highly unstable and that cells are prone to undergo XCI in response to cellular stress induced by culture conditions.

Also, the XCI status of female hPSC has been widely described as being unstable and showing a great variability not only between different cell lines but also between sub-cultures and from different passages of the same line (Liu et al., 2011; Shen et al., 2008; Silva et al., 2008). One of the possible explanations for this diversity in results is that different studies use different markers for XCI. Recently it has been shown that the use of *XIST* expression analysis is not sufficient to determine XCI. A meta-analysis on hPSC, (including both hESC and hiPSC) of the entire set of X-linked genes presented a more reliable identification of the XCI state, resulting in the classification of hPSC into no XCI, and differentiating the class II or class III cells in partial XCI or full XCI (Bruck and Benvenisty, 2011). It is thus important to take a maximum of marks into account to accurately establish the XCI status.

6.3.2 Non-random or skewed XCI in hPSC

Normally, the cells of adult human females display random XCI (Moreira de Mello et al., 2010). In contrast, the pattern of XCI in hPSC shows a tendency to strongly skew towards having one of the X chromosomes inactive while the other is preferentially active, resulting in a non-random XCI pattern (Dvash et al., 2010; Lengner et al., 2010; Liu and Sun, 2009). This bias from the normal 50:50 is also termed skewing. Cultures that present 75-80% of the cells inactivating the same X chromosome are considered as partially skewed, and cultures with 90% or more cells having the same X-chromosome inactive are considered fully skewed. The cells with skewed XCI present monoallelic expression of the X-linked genes. Skewing of the XCI can happen very early during *in vitro* culture of hESC (Liu and Sun, 2009). However, when and how the non-random XCI takes place and is stably maintained is still unknown.

Several hypotheses try to explain the occurrence of non-random XCI. The first is based on the founder effect and proposes that a very limited pool of cells, maybe even a single cell, contribute to the formation of the first colony in hESC (Dvash et al., 2010). The skewing would, in this case, be a result of a genetic bottleneck during the

derivation process. A second hypothesis proposes that the skewing pattern derives from a selective advantage of a specific Xi (Dvash et al., 2010; Lengner et al., 2010; Liu et al., 2011). Furthermore, the transition from the naïve state to the primed state when cells are cultured under physiological oxygen levels also results in skewed XCI. These results support the idea that hESC undergo XCI in response to stress induced by suboptimal culture conditions, and that a clonal selection of the best-adapted cells takes place (Dvash et al., 2010) Finally, it is possible that it is caused by an stochastic event.

The consequences of this skewed XCI remain unclear. In a recent study, the authors found that lines with a skewed XCI present an increased genetic instability on the X-chromosome, with more copy number changes occurring, and with frequent loss of heterozygosity for *XIST* and *XACT* (Luo et al., 2014). This genetic instability specific to the skewed lines adds to the already well-known genetic instability of hPSC (Lund et al., 2012) with its obvious implications for the use of hPSC in a clinical setting.

7 Aims of the study (II)

The aim of this part of the research was to study the XCI pattern in 23 female hPSC. From them, two were hiPSC lines and 21 hESC lines. Of these, 16 hESC lines were derived from whole blastocysts, 2 hESC lines were derived from single blastomeres obtained from 4- and 8-cell stage embryos and one line was parthenogenetic diplodized cell line. Two male hESC lines and one blood sample were included as control cell lines.

The main aims of the work were to investigate:

1. If the line origin has an effect on the XCI pattern
2. If hPSC present a preferential stage of XCI
3. If XCI is random or skewed
4. If there is a parent-of-origin effect when choosing which X-chromosome is inactivated in hPSC
5. If the pattern of XCI could be driven by a selective advantage
6. If this XCI pattern changes during long-term in culture and under different culture conditions
7. If the XCI pattern in undifferentiated hESC is maintained upon differentiation into trophoblast-like cells as well as hESC-derived Osteoprogenitor like cells

-

8 Human pluripotent stem cells rapidly loose X chromosome inactivation marks and progress to a skewed methylation pattern during culture

Mieke Geens*, **Anna Seriola***, Lise Barbé, Josep Santalo, Anna Veiga, Kimberly Dée, Lindsey Van Haute, Karen Sermon[$], Claudia Spits[$]. * Joint first authorship, $ Joint last authorship. Submitted to Stem Cell Reports.

8.1 INTRODUCTION

One of the classic examples of epigenetic gene regulation during development is dosage compensation of the X chromosome in female cells, by a process termed X chromosome inactivation (XCI). The initiation and maintenance of XCI is crucial for embryogenesis and adult cell physiology as many X-linked loci are involved in gene regulation and developmental processes. The X chromosome is enriched in genes related to sexual reproduction, as well as cancer-testis antigen genes (Heard and Disteche, 2006). Also, several loci have been associated with mental retardation, and proper expression of X-linked genes at the right dosage is essential for brain function and social skill development (Skuse, 2005). In addition, disruption of XCI is often found in pathological conditions such as in female cancer cells (Ganesan et al., 2005).

Human pluripotent stem cells (hPSC), both human embryonic stem cells (hESC) and human induced pluripotent stem cells (hiPSC), are regarded as suitable *in-vitro* models to study early human development. As such, they are expected to reliably mimic the early *in vivo* developmental processes, both at the genetic as at the epigenetic level. The study of XCI patterns in female hPSC lines has yielded very variable results. Variability has not only been observed between different lines (Adewumi et al., 2007; Bruck and Benvenisty, 2011; Dvash et al., 2010; Silva et al., 2008), but also between subcultures and different passages of the same line in different studies (e.g. the NIH-approved hESC line H9; (Dhara and Benvenisty, 2004;

Hall et al., 2008; Hoffman et al., 2005; Shen et al., 2008; Silva et al., 2008) or within one study (Hall et al., 2008; reviewed in Lund et al., 2012; Nguyen et al., 2013b). In a model first proposed by Silva et al. (2008), hPSC can be categorized into three classes according to their XCI status (Silva et al., 2008). Class I cells possess two active X chromosomes (XaXa) in the undifferentiated state and are considered closest to ground state pluripotency. Upon differentiation, the cells initiate XCI. This epigenetic state, however, is considered extremely delicate and class I cells rapidly progress to class II, in which the cells already underwent XCI (XaXi) in the undifferentiated state. Class II cells are characterized by the expression of the long non-coding X-inactive specific transcript (XIST) RNA from the X chromosome that will be inactivated. The inactivation of this chromosome is in part achieved by cis-coating with the XIST transcript (Penny et al., 1996). In addition, the inactivated X chromosome has DNA hypermethylation and specific histone modifications, e.g. the presence of histone H3 lysine 27 trimethylation (H3K27me3), histone H4 lysine 20 methylation (H4K20me1) and core histone variant macroH2A1, all co-localized with the Barr body (Brockdorff, 2002). Class III cells, on the other hand, display XCI but lack XIST expression and have a gradual loss of histone modifications and methylation, often coinciding with the erosion of transcription inactivation (Bruck and Benvenisty, 2011; Hall et al., 2008; Lengner et al., 2010; Mekhoubad et al., 2012; Shen et al., 2008; Silva et al., 2008). This last class of XCI is often referred to as a culture-adapted state. Human female somatic cells usually display a random XCI pattern (Moreira de Mello et al., 2010), although age and some specific diseases including several monogenic disorders and cancers may result in a non-random or skewed methylation pattern (Martínez et al., 2005; Plenge et al., 2002). Several studies on XCI in hESC, however, have reported the occurrence of non-random inactivation patterns (Dhara and Benvenisty, 2004; Lengner et al., 2010; Liu and Sun, 2009; Luo et al., 2014; Shen et al., 2008) but left some key questions open that we tried to address in this study.

We initiated this study to investigate the XCI pattern and stability in female hESC lines. Our aim was to determine whether a preferential XCI pattern exists in hESC, whether the XCI pattern changes during long-term culture and whether the XCI pattern in undifferentiated hESC is maintained upon differentiation. Next to conventional inner cell mass (ICM)-derived hESC lines, we also investigated two hESC lines derived from single blastomeres at cleavage stage (Geens et al., 2009) and two hiPSC lines. As several of the hESC used in this study originate from embryos after pre-implantation genetic diagnosis, DNA samples of the male and female gamete

donor were available which provided us with the unique opportunity to identify the origin of the inactivated X chromosome and makes this the first study depicting a full image of the XCI pattern of a large cohort of hESC lines. Our results show that during culture hPSC rapidly loose XCI marks and progress from a XIST-dependent XCI state to a XIST-independent XCI state with loss of repressive histone modifications and often erosion of methylation. We also report a remarkably high incidence of non-random XCI patterns. Moreover, our data show that this skewing of the methylation patterns is independent from the transition to the culture-adapted XCI stage or the origin of the X chromosome and is not occurring as a passenger event, driven by culture take-over due to a chromosomal aberration. These results suggest that XCI skewing is most probably driven by the activation or repression of a specific allele on the X chromosome, conferring a growth or survival advantage to the cells.

8.2 RESULTS

8.2.1 hPSC display significant degree of loss of XCI marks

In this work, we studied the XCI status of 25 hPSC lines (an overview can be found in the supplementary table 1). Next to two male hESC lines, used as negative control, we studied two female hiPSC lines and 21 female hESC lines. Of these, two lines were derived from single blastomeres (bES[13] and VUB26 (Geens et al., 2009)) and one was derived from a partenogenic embryo (pES[12]). Whenever available, the lines were investigated at different time points during long-term culture, with passage numbers ranging from 3 to 144. We also included a normal female control blood sample as reference.

We evaluated the XCI status based on several marks. First, we studied the expression levels of *XIST*, a key factor during the first stages of XCI. We quantified the expression of this transcript in 10 hPSC lines, ranging from passage 8 to 41. In all of these lines *XIST* was either absent or present at levels comparable to that of the male control hESC lines VUB01 and ES[6] (Figure 12a). Next, we differentiated a subset of lines into somatic osteoprogenitor-like (OPL) cells (Mateizel et al., 2008) or towards trophoblast cells (Marchand et al., 2011) to evaluate if differentiation resulted in up-regulation of the transcript(Dvash et al., 2010; Hall et al., 2008). In none of the studied lines, differentiation induced *XIST* expression (Figure 12a).

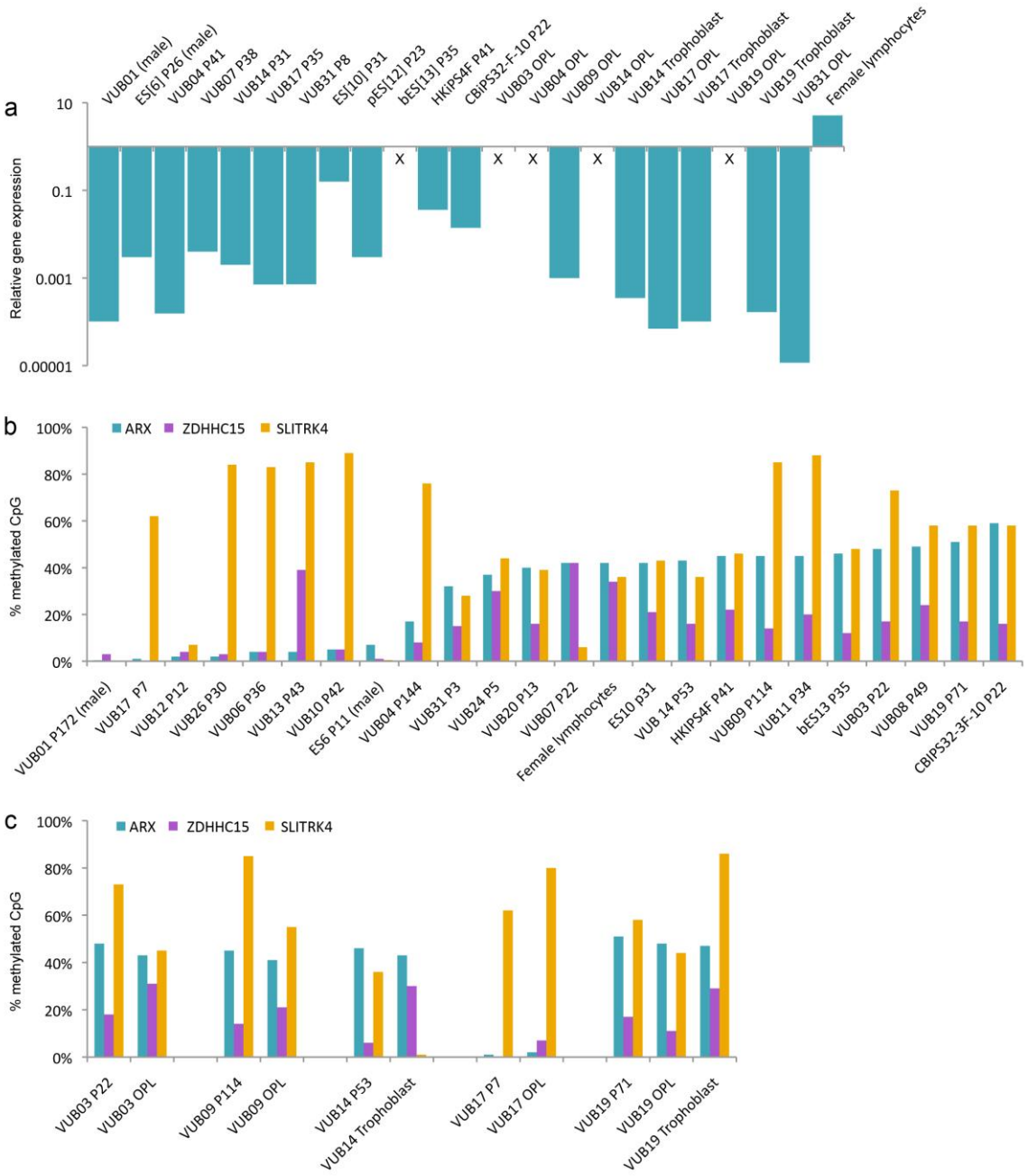


Figure 12. XIST expression and average methylation of XCI associated sites. Figure **(a)** shows the fold change of XIST expression when compared to control female amniocytes. Whereas female lymphocytes express XIST in similar levels as the amniocytes, the transcript is either absent (X) or present at levels similar to those seen in the male controls, VUB01 and ES[6], after several passages of in vitro culture. **(b)** and **(c)** represent the average percentage of methylation for the *ARX*, *ZDHHC15* and *SLITRK4* loci as quantified by MPBS in all hPSC cell lines studied in undifferentiated conditions (a) and upon differentiation into osteoprogenitor like cells (OPLs) or trophoblast (b), respectively.

Secondly, we analyzed the localization and quantity of the repressive histone mark

H3K27me3 by immunofluorescence. We tested 9 lines (with passage numbers ranging between 6 and 83), and found that 8 of these lines did not show the XCI-specific foci of H3K27me3. One of the lines (ES[10]) showed the histone mark in 1% of its cells.

Finally, we studied XCI-specific DNA methylation. For this, we used two approaches. For a fast screening of all samples, we applied methylation-sensitive DNA restriction followed by PCR for the polymorphic aristaless related homeobox gene (*ARX*) located at Xp21. Because of the polymorphic short tandem repeat located in the analysed region, this method cannot only be used to investigate XCI-specific DNA methylation but is also able to fingerprint the methylated X chromosome. Conversely, it only gives information on one specific CpG site in the promoter region of the *ARX* gene. To gain a broader and deeper insight in the DNA methylation patterns, we performed massive parallel bisulphite sequencing (MPBS) on a subset of samples. Next to the ARX locus, we studied two other loci for which the methylation status correlates with XCI status: the *ZDHHC15* promoter region (Xq13.3) and the *SLITRK4* (Xq27.3) promoter region (Bertelsen et al., 2011). These two sites also contain polymorphic short tandem repeats, in analogy to the *ARX* locus. This approach provides data for several CpG sites within one locus, and for hundreds of clones at the same time. This method allowed us to determine and quantify different epi-alleles and analyse the ratios of clones displaying low or high levels of methylation. Furthermore, the high number of analysed clones provides a reliable and in-depth image of the cellular diversity within a specific sample. Finally, and reassuringly, we found a very good correlation between the results obtained using methylation-sensitive DNA restriction and those obtained by MPBS (see table 4).

Table 4. Overview of the methylation patterns in all samples. This table resumes the results of the methylation-sensitive DNA restriction and PCR for the polymorphic marker *ARX*, as well as of the massive parallel bisulphite sequencing (MPBS) for the *ARX*, *ZDHHC15* and *SLITRK4* loci. For MPBS the ratios of hypermethylated epi-alleles at the two X chromosomes are indicated. When donor DNA was available,♂ and ♀ indicate the paternal or maternal allele respectively. Otherwise, the shortest and longest allele are indicated as 1 or 2 respectively. Undifferentiated passages are referenced by the passage number. Differentiated samples are indicated as OPL for osteoprogenitor-like differentiation and as Tropho for trophoblast differentiated cells.

		DNA RESTRICTION[2]	**MASSIVE PARALLEL BISULPHITE SEQUENCING[3]**		
Line	**Passage number[1]**	**ARX**	**ARX**	**ZDHHC15**	**SLITRK4**
Female control	Not applicable	Random methylation	1. 47% 2. 53%	Homozygous	Homozygous
VUB01 (male)	172	No methylation	Hypomethylated	Hypomethylated	Hypomethylated

VUB03	22	Skewed, ♂ methylated	♂. 76% ♀. 24%	1. 20% 2. 80%	1. Hypermethylated 2. Hypermethylated
	139	Skewed, ♂ methylated			
	OPLs	Skewed, ♂ methylated	♂. 74% ♀.26%	1. 20% 2. 80%	1. 50% 2. 50%
VUB04	12	Skewed, ♂ methylated			
	34	Skewed, ♂ methylated			
	144	Skewed, ♂ methylated	♂. Hypomethylated ♀. Hypomethylated	1. Hypomethylated 2. Hypomethylated	1. Hypermethylated 2. Hypermethylated
	OPL	Skewed, ♂ methylated			
VUB06	36	No methylation	1. Hypomethylated 2. Hypomethylated	1. Hypomethylated 2. Hypomethylated	1. Hypermethylated 2. Hypermethylated
VUB07	22	Skewed	1. 27% 2. 73%	1. 30% 2. 70%	1. Hypomethylated 2. Hypomethylated
	88	Skewed			
VUB08	49	Skewed, ♀ methylated	♂. 14% ♀. 86%	1. 92% 2. 8%	1. Hypomethylated 2. Hypomethylated
VUB09	114	Skewed, ♂ methylated	♂. 80% ♀. 20%	1. 47% 2. 53%	1. Hypermethylated 2. Hypermethylated
	OPL		♂. 80% ♀. 20%	1. 37% 2. 63%	1. 63% 2. 36%
VUB10	42	No methylation	1. Hypomethylated 2. Hypomethylated	1. Hypomethylated 2. Hypomethylated	1. Hypermethylated 2. Hypermethylated
	43	No methylation			
VUB11	30	XCI (homozygous)			
	34		Homozygous	Homozygous	1. Hypermethylated 2. Hypermethylated
VUB12	12	No methylation	1. Hypomethylated 2. Hypomethylated	1. Hypomethylated 2. Hypomethylated	1. Hypomethylated 2. Hypomethylated
	36	No methylation			
VUB13	43	Skewed, ♂ methylated	♂. Hypomethylated ♀. Hypomethylated	1. 52% 2. 48%	1. Hypermethylated 2. Hypermethylated
VUB14	21	Skewed, ♂ methylated			
	53	Skewed, ♂ methylated	♂. 93% ♀. 7%	1. 86% 2. 14%	1. 29% 2. 71%
	Tropho	Skewed, ♂ methylated	♂. 86% ♀. 14%	1. 78% 2. 22%	1. 0% 2. 100%
VUB17	7	No methylation	1. Hypomethylated 2. Hypomethylated	1. Hypomethylated 2. Hypomethylated	1. 56% 2. 44%
	10	No methylation			
	50	No methylation			
	OPL	No methylation	1. Hypomethylated 2. Hypomethylated	1. Hypomethylated 2. Hypomethylated	1. Hypermethylated 2. Hypermethylated
	Tropho	No methylation			

VUB19	21	Skewed, ♀ methylated			
	71	Skewed, ♀ methylated	♂. 13% ♀. 87%	1. 91% 2. 9%	1. 35% 2. 65%
	OPL	Skewed, ♀ methylated	♂. 9% ♀. 91%	1. 88% 2. 12%	1. 35% 2. 65%
	Tropho	Skewed, ♀ methylated	♂. 19% ♀. 81%	1. 86% 2. 14%	1. Hypermethylated 2. Hypermethylated
VUB20	10	Partially skewed, ♀ methylated*	♂. 31% ♀. 69%	1. 35% 2. 65%	1. 29% 2. 71%
	13	* different subline	♂. 43% ♀. 57%	1. 44% 2. 56%	1. 43% 2. 57%
	P23	Random methylation			
		Skewed, ♀ methylated			
VUB22	14	Skewed, ♂ methylated			
	40	Skewed, ♂ methylated			
VUB24	5	Random methylation	1. 48% 2. 52%	1. 57% 2. 43%	Homozygous
	41	Skewed			
VUB26	P10	No methylation			
	P16	No methylation			
	P30		1. Hypomethylated 2. Hypomethylated	1. Hypomethylated 2. Hypomethylated	1. Hypermethylated 2. Hypermethylated
VUB31	3	Skewed, ♀ methylated	♂. 48% ♀. 52%	Homozygous	1. 62% 2. 38%
ES[6] (male)	11	No methylation	Hypomethylated	Hypomethylated	Hypomethylated
ES[10]	29	Skewed			
	31	Skewed	1. 26% 2. 74%	1. 27% 2. 73%	Homozygous
pES[12]	12 23	Homozygous Homozygous			
bES[13]	10	Partially skewed	1. 17% 2. 83%	1. 46% 2. 54%	1. 56% 2. 44%
	11	Partially skewed	1. 36% 2. 64%	1. 47% 2. 53%	1. 55% 2. 45%
	25	Skewed	1. 16% 2. 84%	1. 36% 2. 64%	1. 63% 2. 37%
	35	Skewed	1. 17% 2. 83%	1. 59% 2. 41%	1. 61% 2. 39%
HKiPS4F	26	Skewed	1. 83% 2. 17%	1. 51% 2. 49%	1. 68% 2. 46%
	41	Skewed	1. 86% 2. 14%	1. 50% 2. 50%	1. 72% 2. 28%
CBiPS32-3F-10	14	Skewed			
	22	Skewed	1. 13% 2. 87%	Homozygous	1. 30% 2. 70%

[1]Undifferntiated passages are referenced by the passage number. OPL, osteoprogenitor differentiated cells. Tropho, trophoblast differencated cells.

[2]When donor information was available,♂ and ♀ indicate paternal or maternal methylation, respectively.

[3]1 and 2 indicate the two different alleles with the methylation percentage for each of them

First, we screened all lines using the methylation-sensitive restriction of *ARX*. We found that 18 of the 23 female hPSC lines showed methylation at these sites, suggesting that all but five lines (VUB06, VUB10, VUB12 VUB17 and VUB26) presented XCI-linked CpG methylation. We also tested differentiated progeny of five hESC lines (VUB03, VUB04, VUB17, VUB19 and VUB31), and found that the methylation pattern did not change upon differentiation. All the results can be found in the table 4.

In a next step, we studied the hPSC lines and differentiated samples by MPBS for *ARX*, *ZDHHC15* and *SLITRK4*. Figure 1b shows the average methylation for the three loci in all hPSC lines. For the *ARX* locus, we found that the hPSC lines could be broadly categorized in two types: lines showing on average 45% of methylated CpG sites (similarly to a female control sample), and lines showing an average of 3% methylated sites (similarly to the male hPSC lines). One line (VUB04), with 17% methylated CpGs, falls in between these two categories. The average methylation profiles obtained for the *ZDHHC15* locus correlated well with those for *ARX*, but in the hPSC, its overall methylation levels were lower than the levels observed at the *ARX* locus and those of the control sample. The only exception to this consistency was VUB13, which showed hypomethylation for the *ARX* locus but not for *ZDHHC15*. The overall lower methylation levels at the *ZDHHC15* locus could suggest that this locus is easily demethylated and could be a good predictor of erosion of DNA methylation, a phenomenon described earlier in hPSC (Bruck and Benvenisty, 2011; Mekhoubad et al., 2012). However, the *ZDHHC15* locus, located at Xq13.3 is in the region very close to the *XIST* locus, has also been referred to as the "predictive inactive region". According to Bruck and Benvenisty, this region is a more stable region for XCI and loses methylation the latest (Bruck and Benvenisty, 2011). The discordance to our findings may be due to the fact that while Bruck and Benvenisty extrapolated gene-expression profiles to XCI, we studied the methylation itself. Finally, the methylation profiles of *SLITRK4* were comparable to those found for *ARX* and *ZDHHC15*, except for eight lines, which showed very high methylation rates (>70%). *SLITRK4* is a transmembrane protein that suppresses neural growth (Aruga and Mikoshiba, 2003). The hypermethylation of the promotor region suggests transcriptional inactivation of this gene. In line with this, gene expression microarray data shows that generally, hESC show very low levels of this gene (data not shown). In some cases, it is not expressed at all, which concords with the fact that some of the lines show these hypermethylated profiles. Because of this it is likely that this region is not suitable for

a correct determination of XCI in hPSC.

Regarding the experiments involving samples from hESC and their differentiated progeny, in the general our results show that the differentiation process does not change the average methylation patterns of the cells (Figure 1c), in line with the results of the methylation-sensitive restriction. Only in VUB14, we observed a shift from hypomethylation (6%) at the *ZDHHC15* locus in the undifferentiated cells to a normal methylation level (30%) upon trophoblast differentiation. The *SLITRK4* locus, however, showed an opposite pattern (36% to 1% methylation respectively). It is worth noting that differentiation of VUB17, one of the lines showing very low methylation in both the *ARX* and *ZDHHC15* loci, did not restore the XCI-specific methylation to the levels expected in somatic cells.

Taking all the results above into consideration, hPSC show loss of expression of XIST and do not present the expect patterns of histone modification H3K27me3. Two thirds of the lines still presented XCI-specific DNA methylation, while 7 female lines showed methylation levels equivalent to that of male samples. These lines appear to present an advanced state of erosion of X-chromosome methylation. This is supported by the fact that, if the lines had no XCI rather than eroded XCI, *XIST* expression would have been induced upon differentiation (Dvash et al., 2010; Hall et al., 2008; Silva et al., 2008). This was not the case in these lines, as for example in VUB17. One line, VUB04, appears to display eroded DNA methylation, but to a lesser extent than the other seven lines. Overall, no differences were observed between hESC and hiPSC, or for hESC lines derived from individual blastomeres or a parthenogenetic embryo. Finally, the differentiated cells stably inherit the XCI patterns of the undifferentiated progenitors.

To further investigate the erosion of methylation at the cellular level and to determine possible heterogeneity within one hESC line, we determined mono- or bi-allelic gene expression patterns of X-linked genes by Minisequencing of DNA and cDNA. We generated six clonal sub-lines derived from single cells of VUB07. The clonal origin of the lines will result in a non-random, also referred to as skewed, methylation pattern, which is expected to translate at the gene expression level into mon-allelic expression. However, in four out of these six sub-lines, one of the three markers displayed a bi-allelic expression pattern (see Supplementary Table 2). These results confirm the erosion of the methylation pattern of the inactivated X chromosome, leading to a relaxation of the gene expression, as previously described (Bruck and Benvenisty, 2011).

8.2.2 hPSC present a predominantly skewed XCI pattern

The fact that the three loci we studied contain polymorphic short tandem repeats allowed us to distinguish the two X-chromosomes, and investigate differential methylation patterns. In somatic cells, generally, XCI occurs in a random fashion, meaning that in a large cell population, both X-chromosomes have equal chances to be methylated or unmethylated.

Using the methylation-sensitive DNA restriction assay, we found that the *ARX* locus was heterozygous in 21 out 23 hPSC lines, allowing for the identification of the two individual X-chromosomes (Supplementary Table 3).

Five lines showed no signal after restriction (because of hypomethylation, as discussed above), precluding them from further analysis. For all remaining 18 lines we found a striking non-random methylation pattern, at least in the latest passage at which they were analysed. The skewing in XCI appeared so pronounced, we did not find a shifted methylation ratio for the alleles but found that one of the alleles was always fully restricted, suggesting complete absence of methylation (Supplementary Figure 4).

Next, we split the MPBS data according to the length of the short tandem repeat. In our hPSC, 20 out of 21 female lines were heterozygous for *ARX*, 15 for *ZDHHC15* and 16 for *SLITRK4*. VUB11 was homozygous for both *ARX* and *ZDHHC15*, and showed hypermethylation at the *SLITRK4* locus, which made it impossible to differentiate between the two chromosomes in this particular line. Details on the heterozygosity of each line can be found in the supplementary Table 3. We categorized the epi-alleles in hyper-methylated and hypo-methylated based on the number of methylated CpG sites. It is worth noting that, in both the female control sample and the hPSC and their differentiated progeny it was rare to find alleles, in which all sites were methylated,. A graphical overview of the frequencies of the different epi-alleles for each sample and each marker can be found in Supplementary Figure 2. Using the female control as a reference, for the *ARX* and *SLITRK4* loci, epi-alleles were classified as hyper-methylated when 50% or more of the CpG sites were methylated and hypo-methylated when less than 50% of the sites showed methylation. For the *ZDHHC15* locus, due to the overall lower methylation rates, we set the threshold at 30% methylation.

Using this classification, the male samples show 100% hypo-methylated epi-alleles for the three loci, while the female control sample shows a pattern close to 50% hyper-methylated and 50% hypo-methylated alleles (51.7% hyper-methylated and 48.2% hypo-methylated for *ARX*; 43.5% and 56.5% respectively for *ZDHHC15* and 46.3% and 53.7% for *SLITRK4*). Conversely, hPSC showed different types of patterns.

Supplementary Figure 2. Frequencies of the different epi-alleles. This figure gives an overview of the frequencies of all different epi-alleles in the undifferentiated cells as detected by MPBS for *ARX*, *ZDHHC15* and *SLITRK4.* The size of the bubbles is proportional to the number of epi-alleles displaying a specific percentage of methylation at the different CpG sites. Lymphocytes from a female control sample were used as reference value.

Overall, loci that were hypo- or hypermethylated, showed similar methylation profiles for both alleles. On the other hand, loci with normal overall methylation ratios could be divided into two categories: loci suggesting random XCI and loci with a remarkable bias towards hypermethylation in one of the alleles and hypomethylation of the other (skewed XCI). Figure 13 shows the frequencies of hyper and hypomethylated sequences for each of the two X chromosomes.

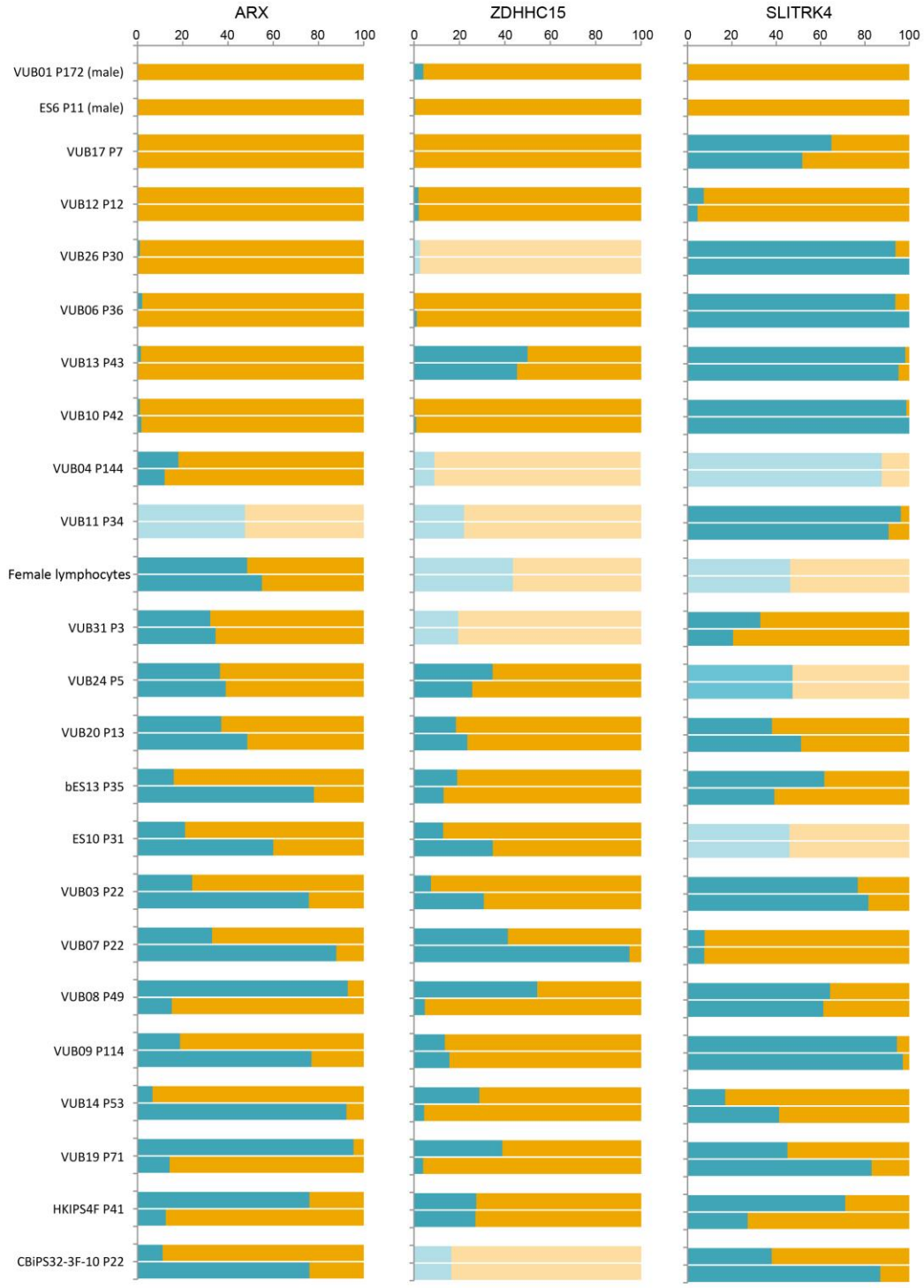


Figure 13. Frequencies of hyper and hypomethylated sequences for each of the two X chromosomes. The ratios of hypermethylated (blue bars) versus hypomethylated (orange bars) epi-alleles are presented for the two different alleles present in the cells and for the 3 markers under investigation: *ARX*, *ZDHHC15* and *SLITRK4*. Discrimination between the alleles of the two X chromosomes was based on heterozygosity for the short tandem repeats within the loci. Samples homozygous for a specific marker are indicated in faint colors. The thresholds for classifying epi-alleles as either hypermethylated or hypomethylated were set at 50% or more methylated CpG sites for the *ARX* and *SLITRK4* loci, and 30% of

methylated CpG sites for the *ZDHHC15 locus.*

Notably, we observed highly similar methylation patterns in differentiated progeny as in the undifferentiated cells they were derived from, indicating that the XCI patterns that are present at the undifferentiated state are stably transferred upon differentiation (see Figure 12c). The only exception was VUB14, in which we observed a loss of methylation at the *SLITRK4* locus, as mentioned earlier.

8.2.3 The loss of XCI marks and XCI skewing occur rapidly during the first passages of in-vitro culture and remains stable during long-term culture

Next, we investigated the evolution of XCI during in vitro culture. We selected the earliest possible passages of our lines, and carried out *XIST* analysis, immunohistochemistry for H3K27me3 and methylation analysis for those samples available.

Monitoring of *XIST* during very early passages of undifferentiated hPSC cultures, revealed that often at this stage *XIST* expression was still present at biologically relevant levels, but that its expression was quickly lost during cell culture (Figure 14a). Similarly, we found that cells at very early passages of the same lines still had foci of the repressive H3K27me3 histone mark, and that these were lost after a few passages (Figure 14b). Remarkably, the immunocytochemistry showed that the cell populations were very heterogeneous, with a mix of positive and negative cells within one passage, and even within one colony (Figure 14c). Overall, the loss of the histone marks and XIST expression correlated in time within the different lines, with individual differences amongst lines.

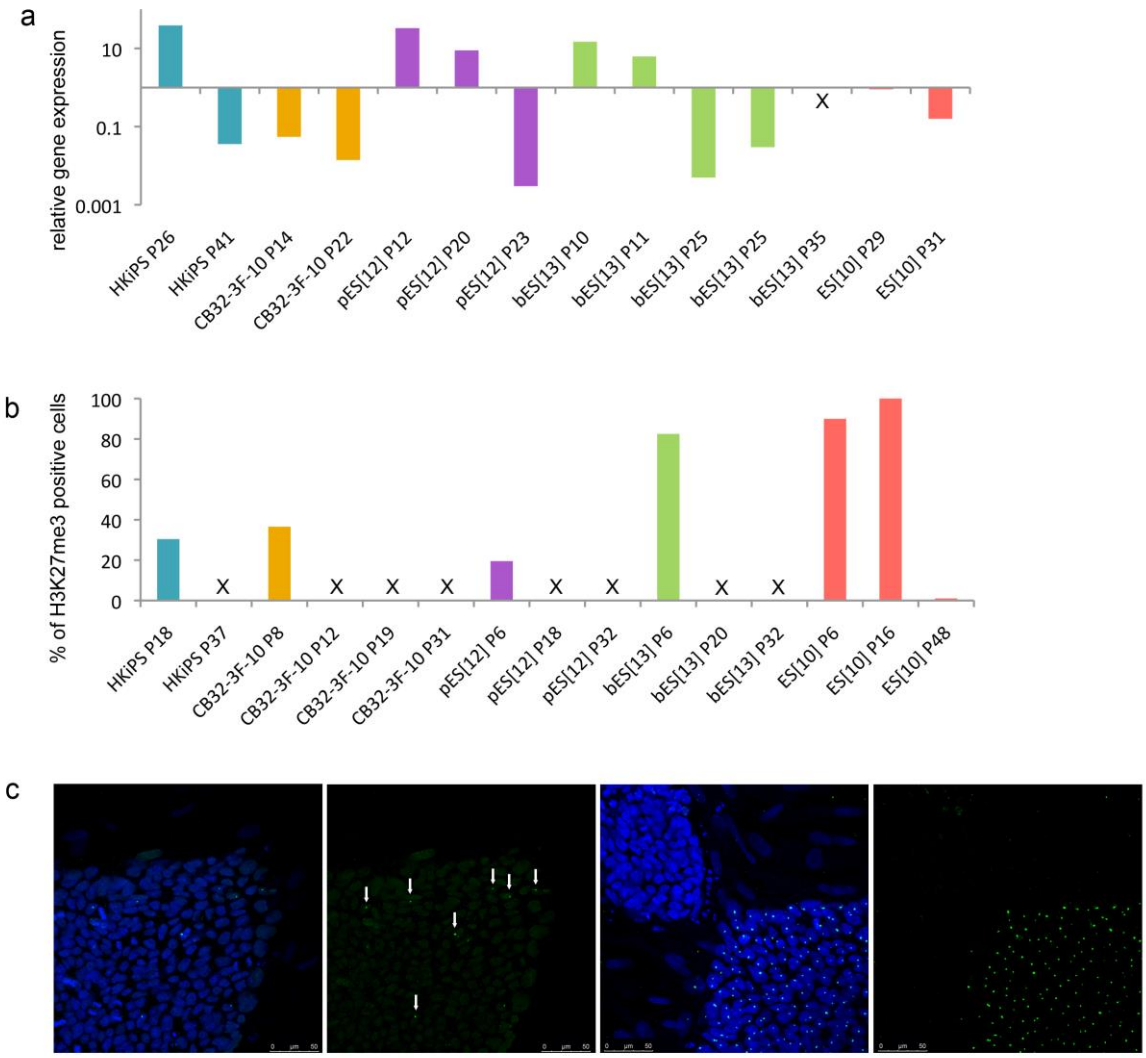


Figure 14. Loss of XCI marks during early *in vitro* culture. (a) *XIST* expression relative to control female amniocytes measured by real time-PCR. Five cell lines were investigated at low (P5-P15), mid (P15-P30) and/or high (>30) passage number. Absence of expression is indicated by (X). (b) Percentage of cells depicting an H3K27me3 nuclear focus by immunofluorescence. For every cell line, at least 200 cells from ten different colonies were counted. 0% is indicated with an (X). (c) Example of immunofluorescence detection of H3K27me3 on hPSC grown on human foreskin fibroblast (used as internal negative control). H3K27me3 was stained in green (Alexa Fluor 488) and the nuclear counterstaining was performed using DAPI. White arrows point at H3K27me3 positive cells in the sample showing only a low percentage of positive cells.

For the study of the methylation patterns, we first analysed a supplementary 21 DNA samples using the methylation sensitive restriction of the *ARX* locus. We added samples in order to cover, whenever possible, early and late passages of each of the cell lines studied. An overview of all the results can be found in the table 4. All together, these results suggested that at an early stage, XCI-specific methylation is random to moderately skewed, but after longer culture shifts towards a non-random,

skewed pattern. This can be seen in bES[13], VUB20 and VUB24 where we observed random methylation at passages 11, 13 and 5 respectively, while at passages 25, 23 and 41 a completely skewed methylation pattern was detected. In several other lines, however, non-random XCI was already set at very early passages after derivation (e.g. P12 in VUB04) and remained stable during extended culture (for instance up to passage 144 in VUB04). While we could not detect a gradual erosion of methylation in the lines in this study, we clearly show that once XCI-specific methylation is lost, also this hypo-methylated pattern remains stable during longer culture (e.g. VUB17, studied at passages 7, 10 and 50; VUB26 studied at passages 10 and 16).

Taken together, these results show that hPSC lose most of their XCI marks very quickly during the first passages in culture, and that there is a transition from a random XCI status to a completely skewed methylation pattern. It appears that this transition from random to skewed XCI does not coincide with the loss of histone marks or *XIST* expression, and overall, the skewing of the methylation patterns occurred earlier in time in culture. For example, ES[10] already displayed a completely skewed methylation pattern at passage 29, whilst at this passage we could still detect high expression of the *XIST* transcript.

In order to reveal whether the skewing of the methylation patterns happens randomly or whether there is a preferential allele to be active/inactive, we made single colony-derived sub-lines from early passage cultures of bES[13] and VUB20. We used these lines because they presented as lines with random to only moderate skewing at very early passages, but progressed to skewed methylation patterns during longer cultures. The rationale was that if this skewing is driven by a selective advantage, each clonal line initiated from a small group of cells with a random XCI pattern would display the same skewed pattern after some time in culture. We generated 10 sub-clonal lines from bES[13], which we monitored for 5 passages, and 8 sub-clones of VUB20, which we monitored for 2 passages. We carried out a first screening of the skewing/methylation using the methylation-sensitive restriction of the *ARX* locus in all samples, and performed MBPS for a subset of these. The results are shown in Figure 15.

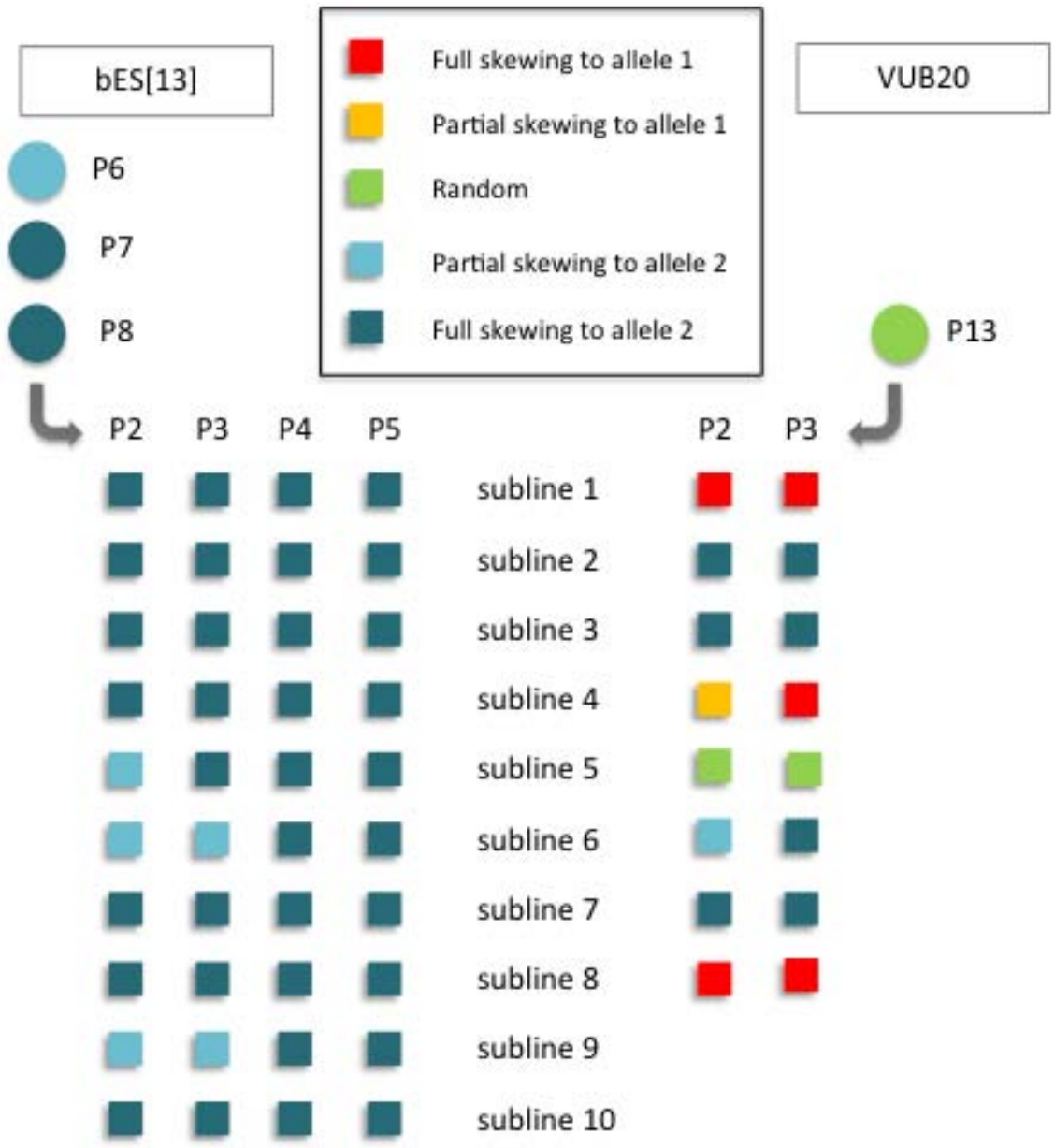


Figure 15. Subline analysis of partially skewed cell lines. Overview of the different methylation patterns as analyzed by methylation-sensitive DNA restriction and PCR for *ARX*. The circles represent an entire cell culture at a given passage. From these cultures, single colony-derived sublines, represented by squares, were created. All lines are color coded according to the legend. bES[13] was expanded until P8 and at which point 10 sublines were created. Each sub-line was passaged until passage 5 (P5). The same subline derivation was performed for VUB20 at P13. Eight sublines were derived and passaged until P3.

When bES[13] culture was reinitiated to perform the single-colony experiments from cells cryopreserved at passage 6, we found that the line showed already a more skewed pattern than the passage 11 previously tested. This pattern was inherited by all sub-lines, which displayed non-random XCI patterns, with inactivation of the same X chromosome. MPBS for the original line and a subset of sub-clones confirmed the

skewed methylation patterns for *ARX* with 68% of methylated epi-alleles at allele 2 at passage 7, increasing to 81% and 82% at passages 8 and 10. Similarly, in the second passages of sub-lines 3, 5 and 6 we found 78%, 84% and 77% of methylated epi-alleles at this allele. For *SLITRK4*, we could also observe a higher proportion of methylated epi-alleles at the same allele in all sub-lines tested, however, for this marker the skewing was much less pronounced (from 52% to 57%). For *ZDHHC15*, we could not detect a preferentially methylated allele, possibly due to the low overall methylation levels. The MPBS results are shown in Figure 16.

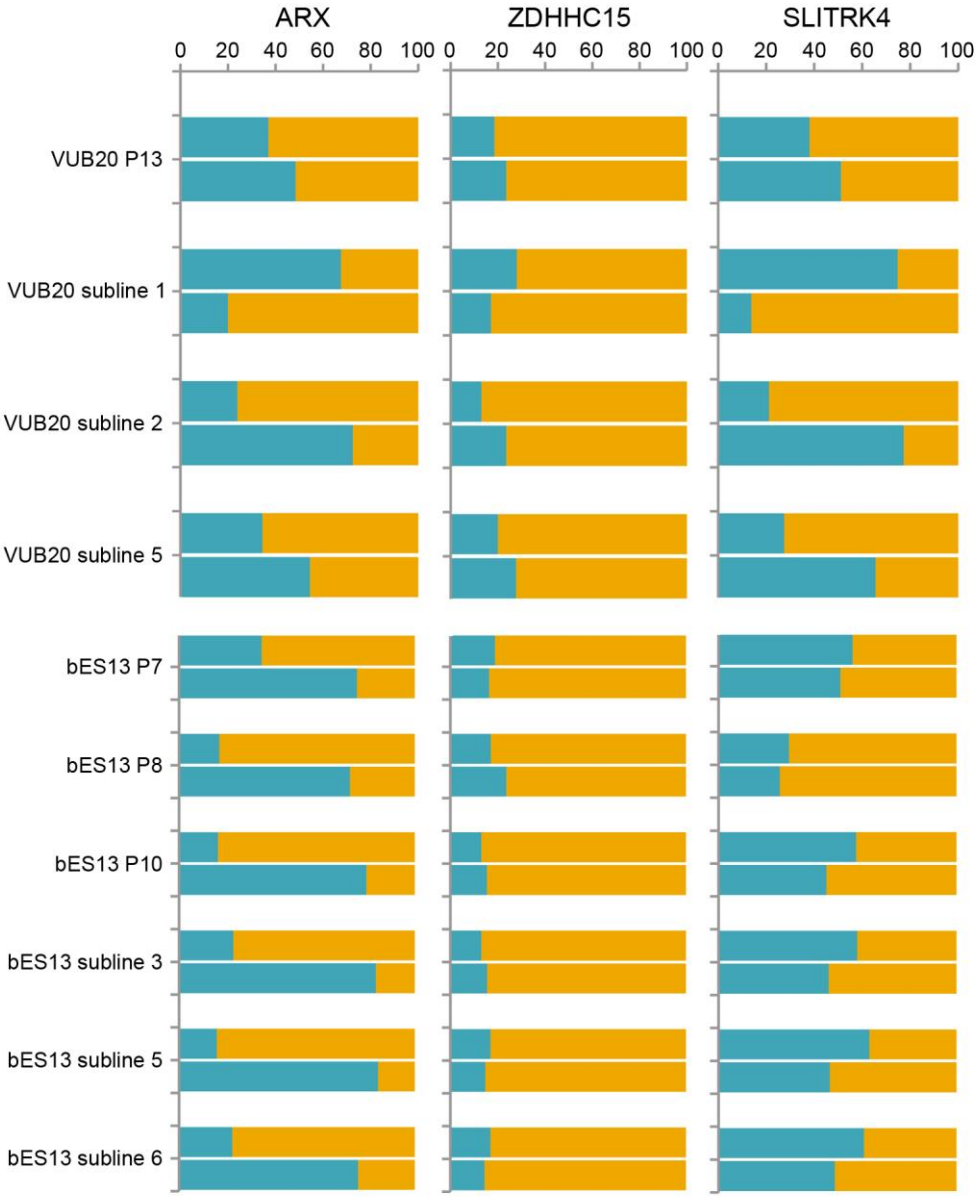


Figure 16. MPBS results for single colony-derived sublines. This figure shows the ratios of hypermethylated (blue bars) versus hypomethylated (orange bars) epi-alleles at the two different alleles for *ARX*, *ZDHHC15* and *SLITRK4* in the original and sublines of bES[13] and VUB20 (the origin of the sub-clones

is depicted in Figure 4). The thresholds for classification as hyper- or hypomethylated was 50% methylated CpG sites for the *ARX* and *SLITRK4* loci, and 30% of methylated CpG sites for the *ZDHHC15* locus.

The sub-lines of VUB20, on the other hand, told a different story. Our screening assay showed that the VUB20 line, which was used to generate the clonal lines (P13), showed a random XCI pattern after two passages, while three out of eight sub-lines displayed skewed methylation for one allele and four sub-lines for the other. Differently, sub-line 5 displayed a random methylation pattern. MPBS on the original line and three sub-lines confirmed these results: while in the original P13 line we could only find a slight overrepresentation of methylated epi-alleles (57%, 56% and 57% at allele 2 for *ARX*, *ZDHHC15* and *SLITRK4* respectively), sub-line 1 displayed high skewing towards allele 1 in all three markers (77%, 62% and 84% for *ARX*, *ZDHHC15* and *SLITRK4* respectively), sub-line 2 was skewed towards allele 2 for all three markers (75%, 64% and 78% for *ARX*, *ZDHHC15* and *SLITRK4* respectively) as well as sub-line 5, however, in this case the skewing was less pronounced (61%, 58% and 70% for *ARX*, *ZDHHC15* and *SLITRK4* respectively). All results are shown in Figure 16 and supplementary table 4.

In summary, although we observe a clear skewing of both bES[13] and VUB20 at later stages, we were able to generate sub-lines of an early passage of VUB20 representing opposite methylation patterns, while we did not succeed to do so for bES[13] due to the presence of skewing right before sub-cloning. Although these results suggest that randomness or not randomness of the skewing may be line-dependent, these results need to be taken with caution as the original lines already displayed different degrees of skewing upon generation of the sub-lines.

8.2.4 XCI skewing is not related to parent of origin and is not a passenger mutation

One possible reason for non-random XCI is a parent-of-origin specific inactivation, as seen in mouse trophectoderm lineage, in which the paternally inherited X chromosome is silenced (Takagi and Sasaki, 1975; van den Berg et al., 2011). We had access to DNA of the donors for ten hESC lines, and studied the parent-of-origin of the X-chromosomes using the short tandem repeat of the ARX locus. We found that six lines had preferentially inactivated the paternally inherited X-chromosome and four lines inactivated the maternally inherited X-chromosome (Supplementary table 5). These results rule out that the mode of inheritance influences the XCI pattern.

Another option is that the skewed methylation pattern reflects the clonal growth and culture take-over of one individual cell driven by a chromosomal abnormality that confers a selective advantage to the cells (Amps et al., 2011; Nguyen et al., 2014). Our cell lines are routinely checked by aCGH to monitor their genetic content. We have previously published normal karyotypes for some of the hESC lines under investigation. For example: VUB03, VUB04, VUB07, VUB14, VUB19, VUB31 have been shown to carry a normal 46,XX karyotype at even later passages than those from which samples were investigated in this study (see Supplementary table 6) (Nguyen et al., 2014; Spits et al., 2008). The skewed methylation patterns that we observed in these lines can thus not be explained as a passenger phenomenon either, leaving the question on the cause of the observed skewed methylation patterns open for now.

8.3 DISCUSSION

In this study, we demonstrated that female hPSC lines display XCI from very early stages of *in vitro* culture. While they rapidly loose XCI-specific marks such as *XIST* expression and XCI-specific histone modifications and often display low levels of DNA methylation, we could not detect any reactivation of the inactivated X chromosome. Also, we found that those lines that retain XCI-specific DNA methylation display a striking non-random methylation pattern, with clear skewing of the methylation on one of the two alleles. While both events seem to happen during the earlier stages of *in vitro* culture, our results suggest that they are happening independently from each other.

Our data suggest that our lines underwent a very fast transition from a *XIST*-dependent XCI state (stage II, (Silva et al., 2008)) just after derivation to a culture-adapted XCI state with loss of XCI-specific epigenetic marks (stage III,(Silva et al., 2008). While the loss of XCI marks has been report (Bruck and Benvenisty, 2011; Hall et al., 2008; Mekhoubad et al., 2012), our results show an extreme extent of erosion, especially at the DNA methylation level, in about one third of the lines under investigation. The effect of the loss of XCI marks on the phenotype of the cells is not completely clear yet. While Anguera et al. compared gene expression levels between class II and class III cells and did not observe any consistent downregulation of X-linked genes (Anguera et al., 2012), in this study, we showed that at the single cell level the eroded XCI marks can result in an erosion of gene repression. A recent study by Vallot et al. demonstrated that not all genes are equally susceptible to X chromosome inactivation (Vallot et al., 2015). Moreover, erosion of XCI marks seems

to be a pluripotency-related phenomenon, as it has never been reported in differentiated somatic cells, apart from highly malignant cancers (Vallot et al., 2015).
The fact that the eroded XCI marks, obtained by the undifferentiated cells can be found back in the differentiated progeny may impact the phenotype of the differentiated cells and thus have important consequences for the applicability of these cells as *in vitro* model for disease or as cell source in regenerative medicine. X chromosome inactivation has important implications during for example neural development (Lister et al., 2013; Nguyen et al., 2014; Wu et al., 2014) and the occurrence of aberrant XCI patterns in several cancers highlights its role in cell cycle progression and apoptosis (Ganesan et al., 2005; Silva et al., 2008). Several groups have suggested that the differentiation potential of hPSC and their potential oncogenic properties can be directly correlated to the XCI pattern of the cells. Cells exhibiting class III XCI form scant and poorly differentiated embryoid bodies (EB) when compared to EBs generated from class I or class II (cells (Silva et al., 2008). Similar results were reported after teratoma formation. While class II cells generated solid tumors showing prominent differentiation into structures recapitulating adult organs and tissues, teratomas derived from class III lines were all cystic with only low degree and complexity of differentiation into the three germ layers (Anguera et al., 2012; Bruck and Benvenisty, 2011; Hall et al., 2008; Mekhoubad et al., 2012). Also, class III hPSC lines exhibited shorter doubling times during *in vitro* culture and displayed a marked upregulation of cancer-associated genes (including *MAGEA2*, *MAGEA6*, *RAB6B*, *ACP5* and *AIF1*) and repression of tumor suppressors (e.g. FTX) than class II and male lines (Anguera et al., 2012).
The transition from class II to class III cells may be explained by sub-optimal culture conditions. It is hypothesized that hESC would have two active X-chromosomes immediately after derivation but will come under pressure to eliminate the presumed growth disadvantage associated with this state during hESC propagation, and therefore initiate XCI (Bruck et al., 2013; Diaz Perez et al., 2012; Hall et al., 2008; Lengner et al., 2010; O'Leary et al., 2012; Silva et al., 2008; Vallot et al., 2015). Our cell lines were cultured on mouse embryonic fibroblasts and in Knock-Out Dulbecco's modified Eagle medium supplemented with 20% Knock-Out Serum Replacement and 4ng/ml bFGF, a commonly used culture condition (Fraga et al., 2011; Vallot et al., 2015). These conditions, however, are certainly not the most optimal as cells cultured in this and other conditions have been shown to display high levels of genomic and epigenomic instability and culture take-over by cells displaying a growth advantage conferred to by a genetic aberration (Lister et al., 2013; Lund et al., 2012; Nazor et

al., 2012; Nguyen et al., 2013b; 2013a). Further optimization of culture conditions may lead to a better control over the XCI pattern. For example, culture in low oxygen, or the addition of antioxidants or small molecules have been shown to stabilize the class I XCI pattern in which cells display two active X chromosomes (Diaz Perez et al., 2012; Lengner et al., 2010).

In this study, we also demonstrated a strikingly high incidence of lines carrying a skewed XCI methylation pattern. In fact, any line that did not display an extended erosion of methylation eventually showed this non-random pattern of hypermethylation of one allele and hypomethylation of the other. The scale and early timing of the occurrence of the skewed XCI patterns that we observe is remarkable. Although the occurrence of non-random XCI patterns in hESC has previously been reported (Liu et al., 2009) , the extent of this phenomenon was, until now, not clear. These non-random patterns are in contrast to female human somatic cells in which usually a random XCI pattern is observed (Moreira de Mello et al., 2010) and also to hESC in ground state XCI which contain two active X chromosomes (XaXa) and undergo random XCI upon differentiation (Dhara and Benvenisty, 2004; Lengner et al., 2010).

We showed that the observed skewed XCI patterns are not a parent-of-origin dependent phenomenon, nor that they are linked to a culture take-over by cells that have acquired a proliferation or survival advantage conferred by a chromosomal aberration (Nguyen et al., 2013a). A high occurrence of non-random methylation patterns could also be expected if hESC lines would originate from only a few cells. If a hESC line has only one or two founder cells, there is a high probability that the daughter cells share the same XCI pattern. However, this could not explain the gradual skewing that we observed in several lines. Moreover, O'Leary *et al.* reported on the existence of a post-ICM intermediate (PICMI) after plating of human blastocyst embryos for hESC derivation (Hall et al., 2008; Lengner et al., 2010; O'Leary et al., 2012; Silva et al., 2008). While this transient epiblast-like structure has undergone XCI, further transformation into hESC resulted into X chromosome reactivation. This would imply that XCI in our lines occurred after derivation and that the number of cells contributing to the line does not influence the observed patterns.

The most plausible reason for the observed skewing may lay in the presumed suboptimal culture conditions. These suboptimal culture conditions probably induce precocious XCI but may also provide an environment in which cells that have activated a specific locus could have an advantage in culture, overgrowing the other cells and thereby inducing a skewed XCI pattern in the culture. We hypothesize that the

activation of certain alleles on the X chromosome may confer a survival or growth advantage to the cells and thereby drives the culture towards a non-random inactivation pattern. Our hypothesis is in concordance with the observation by Lengner *et al*. that hESC lines with two active X chromosomes, obtained through derivation at low oxygen tension, acquire random XCI upon differentiation, while prolonged culture of the same lines at atmospheric oxygen conditions results in skewed XCI patterns in the undifferentiated cells (Lengner et al., 2010). Moreover, the existence of a culture advantage linked to an active locus on the X chromosome could also explain the significant bias found in female versus male hESC lines worldwide (Ben-Yosef et al., 2012). This would require, however, a highly recurrent X-linked polymorphism that alters the expression of a gene that, directly or indirectly, could confer a significant culture advantage to the cells. There is an abundance of X-linked genes that could be possible candidates to confer a culture advantage. The X chromosome harbors for example several proto-oncogenes (e.g. *ELK1* and *ARAF*) and tumor suppressor genes (e.g. *SRPX*, *TFE3*), genes involved in cell cycle progression or apoptosis (e.g. the androgen receptor and its interacting protein NONO) and the group of cancer testis antigens (the MAGE proteins).

Skewing is also observed in iPSC but the mechanism may be different. Upon iPSC reprogramming from somatic cells, the two X chromosomes are reactivated (Kim et al., 2015). However, as the iPSC quickly acquire XCI, there seems to be a preferential inactivation of the X chromosome that was also inactive in the donor somatic (Barakat et al., 2015; Diaz Perez et al., 2012; Lengner et al., 2010; Pomp et al., 2011). In this case, non-random XCI, rather than driven by a culture advantage, may be caused by incomplete reprogramming of the epigenome and clonal growth.

Our results highlight the extent of aberrant XCI in female hPSC. While very variable XCI patterns have been reported by several other groups (Lund et al., 2012; Nguyen et al., 2013b) and hPSC exist in different classes of XCI (Silva et al., 2008), more insight into the XCI status and its stability in human pluripotent stem cells is key to achieve the full potential of hPSC in the future.

8.4 METHODS

8.4.1 Cell culture

hESC and human induced pluripotent stem cells (hiPSC) were used for this study. The hESC lines used in this study were VUB01, VUB03_DM1, VUB04_CF, VUB06, VUB07, VUB08_MFS, VUB09_FSHD, VUB10SCA7, VUB11_FXS, VUB12_APKD, VUB13_FXS,

VUB14, VUB17, VUB19_DM1, VUB20_CMT, VUB22_CF and VUB31_FSHD, derived at the Vrije Universiteit Brussel (VUB) (Mateizel *et al.*, 2006; Mateizel *et al.*, 2010) and ES[6], ES[10], pES[12] and bES[13] derived at the CMR[B]. HKiPS4F and CBiPS32-3F-10 are hiPSC cell lines and were also derived at the CMR[B]. All cell lines used are registered in the EU hPSC registry (http://www.hescreg.eu/). All cell lines are female, except for VUB01 and ES[6] which are male and were used as control.

HESC at VUB were derived and cultured on inactivated CF1 mouse embryonic fibroblasts (MEF) at 37°C in 5% CO_2 and atmospheric O_2 concentrations in hESC medium consisting of knockout D-MEM (Life Technologies, Gent, Belgium) supplemented with 20% Serum Replacement (SR, Life Technologies), 2 mM L-glutamine (Life Technologies), 0.01mM non-essential amino acids (NEAA, Life Technologies), 0.1mM β-mercaptoethanol (Sigma-Aldrich), 4ng/mL human recombinant basic fibroblast growth factor (bFGF; Life Technologies) and 10mM penicillin and streptomycin (Life Technologies), as described earlier (Geens et al., 2009; Mateizel et al., 2006, 2010, 2012). Passaging was performed mechanically.

HPSC cell lines from the CMR[B] were derived on human foreskin fibroblast (HFF) and cultured on HFF or Matrigel® at 37ºC in 5% CO_2 and atmospheric O_2 concentrations in hESC medium consisting of Knockout D-MEM (Life Technologies) supplemented with 20% Knockout Serum Replacement (SR, Life Technologies), 2 mM L-glutamine (Life Technologies), 0.01mM non-essential amino acids (NEAA, Life Technologies), 0.1mM β-mercaptoethanol (Sigma-Aldrich), 4ng/mL human recombinant basic fibroblast growth factor (bFGF; Peprotech) and 10mM penicillin and streptomycin (Life Technologies), when cultured on HFF and using mTeSR®1 when cultured on Matrigel® (BD Matrigel). Cell passaged was performed mechanically with a cell stripper when grown on HFF and enzymatically with 0.05% Trypsin when the cells were grown on Matrigel®.

Single cell-derived hESC colonies were created after at least one hour 10 µM Y-27632 (ROCK inhibitor) treatment. Cells were collected using non-enzymatic cell dissociation buffer (C5914, Sigma-Aldrich, Diegem, Belgium) and dislodging of the colonies by vigorous pipetting. The collected cells were passed through a cell strainer (mesh Ø = 40µm) to remove cell clumps, before being plated at low concentration on new MEF-coated culture dishes in hESC medium.

HESC differentiation into osteogenic progenitor-like cells (OPL) was performed as described by Mateizel *et al* (Mateizel et al., 2008). HESC were collected using 1mg/mL collagenase IV. The cell clumps were resuspended in differentiation medium (hESC medium without bFGF and with foetal calf serum (Life Technologies) instead of SR)

and plated on 0.1% gelatine-coated culture plates. The medium was refreshed every 2 days. After 24 days in culture, the cells were trypsinised and plated on 0.1% gelatine-coated culture dishes. When confluent, cells were harvested with trypsin–EDTA and passaged at a 1:2 - 1:3 ratio.

HESC differentiation into trophoblast was performed according to a previously published protocol (Marchand et al., 2011). In brief, hESC were cultured for 5 days on Matrigel® in hESC medium conditioned for 24 h on MEF (MEF-conditioned medium) supplemented with 4ng/mL bFGF. Cells were dissociated using trypsin–EDTA and plated at 100 000 cells/well in a Matrigel-coated 12-well plate. After 24 hours, the cultures were washed with PBS and MEF-conditioned medium, supplemented with 100ng/mL bFGF was applied. Medium was refreshed daily and cells were collected after 10 days of differentiation.

8.4.2 DNA methylation analysis

To investigate the methylation pattern of our cell lines, we applied methylation-sensitive DNA restriction, followed by PCR for a polymorphic short tandem repeat (STR) in the aristaless related homeobox gene (ARX) located at Xp21

DNA was extracted by phenol-chloroform. Methylation-sensitive restriction was performed on 100ng of DNA in a total volume of 20 µl comprising 2 µl 10x buffer and 1 µl HapII enzyme (1053AH, Takara Biotechnology co, Saint-Germain-en-Laye, France). An additional 1 µl of enzyme was added after the first 5 h of incubation at 37°C, followed by further overnight incubation. In control reactions, the HapII enzyme was replaced by H_2O. For each digestion reaction, a male control sample was included to check whether the digestion reaction was completed successfully.

PCR reactions for the polymorphic loci of aristaless related homeobox (*ARX*), located at Xp21.3, and proprotein convertase subtilisin/kexin type 1 inhibitor (*PCSK1N*), located at Xp11.23 were performed using 20 ng of DNA in a 25 µl total PCR volume containing 10% 10x AmpliTaq PCR buffer II, 2mM MgCl2 (both N808-1067, Life Technologies), 10 pmol of primers, 0.2 mmol dNTP, 2.5% DMSO and 1.25 U of Amplitaq polymerase (N808-1067, Life Technologies). The primers used were: ARX-1: TCCAGAATCTGTTCCAGAGCGTGC, ARX-2: GCTGTGAAGGTTGCTGTTCCTCAT, PCSK1N-1: ATGCGAAGACCATTCCCTCT, PCSK1N-2: GTGCGTGTGATGTGAGGAGA. PCR was performed with the following parameters: 5 min at 96°C, 30 cycles (30 s at 94°C, 30 s at 60°C, 30 s at 72°C), 5 min at 72°C. PCR products were analysed by capillary electrophoresis using the 3730 genetic analyser (Life Technologies) and Gene Mapper 4.0 Software (Life Technologies).

8.4.3 Massive parallel bisulphite sequencing

Massive parallel bisulphite sequencing was performed for three different loci previously shown to be correlated to XCI status (Bertelsen et al., 2011). Apart from the *ARX* locus (Xp21) on which we applied the methylation-sensitive DNA restriction, we also investigated methylation at the *ZDHHC15* locus (Xq13.3) and the *SLITRK4* (Xq27.3) locus. Both loci are located just upstream of the first exons of the respective genes and contain an AC dinucleotide repeat. One microgram of DNA was subjected to bisulphite treatment using the Imprint DNA modification kit (Sigma-Aldrich), following the protocol provided by the manufacturer. After purification with the High pure PCR product purification kit (Roche, Vilvoorde, Belgium) and elution in 50 µL elution buffer, PCR was performed on 5 µL purified DNA with 2.5 U JumpStart Taq DNA Polymerase and a 10X buffer supplied by the manufacturer (Sigma-Aldrich). The PCR mix further contained 0.4 µM forward and reverse primer, 200 µM of each dNTP (GE Healthcare, Diegem, Belgium) and was carried out in a total volume of 25 µL. The primers sets were adapted to be used for a Illumina Nextera-based library preparation: ARX-1: TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GAG GAG TTT TTT AGA ATT TGT TTT AGA G, ARX-2: GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GAT AAA CTT AAA AAA AAC CAT CCT CAC CCT, ZDHHC15-1: TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG GAA AAG GGA AGG AGG GAG TG, ZDHHC15-2: GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GTT TAT TTC TAT CTC AAT CTC TCT A, SLITRK4-1: TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG TTG TTG TAT TTG AAT GTT ATT GTT GTT G, SLITRK4-2: GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GCC CTT TTC TTC TCT TCC AAT CC (Integrated DNA Technologies, Heverlee, Belgium). The PCR program used following cycling parameters: 94°C for 5 min, followed by 40 cycles at 94°C for 30 s, an annealing 60°C for 30 s and extension at 72°C for 50 s followed by a final elongation at 72°C for 7 min.

Library preparation was carried out by PCR using the NEBNext High-Fidelity 2X PCR Master Mix (Bioké) according to manufacturer's instructions, with Illumnia Nextera indexed primers (Illumina), and with the following PCR program: 30 s at 98°C, 15 cycles of 10 s at 98°C, 30 s at 65°C and 30 s at 72°C, followed by the final elongation of 5 min at 72°C. The products are purified using magnetic bead purification and size selection using a 0.7 beads to sample ration (Analis). The libraries are spiked in with a 20% PhiX (Illumina) to compensate for low complexity. The final libraries are loaded on the MiSeq Reagent Nano Kit v2 (500 cycles) according to manufacturer's instructions and sequenced at 2x250 bp (Illumina).

The data analysis was carried out in two steps. First, a Perl script was used to process

the FASTA files for each of the loci. In this step, the sequences were categorized according to the short tandem repeat length. Next, only the files containing the short tandem repeat of the expected length were analysed using the BIQ Analyzer, a free software tool for DNA methylation analysis (Bock et al., 2005).

8.4.4 Gene expression analysis

RNA was extracted using the RNaesy Mini Kit with on-column DNase digest according to the manufacturer's protocol (Qiagen) and was reverse transcribed using the First-strand cDNA Synthesis Kit (GE Healthcare) with the NotI-d$(T)_{18}$ primer.
Relative quantification real-time polymerase chain reaction (Rq-PCR) was performed on the ABI 7500 real-time thermocycler (Life Technologies) with the following cycling parameters: 2 min at 50°C, 10 min at 95°C, 40 cycles of 15 s at 95°C and 1min at 60°C. The final reaction mix of 25 µl per well contained 12.5 µl 2x qPCR MasterMix Plus Low ROX (Eurogentec, Seraing, Belgium), 50 ng of cDNA and the appropriate amounts of primers and probes. For *XIST* and *TSIX*, 1.25 µl 20× Assays-on-demand Gene Expression assay mix (Hs01079824 and Hs03299334 respectively, Life Technologies) was applied; for *GAPDH*, which was used as endogenous control, 900 nM of primer mix and 250 nM of probes were added to the mix (primer 1: ATG-GAA-ATC-CCA-TCA-CCA-TCT-T, primer 2: CGC-CCC-ACT-TGA-TTT-TGG, probe: CAG-GAG-CGA-GAT-CC). Normalization against the endogenous control *GAPDH* using the ΔΔCt method of quantification was applied to relatively quantify *XIST* gene expression between multiple samples. Relative amount of mRNA was calculated as 2-ΔΔCt.

8.4.5 Minisequencing

We applied minisequencing to analyse allele-specific expression of X-linked genes. PCR reactions were performed using 20 ng of DNA or cDNA in a 25 µl total PCR volume containing 10% 10x AmpliTaq PCR buffer II, 2mM MgCl2 (both N808-1067, Life Technologies), 10 pmol of primers, 0.2 mmol dNTP, 2.5% DMSO and 1.25 U of Amplitaq polymerase (N808-1067, Life Technologies), using the following primer sets: *AFF2* (CCTGGAAAGGCATGCTGTAT and CTGTCTCATCTGCCTGCAAC), *ATP7A* (CAGTTTTTCGGAGGCTGGTA and AACAGCATAGGGGGTGTGTC) and *FGF13* (CTCTTGGGGAAGGAGTGTTG and CGCTCACATTGATTTCATGC). PCR was performed using the following cycling parameters: 1 min at 94°C, 35 cycles (30 s at 94°C, 30 s at 62°C, 30 s at 72°C), 7 min at 72°C.
PCR products were purified using the High pure PCR product purification kit (Roche) according to the manufacturer's protocol. A single nucleotide amplified polymorphism

reaction (SNaP) was performed in a reaction mix containing 2 µl H_2O, 2 µl SNaPshot® multiplex ready reaction mix (1107068, Life Technologies), 2 µl sequencing primer (1 mM), and 2 µl of the purified PCR product, using the following cycling parameters: 10 s at 96°C, followed by 25 cycles (10 s at 96°C, 5 s at 50°C, 10 s at 60°C). Following sequencing primers were applied: AFF2 (GTGGAGCAGTGTTTTKACCTTTGATAC), ATP7A (GCATTTGCCTACTCTTTGATTATTCTTCTA) and FGF13 (CAATCCAATTGCTGATGTGATTCATGAATA). Minisequencing products were analysed by capillary electrophoresis using the 3130xl genetic analyser (Life Technologies) and Gene Mapper 4.0 Software (Life Technologies).

8.4.6 Immunostainings

Female amniocytes from amniotic fluid sampling were used as positive control for the presence of XCI. Amniocytes were plated on uncoated Lab-tek chamber slides (Thermo Scientific, Erembodegem, Belgium) in supplemented Chang medium (C100, Alere Medical, Sint-Denijs-Westrem, Belgium). HESC colonies were mechanically isolated and plated on Matrigel® (354277, BD Biosciences, Erembodegem, Belgium)-coated Lab-tek chamber slides in mTeSR® medium (05855, StemCell Technologies, Grenoble, France). The slides were incubated overnight at 37°C, 5% CO2.

Cell cultures were washed three times with PBS (Life Technologies) before fixation with 4% paraformaldehide for 5 min at room temperature with -20°C methanol. Fixed cells were washed three times with PBS followed by blocking of unspecific binding sites by 1 hour incubation in PBS/5% goat serum at room temperature. Primary antibodies against H3K27me3 (Mouse monoclonal IgG3, ab6002, Abcam, Cambridge, UK), H3K9ac (Rabbit polyclonal IgG, ab10812, Abcam) and macroH2A1 (Rabbit polyclonal IgG, ab37264, Abcam) and their corresponding isotype controls were applied at 5 µg/mL and incubated overnight at 4°C. After three 5 min washes with PBS (Life Technologies), the secondary antibodies, Alexa Fluor 546 goat anti-mouse IgG (A-11018, Life Technologies) and Alexa Fluor 488 goat anti-rabbit IgG (A-11070, Life Technologies) were added at 10 µg/mL and incubated at 4°C, for 2 h. Finally the slides were washed three times with PBS, a droplet of Slow-Fade® gold antifade reagent with DAPI (S36936, Life Technologies) was added and a cover glass applied. An Olympus IX2-UCB fluorescent microscope (Olympus, Tokyo, Japan) and CellF software (Olympus) were used for analysis.

Immunostainings for ES[6], ES[10], pES[12] , bES[13], HKiPS4F and CBiPS32-3F-10 were performed following Martí et al 2012 (Marti et al 2013; (DOI: 10.1038/nprot.2012.154) hPSC cell lines were cultured for 5-7 days in SlideFlasks

(brand) coated with Matrigel®. Cell culture medium was removed and colonies were fixed with 4% paraformaldehyde for 20 minutes at room temperature, washed three times with PBS and kept at 4ºC until its use. Antigen retrieval step was performed with Target retrieval solution, at pH=9 10X (Dako, Cat no. S2367). Samples were washed three times with 1x TBS plus 0.5% (vol/vol) Triton X-100 plus and 6% (vol/vol) donkey serum and incubated for 1 hour at room temperature for blocking and permeabilization of the plasma membrane. Primary antibody against H3K27me3 (Anti-trimethyl-Histone H3 (Lys27) Millipore, cat. No. 07-449) was applied at 5 µg/mL with TBS++ (TBS 1X with 0.1% Triton X-100 and 6% of donkey serum) and incubated overnight at 4°C. After 3 washings with TBS++, the secondary antibody Cy2 (anti-rabbit IgG Cy2, abcam cat. No. ab6940) was added and incubate at 37ºC for 2 hours in a humid chamber. Finally, the slides were washed 3 times with 1X TBS for 5 minutes each and incubated with DAPI at a 1:10,000 ratio for 10 minutes at room temperature. Two or three drops of mounting medium were added and a cover glass applied. A Leica confocal microscopy (cat. No. DMI400) with objectives (HC PL APO 10.0X0.30 507902); HCX PL APO x20.0/0.70 IMM UV 11506191); (HCX PL APO CS x40.0/1.25 OIL UV 11506251); (HCX PL APO Lamba blue x63.0/14.0 OIL UV) and lasers (diode/488 nm argon/561nm diode pumped solid state (DPSS), 633 nm HeNe) was used for the analysis.

8.5 SUPPLEMENTARY DATA

Supplementary Table 1. Overview of hPSC lines used in this study.

Cell type	Cell line Name[1]	Karyotype	Origin
hESC	ES6	46,XY	Blastocyst
	VUB01	46,XY	Blastocyst
	bES[13]	46,XX	Single Blastomere
	ES10	46,XX	Blastocyst
	pES[12]	46,XX (parthenogenetic)	Blastocyst
	VUB03_DM1	46,XX	Blastocyst
	VUB04_CF (carrier)	46,XX	Blastocyst
	VUB06	46,XX	Blastocyst
	VUB07	46,XX	Blastocyst
	VUB08_MFS	46,XX	Blastocyst
	VUB09_FSHD	46,XX	Blastocyst
	VUB10_SCA7	46,XX	Blastocyst
	VUB11_FXS	46,XX	Blastocyst
	VUB12_APKD	46,XX	Blastocyst
	VUB13_FXS	46,XX	Blastocyst
	VUB14	46,XX	Blastocyst
	VUB17	46,XX	Blastocyst
	VUB19_DM1	46,XX	Blastocyst
	VUB20_CMT1A	46,XX	Blastocyst
	VUB22_CF	46,XX	Blastocyst
	VUB24_DM1	46,XX	Blastocyst
	VUB26	46,XX	Single Blastomere
	VUB31_FSHD	46,XX	Blastocyst
hiPSC	CBiPS32-3F-10	46,XX	Cord Blood
	HKiPS4F	46,XX	Keratynocytes
Female control		46,XX	Lymphocytes

[1]DM1, myotonic dystrophy; CF, cystic fibrosis; MFS, Marfan syndrome; FSHD, Facioscapulohumeral muscular dystrophy; SCA7, Spinocerebellar ataxia 7; FXS, Fragile X Syndrome; APKD, Adult Polycystic Kidney Disease; CMT1A, Charcot-Marie-Tooth disease, demyelinating type 1ª

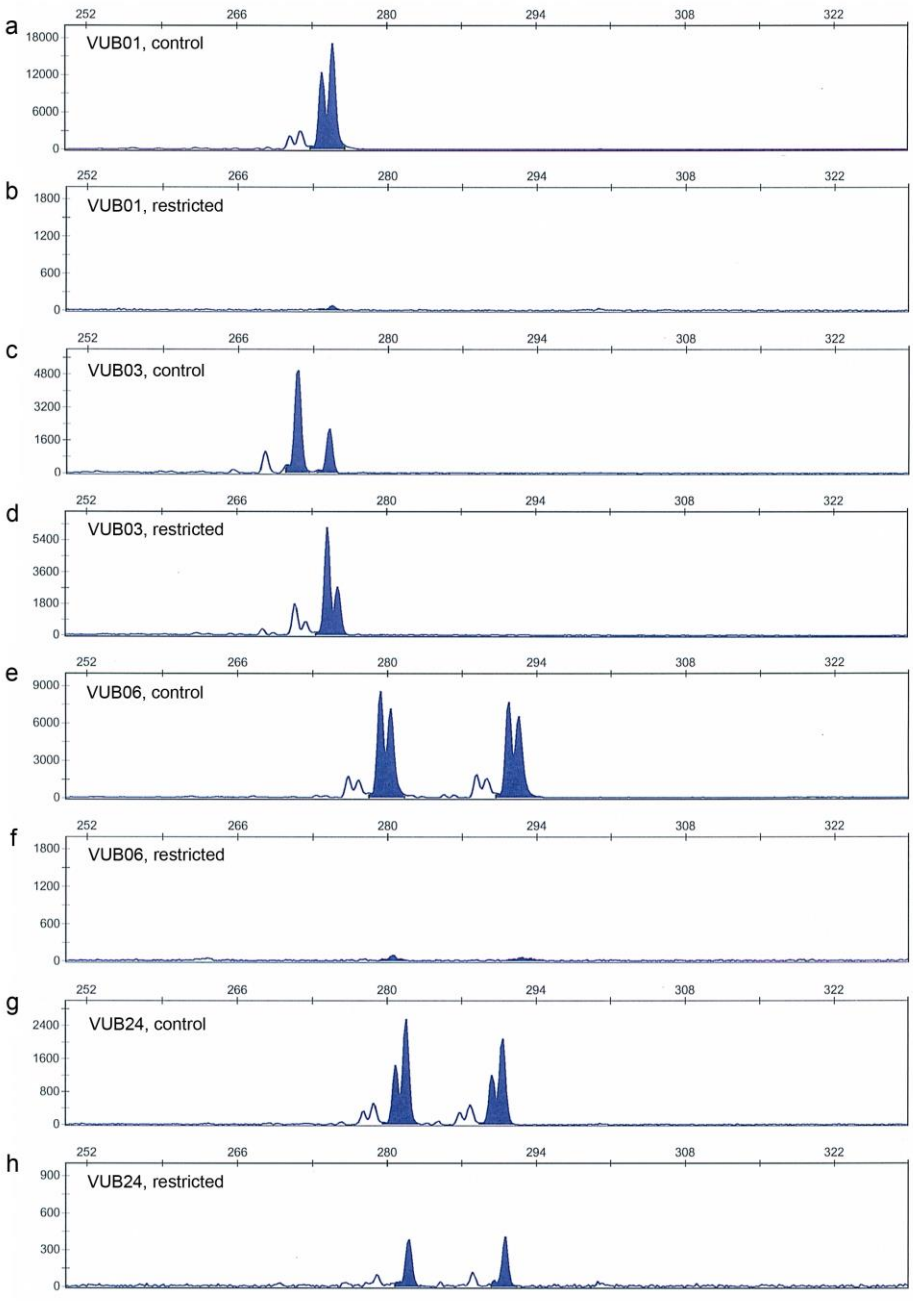
Supplementary Table 2. Determination of mono-/bi-allelic gene expression of X-linked genes. Determination of genotype and allele expression of the X-linked genes *AFF2*, *ATP7A* and *FGF13* and the corresponding XCI pattern in VUB07 and single cell-derived sublines. (P= passage number; S= subline number; SNP= single nucleotide polymorphism).

Cell line	Passage/subline	Gene name	SNP ID	Genotyping	Allel expression	XCI pattern[1]
VUB07	P42/original	AFF2	rs6641482	A/G	A	skewing
		ATP7A	rs2227291	C/G	C	skewing
		FGF13	rs2267628	A/G	A	skewing
	P50/S1	AFF2	rs6641482	A/G	A	skewing
		ATP7A	rs2227291	C/G	G/C	random
		FGF13	rs2267628	A/G	A	skewing
	P50/S4	AFF2	rs6641482	A/G	A	skewing
		ATP7A	rs2227291	C/G	C	skewing
		FGF13	rs2267628	A/G	A	skewing
	P50/S5	AFF2	rs6641482	A/G	A	skewing
		ATP7A	rs2227291	C/G	C	skewing
		FGF13	rs2267628	A/G	A	skewing
	P50/S8	AFF2	rs6641482	A/G	A/G	random
		ATP7A	rs2227291	C/G	C	skewing
		FGF13	rs2267628	A/G	A	skewing
	P50/S9	AFF2	rs6641482	A/G	A/G	random
		ATP7A	rs2227291	C/G	C	skewing
		FGF13	rs2267628	A/G	A	skewing
	P50/S10	AFF2	rs6641482	A/G	A/G	random
		ATP7A	rs2227291	C/G	C	skewing
		FGF13	rs2267628	A/G	A	skewing

Supplementary Table 3. Heterozygosity of hPSC lines. Overview of the heterozygosity for the short tandem repeats at the *ARX*, *ZDHHC15* and *SLITRK4* loci, as determined by PCR, for all lines in this study.

Sample	**ARX**	**ZDHHC15**	**SLITRK4**
bES13	Heterozygous	Heterozygous	Heterozygous
CBIPS32-3F-10	Heterozygous	homozygous	Heterozygous
ES10	Heterozygous	Heterozygous	homozygous
ES6 (male)	single allele	single allele	single allele
female control (lympho)	Heterozygous	homozygous	homozygous
HKIPS4F	Heterozygous	Heterozygous	Heterozygous
pES[12]	homozygous	homozygous	homozygous
VUB01 (male)	single allele	single allele	single allele
VUB03	Heterozygous	Heterozygous	Heterozygous
VUB04	Heterozygous	homozygous	homozygous
VUB06	Heterozygous	Heterozygous	Heterozygous
VUB07	Heterozygous	Heterozygous	Heterozygous
VUB08	Heterozygous	Heterozygous	homozygous
VUB09	Heterozygous	homozygous	Heterozygous
VUB10	Heterozygous	Heterozygous	Heterozygous
VUB11	homozygous	homozygous	Heterozygous
VUB12	Heterozygous	Heterozygous	Heterozygous
VUB13	Heterozygous	Heterozygous	Heterozygous
VUB14	Heterozygous	Heterozygous	Heterozygous
VUB17	Heterozygous	Heterozygous	Heterozygous
VUB19	Heterozygous	Heterozygous	Heterozygous
VUB20	Heterozygous	Heterozygous	Heterozygous
VUB22	Heterozygous	not applicable	not applicable
VUB24	Heterozygous	Heterozygous	homozygous
VUB26	Heterozygous	homozygous	Heterozygous
VUB31	Heterozygous	homozygous	Heterozygous

Supplementary Figure 4. Determination of DNA methylation patterns at the *ARX* locus. This figure gives examples of the different methylation patterns observed at the ARX locus as determined by methylation-sensitive DNA restriction, followed by PCR for the *ARX* gene. The male VUB01 line was used as control for the DNA restriction. While in control DNA a single peak represented the single allele (a), absence of this fragment after DNA restriction indicated that the restriction protocol was optimal (b). In VUB03, a full restriction of the shortest fragment was indicative for non-random hypermethylation of the longest allele (c, d). As in VUB06 both alleles were absent after restriction, this line displayed a hypomethylated pattern (e, f), while the detection of two alleles in VUB24, both before and after the methylation-sensitive restriction, indicated a random methylation pattern.


a
VUB01, control
b
VUB01, restricted
c
VUB03, control
d
VUB03, restricted
e
VUB06, control
f
VUB06, restricted
g
VUB24, control
h
VUB24, restricted
252
266
280
294
308
322

Supplementary Table 4. Allele specific methylation of subcloned lines. Overview of the ratios of hypermethylated epi-alleles relative between the two different chromosomes in the original and single colony-derived sublines of VUB20 and bES[13]. The different alleles are discriminated based on the length of the short tandem repeat and are indicated as 1 and 2.

		DNA RESTRICTION	MASSIVE PARALLEL BISULPHITE SEQUENCING[1]		
Line	**Passage number**	**ARX**	**ARX**	**ZDHHC15**	**SLITRK4**
VUB20	13		1. 43%	1. 44%	1. 43%
			2. 57%	2. 56%	2 .57%
	Sub 1		1. 77%	1. 62%	1. 84%
			2. 23%	2. 38%	2. 16%
	Sub 2		1. 25%	1. 36%	1. 21%
			2. 75%	2. 64%	2. 78%
	Sub 5		1. 39%	1. 42%	1. 29%
			2. 61%	2. 58%	2. 70%
bES[13]	7		1. 32%	1. 53%	1. 52%
			2. 68%	2. 47%	2. 48%
	8		1. 19%	1. 42%	1. 53%
			2. 81%	2. 58%	2. 46%
	10		1. 17%	1. 46%	1. 56%
			2. 83%	2. 54%	2. 44%
	Sub 3.2		1. 22%	1. 46%	1. 56%
			2. 78%	2. 54%	2. 44%
	Sub 5. 2		1. 16%	1. 53%	1. 57%
			2. 84%	2. 46%	2. 42%
	Sub 6.2		1. 23%	1. 54%	1. 56%
			2. 77%	2. 46%	2. 44%

Supplementary Table 6. Parent of origin of the methylated X chromosomes. Results of the methylation analysis for *ARX* using methylation sensitive restriction on ten hPSC lines for those lines displaying skewed methylation patterns.

Sample	ARX methylation-sensitive restriction	XCI pattern
VUB03	Non-random methylation, ♂ methylated	Skewed XCI, ♀ active
VUB04	Non-random methylation, ♂ methylated	Skewed XCI, ♀ active
VUB09	Non-random methylation, ♂ methylated	Skewed XCI, ♀ active
VUB13	Non-random methylation, ♂ methylated	Skewed XCI, ♀ active
VUB14	Non-random methylation, ♂ methylated	Skewed XCI, ♀ active
VUB08	Non-random methylation, ♀ methylated	Skewed XCI, ♂ active
VUB19	Non-random methylation, ♀ methylated	Skewed XCI, ♂ active
VUB20	Non-random methylation, ♀ methylated	Skewed XCI, ♂ active
VUB31	Non-random methylation, ♀ methylated	Skewed XCI, ♂ active
VUB07	Non-random methylation, no donor DNA available	Skewed XCI, active allele unknown
ES[10]	Non-random methylation, no donor DNA available	Skewed XCI, active allele unknown
bES[13]	Non-random methylation, no donor DNA available	Skewed XCI, active allele unknown
HKiPS4F	Non-random methylation, no donor DNA available	Skewed XCI, active allele unknown
CBiPS32-3F-10	Non-random methylation, no donor DNA available	Skewed XCI, active allele unknown

Supplementary Table 7. Genetic content of hPSC lines. Overview of the genetic content of hPSC lines at relevant passages as determined by array comparative genomic hybridization.

Cell line	Passage at which genetic content was studied	Genetic content by array comparative genomic hybridization	Passage at which XCI was studied	XCI pattern as determined by methylation-sensitive DNA restriction
VUB03	30	46,XX	22	skewed
	98, 102	46,XX,dup(20)(q11.21)	139	skewed
VUB04	23, 29	46,XX	12	skewed
	66	46,XX,dup(5)(q14.2q35.3),del(18)(q21.2q23)	34	skewed
	117, 136	46,XX,dup(5)(q14.2q35.3),del(18)(q21.2q23)	144	skewed
VUB06	36	46,XX	36	hypomethylated
VUB07	32	46,XX	22	skewed
	77, 122, 131	46,XX,dup(20)(q11.21)	88	skewed
VUB09	100	46,XX,dup(11)(q24.2q25),dup(20)(q11.21)	114	skewed
VUB11	45	46,XX	30, 34	homozygous
VUB13	43	46,XX, dup(5)(q21.3q35.3), del(18)(q21.2q23)	43	skewed
VUB14	13, 14	46,XX	21, 53	skewed
VUB17	7	46,XX	7, 10, 50	hypomethylated
VUB19	5	46,XX	21	skewed
	32	46,XX	71	skewed
VUB20	7	46,XX,dup(5)(q13.2)	13	random
	18	46,XX,dup(5)(q13.2)	23	skewed
VUB24	10	46,XX	5	random
	31	46,XX	41	skewed
VUB26	10	46,XX,dup(7)(q33qter), del(18)(q23qter)	10, 16, 30	hypomethylated
VUB31	42	46,XX	3	skewed

9 General discussion and future perspectives

Many human diseases are complex and their phenotypes vary among individuals. This can make it difficult to reach a correct understanding of the cause and the involved mechanisms; knowledge that is necessary to develop the best possible treatment. Considering the obvious ethical concerns that human experimentation rise, disease modelling brings the opportunity to explore how a specific disease works in an animal, a computer or in a cell in culture.

Animal disease models are valuable tools for the study of human disease, although they often do not fully reproduce all aspects of the biology of the disease, even if the animals present many of the symptoms associated to the human version of the disease.

Other limitations of experimentation with animal models are the long time required to generate transgenic animals and labour-intensiveness of performing research on them. Animal research is typically expensive and holds ethical implications as well. And more importantly, treatments that are found to work in animals do not necessarily work in humans. One of the solutions to overcome these problems is to develop human models, and in this context, hPSC, in addition to their promising potential in regenerative medicine and cell replacement, they are also an excellent option for *in vitro* human disease research modelling.

Stem cell modelling is one of the most recent and promising technologies for the study of human disease. hPSC ability of unlimited self-renewal and capacity to differentiate into specialized cell types of interest allow for the study of cell-type specific responses to a certain treatment, multiple times and in large samples. This is in contrast to the use of primary cultures of adult somatic cells. These cultures are challenging to be maintained and expanded in culture due to their low proliferation capacity, unless immortalized. In addition, the relevant cells for the study of a certain diseases are not easy or impossible to be cultured or to obtain, such as oligodendrocytes, astrocytes or striatal neurons.

9.1 HPSC AS RESEARCH MODELS FOR HD AND DM1

The primary aim of the first part of the thesis was to carry out the first steps needed

to evaluate the potential of hPSC as cellular models in the study of HD and DM1. At the start of our study, the derivation of a large number of hESC lines carrying different types of monogenic diseases had already been reported worldwide (Mateizel, 2005; Verlinsky et al., 2005), but few research had been done on them. In our work, we used four of such hESC cell lines, derived in our laboratory. Three lines were obtained from PGD embryos diagnosed with DM1 and one line was derived from an HD affected embryo. We hypothesized that these hESC cell lines could be used as *in vitro* cellular models to study the mechanisms underlying TNR instability, and that these cells could recapitulate the behaviour of the TNR *in vivo*.

In a nutshell, we characterized the behaviour of the TNR in the DM1 and HD locus in hESC cell lines carrying expansions in these sites, and its correlation to the expression levels of MMR genes, in both the undifferentiated and differentiated state. For this, we differentiated the cells into osteogenic progenitor-like cells and neural progenitor cells by directed differentiation, and induced spontaneous differentiation by allowing the cells to form teratomas in mice.

9.1.1 DM1

In our hESC DM1 cell lines, we observed a highly unstable CTG repeat from very early passages onwards and observed both expansions and contractions of the TNR, in analogy to what has been seen in DM1 patients.

A recent paper published after our work, studied different hiPSC clones derived from two fibroblast lines obtained from DM1 patients and obtained similar results regarding the instability of the repeat (Du et al., 2013). Each clone came from an individual fibroblast, and was analysed shortly after derivation, which provided an image of the TNR size variability in the fibroblast population, and in later passages, which revealed its instability in culture. The different clones presented high variability in the repeat size, reflecting the somatic instability in the fibroblast sample. Furthermore, in concordance with our results, the hiPSC showed contractions and expansions during culture. Interestingly, they observed that one of the clones with 57 CTG·CAG repeats was stable while two other clones with higher number of repeats, 126 and 773 repeats, presented instability with an expansion rate of 1 and 22 repeat per passage respectively. They established that the repeat expansion rate in hiPSC was length dependent and that between the threshold of 57 and 126 repeats the expansion rate dramatically increased (Du et al., 2013).

When our hESC-DM1 cell lines were differentiated into OPLs, the repeat underwent

stabilization when compared to the undifferentiated source in less than 11 passages (more than 14 weeks of culture). Du et al (2013) also described an overall tendency towards the stabilization of the repeat within 6 weeks of DM1 hiPSC differentiation into embryoid bodies. Similarly, they observed that when DM1 hiPSC were differentiated into neurospheres, the repeat became stable after 2 to 6 weeks of differentiation. (Du et al., 2013). In analogy, stabilization of TNR expansions upon differentiation has also been described in Friedreich Ataxia (Du et al., 2012; Hick et al., 2013).

Next, we investigated the evolution of the gene expression of the MMR genes hMSH2, hMSH3, hMSH6, hPMS1, hPMS2, hMLH1 and hMLH3 during differentiation. We found a significantly higher expression of these genes in undifferentiated cells than in DM1 hESC-derived OPLs cells, hMSH2, hMSH3 and hMSH6 showing the strongest reduction in expression. In addition, the down-regulation of the MMR proteins expression correlated with the stabilization of the repeat. Similar results were observed in the DM1 hiPSC cell lines of Du et al. The MMR expression was similar among different clones and to the hESC control line, and higher than the levels of MMR proteins in the fibroblasts of origin (Du et al., 2013). Transgenic down-regulation of MSH2 induced the stabilization of the expansion. However, they did not observed significant down-regulation of the MMR in DM1 hiPSC compared to neurospheres.

In summary, we observed a significant TNR instability for the DM1 locus in hESC, and that differentiation resulted in a stabilization of the repeat. This stabilization was concommitant with a downregulation of the MMR. Our results have been replicated in hiPSC by Du et al. showing their reproducibility and suggesting they may be extrapolated to other hPSC lines worldwide.

More recently, other groups have engaged in very exciting research on the ability of hiPSC to recapitulate cellular characteristics of DM1. For instance, Xia and collaborators have shown the presence of CUG foci in the nucleus of the hiPSC differentiated into neural cells (Xia et al., 2013). Marteyn et al. have studied the impairment of neurone and synapsis generation (Marteyn et al., 2011) and Denis et al proved a proliferation defect of DM1 hESC-derived neural stem cells (Denis et al., 2013).

Another interesting venue that has been explored is the possibility to carry out genome editing in hPSC carrying a DM1 mutation, with an eye to using these cells in an autologous cell replacement therapy or to generate hPSC lines with the same genetic background except for the DM1 mutation. Xia et al. (2015) performed *in vitro* genome editing to prevent the production of toxic *DMPK* mutant transcripts and eliminate pathogenic nuclear CUG RNA foci. For this, they integrated exogenous polyA

signals upstream of the *DMPK* CTG repeat-expansion. They used TALENs to generate double strand DNA breaks allowing the integration of the polyA tails that blocked the production of the mRNAs with expanded CUG repeats. This prevented the production of toxic RNA and resulted in a phenotype reversal in DM1 hiPS-cells derived stem cells. With this study they provide proof of concept that genome editing can prevent the production of toxic mutant transcripts and reverse the observed diseased cell phenotypes in DM1 stem cells (Xia et al., 2015)

9.1.2 HD

Previous to our work, little work had been done on TNR instability for the HD locus in human cultured cells. Lymphoblasts from HD patients show a stable number of repeats during culture (Cannella et al., 2009) and only a moderate instability was observed in immortalized fibroblasts derived from transgenic HD mice with an expansion of 155 repeats (Manley et al., 1999). Furthermore, one study reported the derivation of two HD-derived hESC cell lines, SI-186 $(CAG)_{37}$ and SI-187 $(CAG)_{51}$. They studied the stability of the mutant allele by small pool PCR and found a stable CAG when in undifferentiated hPSC. Conversely, when they differentiated the cells into forebrain neurons, they observed very modest instability in two out of six analysed pools (Niclis et al., 2009).

In our study, and contrarily to the DM1 locus, the HD repeat in VUB05_HD was very stable in all conditions, both in undifferentiated hESC and cells differentiated into osteogenic progenitor-like cells, teratoma cells and neural progenitors. This is quite in line with the data mentioned above, in which human cultured cells in the general, including hESC show very limited TNR in the HD locus.

Since the publication of our study, several groups have characterized new hESC lines carrying HD expansions, including their differentiation into neuronal lineages (Bradley et al., 2011; Niclis, 2013).

An important point for the progress in this field of research has been the fast evolution of the protocols for hPSC neuronal differentiation. It is now possible to obtain a great number of the relevant cells for the study of HD, namely the medium spiny striatal neurons expressing dopamine and cAMP-regulated phosphoprotein (DARPP32+ cells). The differentiation of hPSC into DARPP32+/GABA+ striatal projection neurons cells was first achieved by Aubry et al. (2008; (Aubry et al., 2008)), but since then, other improved protocols have been described that result in highly pure populations (Ma et al., 2012).

In the follow-up study of the group of Niclis et al. (Niclis, 2013), the authors

performed a deeper study of two previously described hESC cell lines. After extended *in vitro* culture and cellular acquisition of a forebrain GABAergic neuron identity, the authors found that the glutamate signalling was indeed perturbed, which is indicative of Huntington's pathology. They detected low-level instability in neurospheres derived from these hESC lines. These results are not concordant with ours, but can be explained by the fact that the cells we used as differentiated models were not close enough to the cell type affected in the actual disease, and that the protocols for differentiation were not as advanced as in more recent years. Similarly, the group of Jacquet et al. found no TNR instability in three hESC lines carrying an HD expansion, nor in their differentiated progeny (spontaneously *in vivo* differentiated into teratomas and directed differentiation into cardiomyocytes) (Du et al., 2013; Jacquet et al., 2015).

Regarding hiPSC, different studies report different findings. While some studies reported that the HD TNR was stable in both the differentiated and undifferentiated state (Du et al., 2013; Camnasio et al., 2012), others reported instability of the repeat upon differentiation (Consortium, 2012; Du et al., 2013). In the work of Camnasio et al., no change in the length of the pathological CAG in HD-derived hiPS was found during reprogramming, after long-term culture (more than 40 passages in vitro) or differentiation into neurons. Differently, the HDiPSC study described CAG mild instability of the normal and expanded allele for most of the hiPSC cell lines studied with passage and upon differentiation into neural stem cells. One cell line presented stabilization of the repeat in the normal allele in all conditions with just a minor increase (<10%) of the repeat in the expanded allele upon 26 passages as neural stem cells (Consortium, 2012; Du et al., 2012). Taking together the reports from different groups, it appears that some hPSC lines are more prone to CAG expansion *in vitro* than others.

One of the issues our study did not address, but are very important for the use of hPSC as research models, is the evaluation of the cellular impact of the pathogenic mutant allele expression. In this direction, hESC (Niclis, 2013) and hiPSC lines have been analysed for the appearance of HD hallmarks (Castiglioni et al., 2012; Consortium, 2012; Jeon et al., 2012). Several HD phenotypes have been observed in hESC, including transcriptional dysregulation, CAG repeat instability, mutant HTT aggregates, cholesterol biosynthesis perturbation, lysosomal dysfunction and neuronal vulnerability (Niclis et al., 2013). Camnasio et al. found, for instance, that the lysosomal activity from HD-derived iPS was higher compared to control wild type hiPS cell lines. This finding is relevant because both in mouse and rat HD models have been

described to have an increased autophagy associated to clearance of organelles and misfolded proteins such as mHTT (Kegel et al., 2000) This process would contribute to neuron *in vitro* preservation by eliminating accumulation of misfolded proteins.

In 2012, the HD IPSC Consortium carried out an exhaustive cellular characterization of 8 hiPSC, a control healthy line and a miPSC cell line. They studied their viability, pluripotency and differentiation capacity, mitochondrial functionality and neural degeneration. Several phenotypes were observed, including transcriptional dysregulation, mutant HTT aggregates, cholesterol biosynthesis perturbation, lysosomal dysfunction and neuronal vulnerability. They described typical phenotypes induced by the expansion of the CAG repeat, including cell toxicity. These findings are interesting because even if the cells do not show the exact mature pattern of HD affected cells, this shows that hiPSC are indeed very interesting models for the study of the HD disease as well as for drug testing.

Another interesting possibility when working with hESC as research models is to generate HD-affected hESC lines from hESC without an endogenous mutation. For this, a stable expression of cDNA encoding the mutant HTTexon1 fragment containing different lengths of the polyQ tail is induced in the cells by transgenic manipulation. Using this method, the cells have been shown to express mutant *Huntingtin* (mHTT) aggregates and soluble m-HTT dependent neurodegeneration characteristic from HD patients. These models are suitable, for instance, for studies of the mechanism of mHTT toxicity and high-throughput screening targeting either mHTT levels or toxicity. (Lu and Palacino, 2013)

The opposite direction is also true, in which a gene correction can be carried out on hPSC with an HD mutation. This provides a human model with the diseased cells and the corrected cells with the same genetic backgrounds. This could provide information on genetic differences between the diseased cells and the normal ones as well as provide us bringing us one step closer to cell replacement therapies. In one such study, the authors showed that the gene correction resulted in the restoration of the mitochondrial deficits, levels of BDNF and the altered cadherin and TGF b signalling associated to HD phenotype (Mahru et al., 2012; Xia et al., 2015).

Finally, there is the possibility of using hPSC in cell therapy. In that direction, several attempts have been made to correct autologous HD-hiPSC for its further use in treatment. An hiPSC line has been corrected by means of homologous recombination (An et al., 2012; Zwaka et al., 2003) as well as through electroporation (Cannella et al., 2009; Costa et al., 2007) and BAC based homologous recombination (Manley et al., 1999; Song et al., 2010).

9.1.3 Stem cells in the development of therapies for HD and DM1

The current therapies for HD and DM1 patients are based on pharmacological treatment of the three main clinical manifestations of the disease, namely, psychiatric, cognitive and motor abnormalities. The most common situation is that patients are treated with many different drugs, which, along with differences amongst studied populations and outcome measurements, makes the comparison between the efficiency of the different drugs is challenging.

It is key to bear in mind that the most pharmacological treatments can alleviate the symptoms but cannot cure the disease. Conversely, gene therapy and cell-based therapies have been considered as treatment for TNR patients. The main objective in gene therapy is to target the mutant gene and reduce its expression (for instance with RNAi or antisense oligonucleotides (ASOs)) while at the same time keeping the expression of the wild type gene unaltered. Instead, cell based therapies are intended to delay the progress of the disease by giving trophic support to the affected neurons and endogenous repair of brain areas affected by the disease, to directly replace the cells that have been lost and to modulate the inflammation which may be involved in the disease process.

Cells from different sources have been used for the study and treatment of TNR diseases, including foetal tissue cells, neural progenitor cells (NPC) and stem cells (SC). To date, several cell treatments have made it to clinical trials using Adipose Stem Cells (ASC), neural stem cells (NSC) and bone marrow stem cells (BMSC).

HD is spatially restricted to only one cell type, the projection neurons from the striatum, for this reason, cell-based treatments would only need to replace one cell type to treat this disease. In DM1, affected cells are disseminated in different regions of the organism, which makes it more difficult to design cell-based strategies to treat DM1 than HD. In line with this, most therapies have been tested in HD models and patients.

Foetal tissue grafts have been transplanted in clinical trials for HD. The treatment improved the motor and cognitive skills of the patients. However, the health improvements did not last long because the grafts did not get enough vascularization (BachoudLevi et al., 2006). Both murine and human NSC have been transplanted into HD models and showed partial functional recovery of the brain lesion at the site of transplantation (McBride et al., 2004; Niclis et al., 2009; Song et al., 2007; Visnyei et al., 2006).

Adult human stem cells, as for example BMSC and ASC, can be obtained in a

straightforward manner and afterwards differentiated into neural stem cells. BMSC have been shown to produce a neuroprotective effect when transplanted into HD patients. ASC transplanted in transgenic mice and rat models reduced the volumes of the areas with lesion in the brain. The paracrine effect of the adult stem cells is beneficial for the protection and repair, inhibiting the progression of the disease (Choong et al., 2007; Lee et al., 2009) (Im et al., 2010). It is well known that neural adult stem cells provide a neuroprotective effect. Unfortunately these cells are very challenging to culture and to expand *in vitro* for long time. Furthermore, they have a very short life once transplanted.

9.2 X-CHROMOSOME INACTIVATION IN HPSC

In comparison to the mouse, few is known about early human embryonic development. One of the bottlenecks obviously is the scarce and ethically loaded nature of human embryonic material. On the other hand, the direct extrapolation of results from mouse to human may lead to erroneous conclusions. Even though mouse and human share 99% of their genomes, there are still 1% of the human genes that are not represented in mouse. In addition, temporal and spatial human embryonic development is different between species, making it difficult to replicate human diseases and processes in mouse models. Human PSC are emerging as a new valuable alternative to animal modelling, particularly because of their closeness to the early human developmental cells. In this context, hPSCs may be of great use for disease modelling, drug testing, toxicity testing, biomarker identification and for the study of early developmental mechanisms (Colman and Dreesen, 2009; Tabar and Studer, 2014).

One of the earliest processes involved in embryonic development is the X-chromosome dosage compensation in female embryos. The correct chain of events of XCI, from the initiation to the maintenance, is crucial for the correct development of the embryo as well as for the adult cell physiology (Tompkins et al., 2012). Regarding the mechanisms behind XCI, most of the knowledge is based on mouse models. In the human, all research has been carried out on pre-implantation embryos, whilst there is no post implantation model. Here, hPSC could provide an interesting cellular model. As mentioned above, hPSC can represent a proxy of early human developmental cells, and can recapitulate *ex vivo* the XCI process that in animal models happen differently than in the human.

In adult cells, an interesting case is that of cancers, in which the erosion of the epigenetic marks of the inactivated X-chromosome can induce reactivation of a

number of genes that, can lead to other forms of genomic instability typically associated to cancer cells (Agrelo and Wutz, 2010). Also, in cancer cells, it has been shown that the X chromosome accumulates more mutations than any other autosome, as these mutations are tolerated when hitting the inactive chromosome (Richardson et al., 2006).

It is remarkable that cancer cells and hPSC have many common cellular, metabolic and genetic characteristics. For instance, hPSC carry numerous genetic abnormalities, many very similar to those found in cancer cells (Nguyen et al., 2013b). Even though several studies have proved that these mutant hPSC present normal morphology, express all the pluripotency markers and are capable of differentiate, their gene expression pattern is altered and present reduced differentiation capacity and increased malignant potential when compared to normal cell lines (Fazeli et al., 2011; Werbowetski-Ogilvie et al., 2009). Regarding the epigenetics, alterations of the imprinting and the X-chromosome dosage are typical in both pluripotent stem cells and cancer cells. Recent work on hiPSC showed that the loss of XIST is associated with increased expression of oncogenes (Anguera et al., 2012). Is therefore likely to be essential to maintain a stable XCI to avoid the risk of using hPSC with modified XCI both for disease modelling as well as for translation into the clinic. It is important to bear in mind that several studies have reported variability of XCI status among different hPSC cell lines as well as between sub-lines and different passages (Liu et al., 2011; Shen et al., 2008; Silva et al., 2008). Also, de-repression of X-linked genes was associated to 60% of the cultures of pluripotent cells (Bock et al., 2011; Boulting et al., 2011; Nazor et al., 2012). The loss of XCI marks has been shown to be of importance when using hiPSC as research models in the work of (Mekhoubad et al., 2012). The authors showed that female hiPSC lines that suffered erosion of XCI presented gradual de-repression of X-linked genes. In their work, they used hiPSC to model X-linked Lesch-Nyhan (LNS) disease. While cells with proper XCI recapitulated the disease phenotype in culture, hiPSC with eroded XCI did not show any characteristics of the disease phenotype upon differentiation. Subjecting the hPSC to a sequential cycle of reprogramming-differentiation-reprogramming showed that the XCI in eroded cell lines could not be restored to normal XCI after reprogramming.

In our work, we used hPSC as a cellular model for the study and more profound characterization of the XCI in cells close to human early developmental stages. This study also provides insight on the XCI status of a large number of different hPSC, and, based on the culture dynamics of the cells, we bring forward a hypothesis on the mechanisms behind the changes in XCI status observed in the cells.

We observed that hPSC cell lines display varying levels of loss of XCI marks with most of our cell lines (18/23 hESC) presenting a XaXi pattern with a loss of XCI marks, namely XIST and H3K27me3 expression. The loss of the XCI marks occurs rapidly during the first passages of *in vitro* culture. The fast loss of XCI marks has been hypothesized that it could be caused by sub-optimal culture conditions or cellular stress (Lengner et al., 2010; Shen et al., 2008). Evidence in favour of this hypothesis is that hPSC with two active X chromosomes have been derived and maintained in culture under different conditions that avoid precocious XCI ((Lengner et al., 2010; Tomoda et al., 2012).

A second observation in our study is that, whereas adult female cells present a random XCI pattern, all our cell lines showed a fully skewed pattern of XCI at passages over 10 (p>10). We found that partial skewing and random XCI was only present at very early passages after derivation, and the cultures would quickly transition to containing cells all inactivating the same X chromosome (full skewing). When revising the literature, it becomes apparent that this skewed pattern is the most frequently reported state of XCI reported.

In our work, we hypothesized that this culture take-over was due to a selective advantage rather than a found effect or a stochastic event. There is much evidence that hPSC lines from laboratories worldwide acquire genomic changes during in vitro culture, with late passage hPSC being twice as likely to have genomic changes than early passage hPSC (International Stem Cell Initiative et al., 2011). This strongly points at selective pressure of the sub-optimal culture systems as a major contributor to genomic instability in hPSC. Furthermore, suboptimal culture conditions appear to induce XCI, and represent an environment in which cells with a culture advantage could overtake the culture (Dvash et al., 2010; Hall and Lawrence, 2010; Lengner et al., 2010). To evaluate this possibility, we performed two pseudo-clonal growth experiments. We used two early passages of two lines that previous experiments had shown to still contain a random XCI. We then made sub-clones by using clumps of cells from the source culture, with the idea that if the skewed XCI inactivation was due to a selective advantage, after some time in culture, all our sub-clones should display the same skewing. Disappointingly, our results were non-conclusive. One of our source lines showed already a more advanced state of skewing than we hoped for. Hence, the fact that all sub-clones showed the same skewing did not unequivocally proved a selective advantage. In the second experiment, some of the clones showed skewing for one of the chromosomes, others for the other, and one was random.

We did not further pursue these experiments at the time of preparing the manuscript, but these results open the door to further interesting work to prove or disprove the hypothesis. One of the steps to take is to continue the culture of the sub-clones above mentioned. They were analysed only two passages after sub-cloning. If there is a selective advantage of a specific chromosome, longer culture should result in a skewing for this specific X chromosome. A second set of experiments would be to design competition assays.

In these, two cell populations, each with different allele inactivated, will be placed in competition to determine whether one allele takes over the culture. To isolate the cells of each type within one hPSC line, single-cell clones should be made from a line still at a random state, or with moderate skewing. One of the sub-lines could be modified to express GFP. Next, the lines can be mixed in a 50:50 ratio, and the evolution of the fraction of GFP positive cells can be tracked life. If replicated experiments prove that it is always the same X chromosome taking over the culture, this would provide strong evidence in favour of the hypothesis. In a next step, further analysis could be carried out in order to determine a differential gene expression that would provide a hint to the mechanism behind the selective advantage.

Finally, it is obvious that there is still work to be done identifying culture systems that protect hPSC from genome instability in the general, and, as illustrated by our work, from XCI erosion. Until ideal culture systems are in place, it is also imperative to continue monitoring the epigenetic and genetic instability undergone by cells in culture, since they can affect the use of hPSC, both as models for the study of human diseases and for therapeutic applications.

10 Conclusions

The aims of the first part of this thesis were:

1. Characterising the levels of instability of the CAG and CTG repeat during in vitro culture of hESC carrying expansions in the HD and DM1 locus.
2. Investigating the effect of differentiation into osteogenic progenitors-like cells and neural progenitors on the stability of the TNR in both diseases.
3. Investigate the levels of gene-expression of the MMR genes in both differentiated and undifferentiated cells, and its correlation to the behaviour of the TNR.

In our work, we observed a significant TNR instability for the DM1 locus in hESC, and that differentiation resulted in a stabilization of the repeat. This stabilization was concommitant with a downregulation of the MMR. Our results were later replicated in hiPSC by Du et al, showing their reproducibility and suggesting they may be extrapolated to other hPSC lines worldwide (Du et al., 2013).

Regarding the HD repeat, we found it was very stable in all conditions studied, both in undifferentiated hESC and cells differentiated into osteogenic progenitor-like cells, teratoma cells and neural progenitors. This is in line with other studies showing that hESC show very limited TNR in the HD locus. On the other hand, some groups have now reported some instability of this locus in cells differentiated into the neuronal lineage, to cells much closer to the cells that show instability in vivo. This was an experiment we were not able to carry out because of limitations in the differentiation protocols at the time of the study. (Bradley et al., 2011; Niclis, 2013)(Camnasio et al., 2012; Du et al., 2013; Jacquet et al., 2015)

Overall, hPSC appear to be a good in vitro model for the study of both DM1 and HD TNR instability, as the repeat follows in vitro the same patterns as found in vivo, including its dependency of the MMR machinery, particularly in the case of DM1.

The aims of the second part of the thesis were to investigate:

1. If hPSC present a preferential stage of XCI
2. If XCI is random or skewed
3. If there is a parent-of-origin effect when choosing which X-chromosome is inactivated in hPSC
4. If the pattern of XCI could be driven by a selective advantage
5. If this XCI pattern changes during long-term in culture and under different

culture conditions

6. If the XCI pattern in undifferentiated hESC is maintained upon differentiation into trophoblast-like cells as well as hESC-derived Osteoprogenitor like cells

In this study, we investigated the XCI pattern in inner cell mass-derived hESC lines, including one parthenogenetic embryo, hESC lines derived from single blastomeres at cleavage stage (Geens et al., 2009) and two hiPSC lines. Our results show that during culture hPSC rapidly progress from a XIST-dependent to a XIST-independent XCI state with loss of repressive histone modifications and erosion of methylation. We also report a remarkably high incidence of non-random XCI patterns. Moreover, our data show that this skewing of the methylation patterns is independent from the transition to the XIST-independent XCI state, to the parental origin of the X chromosome and that it is not occurring as a passenger event, driven by culture take-over due to a chromosomal aberration. These results suggest that XCI skewing is likely to be driven by the activation or repression of a specific allele on the X chromosome, conferring a growth or survival advantage to the cells.

Taken together, these results suggest caution when using hPSC as early human developmental research models with an eye on XCI. The eroded state of XCI found in many of the hPSC lines, and the frequency of skewed XCI patterns suggests that these cells are not a good proxy to early embryonic cells, at least what XCI is concerned. Conversely, they may still provide an interesting model to study gene function, particularly if the hypothesis of the driver gene is confirmed. Finally, research should be conducted to elucidate the impact of XCI erosion on hPSC lines used in a clinical setting.

11 References

Agrelo, R., and Wutz, A. (2010). ConteXt of change--X inactivation and disease. EMBO Molecular Medicine *2*, 6–15.

Anguera, M.C., Sadreyev, R., Zhang, Z., Szanto, A., Payer, B., Sheridan, S.D., Kwok, S., Haggarty, S.J., Sur, M., Alvarez, J., et al. (2012). Molecular signatures of human induced pluripotent stem cells highlight sex differences and cancer genes. Cell Stem Cell *11*, 75–90.

Aruga, J., and Mikoshiba, K. (2003). Identification and characterization of Slitrk, a novel neuronal transmembrane protein family controlling neurite outgrowth. Molecular and Cellular Neuroscience *24*, 117–129.

Ashizawa, T., Monckton, D.G., Vaishnav, S., Patel, B.J., Voskova, A., and Caskey, C.T. (1996). Instability of the Expanded (CTG)nRepeats in the Myotonin Protein Kinase Gene in Cultured Lymphoblastoid Cell Lines from Patients with Myotonic Dystrophy. Genomics *36*, 47–53.

Aubry, L., Bugi, A., Lefort, N., Rousseau, F., Peschanski, M., and Perrier, A.L. (2008). Striatal progenitors derived from human ES cells mature into DARPP32 neurons in vitro and in quinolinic acid-lesioned rats. Proceedings of the National Academy of Sciences *105*, 16707–16712.

Augui, S., Filion, G.J., Huart, S., Nora, E., Guggiari, M., Maresca, M., Stewart, A.F., and Heard, E. (2007). Sensing X Chromosome Pairs Before X Inactivation via a Novel X-Pairing Region of the Xic. Science *318*, 1632–1636.

Avery, S., Hirst, A.J., Baker, D., Lim, C.Y., and Alagaratnam, S. (2013). BCL-XL mediates the strong selective advantage of a 20q11. 21 amplification commonly found in human embryonic stem cell cultures. Stem Cell Reports.

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.

Barakat, T.S., Ghazvini, M., de Hoon, B., Li, T., Eussen, B., Douben, H., van der Linden, R., van der Stap, N., Boter, M., Laven, J.S., et al. (2015). Stable X Chromosome Reactivation in Female Human Induced Pluripotent Stem Cells. Stem Cell Reports *4*, 199–208.

Bellin, M., Marchetto, M.C., Gage, F.H., and Mummery, C.L. (2012). Induced pluripotent stem cells: the new patient? Nat Rev Mol Cell Biol *13*, 713–726.

Ben-Nun, I.F., and Benvenisty, N. (2006). Human embryonic stem cells as a cellular model for human disorders. Molecular and Cellular Endocrinology *252*, 154–159.

Ben-Yosef, D., Amit, A., Malcov, M., Frumkin, T., Ben-Yehudah, A., Eldar, I., Mey-Raz, N., Azem, F., Altarescu, G., Renbaum, P., et al. (2012). Female sex bias in human embryonic stem cell lines. Stem Cells and Development *21*, 363–372.

Benitez, J., Robledo, M., Ramos, C., Ayuso, C., Astarloa, R., Garcia Yébenes, J., and Brambati, B. (1995). Somatic stability in chorionic villi samples and other Huntington

fetal tissues. Hum Genet *96*, 229–232.

Berletch, J.B., Yang, F., and Disteche, C.M. (2010). Escape from X inactivation in mice and humans.

Bertelsen, B., Tümer, Z., and Ravn, K. (2011). Three new loci for determining x chromosome inactivation patterns. J Mol Diagn *13*, 537–540.

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell *144*, 439–452.

Boulting, G.L., Kiskinis, E., Croft, G.F., Amoroso, M.W., Oakley, D.H., Wainger, B.J., Williams, D.J., Kahler, D.J., Yamaki, M., Davidow, L., et al. (2011). A functionally characterized test set of human induced pluripotent stem cells. Nature Biotechnology *29*, 279–286.

Braam, S.R., Passier, R., and Mummery, C.L. (2009). Cardiomyocytes from human pluripotent stem cells in regenerative medicine and drug discovery. Trends in Pharmacological Sciences *30*, 536–545.

Bradley, C.K., Scott, H.A., Chami, O., Peura, T.T., Dumevska, B., Schmidt, U., and Stojanov, T. (2011). Derivation of Huntington's disease-affected human embryonic stem cell lines. Stem Cells and Development *20*, 495–502.

Brennand, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D., et al. (2011). Modelling schizophrenia using human induced pluripotent stem cells. Nature *473*, 221–225.

Brockdorff, N. (2002). X-chromosome inactivation: closing in on proteins that bind Xist RNA. *18*, 352–358.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature *448*, 191–195.

Brown, C.J., Lafreniere, R.G., Powers, V.E., and Sebastio, G. (1991). Localization of the X inactivation centre on the human X chromosome in Xq 13.

Brown, C.J., and Willard, H.F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. Nature *368*, 154–156.

Bruck, T., and Benvenisty, N. (2011). Meta-analysis of the heterogeneity of X chromosome inactivation in human pluripotent stem cells. Stem Cell Research *6*, 187–193.

Bruck, T., Yanuka, O., and Benvenisty, N. (2013). Human pluripotent stem cells with distinct X inactivation status show molecular and cellular differences controlled by the X-Linked ELK-1 gene. CellReports *4*, 262–270.

Cannavo, E., Marra, G., Sabates-Bellver, J., Menigatti, M., Lipkin, S.M., Fischer, F., Cejka, P., and Jiricny, J. (2005). Expression of the MutL Homologue hMLH3 in Human

Cells and its Role in DNA Mismatch Repair. Cancer Research.

Cannella, M., Maglione, V., Martino, T., Ragona, G., Frati, L., Li, G.-M., and Squitieri, F. (2009). DNA instability in replicating Huntington's disease lymphoblasts. BMC Med Genet *10*, 11.

Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400–404.

Castel, A.L., Cleary, J.D., and Pearson, C.E. (2010). Nat Rev Mol Cell Biol 2010 Castel. Nat Rev Mol Cell Biol *11*, 165–170.

Castiglioni, V., Onorati, M., Rochon, C., and Cattaneo, E. (2012). Induced pluripotent stem cell lines from Huntington's disease mice undergo neuronal differentiation while showing alterations in the lysosomal pathway. Neurobiology of Disease *46*, 30–40.

Cauffman, G., De Rycke, M., Sermon, K., Liebaers, I., and Van de Velde, H. (2009). Markers that define stemness in ESC are unable to identify the totipotent cells in human preimplantation embryos. Human Reproduction *24*, 63–70.

Chang, D.K., Ricciardiello, L., Goel, A., Chang, C.L., and Boland, C.R. (2000). Steady-state regulation of the human DNA mismatch repair system. J. Biol. Chem. *275*, 18424–18431.

Charbonneau, N., Amunugama, R., Schmutte, C., Yoder, K., and Fishel, R. (2014). Evidence that hMLH3 functions primarily in meiosis and in hMSH2-hMSH3 mismatch repair. Cancer Biology & Therapy *8*, 1411–1420.

Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced.

Chen, K.G., Mallon, B.S., McKay, R.D.G., and Robey, P.G. (2014). Human pluripotent stem cell culture: considerations for maintenance, expansion, and therapeutics. Cell Stem Cell *14*, 13–26.

Chong, J.J.H., Yang, X., Don, C.W., Minami, E., Liu, Y.-W., Weyers, J.J., Mahoney, W.M., Van Biber, B., Cook, S.M., Palpant, N.J., et al. (2014). Human embryonic-stem-cell-derived cardiomyocytes regenerate non-human primate hearts. Nature *510*, 273–277.

Choong, P.-F., Mok, P.-L., Cheong, S.-K., Leong, C.-F., and Then, K.-Y. (2007). Generating neuron-like cells from BM-derived mesenchymal stromal cells in vitro. Cytotherapy *9*, 170–183.

Cleary, J.D., and Pearson, C.E. (2003). The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. Cytogenet Genome Res *100*, 25–55.

Cleary, J.D., Tomé, S., López Castel, A., Panigrahi, G.B., Foiry, L., Hagerman, K.A., Sroka, H., Chitayat, D., Gourdon, G., and Pearson, C.E. (2010). Tissue- and age-specific DNA replication patterns at the CTG/CAG-expanded human myotonic dystrophy type 1 locus. Nat Struct Mol Biol *17*, 1079–1087.

Colman, A., and Dreesen, O. (2009). Pluripotent Stem Cells and Disease Modeling.

Cell Stem Cell *5*, 244–247.

Consortium, T.H.I. (2012). Induced Pluripotent Stem Cellsfrom Patients with Huntington's DiseaseShow CAG-Repeat-Expansion-Associated Phenotypes. Stem Cell *11*, 264–278.

Costa, M., Dottori, M., Sourris, K., Jamshidi, P., Hatzistavrou, T., Davis, R., Azzola, L., Jackson, S., Lim, S.M., Pera, M., et al. (2007). A method for genetic modification of human embryonic stem cells using electroporation. Nature Protocols *2*, 792–796.

De Paepe, C., Cauffman, G., Verloes, A., and Sterckx, J. (2013). Human trophectoderm cells are not yet committed. Human ….

De Paepe, C., Krivega, M., Cauffman, G., Geens, M., and Van de Velde, H. (2014). Totipotency and lineage segregation in the human embryo. Mol. Hum. Reprod. *20*, 599–618.

De Temmerman, N., Seneca, S., Van Steirteghem, A., Haentjens, P., Van der Elst, J., Liebaers, I., and Sermon, K.D. (2008). CTG repeat instability in a human embryonic stem cell line carrying the myotonic dystrophy type 1 mutation. Mol. Hum. Reprod. *14*, 405–412.

Dean, N.L., Tan, S.L., and Ao, A. (2006). Instability in the transmission of the myotonic dystrophy CTG repeat in human oocytes and preimplantation embryos. Fertility and Sterility *86*, 98–105.

Delmas, P., Cummings, S.R., Zoog, H.B., Martin, J.S., Austin, M., McClung, M.R., Wang, A., Siris, E.S., Eastell, R., Reid, I.R., et al. (2009). Denosumab for Prevention of Fractures in Postmenopausal Women with Osteoporosis. N. Engl. J. Med. *361*, 756–765.

Denis, J.A., Gauthier, M., Rachdi, L., Aubert, S., Giraud-Triboult, K., Poydenot, P., Benchoua, A., Champon, B., Maury, Y., Baldeschi, C., et al. (2013). mTOR-dependent proliferation defect in human ES-derived neural stem cells affected by myotonic dystrophy type 1. Journal of Cell Science *126*, 1763–1772.

Dhara, S.K., and Benvenisty, N. (2004). Gene trap as a tool for genome annotation and analysis of X chromosome inactivation in human embryonic stem cells. Nucleic Acids Research *32*, 3995–4002.

Diaz Perez, S.V., Kim, R., Li, Z., Marquez, V.E., Patel, S., Plath, K., and Clark, A.T. (2012). Derivation of new human embryonic stem cell lines reveals rapid epigenetic progression in vitro that can be prevented by chemical modification of chromatin. Human Molecular Genetics *21*, 751–764.

Donohoe, M.E., Silva, S.S., Pinter, S.F., Xu, N., and Lee, J.T. (2009). The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *460*, 128–132.

Dragileva, E., Hendricks, A., Teed, A., Gillis, T., Lopez, E.T., Friedberg, E.C., Kucherlapati, R., Edelmann, W., Lunetta, K.L., MacDonald, M.E., et al. (2009). Intergenerational and striatal CAG repeat instability in Huntington's disease knock-in mice involve different DNA repair genes. Neurobiology of Disease *33*, 37–47.

Du, J., Campau, E., Soragni, E., Ku, S., Puckett, J.W., Dervan, P.B., and Gottesfeld,

J.M. (2012). Role of Mismatch Repair Enzymes in GAA·TTC Triplet-repeat Expansion in Friedreich Ataxia Induced Pluripotent Stem Cells. Journal of Biological Chemistry *287*, 29861–29872.

Du, J., Campau, E., Soragni, E., Jespersen, C., and Gottesfeld, J.M. (2013). Length-dependent CTG·CAG triplet-repeat expansion in myotonic dystrophy patient-derived induced pluripotent stem cells. Human Molecular Genetics *22*, 5276–5287.

Dvash, T., Lavon, N., and Fan, G. (2010). Variations of X chromosome inactivation occur in early passages of female human embryonic stem cells. PLoS ONE *5*, e11330.

Ebert, A.D., Yu, J., Rose, F.F., Mattis, V.B., Lorson, C.L., Thomson, J.A., and Svendsen, C.N. (2008). Induced pluripotent stem cells from a spinal muscular atrophy patient. Nature *457*, 277–280.

Edwards, R.G., and Beard, H.K. (1997). Oocyte polarity and cell determination in early mammalian embryos. Mol. Hum. Reprod. *3*, 863–905.

Eggan, K., Akutsu, H., hochedlinger, K., rideout, W., Yanagimachi, R., and jaenisch, R. (2000). X-Chromosome inactivation in cloned mouse embryos. *290*, 1578–1581.

Eguizabal, C., Montserrat, N., Veiga, A., and Belmonte, J. (2013). Dedifferentiation, Transdifferentiation, and Reprogramming: Future Directions in Regenerative Medicine. Seminars in Reproductive Medicine *31*, 082–094.

Eiges, R., Schuldiner, M., Drukker, M., and Yanuka, O. (2001). Establishment of human embryonic stem cell-transfected clones carrying a marker for undifferentiated cells. Current Biology.

Eiges, R., Urbach, A., Malcov, M., Frumkin, T., Schwartz, T., Amit, A., Yaron, Y., Eden, A., Yanuka, O., Benvenisty, N., et al. (2007). Developmental Study of Fragile X Syndrome Using Human Embryonic Stem Cells Derived from Preimplantation Genetically Diagnosed Embryos. Cell Stem Cell *1*, 568–577.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature.

Fazeli, A., Liew, C.-G., Matin, M.M., Elliott, S., Jeanmeure, L.F.C., Wright, P.C., Moore, H., and Andrews, P.W. (2011). Altered patterns of differentiation in karyotypically abnormal human embryonic stem cells. Int. J. Dev. Biol. *55*, 175–180.

Flach, G., Johnson, M.H., Braude, P.R., Taylor, R.A., and Bolton, V.N. (1982). The transition from maternal to embryonic control in the 2-cell mouse embryo. The EMBO Journal *1*, 681–686.

Foiry, L., Dong, L., Savouret, C., Hubert, L., Riele, te, H., Junien, C., and Gourdon, G. (2006). Msh3 is a limiting factor in the formation of intergenerational CTG expansions in DM1 transgenic mice. Hum Genet *119*, 520–526.

Fortune, M.T., Anvret, M., Vassilopoulos, C., Tan, S.L., Coolbaugh, M.I., Ahlberg, G., Siciliano, M.J., Ao, A., Monckton, D.G., Grandell, U., et al. (1993). Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. Human Molecular Genetics *2*, 1397–1400.

Fortune, M.T., Vassilopoulos, C., Coolbaugh, M.I., Siciliano, M.J., and Monckton, D.G.

(2000). Dramatic, expansion-biased, age-dependent, tissue-specific somatic mosaicism in a transgenic mouse model of triplet repeat instability. Human Molecular Genetics *9*, 439–445.

Fraga, A.M., Souza de Araújo, É.S., Stabellini, R., Vergani, N., and Pereira, L.V. (2011). A survey of parameters involved in the establishment of new lines of human embryonic stem cells. Stem Cell Rev *7*, 775–781.

G Jansen, P.W.M.C.W.N.H.S.L.V.C.H.H.B.B.W. (1994). Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in (CTG)n repeat variation and selection against extreme expansion in sperm. Am. J. Hum. Genet. *54*, 575.

Gacy, A.M., Goellner, G.M., Spiro, C., Chen, X., and Gupta, G. (1998). GAA instability in Friedreich's ataxia shares a common, DNA-directed and intraallelic mechanism with other trinucleotide diseases. Molecular Cell.

Gacy, A.M., Goellner, G., Juranić, N., Macura, S., and McMurray, C.T. (1995). Trinucleotide repeats that expand in human disease form hairpin structures in vitro. Cell *81*, 533–540.

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. Nature *504*, 282–286.

Ganesan, S., Richardson, A.L., Wang, Z.C., Iglehart, J.D., Miron, A., Feunteun, J., Silver, D., and Livingston, D.M. (2005). Abnormalities of the inactive X chromosome are a common feature of BRCA1 mutant and sporadic basal-like breast cancer. Cold Spring Harb. Symp. Quant. Biol. *70*, 93–97.

Geens, M., Mateizel, I., Sermon, K., De Rycke, M., Spits, C., Cauffman, G., Devroey, P., Tournaye, H., Liebaers, I., and Van de Velde, H. (2009). Human embryonic stem cell lines derived from single blastomeres of two 4-cell stage embryos. Human Reproduction *24*, 2709–2717.

Gerecht-Nir, S., and Itskovitz-Eldor, J. (2004). The promise of human embryonic stem cells. Best Practice & Research Clinical Obstetrics & Gynaecology *18*, 843–852.

Giorgetti, A., Montserrat, N., Aasen, T., González, F., Rodriguez-Piza, I., Vassena, R., Raya, A., Boué, S., Barrero, M.J., Corbella, B.A., et al. (2009). Generation of induced pluripotent stem cells from human cord blood using OCT4 and SOX2. Cell Stem Cell *5*, 353–357.

Giorgetti, A., Montserrat, N., Rodriguez-Piza, I., Azqueta, C., Veiga, A., and Belmonte, J.C.I. (2010). Generation of induced pluripotent stem cells from human cord blood cells with only two factors: Oct4 and Sox2. Nature Protocols *5*, 811–820.

Gomes-Pereira, M., Fortune, M.T., Ingram, L., McAbney, J.P., and Monckton, D.G. (2004). Pms2 is a genetic enhancer of trinucleotide CAG·CTG repeat somatic mosaicism: implications for the mechanism of triplet repeat expansion. Human Molecular ….

Gonitel, R., Moffitt, H., Sathasivam, K., Woodman, B., Detloff, P.J., Faull, R.L.M., and Bates, G.P. (2008). DNA instability in postmitotic neurons. Proc. Natl. Acad. Sci. U.S.a. *105*, 3467–3472.

González, F., Boué, S., and Belmonte, J.C.I. (2011). Methods for making induced pluripotent stem cells: reprogramming à la carte. Nat Rev Genet *12*, 231–242.

Goula, A.-V., Berquist, B.R., Wilson, D.M., Wheeler, V.C., Trottier, Y., and Merienne, K. (2009). Stoichiometry of base excision repair proteins correlates with increased somatic CAG instability in striatum over cerebellum in Huntington's disease transgenic mice. PLoS Genet *5*, e1000749.

Halabi, A., Ditch, S., Wang, J., and Grabczyk, E. (2012). DNA mismatch repair complex MutSβ promotes GAA·TTC repeat expansion in human cells. Journal of Biological Chemistry *287*, 29958–29967.

Hall, L.L., and Lawrence, J.B. (2010). XIST RNA and architecture of the inactive X chromosome: implications for the repeat genome. Cold Spring Harb. Symp. Quant. Biol. *75*, 345–356.

Hall, L.L., Byron, M., Butler, J., Becker, K.A., Nelson, A., Amit, M., Itskovitz-Eldor, J., Stein, J., Stein, G., Ware, C., et al. (2008). X-inactivation reveals epigenetic anomalies in most hESC but identifies sublines that initiate as expected. J. Cell. Physiol. *216*, 445–452.

Hanley, S.C., Assouline-Thomas, B., Makhlin, J., and Rosenberg, L. (2011). Epidermal growth factor induces adult human islet cell dedifferentiation. J. Endocrinol. *211*, 231–239.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601.

Heard, E., Mongelard, F., Arnaud, D., and Avner, P. (1999). Xist yeast artificial chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies.

Hochedlinger, K., and Jaenisch, R. (2006). Nuclear reprogramming and pluripotency. Nature *441*, 1061–1067.

Hochedlinger, K., and Plath, K. (2009). Epigenetic reprogramming and induced pluripotency. Development *136*, 509–523.

Hoffman, L.M., Hall, L., Batten, J.L., Young, H., Pardasani, D., Baetge, E.E., Lawrence, J., and Carpenter, M.K. (2005). X-inactivation status varies in human embryonic stem cell lines. Stem Cells *23*, 1468–1478.

Hotta, A., and Ellis, J. (2008). Retroviral vector silencing during iPS cell induction: an epigenetic beacon that signals distinct pluripotent states. J. Cell. Biochem. *105*, 940–948.

Im, W., Lee, S.-T., Park, J.-E., Oh, H.J., Shim, J., Lim, J., Chu, K., and Kim, M. (2010). Transplantation of patient-derived adipose stem cells in YAC128 Huntington's disease transgenic mice. PLoS Curr *2*.

International Stem Cell Initiative, Amps, K., Andrews, P.W., Anyfantis, G., Armstrong, L., Avery, S., Baharvand, H., Baker, J., Baker, D., Munoz, M.B., et al. (2011). Screening ethnically diverse human embryonic stem cells identifies a chromosome 20 minimal amplicon conferring growth advantage. Nature Biotechnology *29*, 1132–

1144.

Iyer, R.R., Pluciennik, A., Burdett, V., and Modrich, P.L. (2006). DNA mismatch repair: functions and mechanisms. Chemical Reviews.

Jacquet, L., Neueder, A., Földes, G., Karagiannis, P., Hobbs, C., Jolinon, N., Mioulane, M., Sakai, T., Harding, S.E., and Ilic, D. (2015). Three Huntington's Disease Specific Mutation-Carrying Human Embryonic Stem Cell Lines Have Stable Number of CAG Repeats upon In Vitro Differentiation into Cardiomyocytes. PLoS ONE *10*, e0126860.

Jeon, I., Lee, N., Li, J.-Y., Park, I.-H., Park, K.S., Moon, J., Shim, S.H., Choi, C., Chang, D.-J., Kwon, J., et al. (2012). Neuronal properties, in vivo effects, and pathology of a Huntington's disease patient-derived induced pluripotent stem cells. Stem Cells *30*, 2054–2062.

Jiricny, J. (2006). The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol *7*, 335–346.

Jopling, C., Boué, S., and Belmonte, J.C.I. (2011). Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. Nat Rev Mol Cell Biol *12*, 79–89.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature *458*, 771–775.

Kang, S., Jaworski, A., Ohshima, K., and Wells, R.D. (1995). Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. Nat Genet *10*, 213–218.

Kegel, K.B., Kim, M., Sapp, E., McIntyre, C., Castaño, J.G., Aronin, N., and DiFiglia, M. (2000). Huntingtin expression stimulates endosomal-lysosomal activity, endosome tubulation, and autophagy. J. Neurosci. *20*, 7268–7278.

Kennedy, L. (2003). Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. Human Molecular Genetics *12*, 3359–3367.

Khajavi, M., Tari, A.M., Patel, N.B., Tsuji, K., Siwak, D.R., Meistrich, M.L., Terry, N.H.A., and Ashizawa, T. (2001). "Mitotic drive" of expanded CTG repeats in myotonic dystrophy type 1 (DM1). Human Molecular ….

Kim, D., Kim, C.-H., Moon, J.-I., Chung, Y.-G., Chang, M.-Y., Han, B.-S., Ko, S., Yang, E., Cha, K.Y., Lanza, R., et al. (2009). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell Stem Cell *4*, 472–476.

Klimanskaya, I., Chung, Y., Becker, S., Lu, S.-J., and Lanza, R. (2006). Human embryonic stem cell lines derived from single blastomeres. Nature *444*, 481–485.

Kovtun, I.V., and McMurray, C.T. (2008). Features of trinucleotide repeat instability in vivo. Cell Research *18*, 198–213.

Kriks, S., Shim, J.-W., Piao, J., Ganat, Y.M., Wakeman, D.R., Xie, Z., Carrillo-Reid, L., Auyeung, G., Antonacci, C., Buch, A., et al. (2011). Dopamine neurons derived from

human ES cells efficiently engraft in animal models of Parkinson's disease. Nature.

Kroon, E., Martinson, L.A., Kadoya, K., Bang, A.G., Kelly, O.G., Eliazer, S., Young, H., Richardson, M., Smart, N.G., Cunningham, J., et al. (2008). Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. Nature Biotechnology *26*, 443–452.

Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. Annual Review of Biochemistry *74*, 681–710.

Lancaster, M.A., Renner, M., Martin, C.A., and Wenzel, D. (2013). Cerebral organoids model human brain development and microcephaly. Nature.

Lee, J.T., and Bartolomei, M.S. (2013). X-inactivation, imprinting, and long noncoding RNAs in health and disease. Cell.

Lee, S.-T., Chu, K., Jung, K.-H., Im, W.-S., Park, J.-E., Lim, H.-C., Won, C.-H., Shin, S.-H., Lee, S.K., Kim, M., et al. (2009). Slowed progression in models of Huntington disease by adipose stem cell transplantation. Ann. Neurol. *66*, 671–681.

Lengner, C.J., Gimelbrant, A.A., Erwin, J.A., Cheng, A.W., Guenther, M.G., Welstead, G.G., Alagappan, R., Frampton, G.M., Xu, P., Muffat, J., et al. (2010). Derivation of Pre-X Inactivation Human Embryonic Stem Cells under Physiological Oxygen Concentrations. Cell *141*, 872–883.

Lerou, P.H., Yabuuchi, A., Huo, H., Takeuchi, A., Shea, J., Cimini, T., Ince, T.A., Ginsburg, E., Racowsky, C., and Daley, G.Q. (2008). Human embryonic stem cell derivation from poor-quality embryos. Nature Biotechnology *26*, 212–214.

Lessing, D., Anguera, M.C., and Lee, J.T. (2013). X chromosome inactivation and epigenetic responses to cellular reprogramming. Annu Rev Genomics Hum Genet *14*, 85–110.

Lin, Y., and Wilson, J.H. (2009). Diverse effects of individual mismatch repair components on transcription-induced CAG repeat instability in human cells. DNA Repair *8*, 878–885.

Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905.

Liu, W., and Sun, X. (2009). Skewed X chromosome inactivation in diploid and triploid female human embryonic stem cells. Human Reproduction *24*, 1834–1843.

Liu, W., Yin, Y., Jiang, Y., Kou, C., Luo, Y., Huang, S., Zheng, Y., Li, S., Li, Q., Guo, L., et al. (2011). Genetic and epigenetic X-chromosome variations in a parthenogenetic human embryonic stem cell line. *28*, 303–313.

López Castel, A., Tomkinson, A.E., and Pearson, C.E. (2009). CTG/CAG repeat instability is modulated by the levels of human DNA ligase I and its interaction with proliferating cell nuclear antigen: a distinction between replication and slipped-DNA repair. Journal of Biological Chemistry *284*, 26631–26645.

Lu, B., and Palacino, J. (2013). A novel human embryonic stem cell-derived Huntington's disease neuronal model exhibits mutant huntingtin (mHTT) aggregates

and soluble mHTT-dependent neurodegeneration. The FASEB Journal 1–10.

Lund, R.J., Närvä, E., and Lahesmaa, R. (2012). Genetic and epigenetic stability of human pluripotent stem cells. Nat Rev Genet *13*, 732–744.

Luo, Y., Li, J., Zhu, D., Fan, Y., Li, S., and Sun, X. (2014). High-resolution chromosomal microarray analysis of early-stage human embryonic stem cells reveals an association between X chromosome instability and skewed X inactivation. Cell & Bioscience *12*, 429–442.

LYON, M.F. (1962). Sex chromatin and gene action in the mammalian X-chromosome. Am. J. Hum. Genet. *14*, 135–148.

Ma, L., Hu, B., Liu, Y., Vermilyea, S.C., Liu, H., Gao, L., Sun, Y., Zhang, X., and Zhang, S.-C. (2012). Human Embryonic Stem Cell-Derived GABA Neurons Correct Locomotion Deficits in Quinolinic Acid-Lesioned Mice. Cell Stem Cell *10*, 455–464.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., et al. (2007). Directly Reprogrammed Fibroblasts Show Global Epigenetic Remodeling and Widespread Tissue Contribution. Cell Stem Cell *1*, 55–70.

Mahru, A.C., Zhang, N., Scott, G., Montoro, D., Wittkop, T., Mooney, S., Melov, S., and Ellerby, L.M. (2012). Genetic Correction of Huntington's Disease Phenotypes in Induced Pluripotent Stem Cells. Stem Cell *11*, 253–263.

Mak, W., Nesterova, T.B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A.P., and Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. Science *303*, 666–669.

Manley, K., Lia, A.S., Shirley, T.L., Seznec, H., Flaherty, L., Hofmann-Radvanyi, H., Messer, A., Radvanyi, F., Duros, C., Saquet, C., et al. (1998). Somatic instability of the CTG repeat in mice transgenic for the myotonic dystrophy region is age dependent but not correlated to the relative intertissue transcription levels and proliferative capacities. Human Molecular Genetics *7*, 1285–1291.

Manley, K., Pugh, J., and Messer, A. (1999). Instability of the CAG repeat in immortalized fibroblast cell cultures from Huntington's disease transgenic mice. Brain Res. *835*, 74–79.

Marchand, M., Horcajadas, J.A., Esteban, F.J., McElroy, S.L., Fisher, S.J., and Giudice, L.C. (2011). Transcriptomic signature of trophoblast differentiation in a human embryonic stem cell model. Biology of Reproduction *84*, 1258–1271.

Marchetto, M.C.N., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H., and Muotri, A.R. (2010). A Model for Neural Development and Treatment of Rett Syndrome Using Human Induced Pluripotent Stem Cells. Cell *143*, 527–539.

Marra, G., and Jiricny, J. (2005). DNA mismatch repair and colon cancer. Advances in Experimental Medicine and Biology *570*, 85–123.

Marteyn, A., Maury, Y., Gauthier, M.M., Lecuyer, C., and Vernet, R. (2011). Mutant Human Embryonic Stem Cells Reveal Neurite and Synapse Formation Defects in Type 1 Myotonic Dystrophy. Cell Stem Cell.

Martí, M., Mulero, L., Pardo, C., Morera, C., Carrió, M., Laricchia-Robbio, L., Esteban, C.R., and Belmonte, J.C.I. (2013). Characterization of pluripotent stem cells. Nature Protocols *8*, 223–253.

Martínez, R., Bonilla-Henao, V., Jiménez, A., Lucas, M., Vega, C., Ramos, I., Sobrino, F., and Pintado, E. (2005). Skewed X inactivation of the normal allele in fully mutated female carriers determines the levels of FMRP in blood and the fragile X phenotype. Mol. Diagn. *9*, 157–162.

Martorell, L. (1997). Somatic instability of the myotonic dystrophy (CTG)n repeat during human fetal development. Human Molecular Genetics *6*, 877–880.

Mascetti, V.L., and Pedersen, R.A. (2014). Naiveté of the human pluripotent stem cell. Nature Biotechnology *32*, 68–70.

Masui, O., Bonnet, I., Le Baccon, P., Brito, I., Pollex, T., Murphy, N., Hupé, P., Barillot, E., Belmont, A.S., and Heard, E. (2011). Live-Cell Chromosome Dynamics and Outcome of X Chromosome Pairing Events during ES Cell Differentiation. Cell *145*, 447–458.

Mateizel, I. (2005). Derivation of human embryonic stem cell lines from embryos obtained after IVF and after PGD for monogenic disorders. Human Reproduction *21*, 503–511.

Mateizel, I., De Becker, A., Van de Velde, H., De Rycke, M., Van Steirteghem, A., Cornelissen, R., Van der Elst, J., Liebaers, I., Van Riet, I., and Sermon, K. (2008). Efficient differentiation of human embryonic stem cells into a homogeneous population of osteoprogenitor-like cells. Reproductive BioMedicine Online *16*, 741–753.

McBride, J.L., Behrstock, S.P., Chen, E.-Y., Jakel, R.J., Siegel, I., Svendsen, C.N., and Kordower, J.H. (2004). Human neural stem cell transplants improve motor function in a rat model of Huntington's disease. J. Comp. Neurol. *475*, 211–219.

McCracken, K.W., Catá, E.M., Crawford, C.M., Sinagoga, K.L., Schumacher, M., Rockich, B.E., Tsai, Y.-H., Mayhew, C.N., Spence, J.R., Zavros, Y., et al. (2014). Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. Nature *516*, 400–404.

McMurray, C.T. (1999). DNA secondary structure: A common and causative factor for expansion in human disease. Proceedings of the National Academy of Sciences *96*, 1823–1825.

McMurray, C.T. (2010). Mechanisms of trinucleotide repeat instability during human development. Nat Rev Genet *11*, 786–799.

Mekhoubad, S., Bock, C., de Boer, A.S., Kiskinis, E., Meissner, A., and Eggan, K. (2012). Erosion of dosage compensation impacts human iPSC disease modeling. *10*, 595–609.

Messer, A., Manley, K., Shirley, T.L., and Flaherty, L. (1999). Msh2 deficiency prevents in vivo somatic instability of the CAG repeat in Huntington disease transgenic mice - Nature Genetics. Nat Genet *23*, 471–473.

Mirkin, S.M. (2007). Expandable DNA repeats and human disease. *447*, 932–940.

Modrich, P. (2006). Mechanisms in eukaryotic mismatch repair. J. Biol. Chem. *281*, 30305–30309.

Monckton, D.G., Kovtun, I.V., Wong, L.J., McMurray, C.T., Ashizawa, T., and Caskey, C.T. (2001). Trinucleotide expansion in haploid germ cells by gap repair. Nat Genet *27*, 407–411.

Monkhorst, K., Jonkers, I., Rentmeester, E., Grosveld, F., and Gribnau, J. (2008). X inactivation counting and choice is a stochastic process: evidence for involvement of an X-linked activator. Cell *132*, 410–421.

Moore, H., Greenwell, P.W., Liu, C.P., Arnheim, N., and Petes, T.D. (1999). Triplet repeats form secondary structures that escape DNA repair in yeast. Proceedings of the National Academy of Sciences *96*, 1504–1509.

Moreira de Mello, J.C., de Araújo, E.S.S., Stabellini, R., Fraga, A.M., de Souza, J.E.S., 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.

Nahhas, F.A., Garbern, J., Krajewski, K.M., Roa, B.B., and Feldman, G.L. (2005). Juvenile onset Huntington disease resulting from a very large maternal expansion. Am. J. Med. Genet. A *137A*, 328–331.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2007). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nature Biotechnology *26*, 101–106.

Nakatani, R., Nakamori, M., Fujimura, H., Mochizuki, H., and Takahashi, M.P. (2015). Large expansion of CTG•CAG repeats is exacerbated by MutSβ in human cells. Sci Rep *5*, 11020.

Navarro, P., Oldfield, A., Legoupi, J., Festuccia, N., Dubois, A., Attia, M., Schoorlemmer, J., Rougeulle, C., Chambers, I., and Avner, P. (2011). Molecular coupling of Tsix regulation and pluripotency. Nature *468*, 457–460.

Nazor, K.L., Altun, G., Lynch, C., Tran, H., Harness, J.V., Slavin, I., Garitaonandia, I., Müller, F.-J., Wang, Y.-C., Boscolo, F.S., et al. (2012). Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. Cell Stem Cell *10*, 620–634.

Nguyen, H.T., Geens, M., and Mertzanidou, A. (2013a). Gain of 20q11. 21 in human embryonic stem cells improves cell survival by increased expression of Bcl-xL. Molecular Human ….

Nguyen, H.T., Geens, M., and Spits, C. (2013b). Genetic and epigenetic instability in human pluripotent stem cells. Hum. Reprod. Update *19*, 187–205.

Nguyen, H.T., Markouli, C., Geens, M., Barbé, L., Sermon, K., and Spits, C. (2014). Human embryonic stem cells show low-grade microsatellite instability. Mol. Hum. Reprod. *20*, 981–989.

Nichol Edamura, K., Leonard, M.R., and Pearson, C.E. (2005). Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells. Am. J. Hum.

Genet. *76*, 302–311.

Nichols, J., and Smith, A. (2009). Perspective. Stem Cell *4*, 487–492.

Niclis, J.C., Trounson, A.O., Dottori, M., Ellisdon, A.M., Bottomley, S.P., Verlinsky, Y., and Cram, D.S. (2009). ArticleHuman embryonic stem cell models of Huntington disease. Reproductive BioMedicine Online *19*, 106–113.

Niclis, J.C. (2013). Characterization of forebrain neurons derived from late-onset Huntington's disease human embryonic stem cell lines. 1–13.

Niclis, J.C., Pinar, A., Haynes, J.M., Alsanie, W., Jenny, R., Dottori, M., and Cram, D.S. (2013). Characterization of forebrain neurons derived from late-onset Huntington's disease human embryonic stem cell lines. Front Cell Neurosci *7*, 37.

Numasawa, Y., Kimura, T., Miyoshi, S., Nishiyama, N., Hida, N., Tsuji, H., Tsuruta, H., Segawa, K., Ogawa, S., and Umezawa, A. (2011). Treatment of human mesenchymal stem cells with angiotensin receptor blocker improved efficiency of cardiomyogenic transdifferentiation and improved cardiac function via angiogenesis. Stem Cells *29*, 1405–1414.

O'Leary, T., Heindryckx, B.O.R., Lierman, S., van Bruggen, D., Goeman, J.J., Vandewoestyne, M., Deforce, D., de Sousa Lopes, S.M.C., and De Sutter, P. (2012). Tracking the progression of the human inner cell mass during embryonic stem cell derivation. Nature Biotechnology 1–6.

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.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313–317.

Owen, B.A.L., Yang, Z., Lai, M., Gajek, M., Badger, J.D., Hayes, J.J., Edelmann, W., Kucherlapati, R., Wilson, T.M., and McMurray, C.T. (2005). (CAG)n-hairpin DNA binds to Msh2–Msh3 and changes properties of mismatch recognition. Nat Struct Mol Biol *12*, 663–670.

Panigrahi, G.B., Slean, M.M., Simard, J.P., Gileadi, O., and Pearson, C.E. (2010). Isolated short CTG/CAG DNA slip-outs are repaired efficiently by hMutSbeta, but clustered slip-outs are poorly repaired. Proceedings of the National Academy of Sciences *107*, 12593–12598.

Pardo, M., Martorell, L., Lang, B., Gámez, J., Yu, L., Cayuela, M.L., Prosser, H., Gould, F.K., Bradley, A., McAbney, J.P., et al. (2004). Germline mutational dynamics in myotonic dystrophy type 1 males: allele length and age effects. *62*, 269–274.

Park, I.-H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2007). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

Pearson, C.E. (2003). Slipping while sleeping? Trinucleotide repeat expansions in germ cells. Expandable DNA Repeats and Human Disease. *9*, 490–495.

Pearson, C.E., and Sinden, R.R. (1996). Alternative Structures in Duplex DNA Formed

within the Trinucleotide Repeats of the Myotonic Dystrophy and Fragile X Loci. Biochemistry *35*, 5041–5053.

Pearson, C.E., Edamura, K.N., and Cleary, J.D. (2005). Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet *6*, 729–742.

Plenge, R.M., Stevenson, R.A., Lubs, H.A., Schwartz, C.E., and Willard, H.F. (2002). Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. Am. J. Hum. Genet. *71*, 168–173.

Pomp, O., Dreesen, O., Leong, D.F.M., Meller-Pomp, O., Tan, T.T., Zhou, F., and Colman, A. (2011). Unexpected X Chromosome Skewing during Culture and Reprogramming of Human Somatic CellsCan Be Alleviated by Exogenous Telomerase. Stem Cell *9*, 156–165.

Pook, M. (2012). DNA methylation and trinucleotide repeat expansion diseases. DNA Methylation—From Genomics to Technology.

Rama, P., Matuska, S., Paganoni, G., Spinelli, A., De Luca, M., and Pellegrini, G. (2010). Limbal stem-cell therapy and long-term corneal regeneration. N. Engl. J. Med. *363*, 147–155.

Rashid, S.T., Corbineau, S., Hannan, N., Marciniak, S.J., Miranda, E., Alexander, G., Huang-Doran, I., Griffin, J., Ahrlund-Richter, L., Skepper, J., et al. (2010). Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. J. Clin. Invest. *120*, 3127–3136.

Raya, A.N., agrave, I.R.I.G.-P., Navarro, S., Richaud-Patin, Y., Guenechea, G., s, A.S.A.N.-D.E., Consiglio, A., Bueren, J., and Belmonte, J.C.I.U.A. (2010). A protocol describing the genetic correction ofsomatic human cells and subsequent generationof iPS cells. Nature Protocols *5*, 647–660.

Richardson, A.L., Wang, Z.C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J.D., Livingston, D.M., and Ganesan, S. (2006). X chromosomal abnormalities in basal-like human breast cancer. Cancer Cell *9*, 121–132.

Rubin, L.L. (2008). Stem Cells and Drug Discovery: The Beginning of a New Era? Cell *132*, 549–552.

Savouret, C. (2003). CTG repeat instability and size variation timing in DNA repair-deficient mice. The EMBO Journal *22*, 2264–2273.

Savouret, C., Garcia-Cordier, C., Megret, J., Riele, te, H., Junien, C., and Gourdon, G. (2004). MSH2-Dependent Germinal CTG Repeat Expansions Are Produced Continuously in Spermatogonia from DM1 Transgenic Mice. Molecular and Cellular Biology *24*, 629–637.

Schmid, V., Alder, H., Plickert, G., and Weber, C. (1988). Transdifferentiation from striated muscle of medusae in vitro. Cell Differ. Dev. *25 Suppl*, 137–146.

Schulz, E.G., Meisig, J., Nakamura, T., Okamoto, I., Sieber, A., Picard, C., Borensztein, M., Saitou, M., Blüthgen, N., and Heard, E. (2014). The Two Active X Chromosomes in Female ESCs Block Exit from the Pluripotent Stateby Modulating the ESC Signaling Network. Stem Cell *14*, 203–216.

Schwartz, S.D., Regillo, C.D., Lam, B.L., Eliott, D., Rosenfeld, P.J., Gregori, N.Z., Hubschman, J.-P., Davis, J.L., Heilwell, G., Spirn, M., et al. (2015). Human embryonic stem cell-derived retinal pigment epithelium in patients with age-related macular degeneration and Stargardt's macular dystrophy: follow-up of two open-label phase 1/2 studies. Lancet *385*, 509–516.

Sermon, K., Goossens, V., Seneca, S., LISSENS, W., De Vos, A., Vandervorst, M., Van Steirteghem, A., and Liebaers, I. (1998). Preimplantation diagnosis for Huntington's disease (HD): clinical application and analysis of the HD expansion in affected embryos. Prenat. Diagn. *18*, 1427–1436.

Shelbourne, P.F., Keller-McGandy, C., Bi, W.L., Yoon, S.R., Dubeau, L., Veitch, N.J., Vonsattel, J.P., Wexler, N.S., The US-Venezuela Collaborative Research Group, Arnheim, N., et al. (2007). Triplet repeat mutation length gains correlate with cell-type specific vulnerability in Huntington disease brain. Human Molecular Genetics *16*, 1133–1142.

Shen, Y., Matsuno, Y., Fouse, S.D., and Rao, N. (2008). X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations.

Shiba, Y., Fernandes, S., Zhu, W.-Z., Filice, D., Muskheli, V., Kim, J., Palpant, N.J., Gantz, J., Moyes, K.W., Reinecke, H., et al. (2012). Human ES-cell-derived cardiomyocytes electrically couple and suppress arrhythmias in injured hearts. Nature *489*, 322–325.

Silva, S.S., Rowntree, R.K., Mekhoubad, S., and Lee, J.T. (2008). X-chromosome inactivation and epigenetic fluidity in human embryonic stem cells. Proceedings of the National Academy of Sciences *105*, 4820–4825.

Slean, M.M., Panigrahi, G.B., Ranum, L.P., and Pearson, C.E. (2008). Mutagenic roles of DNA "repair" proteins in antibody diversity and disease-associated trinucleotide repeat instability. DNA Repair *7*, 1135–1154.

Song, H., Chung, S.-K., and Xu, Y. (2010). Modeling Disease in Human ESCs Using an Efficient BAC-Based Homologous Recombination System. Stem Cell *6*, 80–89.

Song, J., Lee, S.-T., Kang, W., Park, J.-E., Chu, K., Lee, S.-E., Hwang, T., Chung, H., and Kim, M. (2007). Human embryonic stem cell-derived neural precursor transplants attenuate apomorphine-induced rotational behavior in rats with unilateral quinolinic acid lesions. Neurosci. Lett. *423*, 58–61.

Spits, C., Mateizel, I., Geens, M., Mertzanidou, A., Staessen, C., Vandeskelde, Y., Van der Elst, J., Liebaers, I., and Sermon, K. (2008). Recurrent chromosomal abnormalities in human embryonic stem cells. Nature Biotechnology *26*, 1361–1363.

Stadtfeld, M., Masaki, N., Utikal, J., gordon, W., and Hochedlinger, K. (2008). Induced Pluripotent Stem Cells Generated Without Viral Integration. Science *322*, 943–945.

Starmer, J., and Magnuson, T. (2008). A new model for random X chromosome inactivation. Development *136*, 1–10.

Sterneckert, J.L., Reinhardt, P., and Schöler, H.R. (2014). Investigating human disease using stem cell models. Nat Rev Genet *15*, 625–639.

Suzuki, K., Mitsutake, N., Saenko, V., Suzuki, M., Matsuse, M., Ohtsuru, A., Kumagai, A., Uga, T., Yano, H., Nagayama, Y., et al. (2011). Dedifferentiation of human primary thyrocytes into multilineage progenitor cells without gene introduction. PLoS ONE *6*, e19354.

Szabo, E., Rampalli, S., Risueño, R.M., Schnerch, A., Mitchell, R., Fiebig-Comyn, A., Levadoux-Martin, M., and Bhatia, M. (2010). Direct conversion of human fibroblasts to multilineage blood progenitors. Nature *468*, 521–526.

T Ashizawa, (1994). Characteristics of Intergenerational Contractions of the CTG Repeat in Myotonic Dystrophy. Am. J. Hum. Genet. *54*, 414.

Tabar, V., and Studer, L. (2014). Pluripotent stem cells in regenerative medicine: challenges and recent progress. Nat Rev Genet *15*, 82–92.

Tachibana, M., Amato, P., Sparman, M., Gutierrez, N.M., Tippner-Hedges, R., Ma, H., Kang, E., Fulati, A., Lee, H.-S., Sritanaudomchai, H., et al. (2013). Human embryonic stem cells derived by somatic cell nuclear transfer. Cell *153*, 1228–1238.

Takagi, N., and Sasaki, M. (1975). Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. Nature *256*, 640–642.

Takahashi, K., and Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. Cell *131*, 861–872.

Takano, H., Onodera, O., Takahashi, H., Igarashi, S., Yamada, M., Oyake, M., Ikeuchi, T., Koide, R., Tanaka, H., Iwabuchi, K., et al. (1996). Somatic mosaicism of expanded CAG repeats in brains of patients with dentatorubral-pallidoluysian atrophy: cellular population-dependent dynamics of mitotic instability. The American Journal of Human Genetics *58*, 1212–1222.

Takasato, M., Er, P.X., Becroft, M., Vanslambrouck, J.M., Stanley, E.G., Elefanty, A.G., and Little, M.H. (2014). Directing human embryonic stem cell differentiation towards a renal lineage generates a self-organizing kidney. Nat. Cell Biol. *16*, 118–126.

Teklenburg, G., Weimar, C.H.E., Fauser, B.C.J.M., Macklon, N., Geijsen, N., Heijnen, C.J., Chuva de Sousa Lopes, S.M., and Kuijk, E.W. (2012). Cell lineage specific distribution of H3K27 trimethylation accumulation in an in vitro model for human implantation. *7*, e32701.

Telenius, H., Kremer, B., Goldberg, Y.P., Theilmann, J., Andrew, S.E., Zeisler, J., Adam, S., Greenberg, C., Ives, E.J., Clarke, L.A., et al. (1994). Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. Nat Genet *6*, 409–414.

TEMMERMAN, N., Sermon, K., Seneca, S., DERYCKE, M., Hilven, P., LISSENS, W., VANSTEIRTEGHEM, A., and Liebaers, I. (2004). Intergenerational Instability of the Expanded CTG Repeat in the DMPK Gene: Studies in Human Gametes and Preimplantation Embryos. The American Journal of Human Genetics *75*, 325–329.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L.,

Gardner, R.L., and McKay, R.D.G. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. Nature *448*, 196–199.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., and Waknitz, M.A. (1998). Embryonic stem cell lines derived from human blastocysts. Science.

Thornton, C.A., Johnson, K., and Moxley, R.T. (1994). Myotonic dystrophy patients have larger CTG expansions in skeletal muscle than in leukocytes. Ann. Neurol.

Tomé, S., Holt, I., Edelmann, W., Morris, G.E., Munnich, A., Pearson, C.E., and Gourdon, G. (2009). MSH2 ATPase domain mutation affects CTG*CAG repeat instability in transgenic mice. PLoS Genet *5*, e1000482.

Tomoda, K., Takahashi, K., Leung, K., Okada, A., Narita, M., Yamada, N.A., Eilertson, K.E., Tsang, P., Baba, S., White, M.P., et al. (2012). Short Article. Stem Cell *11*, 91–99.

Tompkins, J.D., Hall, C., Chen, V.C.-Y., Li, A.X., Wu, X., Hsu, D., Couture, L.A., and Riggs, A.D. (2012). Epigenetic stability, adaptability, and reversibility in human embryonic stem cells. Proceedings of the National Academy of Sciences *109*, 12544–12549.

Usdin, K., and Woodford, K.J. (1995). CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. Nucleic Acids Research *23*, 4202–4209.

Vallot, C., and Rougeulle, C. (2014). Long non-coding RNAs and human X-chromosome regulation. RNA Biology *10*, 1262–1265.

Vallot, C., Huret, C., Lesecque, 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 *45*, 239–241.

Vallot, C., Ouimette, J.-F., Makhlouf, M., Féraud, O., Pontis, J., Côme, J., Martinat, C., Bennaceur-Griscelli, A., Lalande, M., and Rougeulle, C. (2015). Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. Cell Stem Cell *16*, 533–546.

Van de Velde, H., Cauffman, G., Tournaye, H., Devroey, P., and Liebaers, I. (2008). The four blastomeres of a 4-cell stage human embryo are able to develop individually into blastocysts with inner cell mass and trophectoderm. Human Reproduction *23*, 1742–1747.

van den Berg, I.M., Galjaard, R.J., and Laven, J. (2011). XCI in preimplantation mouse and human embryos: first there is remodelling… - Springer. Hum Genet.

van den Berg, I.M., Laven, J.S.E., Stevens, M., Jonkers, I., Galjaard, R.-J., Gribnau, J., and van Doorninck, J.H. (2009). X Chromosome Inactivation Is Initiated in Human Preimplantation Embryos. The American Journal of Human Genetics *84*, 771–779.

van den Broek, W.J.A.A., Nelen, M.R., Wansink, D.G., Coerwinkel, M.M., Riele, te, H., Groenen, P.J.T.A., and Wieringa, B. (2002). Somatic expansion behaviour of the (CTG)n repeat in myotonic dystrophy knock-in mice is differentially affected by Msh3 and Msh6 mismatch-repair proteins. Human Molecular Genetics *11*, 191–198.

Vassena, R., Boué, S., González-Roca, E., Aran, B., Auer, H., Veiga, A., and Izpisúa Belmonte, J.C. (2011). Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. Development *138*, 3699–3709.

Veiga, A., Calderon, G., Barri, P.N., and Coroleu, B. (1987). Pregnancy after the replacement of a frozen-thawed embryo with> 50% intact blastomeres. Human Reproduction.

Villa-Diaz, L.G., Ross, A.M., Lahann, J., and Krebsbach, P.H. (2012). Concise Review: The Evolution of human pluripotent stem cell culture: From feeder cells to synthetic coatings. Stem Cells *31*, 1–7.

Visnyei, K., Tatsukawa, K.J., Erickson, R.I., Simonian, S., Oknaian, N., Carmichael, S.T., and Kornblum, H.I. (2006). Neural progenitor implantation restores metabolic deficits in the brain following striatal quinolinic acid lesion. Exp. Neurol. *197*, 465–474.

Wang, Q.T., Piotrowska, K., Ciemerych, M.A., Milenkovic, L., Scott, M.P., Davis, R.W., and Zernicka-Goetz, M. (2004). A genome-wide study of gene activity reveals developmental signaling pathways in the preimplantation mouse embryo. Developmental Cell *6*, 133–144.

Ware, C.B., Nelson, A.M., Mecham, B., Hesson, J., Zhou, W., Jonlin, E.C., Jimenez-Caliani, A.J., Deng, X., Cavanaugh, C., Cook, S., et al. (2014). Derivation of naive human embryonic stem cells. Proceedings of the National Academy of Sciences *111*, 4484–4489.

Werbowetski-Ogilvie, T.E., Bossé, M., Stewart, M., Schnerch, A., Ramos-Mejia, V., Rouleau, A., Wynder, T., Smith, M.-J., Dingwall, S., Carter, T., et al. (2009). Characterization of human embryonic stem cells with features of neoplastic progression. Nature Biotechnology *27*, 91–97.

Wheeler, V. (1999). Length-dependent gametic CAG repeat instability in the Huntington's disease knock-in mouse. Human Molecular Genetics *8*, 115–122.

Wohrle, D., Kennerknecht, I., Wolf, M., Enders, H., Schwemmle, S., and Steinbach, P. (1995). Heterogeneity of DM kinase repeat expansion in different fetal tissues and further expansion during cell proliferation in vitro: evidence for a causal involvement of methyl-directed DNA mismatch repair in triplet repeat stability. Human Molecular Genetics *4*, 1147–1153.

Wu, H., Luo, J., Yu, H., Rattner, A., Mo, A., Wang, Y., Smallwood, P.M., Erlanger, B., Wheelan, S.J., and Nathans, J. (2014). Cellular resolution maps of X chromosome inactivation: implications for neural development, function, and disease. Neuron *81*, 103–119.

Wutz, A. (2011). Gene silencing in X-chromosome inactivation:advances in understanding facultative heterochromatin formation. 1–13.

Xia, G., Gao, Y., Jin, S., Subramony, S.H., Terada, N., Ranum, L.P.W., Swanson, M.S., and Ashizawa, T. (2015). Genome Modification Leads to Phenotype Reversal in Human Myotonic Dystrophy Type 1 Induced Pluripotent Stem Cell-Derived Neural Stem Cells. Stem Cells *33*, 1829–1838.

Xia, G., Santostefano, K.E., Goodwin, M., Liu, J., Subramony, S.H., Swanson, M.S.,

Terada, N., and Ashizawa, T. (2013). Generation of neural cells from DM1 induced pluripotent stem cells as cellular model for the study of central nervous system neuropathogenesis. Cell Reprogram *15*, 166–177.

Xu, N., Tsai, C.-L., and Lee, J.T. (2006). Transient homologous chromosome pairing marks the onset of X inactivation. Science *311*, 1149–1152.

Yamashita, A., Morioka, M., Kishi, H., Kimura, T., Yahara, Y., Okada, M., Fujita, K., Sawai, H., Ikegawa, S., and Tsumaki, N. (2014). Statin treatment rescues FGFR3 skeletal dysplasia phenotypes. Nature *513*, 507–511.

Yang, Z., Lau, R., Marcadier, J.L., Chitayat, D., and Pearson, C.E. (2003). Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells. Am. J. Hum. Genet. *73*, 1092–1105.

Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature *453*, 519–523.

Yoon, S.-R., Dubeau, L., de Young, M., Wexler, N.S., and Arnheim, N. (2003). Huntington disease expansion mutations in humans can occur before meiosis is completed. Proc. Natl. Acad. Sci. U.S.a. *100*, 8834–8838.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. Science *318*, 1917–1920.

Zhang, X., Stojkovic, P., Przyborski, S., Cooke, M., Armstrong, L., Lako, M., and Stojković, M. (2006). Derivation of human embryonic stem cells from developing and arrested embryos. Stem Cells *24*, 2669–2676.

Zhou, H., Wu, S., Joo, J.Y., Zhu, S., Han, D.W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y., et al. (2009). Generation of induced pluripotent stem cells using recombinant proteins. Cell Stem Cell *4*, 381–384.

12 Curriculum Vitae

I studied Biology at the Universitat Autonoma de Barcelona (Spain) and graduated in 2006. After that, I did an internship at the Weill Medical College of Cornell University (New York) were I worked at the Embryology lab under Dr. Gianpiero Palermo's supervision. In November 2007, I joined the EMGE team as a stem cell lab technician and to work for the European Union FP7 programme supported StemHD Project, on the study of the trinucleotide repeat instability in human embryonic stem cells. Later, in 2008 I started a joined PhD between the UAB and the VUB, focused on genetic and epigenetic instability of hESC and its impact on the translation to clinic. Once I moved back to Spain in 2012, I worked for 1 year in the private sector at the pharmaceutical company Grifols as a publications assistant in the scientific department. From 2013 until recently, I was working as research assistant at the Center for Regenerative Medicine of Barcelona (CMRB, Spain) carrying out tasks for the European Human Pluripotent stem cell registry as well as doing research on human pluripotent stem cell differentiation into retinal pigmented epithelium and photoreceptors.

PEER-REVIEWED PUBLICATIONS

1) Geens M*, **Seriola A***, Barbé L, Santalo L, Veiga A, Dée K, Van Haute L, Sermon K, Spits C. *Joint first author. *Human pluripotent stem cells rapidly progress to a culture-adapted X chromosome inactivation state with highly skewed, eroded methylation patterns.* (2015) Submitted to Stem Cell Reports IF_{2015}: 5.365.

2) Riera M., **Seriola A.**, Fontrodona L, Salas A, Mora D, Muñoz Y, Albert S, Ramos D, Martí M, Villegas M, Zapata M, Ruberte J, Veiga A, Arumí J. *Cell therapy with pluripotent stem cells differentiated to retinal pigmented epithelium and photoreceptors cells in a rat model of Retinitis pigmentosa*. In preparation for submission to Experimental Neurology. IF_{2014}: 4.696.

3) Kurtz A, Stacey G, Kidane L, **Seriola A**, Stachelscheid H, Veiga A. *Regulatory insight into the European human pluripotent stem cell registry.* Stem Cells Dev. 2014 Dec;23 Suppl 1:51-5. doi: 10.1089/scd.2014.0319. IF_{2013}: 4.202

4) **Seriola A***, Spits C*, Simard JP, Hilven P, Haentjens P, Pearson CE and Sermon K *Joint first author. *Huntington's and myotonic dystrophy hESCs: down-regulated trinucleotide repeat instability and mismatch repair machinery expression upon differentiation*. Human Molecular Genetics (2011) **20**, 176-185. IF_{2011}: 7.636

ORAL AND POSTER PRESENTATION IN SCIENTIFIC MEETINGS

1) **A. Seriola**, M. Geens,L. Barbé, J. Santalo, A. Veiga, C. Spits, K. Sermon. *Is skewed X chromosome inactivation in human embryonic stem cells driven by a culture advantage?.* ISSCR 2015 (W-1371)

2) **Seriola A**, Spits C, Mateizel I, De Temmerman N, Hilven P, Van der Elst J, Liebaers I, Sermon K (2008) *Trinucleotide repeat instability in Human Embryonic Stem Cells and its relationship to the Mismatch repair machinery. ESHRE (O-273)*

3) **Seriola A**, Spits C, Mateizel I, De Temmerman N, Hilven P, Van der Elst J, Liebaers I, Sermon K (2008) *Trinucleotide repeat instability in Human Embryonic Stem Cells and its relationship to the Mismatch repair machinery.* BeSHG 2008 (P-88)

4) Neri QV, **Seriola A**, Takeuchi T, Rosenwarks Z, Palermo GD (2007) *Correlation of ESC derived PGCs to their counterpart in 6.5 DPC embryos.* ASRM 2007.

5) Feliciano M, Neri QV, **Seriola A**, Takeuchi T, Rosenwaks Z, Palermo GD. (2007) *In vitro* differentiation of embryonic stem cells into male germ cells. Human Reproduction ESRHE 2007. (O-169)

TEACHING ACTIVITIES

Lecturer on Human Pluripotent stem cells for the master course "Biologia de la Reproducció I Tècniques de Reproducció Humana Assistida" (Biology of Reproduction and assisted reproduction techniques).