Molecular characterization of pluripotency in embryos and embryonic stem cells

Memòria presentada per Josep Pareja Gómez per a optar al grau de doctor
pel Departament de Ciències de la Salut i de la Vida
de la Universitat Pompeu Fabra

Tesis dirigida per el Dr. Nikica Zaninovic i la Dra. Anna Veiga

Universitat Pompeu Fabra

Dr. Nikica Zaninovic	Dra. Anna Veiga	Josep Pareja Gómez
Director	Directora	Doctorand

Between the fifth and the tenth day the lump stem cells differentiates into the overall building plan of the embryo and its organs. It is a bit like a lump of iron turning into the space shuttle. In fact it is the profoundest wonder we can still imagine and accept, and at the same time so usual that we have to force ourselves to wonder about the wondrousness of this wonder.

Miroslav Holub.

“Another brick in the wall”

Tesi Villanueva ambiguously quoting Pink Floyd

Agraïments / Agradecimientos / Aknowledgements

Aquesta tesi ha costat déu i ajuda, però com que jo sóc ateu, intentaré enumerar l'ajuda: totes les persones que han fet possible que finalment aquest treball vegi la llum. Però són tantes, que per escriure aquests agraïments podria trigar gairebé tant com he trigat a escriure el que ve després.

First, I would like to thank Dr. Zev Rosenwaks and my PhD advisor at the CRMI, Dr. Nikica Zaninovic for his generosity and support. My gratitude for allowing me to fulfill my scientific curiosity granting me the support to do it in, undoubtedly, the best embryology lab of the world. Thanks Nikica for encouraging me to always go scientifically a step further of what I had ever possibly imagined .

Estic en deute amb a la Dra. Anna Veiga per la seva contribució fonamental (en el sentit més estricte de la paraula) a que aquesta tesi sigui una realitat. Primer recolzant-me en tot moment en demanar la beca Fulbright (amb tota la paperassa que va haver de fer), però sobretot, després per el seu suport científic (i moral) en la configuració d'aquest manuscrit quan estava una mica perdut. Anna: moltes, moltes gràcies, per la teva generositat i la teva experiència científica i no científica.

Gracias a la fundación Fulbright por haber financiado parte de mi estancia en el CRMI.

Thanks to Lucinda Veeck-Gosden and Dr. Roger Gosden at the CRMI for opening the doors of the Embryology Lab for me.

Thanks to Dr. Shahin Rafii for the generosity of letting me into his lab to see how cutting-edge science is cooked. Thanks to Dr. Daylon James for the patience of teaching me how to do stuff!

I would like to thank to Dr. Jiying Hao, for her generosity and for contributing decisively with her effort to this research.

Mirant retrospectivament, també vull agrair al Dr. Baldo Oliva haver confiat en mi alhora de configurar el seu grup al GRIB. La meva estada al SBL va ser l'inici de la meva formació com a científic i, al final, saber bioinformàtica ha estat fonamental per desenvolupar part del treball d'aquesta tesi.

I encara mirant més retrospectivament, gràcies a l'amic (Dr.) Gabriel Gil per ser el primer que va mostrar-me el bon camí! Ja saps que em distrec fàcilment!!!

Gràcies a la Fina Gómez (Fingom!) per ser com ets, per haver-me ensenyat tantes coses tant a dintre del laboratori com a fora, per la teva generositat personal i la teva amistat. Estic en deute amb tu. Grazie a gli amici italiani: Fulvia Livralon ed il Dottore Massimo Ciaponi, grazie da vero per la vostra amicizia.

Gràcies a la gent de Dexeus: A la Montse Boada,a la Bego, a la Itziar i la Mª José i les altres embriòlogues per ensenyar-nos als màsters el seu savoir-faire dins el laboratori d'embriologia.

Thanks to everyone in the embryology lab at the CRMI: thanks to Elisabeth, Dina, Chris, Rich, June, Carol, Jason and Margarita because I can't imagine the lab without you. Thanks to Hong for her kindness and her permanent smile. Thanks to Anaya, for being always helpful and a great company. Thanks to Bob Clarke and Rosemary Berrios for their humanity and for being ready to help whenever they were needed. Thanks to Marco Toschi, for his companionship and understanding of what being a legal alien in New York is. Thanks to Evangeline Iwu, because despite being a pain to her with all my demands, she made me feel like her golden boy: you're my golden girl. Thanks for your friendship. A special, extra-extra-large (very very extra large) thanks to my dear friend Myriam Jackson: you're right, the lab wouldn't work without you! Myriam, gracias por su amistad, el cariño que me dio (y me da), por considerarme su darling, y sobretodo por su sentido del humor que nos hace tan felices. No se imagina cuánto la echo de menos. Thanks to my dear friend Naomi Lourie, for the long talks -eating popcorn- when everybody had left the lab (exept for LV) and for making me feel close to her in the middle of the ups and downs of the big city. I miss you.

Gràcies a tots els meus amics i amigues de la vella guàrdia del GRIB! Gràcies al Pep Abril, a l'Òscar i a l'Alfons, per la seva paciència y generositat amb aquest pobre usuari. Gràcies a la Montse (la Barbany) per la seva amistat i per compartir amb mi barracons, linux, i algoritmes! Ets un exemple de passió científica i de passió maternal! Gràcies a en Fabien per ser un model de perseverància i un amic. Gràcies a en Robert per la seva proximitat, bonhomia i amistat. Gràcies a en Loris i la Meritxell, els kamikazes més apassionats que conec i uns grans amics. Gràcies a la Rut per haver-me deixat veure-la de ben a propet, per tots els matins asseguts mirant despuntar el sol davant el mar amb un te i la seva companyia, i per la seva amistat. Gracias a Cristina Dezi, por su inagotable capacidad de dar amor, por haberme dejado ser su abrazador oficial y amigo suyo. Te echo de menos!!! A Eduardo por ser el amigo con el mejor cuerpo de todos los que tengo... y de

corazón más grande también. A en Genís per ensenyar-me a menjar xampinyons crus, i per deixar-se apropar vencent la por i regalar-me la seva amistat. Thanks to Noura, for sharing her wit, her irony and joy with us, and for her friendship. A Natalia, por ese cariño a flor de piel que tiene y la poesia que la envuelve. A la Bet, per ser l'amiga que ÉS permanentment, incondicional, generosa i per ser sempre a 3mm. A Ramón por ser modelo a seguir y amigo SIEMPRE, dando a manos llenas todo lo que tiene, y más.

Gràcies als meus amics de Sabadell, els de tota la vida. A en Marcel per tots els moments d'amistat compartits quan érem tan inconscients i bèsties. A Dani por haber sido siempre referencia personal, por el cariño de una amistad de tantos años. A Virginia por haber sido apoyo incondicional en un momento confuso de mi vida, por desafiarme a que me pusiera camisetas de colores chillones, por su sabiduría y su amistad sincera. A Dani y Laura, a Jose y Marta, por su amistad y porque que cada reencuentro es una fiesta. A Xavi Aznar, por las afinidades y las neuras compartidas además de su amistad. A Carlos por las cosas desconocidas que decidimos compartir y por su amistad. A Ana Mari, por ser como es, por el cariño que me ha brindado siempre, pase lo que pase.

Gracias a Clara, por ser la primera persona en demostrarme qué es la amistad sin condiciones, por ayudarme a ser mejor persona, y por los muchos buenos momentos compartidos durante tantos años.

Gracias a mis niñas de Heidelberg. Ich wurde mich bei meiner liebsten Freundin Pepita (Esther Schindler) bedanken dafür, dass sie mich glücklich meine 3 Monaten in Heidelberg gemacht hat: Danke für deinen riesigen Herz, und entschuldige mich, denn ich habe nicht geschrieben so oft wie ich es mir gewünscht hätte. A Ana Paeffgen, por irradiar alegría, por regalarme su cariño el primer segundo después de conocernos en la Hauptstrasse y por brindarme su amistad desde ese instante. A Sílvia Díez, por los gloriosos momentos compartidos, y porque sin tí, tu cariño y amistad, probablemente esta tesis no se hubiese podido realizar. Gracias por ser la mejor sparring emocional del mundo. Os echo de menos! (Ich vermisse euch!).

Gràcies als meus amics veterinaris, Marcel, Oscar, Maribel que m'han fet sentir com si jo també fos un d'ells. A l'Aida, perquè tot i no ser veterinària és referent en bricolatge, serenitat, bona conducció i carinyu. A l'Alhelí, per ser la veterinària més petita (i amb el cabell més estrany) però també la que te el cor més gros! Gràcies per la teva amistat.

Gràcies a en Jordi (Font) Hereu per la seva amistat des dels dies de l'Autònoma i per ser sempre tot cor i comprensió.

Gràcies als tots els meus companys al Consorci Sant Gregori per haver-me fet sentir un més i haver-se'm apropat. També vull donar les gràcies a tots els usuaris amb els que he tractat, per ensenyar-me què és la humanitat.

Gràcies a l'Ignasi per ser referent en eficàcia i eficiència (diferenciant el significat, hehe) i tantes altres coses i per la seva generositat sense límits i amistat sincera. Gràcies a en Toni, per haver-nos acollit com a amics sense pensar-s'hi ni un moment.

Gràcies a la Marta Sànchez (Martins), per tot el que he après de tu, rigut amb tu i el que hem compartit. Gràcies per la teva dedicació i rigor professionals. Gràcies per la teva amistat generosa, sincera i riallera: saps que sempre voldré ser com tu quan sigui gran, fins i tot, quan sigui gran.

Gràcies a les meves amigues i amics gironines i gironins per haver-me fet sentir que (casi) sóc de Girona i que no els fa res que sigui d'allà baix (haha): a la Meritxell i en Jordi per la combinació impagable de rigor i ironia que fan com a parella; a la Rosor per ser tan estupenda i el carinyu que ens dones, a en David per les seves recomanacions i els seus acudits dolents; a la Maria per la seva proximitat i a en Fermin per la seva bonhomia; a la Rubio per sempre comptar amb nosaltres; a la Llobe per el seu carinyu i empatia, i a en Miki perquè encara que no cregui en el peak-oil, encara li continuo sent amic; a l'Esther i en Prufi perquè sou tan propers i el bon rollo que despreneu; a la Laia P per haver-me fet sentir un més de la colla; a la Pifa, per la seva alegria contagiosa. A l'Imma perquè al seu cor enorme en què tot i tothom hi cap, m'hi ha fet un lloc en el que em sento tan i tan a gust. A la Ivet, per haver-me adoptat, per preocupar-se sempre, per la seves atencions, carinyu, generositat enorme i la seva amistat. A l'Anna Parés, la revelació de Sant Gregori, bessona emocional i amiga.

Als meus amics del Via Fora. A l'Ari i en Sergi, per els moments compartits i per la vostra amistat. A la Núria Chapinal, per les afinitats electives i pel teu carinyu i amistat. A la Sònia per ser emocional, transparent, sincera i amiga. A en Xavi(er), per la seva immensa generositat i amistat, i perquè tot i que alguna vegada pugui haver semblat el contrari (tontet!), ets un referent per mi. A la Mercè, perquè tot i el poc temps que fa que ens coneixem has demostrat ser tot cor. A la Vicky, el meu huracà argentí particular, per la teva amistat sincera, generosa, intensa, descarnada, i per donar

un nou significat al concepte empatia. Gràcies a en Joel, germà i company de tants moments impagables i importants en aquests últims anys. Gracies de debò per la teva amistat.

Thanks to my New York boys: gràcies a l'Èric, per tot el que vam passar junts en mig de la gran ciutat, per cuidar-me, per preocupar-se sempre i per la seva amistat. Gràcies a en Roger, per la connexió immediata, per la complicitat i amistat sincera, a més de per la muntanya que encara hem de començar a fer junts.

Thanks to my New York girls: Gracias a Carolina (la Pola), por ser la alegría de la huerta y por casi llevarme al huerto, además de por tu amistad. Gracias a Laura, todo-corazón, por preocuparse siempre cuando llegaba a casa a altas horas después de las maratones de laboratorio de la última época, por su proximidad y su amistad. Gràcies a la Lara, per què tot i que em va costar una mica, em va deixar apropar-m'hi per donar-me la seva amistat i la recompensa ha estat enorme. Gràcies a en Crístian (nene, com a en Kurt el dia de la competi, no et puc posar a cap altre puesto, verdaaat?) per ser l'heroi del dia, model a seguir (seguim-lo tots!), divertit, amb un miocardi fucsia que no li cap al pit, i sobretot, amic. Gràcies a la Brigitte, per la comprensió instantània només mirar-la, per la seva força vital i la seva amistat. Thanks to Carrie, for her friendship and for sharing with me such precious moments in New York.

Gracias a Tesi por ser capaz de verme antes de que yo la viese y perseverar, por el email de los dialectos del Noruego, por todos y cada uno de los momentos que han venido después de ese, en los que me ha hecho el inmenso regalo de ser mi amiga. Porque si no nos hubiésemos encontrado, seria huérfano de ti y no lo sabría.

Gràcies a la meva nova família: als meus sogres Aurora i Ramon, per haver-me fet sentir un fill més i per la seva generositat. Als meus cunyats Josep, Judit i Laia, per fer-me sentir germà i a les meves nebodes i nebot, Tuies, Afra i Grau, per deixar-me ser el seu oncle Pitu. Gràcies per fer-me sentir a casa cada dia que sóc amb vosaltres.

Gracias a mis hermanas, hermano, cuñados y cuñada: Mª José y Alfonso, Kiko y Pepi y en especial a mis segundos padres, Conchi y Francisco, a todos, gracias por haber sido mi referencia en todo momento y por vuestro apoyo y amor incondicional desde que nací. Gracias a todos mis sobrinos! A los grandes, Sergio y Ana, Núria y Juanma y mis niñas y niños (aunque algunas ya no lo son tanto), Cristina, Sílvia, Rubén, Neus (alias Joselita), Ingrid, Sergi y Laura, por lo importantes que sois para

mi y porque cada día me hacéis sentir más orgulloso de ser vuestro tío.

A mis padres, por haberme dado todas las oportunidades, haberme enseñado el camino y haberme querido sin condiciones. A mi padre, por haber sabido entender y ser paciente. A mi madre, porque aunque ya no está, sé que sería la persona mas feliz del mundo en estos momentos y porque la echo de menos. Esta tesis está dedicada a su memoria.

A l'Oriol, per ser sempre, sempre amb mi, a 2nm o a 7000 km de distància, per sostenir-me quan les cames m'han fet figa, per haver cregut sempre en mi, fins i tot quan jo mateix no ho feia, per ser el científic que sempre voldré ser, per haver-me donat serenitat i perspectiva, per la teva generositat sense límits, per ser amic i amat. Aquesta tesi també està dedicada a tu.

Table of Contents

1. List of Abbreviations

ASC	Adult Stem Cells
bFGF	Basic Fibroblast Growth Factor
BL	Blastocyst
BSA	Bovine Serum Albumin
EB	Embryoid Body
EpiSC	epiblastic stem cells
ECC	Embryonal Carcinoma Cells
EGC	Embryonic Germ Cells
ESC	Embryonic Stem Cells
FBS	Fetal Bovine Serum
hESC	Human Embryonic Stem Cells
H3K4me3	Histone 3 Lysine 4 trimethylated
H3K27me3	Histone 3 Lysine 27 trimethylated
ICSI	Intracytoplasmatic Sperm Injection
ICM	Inner Cell Mass
iPSC	Induced Pluripotency Stem Cells
KO-SR	Knock-Out Serum Replacement
LIF	Leukemia Inhibitory Factor
LOH	Loss of Heterozygosity
MEF	Murine Embryonic Fibroblasts
mESC	Murine Embryonic Stem Cells
MII	Metaphase II
NF	Normalization Factor
PcG	Polycomb Group of proteins
PCR	Polymerase Chain Reaction
PGD	Prenatal Genetic Diagnosis
PGS	Prenatal Genetic Screening
PE	Primitive Endoderm
pESC	parthenogenetic embryonic stem cells
PGC	Primordial Germ Cells
PN	Pronucleus
PRC	Polycomb Repressive Complex
RT-qPCR	Real time quantitative PCR
SCF	Stem Cell Factor
TE	Trophectoderm
TPAMMK	TaqMan PreAmp Master Mix kit
UCB	Umbilical Cord Blood
UPD	Uniparental Dysomy
ZGA	Zygotic Genome Activation
ZP	Zona Pellucida

2. Introduction

Life cycle is a continuum: in mammals, the generation of a zygote after fertilization (by fusion of an oocyte and a spermatozoon) triggers the development of a new organism that will be able to produce new gametes that can start the cycle again. In this circularly repetitive process, the zygote stands out as a unique cell in an essential aspect: it has the potential to generate all cells and tissues necessaries (embryonic and extraembryonic) to form the new organism when implanted at the uterus. This potential is contained in the genetic material inherited from the two gametes it arises from. Nevertheless, the uniqueness of the zygote lies in being the cell initiating the tight and precise genome transcriptional and translational regulation program that will be inherited and progressively modified by its cellular progeny and that is responsible for the differentiating forces that will give rise to all the tissues of a new individual, including gametes.

2.1. Preimplantation embryonic development

The temporal relationships between developmental events described here and the stage of their occurrence within development may differ between species. In this section we describe preimplantation embryogenesis using human as a reference. However the principal events are conserved throughout mammalian species.

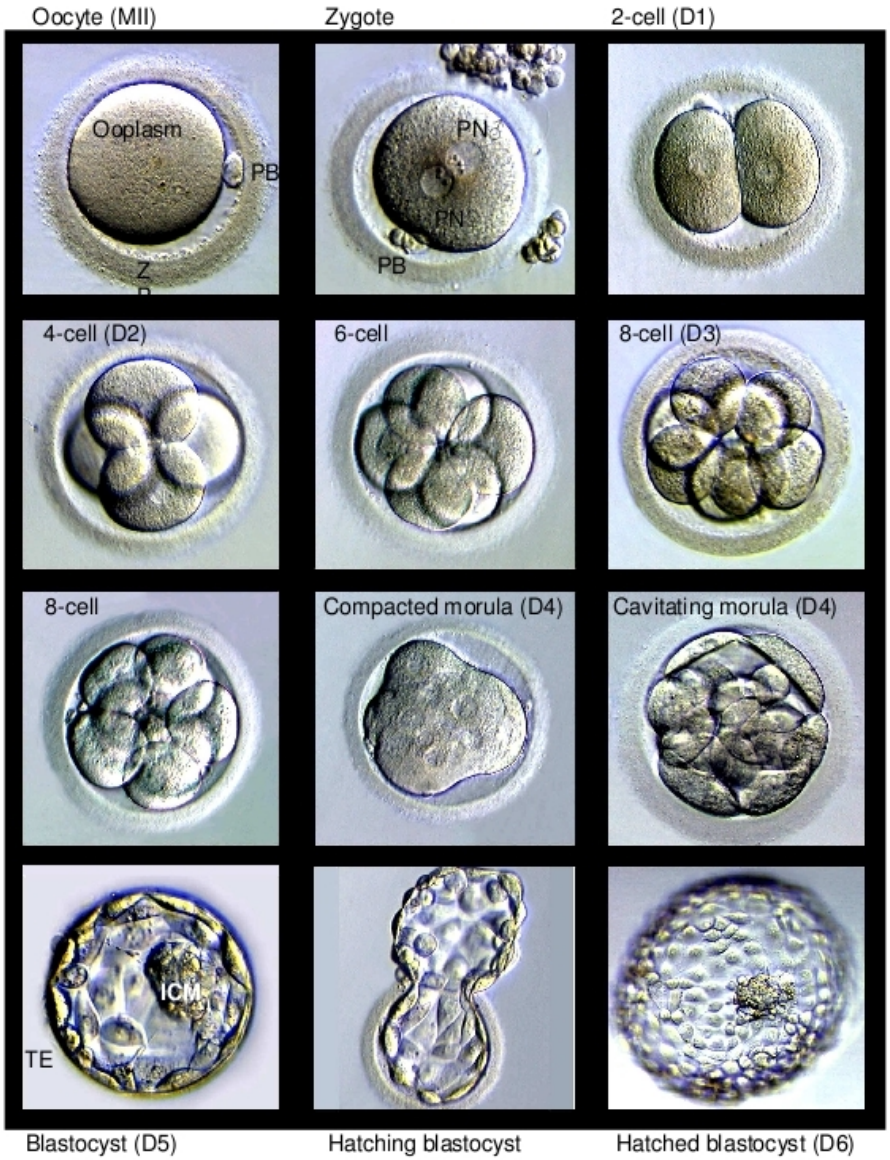
2.1.1 Fertilization

Fertilization occurs when a spermatozoon gets through the cumulus oophorus to contact the *zona pellucida* (ZP) that surrounds the oocyte. This contact induces an acrosomal reaction whereby the spermatozoon releases the contents of its acrosome through the loss of the outer acrosome membrane, including enzymes that facilitate the penetration of the sperm through the ZP into the narrow periviteline space and the fusion with the oocyte plasma membrane. This process results in the exposure of the inner acrosome membrane of the sperm that is necessary for fusion with the oocyte plasma membrane. The equatorial segment of the sperm head is first to attach to the plasma membrane and sperm incorporation occurs rapidly after membrane fusion. Only acrosome-reacted sperm are believed to be able to fuse with the oocyte membrane. This initial step is bypassed by intracytoplasmic sperm injection (ICSI), the injection of a single spermatozoon into the mature (MII) oocyte to assist the fertilization process, a procedure that now accounts for more that half of all IVF procedures[1].

The fusion of gametes at fertilization invokes a cascade of events that initiates oocyte activation. The oocyte completes its second meiotic division: the pairs of chromatids in the chromosomes split at their centromeres and segregate on the spindle to the oocyte or second polar

body. In this way, a haploid number of chromosomes is inherited by the oocyte.

Within 6 to 12 hours after oocyte and sperm fusion, male and female pronuclei are formed from the sperm and oocyte chromatin decondensation and form the nuclear membranes. The stage at which pronuclei are visible is termed the pronuclear stage and the fertilized egg is now defined as a zygote. During the pronuclear phase, DNA synthesis within male and female pronuclei begins synchronously at about 12 hours after sperm/oocyte fusion. In the human, the fertilization ends with the initiation of the first (mitotic) cleavage (Fig. 1). The sperm centrosome controls the first mitotic divisions after fertilization has taken place[2].



PN♂ = male pronucleus; PN♀ = female pronucleus; TE = trophectoderm; ICM = Inner cell mass

Figure 1. Morphological appearance of early stages of human conception in vitro: Oocyte, zygote, cleavage embryos (2-8 cell), morula, cavitating morula, blastocyst and hatching and hatched blastocyst, from day 1 (D1) to day 6 (D6) (adapted from Veeck and Zaninovic, 2003[3]).

2.1.2 Early cleavage

Cleavage of the human embryo involves a series of mitotic divisions, the first being the longest interval (~24 hr) and then every 12 to 18 hours, without any discernible increase in its overall mass.

All data regarding early human embryonic development have been obtained from IVF as these stages are not observable in in vivo situations. The cell doubling time in human embryos between days 2 and 6 has been reported to be 24-31 hours, with accelerated doubling noted after the first two divisions[4]. On a practical basis, two-cell embryos are observed any time after 20 hours postinsemination (D1), usually around 24 hours, and may persist until 42 hours postinsemination. Four-cell embryos are normally observed between 36 and 60 hours postinsemination (D2). Eight-cell stages are not generally seen until after 54 hours (D3), but usually before 72 hours (Fig. 2). Individual blastomeres remain distinct and are totipotent (capable of developing independently to form a new organism) until the eight-cell stage, when changes occur in the structure and properties of their plasma membranes, cytoplasm and gene expression. Symmetrical cleavage to the 8-cell stage has been identified as a favorable observation (with regards to implantation rates) in the human, but often proves to be unreliable for predicting implantation success when used as a single analytic parameter.

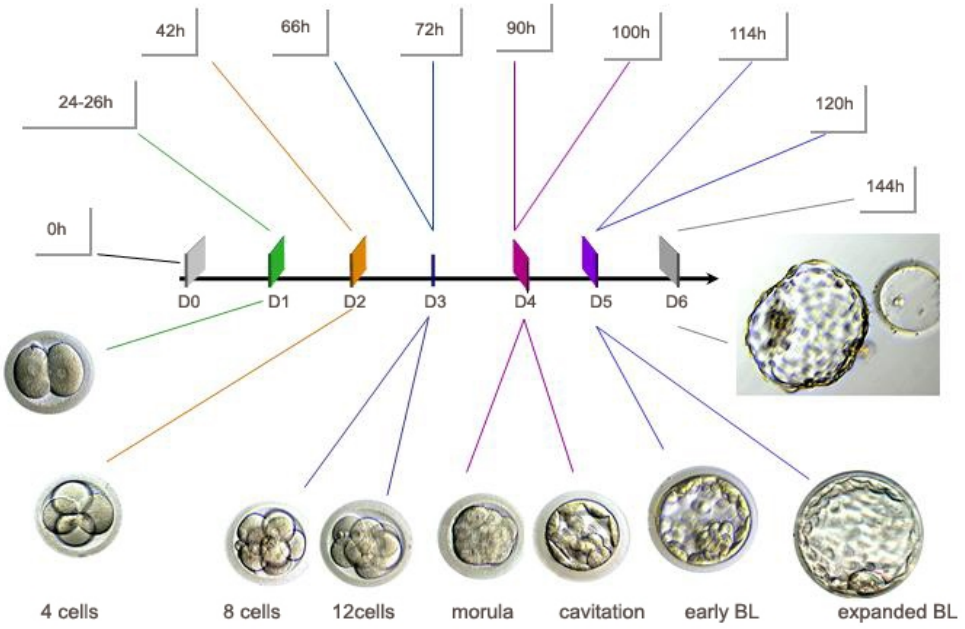


Figure 2 Time line of human embryo development in vitro from insemination/ICSI (D0) to hatched BL (D6) (from Veeck and Zaninovic 2003)[3]

2.1.3 Compaction and cavitation

Cells at the 8-cell stage on day 3-4 become less distinct as they adhere more tightly to one another during the process of compaction. Compaction results from the formation of tight intercellular junctions which cause blastomeres to become closely apposed to each other (Fig. 2). During compaction more cell divisions occur and this stage is also referred to as the morula stage 1. After completion of compaction and formation of tight junctions between blastomeres, a fluid-filed extracellular cavity develops. The cavity is formed by active pumping of ions from extraembryonic fluid to the inside of the embryo by sodium/potassium ATPase in the outer cells, which is followed by the net flow of water along its chemical potential gradient. Within one to two days, this cavity increases to maximum size and is referred to as the blastocoel.

2.1.4 Blastocyst formation

After formation of the blastocoel, the embryo is known as a blastocyst (BL). During cavitation, cell allocation and differentiation occur with the resulting formation of various cell types that are well-defined in the blastocyst, namely the inner cell mass (ICM) and trophectoderm (TE)[3] (Fig. 1). The timing of the blastocyst formation is variable among mammalian species (in humans it is normally formed by day 5) and may display different morphology and developmental stages (early to late BL) prior to implantation time (in humans day 6-7). To be able to implant, BL need to hatch through a hole in the ZP (Fig. 1).

During blastocyst development, two morphologically distinguishable cell types are formed: ICM and TE. The first morphological evidence of such differentiation is observed during embryo cavitation, with the initial positioning of the cells.

The ICM of murine embryos, blastocyst is composed of two basic cell types: pluripotent and primitive endoderm based on a differential expression of key transcription factors involved in lineage commitment. In humans, this clear distinctive expression patterns have not been described and little is known about regulation at this stage of development. In mouse, primitive endoderm cells in the ICM represent differentiated extraembryonic cells, derived from pluripotent ICM cells, which form the extraembryonic primitive endoderm (PE) and yolk sac later in post-implantation embryo development[5]. PE is an active participant in early embryonic patterning, providing signals that specify the anterior-posterior axis of the early embryo and are involved in the induction of the yolk sac hemangioblast (a multipotent cell, common precursor to hematopoietic and endothelial cells)[6].

2.2. Developmental potential in vivo and in vitro

2.2.1. Developmental potential concept

Behind the genesis of a new organism from a zygote lies the concept of developmental potential. This is a complex notion applied to cells that is biologically based on the molecular mechanisms that confer an unspecialized and open cellular identity and the ability to become any specialized cell type in a organism. At conception, the developmental potential is at its maximum possible and from then on, this potential is gradually lost as development progresses to the generation of the post-natal organism due to the steady differentiation of cell populations originated during the process. Cell potentiality is classified according to differences in their differentiation abilities (table 1).

Name	Potential
Totipotent	Ability to form all lineages of the organism (embryonic and extraembryonic). In mammals, only the zygote and the first cleavage blastomeres are totipotent.
Pluripotent	Ability to form all lineages of the adult organism. ICM cells, embryonic stem cells (ESC), embryonic germ cells (EGC), embryonic carcinoma cells (ECC), induced-pluripotency stem cells (iPSC)
Multipotent	Ability to form multiple cell types of one lineage. Adult stem cells
Unipotent	Cells form one cell type. They still have the ability of self-renewal, distinguishing them from terminally differentiated cells. Intestinal stem cells.

Table 1. Cells are classified in four different categories depending on the potency to form the cells and tissues in an organism.

The categorical classification of developmental potential showed on table 1 relies on defined criteria to include different cell lineages on one group or another. However, during embryogenesis, cells transit from one category to the other in a not so categorical way, but rather undergoing progressive changes. Whether a specific cell type belongs to one group or another is determined by experimental evidence generated *in vivo*, and especially *in vitro*, where their developmental abilities can be tested in isolation. In fact, in the attempt to study developmentally potential, different mammalian cells with pluripotent, multipotent and unipotent developmental potential have been isolated in vitro.

In adult organisms, differentiated cell lineages coexist with multipotent (e.g. hematopoietic

stem cells) and unipotent cell lineages (e.g. intestinal stem cells,). Multipotent and unipotent cells share a self renewal capability not present in differentiated cells. However, no cell lineage retains putative pluripotent capacity in a adult organisms. Thus, since *in vivo* pluripotent developmental potential is a transient state of the cells in their journey towards lineage determination, *in vitro* isolation and maintenance of pluripotent cells implies a biological paradox. Pluripotent cells explanted in vitro have shown to retain the capacity to generate all the differentiated functional cell types when reintroduced in an embryo. However these cells display an indefinite capacity to self-renew *in vitro* under certain culture conditions while the *in vivo* capacity of self-renewal of embryonic pluripotent cells is limited to the temporal window previous to irreversible lineage commitment.

Here we describe the cell types in the four different developmental potential categories detailed in table 2, with special attention to the similarities and differences between the cells observed in vivo and those obtained in vitro (a schematic view of these cells can be seen in Fig. 3).

Assay	Experimental Approach	Limitations
In vitro differentiation	Induced differentiation of a differentiation specific markers	The expression for differentiation markers does not test for functionality; marker expression can be due to cellular stress response
Teratoma formation	Induction of tumors demonstrating the potential to generate differentiated cell types of various lineages	Does not test for the ability of cells to promote normal development
Chimera formation	Contribution of cells to normal development following injection into host blastocyst	Host-derived cells in chimera may complement cell non-autonomous defects. Cannot be used with human samples.
Germline contribution	Ability of test cells to generate functional germ cells and offspring	Excludes genetic but not epigenetic defects that could interfere with development. Cannot be used with human samples.
Tetraploid complementation	Injection of test cells into 4n host blastocyst.	Does not test for the ability to form trophectoderm (placental) lineage. Cannot be used with human samples.

Table 2. Assays used to determine the potency of a given cell type.

The view of the different pluripotent cell types obtained *in vitro* from embryos at different developmental stages is illustrative of the fact that progressive steps towards lineage commitment do not affect developmental potential in an absolute way, but rather in a more gradual fashion. In addition, it is important to emphasize that cells adapted to proliferate in vitro represent only a proxy for the in vivo situation and may, at best, approximate the properties of cells in the embryo[7,8]. Consequently, concepts such as pluripotency, multipotency, or differentiation of cultured cells rely on operational criteria and are typically assessed by different functional standards. A descriptive view of these functional tests is presented in table 2. The least stringent functional assay for the developmental potential of a cultured cell is in vitro differentiation followed, with increasing stringency, by the generation of teratomas (germ cell tumors), chimera formation, and germ line contribution. The most rigorous test for developmental potency is the injection of cells into 4n host blastocysts[9,10], which results in animals composed only of the injected donor cells (‘‘all ES’’ embryos or animals) rather than a chimeric composite of injected and host-derived cells.

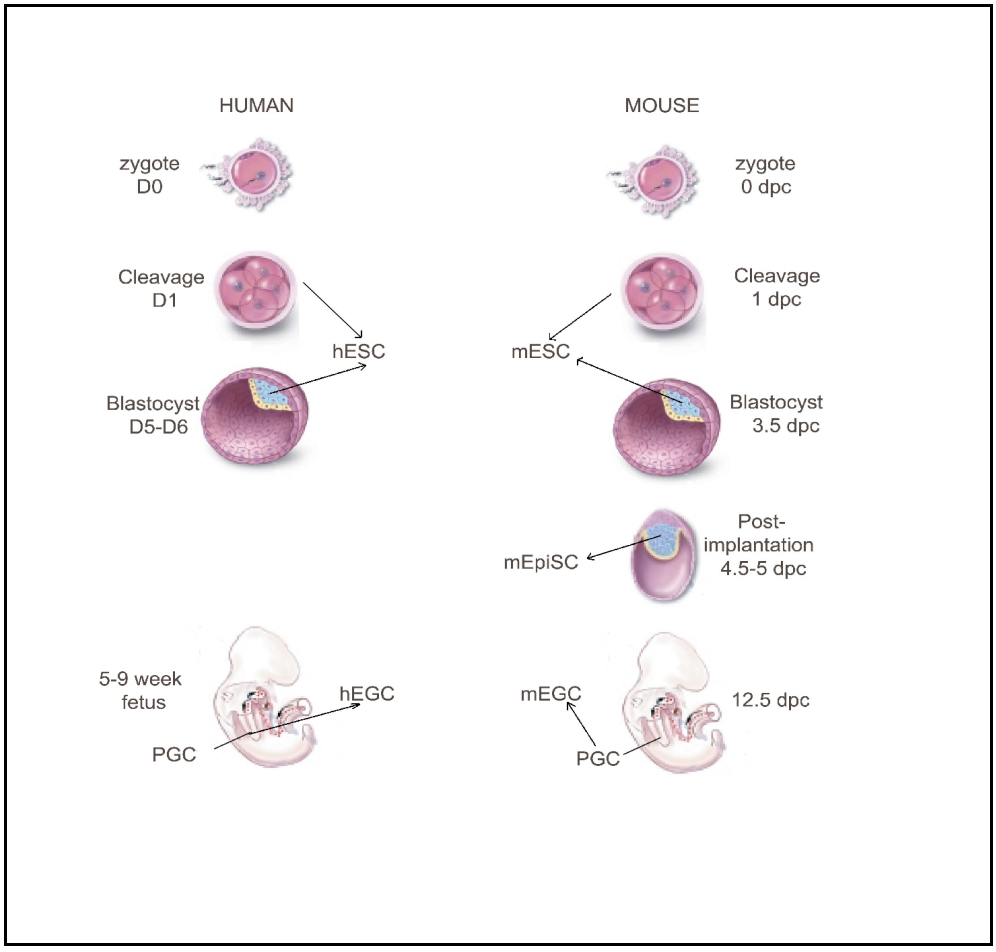


Figure 3. ESC, EpiSC and EGC are three different types of pluripotent cells in vitro. They originate from different embryonic stages in vivo. Thus, different types of pluripotent cells can be derived from totipotent blastomeres, pluripotent ICM cells from preimplantation and postimplantation

blastocyst (the latter only in mouse and rat), and from PGC. Dpc: days post coitum.

2.2.2. Totipotent cells in vivo

In mammals only the zygote and early blastomeres (up to 8 cell embryos) have been shown to be totipotent and can generate the whole organism including extraembryonic tissues. Indeed, upon segregation, when isolated blastomeres of 2-cell, 4-cell and 8-cell embryos are cultured in vitro they have the ability to form a functional blastocyst. However, despite this ability, a proportion of the total number of blastomeres do not reach the blastocyst stage for reasons that are unknown. In the blastomere-derive blastocysts the ratios of inner cell mass to trophectoderm (TE) and ICM to total cells are equivalent to those of intact control blastocysts, the total cell numbers are reduced in a proportional way to the cleavage stage it was obtained from. These blastomere-derived blastocysts have been shown to be capable of full term development of normal individuals in several species, including rabbit, mouse and sheep (table 3).

Embryonic stage	**Functional blastocyst**	**To term development**
2-cell embryo	Mouse, sheep, rabbit, cow, pig, goat, horse, human	Mouse, sheep, rabbit, cow, pig, horse, goat, human
4-cell embryo	Mouse, sheep, rabbit	Sheep, rabbit
8-cell embryo	Sheep, rabbit	Sheep, rabbit

Table 3 Totipotency of blastomeres from 2-cell, 4-cell and 8-cell embryos from different species is exemplified in the findings described in this table. Functional blastocyst refers to the formation of a blastocyst-like structure in which all the lineages are present: epiblast, PE and TE. References for the different species: human[11], mouse[12], sheep[13], rabbit[14,15], cow[16,17], pig[18], horse[19] and goat[20].

2.2.3 Pluripotent cells in vivo

2.2.3.1 ICM cells

After embryonic compaction and cavitation, the blastocyst is originated after the two lineage restriction events that form two differentiated cell populations: trophoblast (that form the trophectoderm), the ICM. The newly formed ICM is a cluster of unspecialized cells in the blastocoelic cavity limiting with the trophoblast. At this point, the ICM cells are no longer able to generate extraembryonic tissues but they are able to generate the entire fetus when transferred into a receptive uterus. Single murine ICM cells isolated at this stage and microinjected into a host

blastocyst, have been shown to contribute to all embryonic lineages[21] showing their pluripotent developmental ability.

2.2.3.2 Epiblast cells

After implantation, the embryonic pluripotent cells originate the epiblast. At this point, these cells undergo dramatic expansion and morphogenesis, transforming form an unstructured cell mass into a columnar epithelium. The epithelialized epiblast (cylindrical in rodents, discoid in other mammals) is subjected to potent lineage-specifying signals form adjacent extraembryonic tissues. It is at this point that epiblastic cells start to undergo a progressive lineage commitment process by triggering the molecular mechanisms that will end up in the irreversible lineage determination during gastrulation. It during this period of time when epiblastic cells have a dual nature caused by the oscillatory expression of synergistically and antagonistically acting transcription factors, on one hand those differentiation-inducing and on the other hand, the pluripotency-promoting ones.

Initial evidence of the pluripotency of the postimplantation epiblast came from studies where isolated mouse epiblast was transplanted to ectopic sites in host animals. The resulting data provided evidence that the epiblast was capable of generating tissues from each of the three germ layers[22] . Clonal fate mapping studies subsequently revealed that single, labeled cells within the post-implantation epiblast could contribute to all embryonic tissues, including the germ lineage[23,24]. Despite this fact, the experiments also revealed a regional predisposition within the postimplantation epiblast in which cells in defined regions of the embryo (anterior, posterior, etc.) normally generate specific fates. The potential of epiblast cells, however, appears to be labile in that isolated cells that are transplanted to a different region of a host epiblast adopt the predicted developmental fate of the area to which they are relocated[25]. These data show that post-implantation epiblast cells are pluripotent and subject to regional influences to allocate them to differentiated lineages.

2.2.3.3 Primordial Germ Cells (PGC)

Primordial germ cells are found as well in the post-implantation mammalian embryo and represent the embryonic precursors of the adult gametes in the gonads of both sexes. They originate from the early post-implantation epiblastic cells and escape from the molecular differentiating pathways that affect the rest of the cells in the epiblast during gastrulation. In fact, avoidance of the somatic differentiation process seems to be related to the proximity of the extraembryonic ectoderm and visceral endoderm cells after implantation[23], because transplantation of cells from other parts of the epiblast to this location gives rise to germ cells, suggesting that inter-cellular signaling is

essential for defining PGC identity. Indeed, this external cues seem to induce specific gene expression in these epiblastic cells that will identify them as PGC[26] Right after implantation (5.5 dpc in mouse, by 2nd week in human) around a few PGC (50 in mouse, 100 in human) start to migrate through the hindgut and dorsal mesentery to arrive in the genital ridge (by 10.5 dpc, 6th week in human), having undergone an extensive proliferative activity during the move. In addition, while migrating, PGC progressively erase DNA methylation marks in the genome that will allow the constitution of a new imprinting pattern during the subsequent gamete formation. Up to this point PGC have kept refractory to all the somatic differentiating cues. However, once they reach the genital ridge, PGCs take up residence in a unique somatic micro-environment, know as niche, in which PGC-soma interactions provide the necessary signals that regulate the balance between self-renewal and differentiation that is needed for proper progression through gametogenesis. In this context, sex-determining molecular signals will induce the differentiation of PGC into spermatiogonial or oogonial cells, implying the acquisition of the unipotent identity.

2.2.4. Pluripotent cells in vitro

Different types of pluripotent cells have been isolated and maintained in vitro. In this section we describe their characteristics and explicit the similarities and differences with the *in vivo* counterparts they originate from.

2.2.4.1. Embryonic Carcinoma Cells (ECC)

The concept of pluripotent embryonic stem cells in vitro arose from pioneering work with mouse and human teratocarcinomas in the 1950s. Teratocarcinomas are malignant germ cell tumors that can be found both in gonads (eg. testicular germ cell tumors) or extra-gonadal sites[27]. The discovery that the incidence of spontaneous testicular teratomas in mouse strain 129 was as high as 1%-10% [28,29] made them amenable to experimental analysis for the first time. The earlier reports evidenced that these cancerous cells originate from PGC at different stages of development that had escaped from the unipotent fate determination and are transformed by unknown mechanisms to give rise to a group of malignant cells. An important finding was that these tumors comprise both an undifferentiated cellular component and a differentiated cellular component that can include tissues from the three germ layers. Subsequent experimental data showed that their undifferentiated component allows these tumors to be serially transplanted between mice, as these cells function as stem cells for the other differentiated tumor components. When these pluripotent cells PGC-derived malignant cells were isolated from the incipient tumors[30], they were shown to be capable of both unlimited self-renewal and multilineage differentiation in vitro and were referred to as

embryonic carcinoma (EC) stem cells. Mouse ECC lines were established in vitro the early 1970s[31] and were widely studied as “in vitro caricatures of development” as they could be cultured in sufficient quantities to perform experiments that would have been impossible with mammalian embryos. Indeed, mouse ECC where shown to express antigens and proteins similar to cells present in the ICM[32,33], which led to the concept that ECC are in vitro counterpart of pluripotent cells in the ICM[34]. Indeed, ECC were shown to contribute to various somatic cell types upon injection into mouse blastocysts[35]. However, most of ECC lines have limited developmental potential due to the accumulation of genomic alterations during teratocarcinoma formation and growth[36]. Indeed, when human ECC lines were derived in 1977[37] these cells were shown to be highly aneuploid, which likely accounts for their inability to differentiate into a wide range of somatic cell types, and which limits their utility as an in vitro model of human development. Interestingly, human ECC showed differences in cell-surface markers when compared with murine ECC. For example, SSEA-1, a cell surface marker specifically expressed on mouse ECC, is absent on human ECC, while SSEA-3, SSEA-4, TRA-1-60, and TRA-1-81 are absent on mouse ECC but are present on human ECC[38]

Despite the interference introduced by the neoplasic transformation process, these experiments established the existence of pluripotent cells and also provided the intellectual framework for the concept of stem cells to develop. This was also the first experimental demonstration of cancer stem cells, anticipating the current intense interest in this field.

2.2.4.2 Embryonic stem cells (ESC)

Embryonic stem cells were first derived in vitro from ICM cells of a mouse blastocyst simultaneously by Evans & Kaufman[39] and Martin[40]in 1981. From then on, ESC have been considered a self-renewing “frozen-in-time” version of the ICM cells they are derived from. Under appropriate conditions they exhibit unlimited self-renewal capacity while retaining the attributes of preimplantation epiblast identity and potency. Specially, when returned to the blastocyst, ESC are readily incorporated into epiblast and re-enter embryonic development to produce functional soma and germ cells (Bradley et al 1984). Moreover, when the 4n embryo complementation technique was available, mESC were shown to be able to from an entire organism[9,10].

The appropriate conditions to derive and maintain mESC in vitro originally included the use of mitotically inactivated cell feeder layers (murine embryonic fibroblasts, MEF) and serum-containing medium (that had been used first for the isolation of ECC lines). Later, medium that was “conditioned” by the co-culture with various cells that was found to be able to sustain ESC in the absence of feeders and serum. The fractionation of conditioned medium led to the identification of

leukemia inhibitory factor (LIF) as a cytokine that sustains mESC culture in vitro[41,42]. In serum-free medium, LIF alone is insufficient to prevent mouse ES differentiation, but in combination with BMP4 (bone morphogenic factor 4, a member of TGFβ superfamily) mESC are sustained[43].

Interestingly different murine strains show differences in their permissiveness to be derived into ESC. As a result, it has been suggested that the efficiency of mESC derivation is strongly influenced by genetic background.

There was a considerable delay between the derivation of mESC in 1981 and the derivation of hESC in 1998[44]. This delay was primarily due to species-specific ESC differences and sub-optimal human embryo culture-media. hESC maintain the developmental potential to contribute to advanced derivatives of all three germ layers in teratomas, even after clonal derivation[45]. Mitotically inactivated fibroblast feeder layers and serum-containing medium were used in the initial derivation of hESC, essentially the same conditions used for the derivation of mESC prior to the identification of LIF. However, it now appears largely to be a lucky coincidence that fibroblast feeder layers support both mouse and hESC, as LIF does not support hESC undifferentiated in vitro culture. What is more, LIF, as well as BMP4 promotes hESC differentiation. Contrarily, basic fibroblast growth factor (bFGF) allows the clonal growth of hESC on MEF in the presence of commercially available serum replacement[45]. Furthermore, in contrast to mESC, both activin and TGFβ have strong positive effects on undifferentiated proliferation of hESC in the present of low concentrations of bFGF, and based on inhibitor studies it has been suggested that TGFβ/Activin signaling is essential for hESC self-renewal[46-48]. Although other growth factors have been reported to have a positive effect on hESC growth including Wnt[49], IGF1[50] and heregulin[51], there are clearly additional important signaling pathways to be identified.

Remarkably, both human and mESC have been derived from earlier to blastocyst embryonic stages, including dissociated blastomeres of cleavage-stage embryos (4-cell and 8-cell)[52,53], whole preblastocyst embryos[54], morulae and blastocysts. It is not yet known whether pluripotent cell lines derived from these various sources have any consistent developmental differences or whether they have an equivalent potential. However, the derivation of pluripotent ESC from totipotent cells suggests either that the specific conditions to keep totipotent cells in vitro are elusive or that the pluripotent state has a greater stability than the totipotent state.

In spite of the time and knowledge accumulated from that initial murine and human derivation experiments, only putative ESC from a few other mammalian species have been derived to date, including rhesus monkeys[55] (1995), marmoset monkeys[55] (1996), and more recently

rat[56](2008). ESC from all these species have been extensively cultured displaying self-renewal capacity and the ability contribute to all germ lineages in chimeric animals (including germ line transmission in mouse and rat). ESC derivation from dog[57](2009) and rabbit[58,59] (2007,2009) preimplantation embryos have been reported recently, showing stable self-renewal and pluripotent developmental potential in form of teratoma formation, but no chimeric animal has been reported to date. Nevertheless, many attempts have been made to derive ESC from other mammalian species[60-64], including hamster, cow, buffalo, sheep, pig, goat, horse, and cat. However, none of the ES-like cells derived from these species adhere to the criteria set forth by mouse and hESC research to be bona fide ESC lines. These ES-like cell lines originated from explanted ICM cells cannot be maintained undifferentiated during long-term culture. Appropriate culture conditions need to be elucidated. Furthermore, thorough characterization of their pluripotency markers and developmental potential is necessary.

2.2.4.3 Murine epiblastic stem cells (mEpiSC)

Recently, murine epiblastic pluripotent stem cells (mEpiSC) were isolated from the cylinder-stage epiblast of postimplantation mouse embryos (5.5-6.5 dpc)[65,66]. mEpiSC differ significantly from mESC but share key features with hESC. For example, mEpiSC derivation failed in the presence of LIF the factor required for the derivation and self-renewal of mESC. In contrast, similar to human ESC, bFGF signaling appears to be critical for EpiSC derivation. In addition, gene expression by mEpiSC closely reflects their post-implantation epiblast origin and is distinct from mESC. Nevertheless, mEpiSCs do share the two key features characteristic of mESC: self-renewal in vitro and multi-lineage differentiation capacity. Unlike mESC, mEpiSC show little or no capacity to colonize developing embryos when introduced into blastocyst. However, they do form multidiferentiated teratomas when injected into adult mice, however, showing high degree of multilineage differentiation.

2.2.4.4 Embryonic germ cells (EGC)

Despite the evidence that teratocarcinomas were derived from primordial germ cells[67] (PGCs), it was not until 1992 that pluripotent stem cells (embryonic germ cells, EGC) were successfully derived from murine PGC directly in vitro[68,69]. In contrast to mESC, the initial derivation of mouse EGC requires a combination of stem cell factor (SCF), LIF, and FGF in the presence of a feeder layer. In culture, EGC are morphologically indistinguishable from mESC and gene expression profile of murine EGC is strikingly similar to that of ESC. And similar to ESC, upon blastocyst injection, they can contribute extensively to chimeric mice including germ cells[70,71].

Unlike ESC, however, EGC retain some features of the original PGCs, including genome-wide demethylation, erasure of genomic imprints, and reactivation of X-chromosomes[70,72], the degree of which likely reflects the developmental stages of the PGCs from which they are derived[73]. However, in vitro derivation of EGC forces a step of dedifferentiation of PGC to convert germ cells committed to become unipotent into pluripotent stem cells. Thus, while PGC represent a unipotent lineage, their genome regulation is closer to that of the ICM of blastocysts, which explains why EGC are like ESC and unlike EpiSC.

The derivation of human EGC was reported in 1998[74], but in spite of efforts by several groups, their long-term proliferative potential appeared to be limited[75]. Early passage human EGC have been reported to differentiate into multiple lineages in vitro, but this has yet to be demonstrated from a clonally derived, stable cell line, nor to date have any human EGC lines been reported to form teratomas. Besides having different growth factor requirements from human ESC, human EGC have a very distinct morphology and express SSEA-1, a cell-surface marker absent on human ESC but present on early human germ cells. Human ECC are also germ cell-derived and share markers and the basic morphology of human ESC, so these differences suggest that a final step in converting these human germ cell lines to a proliferative cell comparable with human ESC/ECC cells is still missing. The properties of the human EGC lines reported to date suggest fundamental species-specific differences between the early germ cell biology of mice and humans and suggest that a human counterpart truly comparable with mouse EGC has yet to be derived.

2.2.5. Multipotent stem cells in vivo and in vitro

Gastrulation is the defining feature of metazoan development that serves to distribute seemingly equivalent pluripotent cells to specific fates. The three embryonic germ layers generated during gastrulation from the pluripotent epiblast, including ectoderm, mesoderm, and definitive endoderm, contain the multipotent stem cells required to build all of the tissues of the developing organism. These stem cells play a critical role in the establishment of embryonic tissues during development and in some cases are retained into adulthood. These cells, generically called adult stem cells (ASC), retain the ability to self-renew and are lineage committed. Thus, despite no longer being pluripotent, ASC keep a developmental potential that confers the ability to originate more than one terminally differentiated cell types (multipotent stem cells) or one terminally differentiated cell type (unipotent stem cells). Differences in developmental potential and lineal relationships between stem and progenitor subsets establish a hierarchical structure for primitive stem cell compartments. Stem cells reside at the top of the hierarchy and give rise to multipotent progenitors

(which have lost the ability to self renew) that in turn give rise to progenitors with more restricted lineage potential. Although best exemplified in the hematopoietic system, the existence of hierarchical relationships between stem and progenitor cells is emerging as a common feature of other tissue-specific stem cell compartments as well[76].

ASC ensure that tissues can be functionally sustained throughout the lifetime of the organism. These remarkable regenerative capacity place stem cells in an elite class of essential cells of living organisms. Some tissues of the organism, such as those in the brain and skeletal muscle, have very little turnover and are well protected, whereas others turnover constantly. Even though the intrinsic properties of ASC are likely to be similar across tissues, each tissue has its own requisites for homeostasis and regeneration. We lose over 20 billion cells a day, requiring constant replenishment to stay alive. More than a billion of these lost cells come from our blood, necessitating a reservoir of constantly renewing hematopoietic stem cells[77]. The intestinal epithelium also undergoes constant turnover, taking only 3–5 days for undifferentiated cells at the bottom of the invaginating crypt to proliferate and differentiate into the enterocytes, goblet cells, or enteroendocrine cells of the absorptive villus[12] . Analogously, every 4 weeks, we have a brand new epidermis as cells in the basal layer terminally differentiate and are shed from the skin surface[78].

However, in vitro, the study of adult stem cells is hampered by a number of factors including low frequency, dependence on functional assays, lack of clear methods to identify them and the inability to maintain their ‘stemness’ in culture. Moreover, implicit in the generative and regenerative functions of tissue-specific adult stem cells is the proper localization of these precursors, which is essential to build organs and tissues during development and to promote localized tissue defense and repair after damage. New studies provide increasing support for the notion that stem cells in vivo require inputs from particular defined microenvironments, or ‘‘niches,’’ which support their unique stem cell functions[79]. Long-term maintenance of stem cells, therefore, requires their migration to and engraftment within supportive stem cell niches, which are conditions that are difficult to reproduce in vitro. Moreover, It has been suggested that in vitro culture environments sometimes alter the patterning of cells in ways that modify their fates and even their developmental potentials[80]. These concerns apply to other mammalian adult stem cells that have been identified and studied primarily based upon their behavior in culture or after expansion in culture. For that reason, isolation and maintenance of adult stem cells in vitro implies a greater challenge than stem cells with greater developmental potential.

2.3 Molecular basis of the developmental potential and differentiation

Mammalian development and cell fate determination in vivo is process characterized by phases of reprogramming and differentiation that require close coordination between genetic and epigenetic events. The cellular context in form of external signaling molecules together with interactions among neighboring cells induce appropriate transcriptional, post-transcriptional and epigenetic responses that activate distinct developmental programs in time and space while ensuring heritability of the phenotypic state to the progeny of cells. Thus, on the whole, cells in a developing organism progressively acquire an distinctive array of epigenetic marks (by a process known as epigenetic programming) that are important for cell specification because they establish a heritable memory of cellular states and differentiation pathways. This balance between external factors and internal regulators determine the genesis and the subsequent propagation of the developmental potential through cell division (Fig. 4).

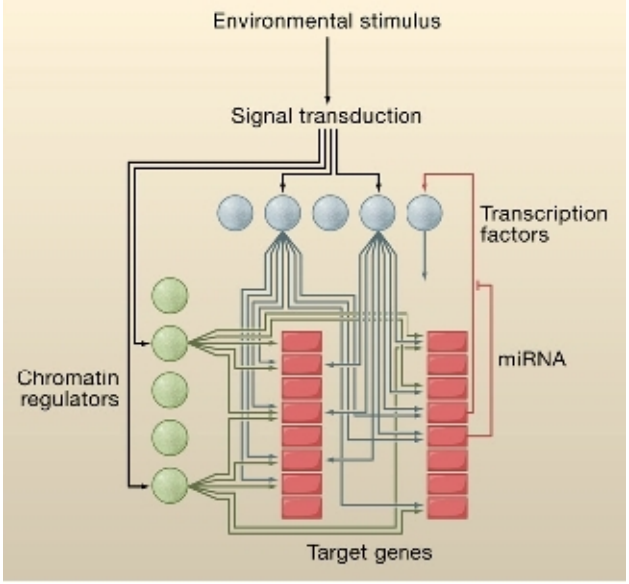


Figure 4 . Complex regulatory network influencing the cell identity (from Jaenisch R, 2008[81]) .

2.3.1. Totipotency

During the short and temporally unique period of the embryogenesis that starts right after fertilization, transcriptional activity has no apparent role in differentiation but rather they ensure the switch from reliance on maternally provided transcripts to active zygotic transcription (ZGA). Even at these early stages, many factors control transcription: specific transcriptional regulators, regulatory RNAs and epigenetic remodelling. The degradation of many maternal RNAs relies on members of the RNA-induced silencing complex in mouse embryos. In addition, the first, so-called minor phase of ZGA requires specific transcriptional regulators, such as transcription intermediate

factor 1α (Tif1α) and proteins of the nucleosome remodelling complex subunit, that are enriched in the sperm pronucleus. Specifically, unlike the female pronucleus, the male pronucleus is extensively remodeled, replacing protamines with histones and undergoing paternal-specific active demethylation of DNA[82-84]. Subsequent cleavage divisions (up to morula stage) are characterized by further passive DNA demethylation and the disappearance of the transcriptionally repressive marks of H3K9me2 and H3K9me3.

Thus, growing evidence for the involvement of nucleosome remodelling complexes in the early stages of development points to the importance of epigenetic regulation of chromatin and a need for an understanding of the role of such complexes in developmental fate decisions. This molecular data in addition to experiments testing the developmental potential of isolated blastomeres and experiments deriving pluripotent ESC from 4 to 8 cell cleavage stage embryos suggest that totipotency is a transient and unstable state that seems to be lost around the third cleavage division in mouse embryos. At this point, embryonic cells commence a gradual specification towards trophectoderm and inner cell mass, that entails the first cell fate decision during embryogenesis.

2.3.2 Lineage commitment in the preimplantation embryo

The configuration of the blastocyst by 3.5 dpc in mouse and by day 5 after fertilization in human implies the formation of three distinctive cellular population: the trophoblast, the epiblast and the primitive endoderm. Molecularly, the formation of the blastocyst implies two consecutive cell-fate decisions in the embryonic cells forming the developing mammalian embryo. The first fate decision implies the conformation of two cell populations. After morula stage, cell positioned in the inside retain pluripotency and cells on the outside develop into extraembryonic trophectoderm[85-87]. This first set of extraembryonic cells will support the development of the embryo in the uterus and provide signaling sources to pattern the embryo before gastrulation. The generation of inside cells requires outer cells to divide in an orientation such that one daughter cell is directed inwards previously to morula formation. Once these populations of cells are set apart, inner cells develop a stable regulatory circuit in which the OCT4 (Pou5f1)[88,89], SOX2[90], NANOG[91,92]and SALL4[93,94] transcription factors promote pluripotency maintenance. By contrast, in outside, trophectoderm destined cells, transcription factors such as TEAD4[95,96], CDX2[97-99] and EOMES[100] become upregulated. Reciprocal repression of trophectoderm targets by OCT4, SOX2 and NANOG in the pluripotent lineage[101,102]. However, before cavitation, pluripotency factors OCT4, SOX2 and NANOG are present in both inside and outside cells. It is upon cavitation that expression of SOX2

and NANOG is rapidly downregulated by CDX2 and in its absence, the trophectoderm is specified. OCT4 expression is downregulated in a slower way, being express in trophectoderm cells until full blastocyst expansion.

Once the blastocoelic cavity is formed, the second cell fate decision takes place: cells of the ICM that are in contact with this cavity are set aside to form the second extraembryonic tissue, the primitive endoderm (PE). The rest of the ICM cells escape this second differentiation process, express pluripotency genes and become progenitors for all cells of the future organism. PE differentiation necessitates the activation of the Gata4 and Gata6 transcription factor genes[103,104], and perhaps of genes encoding other factors yet to be discovered. These transcription factors are proposed to antagonize the expression of pluripotency transcription factors, such as Nanog[105]. Following Gata4 and Gata6 expression, proteins required for PE integrity become upregulated[106,107].

2.3.3 Crucial factors necessary to maintain pluripotency

The combination of global detection methods for transcription factor target genes and epigenetic modifications has revealed the existence of an intriguing interplay between pluripotency factors and epigenetic modifiers. In fact, the dynamic balance between these two regulatory systems probably forms the basis for the pluripotent state.

2.3.3.1 The core transcriptional circuitry of pluripotency: Oct4, Sox2 and Nanog

Although pluripotency signaling pathways differ between mice and humans[108], the core transcriptional circuitry formed by OCT4, NANOG and SOX2 seems to be remarkably conserved[109-114]. Indeed, genome wide-studies have highlighted the colocalization of these three TF in murine and human ESC chormatin[115,109]. This central core module is enhanced by positive feedback and feedforward loops. Particularly, OCT4 and SOX2 form a heterodimer that positively regulates the expression of *Oct4*, *Sox2* and *Nanog*[115,109,116,117]. In addition, NANOG also interacts directly with OCT4[118] and positively regulates the expression of all three genes[115]. Thus, these three transcription factors regulate their own and each other expression in a highly coordinated manner, involving positive protein–protein and protein–DNA feedback loop interactions. Indeed, multiple additional transcription factors have been found to colocalize with them. Furthermore, all three transcription factors promote the transcription of pluripotency-promoting genes such as *Sall4*, *Tcl* (T cell leukaemia/lymphoma), *Tbx3*, *Rest*, *Zic3*, *Hesx1* (homeobox expressed in ESC 1), *Stat3* (signal transducer and activator of transcription 3), *Rex1* (also known as Zpf42), *Tcf3* and *Dax1* [115]. These TF have been found to colocalize with OCT4, SOX2 and NANOG in the maintenance of the pluripotent state. In addition, experiments using ESC have shown that the pluripotency triad of TF

repress the expression of the genes involved in lineage commitment. These repressed genes include: *Hand1* (heart and neural crest derivatives-expressed 1), eomesodermin (*Eomes*) (both involved in trophectoderm development); *Lhx5* (LIM homeobox 5), *Otx1* (orthodenticle homologue 1), *Hoxb1* (all involved in ectoderm development); *Myf5* (myogenic factor 5), *T* (brachyury protein homologue), *Gsc* (goosecoid) (all involved in mesoderm development); and *Foxa2* (forkhead box A2) and *Gata6* (GATA-binding protein 6) (both involved in endoderm development).

Thus, OCT4, SOX2 and NANOG are central to the maintenance of pluripotent cellular identity; appropriate expression of this protein trio holds the cell in a pluripotent self-renewing state by activating other pluripotency-specific genes and repressing genes that are associated with lineage commitment.

2.3.3.2. Cell cycle in pluripotent cells

Relative to most somatic cells, ESC divide very rapidly. Undifferentiated mESC transit the cell cycle once every 8–12 h, depending on the cell line and cultivation conditions. Human ESC take approximately 15–30 h to transit this cycle[119,120]. Irrespective of the timing differences, both mouse and human ESC have a very short G1 phase. In mouse it is roughly 1–2 h[121] and in primate and human cells it takes about 2.5–3.0 h [119]. In contrast, mouse embryonic fibroblasts transit the G1 phase in 6–12 h and many adult cells take even longer. In somatic cells, passage from G1 into the S phase normally requires mitogen activated cyclin dependent kinases (Cdk) 4 and 6, cyclins (D and E), and members of the retinoblastoma (Rb) tumor suppressor protein family, including pRb, p107, and p130. In hESC, pRb, p107 an p130 are hyperphosphorilated and inactive , which allows E2F-responsive genes to be transcribed independently of cell cycle phase.

The only cell cycle regulators that show cell cycle-dependent expression in undifferentiated mESC are Cdk1 and cyclin B1, both of which show regulation during the G2 phase of the cell cycle[122]. An uncoupled G2 mitotic-spindle checkpoint, which normally helps maintain chromosomal integrity during cell divisions, does not however initiate apoptosis in mouse and human ESC as it does in somatic cells; consequently, ESC fail to undergo mitosis and develop polyploidy. The absence of a robust checkpoint at this phase of the cell cycle is a likely source of karyotypic abnormalities in ESC and their derivatives[123].

Differentiation of pluripotent ESC leads to progressive up-regulation of D cyclins[121], decreased Cdk activity and regulation by the Rb pathway[124]. Moreover, proliferating mESC contain very little of the truncated form of cyclin A2, which binds to and activates cyclin-dependent kinase 2 (CDK2). Importantly, changes in these cell cycle regulators with differentiation occur prior to loss

in OCT4, NANOG, or SOX2 protein abundance, suggesting that the absence of robust cell cycle checkpoints may be critical to the maintenance of pluripotency. Concomitant with the establishment of robust cell cycle checkpoints is the enhanced potential for apoptosis and other forms of cell death[123].

In contrast to mouse, a number of essential regulators show cell cycle dependent levels of expression in human ESC. NANOG regulates S-phase entry in hESCs via transcriptional regulation of cell cycle regulatory components[125]. Irrespective of any species differences, cell cycle regulation in ESC is fundamentally different from that of other somatic cell types, and an understanding of the processes that maintain the unique features of pluripotent cell cycle regulation appear critical to understanding pluripotency.

2.3.3.3 Epigenetic regulation of the pluripotency networks

In addition to the triad of TF, and the check-point free cell cycle, chromatin organization and epigenetic modifications are also key elements of the pluripotent cellular identity. Gene repression mediated by the polycomb-group (PcG) proteins and the conferred H3K27me3 mark is required for ESC pluripotency and plasticity during embryonic development[126,127]. Chromatin immuno-precipitation studies have shown an unexpected but potentially key concept in the biology of ESC pluripotency: genes that are repressed in ESC but are required for later differentiation are marked by bivalent H3K4me3 and H3K27me3 domains that make them poised for activation [128-130]. In fact, the H3K4me3 and H3K27me3 marks can effectively discriminate genes that are expressed (H3K4me3), poised for expression (H3K4me3 and H3K27me3) or stably repressed (H3K27me3), and therefore reflect the cell state and lineage potential[130]. Approximately one-third of genes, however, are not marked by either H3K4me3 or H3K27me3. Interestingly, these genes are mostly repressed in ESC due to a tendency to be marked by DNA methylation, which is therefore a complementary mechanism to histone modifications to ensure appropriate gene expression and heritable gene repression simultaneously [131-133].

Epigenetic modifiers that confer gene silencing have been shown to be bound by OCT4, NANOG, SOX2 and/or some of the pluripotency promoting TF. In fact, one of them, SALL4 has been shown to have a unique role in the cross-talk between the pluripotency-promoting TF and the epigenetic modifiers as it binds cooperatively with polycomb-repressive complexes 1 and 2 (PRC1 and PRC2) to trimethylate H3K27me3 at some loci (but does not require PRC1 or PRC2 to bind to other sites[134]). This observation shows a direct interaction between a TF and epigenetic modifiers in the maintenance of pluripotency. Other proteins involved in epigenetic modification machinery that

have been shown to bind OCT4, NANOG, SOX2 and/or SALL4 include proteins involved in the establishment of H3K27me3 marks (SUZ12 and EZH2[135-137]), histone-methyltransferases (EHMT1, G9A[137-139]), DNA-methyltransferases (DNMT3A, DNMT3B[140]).

2.3.4. Induction of pluripotency: iPS cells

The stepwise acquisition of epigenetic marks culminating in the fixation of lineage fate by differential DNA methylation of gatekeeper genes forms the molecular basis for Waddington's concept of the canalization of developmental potential[141]. Waddington compared the path of a cell lineage towards terminal differentiation with a ball traveling downwards along branching valleys (Fig 4); once in its final valley, the ball cannot easily cross the mountain into neighboring valleys or return to the beginning[141]. This canalization explains how cellular differentiation pathways become stable and potentially irreversible. It also shows the unidirectionality of differentiation that directs developmental progression and prevents teratogenesis during growth.

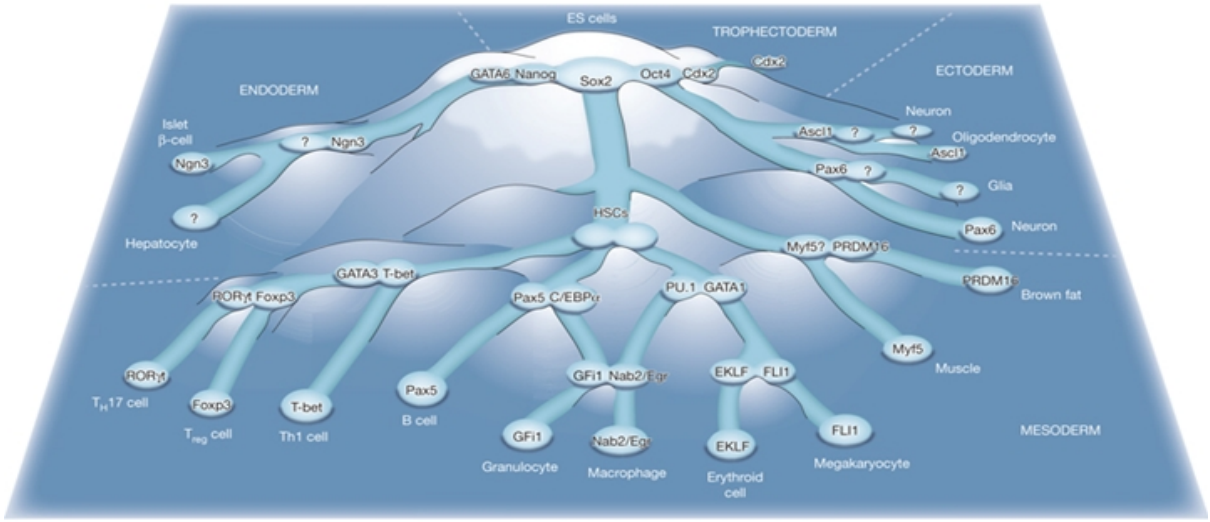


Figure 5 Representation of developmental potential analogy as gravitational potential energy suggested by Waddington. The path to the least developmental potential (differentiated cell lineages) is full of meta-stable states represented by local valleys at different "heights". Thus, pluripotent state is seen as a meta-stable state on the top of the "mountain" of the developmental potential (from Graf T and Enver T, 2009[142]).

However, in 2006, Yamanaka and colleagues reported the ability of fully differentiated cell types to derive induced pluripotent stem (iPS) cells[143]. This experiments implied the groundbreaking reversion of this strict developmental progression, showing the possibility to create bidirectional developmental pathways by experimental manipulation[144,143,145,146]. Thus, temporally limited

overexpression of the four transcription factors OCT4, KLF4, SOX2 and MYC is sufficient to reactivate endogenous pluripotency genes and to regain a developmental potency similar to that of ESC. Other combinations of factors that have been successfully used for the derivation of iPS cells[147].

In addition, efficiency of iPS cell derivation can be enhanced in the presence of inhibitors of histone deacetylases, the histone methyltransferase G9A and DNA methylation[148-150], and these components can substitute for some reprogramming factors, a fact that reflects the necessity for extensive epigenetic reprogramming in the iPS cell generation process. In efforts to understand how such developmental reversion can be achieved in principle, recent computational modeling has produced important insights[151,152]. In these models, transcriptional feedback loops create molecular switches that result in successive gene restrictions that correspond to a controlled differentiation cascade. Although cells are robustly resistant to reprogramming, these models also predicts that amplification of low-level variations in expression (also called “transcriptional noise”) may be sufficient to trigger the reactivation of the core pluripotency switch. The important insights from such studies are that reprogramming towards pluripotency depends on the positive interference of several factors, each with a certain noise amplitude of expression, thus explaining the stochasticity of experimental reprogramming as well as the certain extent to which reprogramming factors can be exchanged. It is also becoming clear that the reversion of a cell into a state of increased potency requires temporally higher levels of gene expression, equivalent to an activation energy, than those required once that state is reached. It is even the case that cells on these meta-stable states can tolerate complete (such as NANOG in ESC[153]) or temporary (as in stochastically expressed genes) absence of some factors that are essential to establish this state in the first instance without loss of developmental potency. With better knowledge of the precise molecular processes during reprogramming, it will be possible to define the exact composition of the factors that are required for each step.

2.4. Derivation of human embryonic stem cells

The aim of the derivation is to induce self-renewable clonal expansion of the pluripotent cells in preimplantation embryo in vitro. Since pluripotent cells are the less exposed lineage of the blastocyst, if they are to be expanded, the derivation method needed to overcome the obstacles to ensure direct contact of these cells with the specific culture conditions. The protocols used to ensure this contact and derive embryonic pluripotent cells in vitro are varied and a no consensus about the optimal procedure has been achieved after twelve years since the first human ESC line was

derived[154]. Indeed, despite the increasing number of newly-derived hESC lines worldwide in recent years (around 600 have been registered to date according to the European Registry[155]) derivation remains a partially inefficient procedure (published derivation efficiencies range 8.5 to 58%[154]). The variable success rates depend on several factors that could be divided into those depending on the biological characteristics of the embryo, and those depending on the derivation procedure.

2.4.1 Sources of human embryos for hESC derivation

2.4.1.1 Normal surplus human embryos

IVF patients are the main source of embryos for hESC derivation, donating both surplus embryos[44,156] and embryos with poor morphological quality (slow developing, suboptimal morphology)[157-159]. Human ESCs have been generated from normal diploid embryos. Human embryos are available by consenting IVF patients donating either their fresh (immediately after transfer) or frozen (cryopreserved embryos that are not used for transfer) embryos.. Additionally, the success of the freezing-thawing techniques will determine the number and quality of embryos available for hESC derivation. Despite those variations and limitations it is clear that derivation of the hESC is equally successful when using either fresh or frozen human embryos[160].

2.4.1.2 Preimplantation genetic diagnosis (PGD) and preimplantation genetic screening (PGS) embryos

PGD is a procedure originally conceived to avoid the use of embryos obtained by IVF potentially affected by parentally inherited genetic conditions. PGS is a specific type of PGD in which aneuploid embryos are discarded. Both procedures that involve the biopsy of one or two blastomeres at the 6-8 cell stage (D3) or blastocyst stage (D5) in order to perform fluorescent *in situ* hybridization (FISH) analysis (PGD and PGS) or PCR analysis (PGD). If these analyses reveal an abnormality, the embryos are susceptible to be donated for research and thus, be used for ESC derivation. Embryos donated after PGS analysis, including those that fertilized normally but carry chromosomal abnormalities comprise a large group available for hESC research. Previous attempts to generate stem cell lines from such embryos showed the derivation of euploid stem cell lines was possible[161,162]. Interestingly, a “self-correction” of chromosomally abnormal embryos in ESC culture has been suggested, where ESC lines were derived from aneuploid embryos[163]. One possible explanation for “self-correction” is trisomic zygote rescue, where an extra chromosome is lost during mitosis by and duplication of a single chromosome may occur by uniparental disomy (UPD). Besides triploid rescue, UPD can also occur in cases with monosomy rescue, where haploid chromosomes can be duplicated.

Human ESC can also be derived from embryos carrying specific monogenic genetic diseases. These lines represent an excellent in vitro model of specific genetic diseases. Several hESC lines have been derived for diseases such as Huntington's Disease, Cystic Fibrosis, Fanconi Anemia, Myotonic Dystrophy and others[164].

2.4.1.3. Discarded human embryos

In 1-3% of the cases, 1PN and 3 PN embryos can be formed during in vitro development[165] after IVF. These embryos can be used for hESC derivation as they are not to be used fro transfer. A diploid hESC line was derived from an apparently mononuclear human zygote (showing one pronucleus (PN) after insemination)[166], probably due to the fact that an asynchrony in PN formation lead to the wrong classification of the zygote. In fact zygotes with a single pronucleus were reported to be diploid in the majority of the cases[167].

2.4.2 Derivation methods from human blastocysts.

Embryos are cultured in vitro until blastocyst stage, which in humans occurs between day 5 and 6 after fertilization. ESC are normally derived at this stage, despite some lines have been derived from blastocysts cultured between 7-9 days after fertilization.

There are three main methods that can be used to derive ESC from blastocysts: a) whole embryo culture with subsequent isolation of the ICM outgrowth; b) immunosurgical ICM isolation; and c) mechanical ICM isolation.

2.4.2.1 Whole embryo culture

The easiest derivation method, whole embryo culture involves placement of the whole ZP-free blastocyst on a feeder cell layer[168]. If successful, the ICM cells grow together with the TE cells as a monolayer. A significant difficulty associated with this method is that the TE outgrowth can suppress the ICM growth into ESC or by induction of ICM differentiation[169]. Whole embryo culture is the method with the lowest derivation rates (8.5-10%)[154] due to the mentioned difficulties.

2.4.2.2 ICM isolation by immunosurgery

It is the most common method for derivation, was developed in 1975 for the study of mouse development[170] and later used for derivation of mESC[39,40] and hESC[44]. It includes the pretreatment of the zona pellucida-free blastocyst with anti-human serum followed by lysis of the trophectoderm (TE) cells with a guinea-pig complement[170,171]. An advantage of this method is the complete ICM isolation with subsequent plating on the feeder cells, free of TE cells. This method can only be used

if the TE layer is intact and the ICM is confined in the blastocoelic cavity, because the complement would destroy any exposed ICM cell. This method has shown a higher range of derivation rates than the whole embryo method, ranging 14-35%[154].

2.4.2.3 Mechanical isolation of the ICM

Mechanical ICM isolation is an alternative to get rid of the TE. There are different methods to mechanically disrupt the blastocyst to separate the ICM from the TE. The earliest strategy used flexible metal needles with sharpened tips, 0.125 mm in diameter to open the zona pellucida and extract the ICM under a stereomicroscope[172]. A drawback of this method is the practical impossibility of removing all the TE cells without damaging the ICM. The second strategy implemented to mechanically isolate the ICM is the use of glass needles and micromanipulators assisted by a non-contact laser. If the ICM is clearly defined, this method allows a higher precision in the isolation of the pluripotent cells[173,174].

Mechanical ICM isolation is also known as partial embryo culture method due to the subsequent culture of the separated ICM area on feeder cells. This strategy has been shown to be the most successful to derive stem cells, with derivation rates as high as 58%[154].

2.4.3 Derivation of hESC from pre-blastocyst human embryos

Derivation of hESC lines was first achieved using single cleavage stage embryo blastomeres in 2006 by Lanza and colleagues[175,176] and since then, it has been reproduced by other groups[177]. On of the purposes this group was to overcome the necessary embryo destruction that is concomitant to the standard hESC derivation. Thus, hESC derivation is possible from embryonic cleavage stages, when cells are totipotent. This method consists of a single blastomere biopsy and the subsequent co-culture of the blastomere with established hESC that confer the optimal milieu for the blastomere growth. This method has been used to derive stem cells from embryos that after biopsy and uterus replacement, gave rise to healthy newborns. Thus, these hESC lines could be used in a cellular therapy without any risk of immunological rejection.

Derivation of hESC lines from day 4 compacted morulae has been described by Sretlchenko and colleagues[178]. Using this technique the whole denuded (ZP-free) embryo is seeded on the feeder cell layer. Since the first cellular differentiation has not taken place at morula stage, a certain proportion of outgrowths were composed with trophectodermal cells, which impaired the ESC derivation efficiency. This procedure was used later by Tesar and colleagues, who had derived murine ESC from different pre-blastocyst embryonic stages, including morula[179].

2.4.4 hESC culture methods

Human ESC were originally cultured on mitotically inactivated murine embryonic fibroblasts (MEF) feeder layers and serum-containing medium[44]. Considering that culture conditions are essential to arrest the natural tendency of pluripotent cells towards differentiation, this was a fortunate finding due to the little knowledge about the factors influencing the maintenance of pluripotency and self-renewal.

Since then, a lot of efforts have been put in the optimization the ESC culture conditions. Most of improvements have consisted in finding culture mediums with defined compositions that allow to control and study the factors influencing the self-renewal and pluripotency maintenance in vitro. This approach can also help to identify those components in non-defined media that promote differentiation. In addition, the anticipated use of hESC for cellular therapies has prompted the necessity of xenobiotic-free derivation and culture conditions for these pluripotent cells.

2.4.4.1 Culture substrates: feeder cells and protein matrices

The original culture method implied the use of a feeder cell layer that conferred a physical and paracrine support for the cells to grow. Feeder cell layers have been shown to produce growth factors that contribute to the maintenance of the pluripotency of the hESC lines. These factors are released into the cellular milieu[180] that can be used as conditioned-medium to grow the cells in feeder-free conditions. When hESC are cultured on a feeder-cell layer, colonies tend to be rounded and compact.

There are several types of feeder cells onto which hESC can be cultured. As mentioned above, the first hESC were derived using mitotically inactivated MEF [44,181]. It was later that human fibroblast cultured from newborn foreskin biopsies were used as feeder layer for derivation and culture of hESC[182], reducing the xenobiotic presence of murine feeder cells. In 2005, several groups reported the culture of hESC on autogeneic fibroblasts, which had been differentiated from hESC themselves[183-185]. This methodology minimizes the possible genomic cross-contamination between the feeder cells and the stem cells and reduces the hypothetical transmission of pathogenic agents. Other cellular types that have been described to support hESC culture human fetal lung fibroblasts[186], human umbilical cord blood (UCB)-derived fibroblast-like cells[187], mesenchymal stem cells[185], human endometrial cells[188], and cells from the fetal liver stroma[189]. From these reports we can conclude that the factors that support undifferentiated hESC growth can be found in various tissues and are neither species-specific nor tissue-specific.

The alternative to the co-culture of hESC with feeder-cell layers is the use on protein matrices. This is a very extended way of culturing hESC that avoids cross-contamination of any kind with other cells, which eases the handling of the hESC cultures and their characterization. Different protein matrices have been described to support hESC culture and derivation. There are two groups of these matrices, depending on the complexity of their composition. In the first group, we find the matrices with a complex composition and on the other group those with defined composition. The main exponent of the complex matrices group is Matrigel, which is the commercial name of the extracellular matrix secreted by a mouse sarcoma cell line that is composed including laminin, collagen IV and other undetermined factors[190,191]. On the second group we find different type of purified glycoproteins that polymerize forming a matrix to support hESC cultures: laminin[191,192] and a defined mix of laminin, fibronectin and vitronectin, plus collagen type IV[193]have been used. Despite their component definition, glycoproteins used in this matrices can be isolated from human or animal samples, which implies either the use of xenobiotic materials (isolated from animal source) or the use of potential source of pathogenic agents (isolated from human source). Thus, the recent report about the successful use of recombinant vitronectin to culture hESC leads the way of a feeder and xenobiotic-free hESC culture[194].

2.4.4.2 Culture media

The original medium used to derive the first hESC line contained fetal bovine serum (FBS) as a source of protein. Despite it could support the derivation and early culture, it was soon realized that it also promoted the differentiation of the cells. As a consequence, a new generation of mediums were developed using defined basic mediums developed for other type of cells and purified protein mixtures of animal origin, which allows a higher control and standardization in comparison to raw FBS. Thus, the most widely used protein-source for medium supplementation after FBS was the commercially available *Knockout Serum Replacement* (KO-SR) (Invitrogen)[195]that contains bovine serum albumin (BSA) and other components which are claimed to be defined by the company. This serum substitute allows to keep stable undifferentiated culture of hESC.

However, the drawback of the mentioned serum-substitute is the animal origin of its components. Therefore, many groups are trying to set up culture conditions that are free of any animal component and that minimizes the possibility of contamination with human pathogen. Thus, chemically defined, xenobiotic-free and recombinant components are being tested to derive and maintain hESC in culture. In 2006 the first hESC lines were derived in xeno-biotic free conditions[196]

using human serum instead of KO-SR. Again, a high tendency towards differentiation was observed in similar terms to those observed using FBS. Recent reports describe the use of new mediums and supplements free of xenobiotic components. All the commercial brands that manufacture these mediums are secretive about their composition for obvious reasons. Therefore, another approach used is the manufacture of chemically defined mediums, in which an exact account of all the components and their proportion in the medium is known. They allow to better define the growth factors (and the concentration) that are needed to keep pluripotency and to direct differentiation. Several chemically defined mediums have been reported to successfully sustain hESC culture[192,197-199]

In order to optimize the human ESC culture conditions, efforts were made to identify components released by the feeder cells needed for hESC self renewal and the molecular effects they exerted on the cells. Analysis of the protein composition of feeder-conditioned media of animal (MEF) and human origin revealed that in contrast to mESC, FGF and TGF/Activin/Nodal signaling pathways are of central importance to the self renewal of human ESC. Basic FGF allows the clonal growth of human ESC on fibroblasts in the presence of KO-SR[45]. At higher concentrations, bFGF allows feeder independent growth of human ESC cultured in the same serum replacement[184,200,201] . The mechanism through which these high concentrations of bFGF exert their functions is incompletely known, although one of the effects is suppression of BMP signaling[201]. In addition, it was found that serum and KO-SR have significant BMP-like activity, which is sufficient to induce differentiation of human ESC, and conditioning this medium on fibroblasts cell cultures reduces this activity, which confers the feeder the quality of withdrawing the differentiation-promoting factors in the medium. At moderate concentrations of bFGF (40 ng/mL), the addition of noggin or other inhibitors of BMP signaling significantly decreases background differentiation of human ESC. At higher concentrations (100 ng/mL), bFGF itself suppresses BMP signaling in human ESC to levels comparable with those observed in fibroblast conditioned medium, and the addition of noggin no longer has a significant effect. Suppression of BMP activity by itself is insufficient to maintain human ESC[201], thus additional roles for bFGF signaling exist. Evidence suggests that bFGF up-regulates the expression of TGF ligands in both feeder cells and human ESC, which, in turn, could promote human ESC self-renewal[202] . Human ESC themselves produce FGFs, which appear insufficient for low-density cell culture but can maintain high-density cultures for variable periods. Inhibition of FGFRs by small molecules causes differentiation of human ESC[203] , suggesting the involvement of FGFRs. The required downstream events, however, are still not well understood, but some evidence implicates activation of the ERK and PI3K pathways[204,205]. Both Activin and TGF

have strong positive effects on undifferentiated proliferation of human ESC in the presence of low or modest concentrations of FGFs, and based on inhibitor studies, it has been suggested that TGF/Activin signaling is essential for human ESC self-renewal[46,206,48].

2.4.5 hESC expansion methods

Human ESC cellular expansion is a required step in order to obtain a sufficient cellular population to plan any expression or genomic analysis, as well as any possible cellular therapy. In contrast to the mESC, where clonal propagation is a standard method of culture, it seems that cell-cell interaction is critical for efficient hESC expansion, where the loss of gap junctions between hESC can increase apoptosis and inhibit growth[207]. Thus, hESC need to be replated in clumps containing between 20-200 cells.

There are two main strategies to expand hESC, the mechanic and the enzymatic. To mechanically expand a hESC culture, colonies have to split into small pieces containing between the appropriate number of cells and transfer them into a new and freshly conditioned culture dish. This method is preferentially used when hESC are cultured on a feeder-cell layer, and colonies are more dense and compact. The enzymatic expansion of the hESC colonies consists in the use of proteolytic enzymes to disaggregate the colonies into single cells or small aggregates of up to 20-40 cells. This method is preferentially used when hESC are cultured on protein substrates due to the tendency of the cells to grow in less dense colonies in such conditions, which hampers use of the mechanical method. Enzymes normally used to disaggregate colonies are trypsin, collagenase or dispase. The mechanical expansion is normally done manually and is more labor-intensive and time consuming than the enzymatic expansion method. However, some authors have found a possible connection between the enzymatic method of expansion and a higher rate of genomic instability in the hESC in comparison with those mechanically expanded[208].

2.4.6 Genomic integrity of hESC lines

Human ESC have been successfully cultured for extended periods and through numerous passages while maintaining a normal diploid karyotype[156,45]. However, hESC lines can develop an abnormal karyotype after long-term culture. Long-term culture of mESC can lead to a decrease in pluripotency and aneuploidy which is the major cause of failure to differentiate to all tissues of the adult chimera, including the germline[209]. In hESC, it was observed that karyotypic changes usually involved addition of a chromosome 12 (chr12) and on 17q (chr17q) and to a lesser extent chromosome X[208,210,211]. Chromosomal changes in hESC have appeared in multiple cell lines and at many different laboratories, mostly after extended passages (over 30)[208]. In addition, 16 out of 30

hESC lines that were initially diploid showed chromosomal instability when cultured using the same conditions in the same laboratory[208].

Interestingly, human embryonic carcinoma cells are typically aneuploid, with trisomies of chr12 and chr17q[212,213]. Furthermore, the pluripotency gene NANOG is located on chr12p and this region is frequently amplified in testicular germ cell tumors[214]. Overexpression of the NANOG gene, promoting self-renewal, may provide cells with an advantage in adapting to culture conditions as aneuploid hESC have a tendency to grow faster. Most chromosomal abnormalities, aneuploidies in particular, can be a reflection of the progressive adaptation of pluripotent hESC to culture conditions. It has been proposed that the chromosomal changes observed in hESC in vitro reflect in vivo tumorigenic events. Specific culture conditions may contribute to chromosomal instability such as the method for passaging cells (mechanical versus enzymatic)[215].

A recent report describes that other than the numerical aberrations detected after extended in vitro culture, subchromosomal aberrations accumulate in the genome of the hESC lines[216]. Specifically, amplifications and loss of heterozygosity (LOH) seem to be the most abundant events occurring in hESC cultured in vitro.

3. Objectives

1. To determine the expression stability of seven widely use reference genes in human an murine embryos and stem cells. Gene expression stability assessment is to be carried out testing the faithfulness of a commercially available preamplification method to increase the starting amount of cDNA without disturbing the original transcript representativity.

2. To assess molecularly and functionally the developmental potential of individual mouse blastomeres from 2-cell, 4-cell and 8-cell embryos.

3. To analyze the genomic stability of human embryonic stem cells lines in culture over extended in vitro culture.

4. To determine whether there is any molecular evidence of a hypothetical chromosomal self-correction in hESC lines derived from PGS-diagnosed aneuploid embryos.

4. Results

4.1 Validation of transcript preamplification method by selection of reference genes for RT-PCR in mouse embryos and stem cells

Validation of transcript preamplification method by selection of reference genes for RT-PCR in mouse embryos and stem cells

Josep Pareja-Gomez[1*], Oriol Vidal[2], Anna Veiga[3] , Nikica Zaninovic[1] and Zev Rosenwaks[1].

1. Ronald O. Perelman and Claudia Cohen Center for Reproductive Medicine, Weill Cornell Medical College, New York, New York, USA.
2. Biology Department, Girona University, Catalonia, Spain.
3. Center for Regenerative Medicine in Barcelona, Catalonia, Spain.

*Corresponding author

Formatted to be submitted to:
Reproduction (ISSN: 1470-1626)

Abstract

Both preimplantational embryogenesis and induced differentiation of ESC are transcriptionally dynamic systems. In order to study gene expression in these systems, little amounts of total mRNA pose a technical challenge that can be overcome by transcript amplification. In addition, regardless expression assay used, transcript abundance normalization is a necessary step that is commonly performed by comparison of every tested transcript to one or more stably expressed reference genes to make reliable comparison between samples. Here we have used total mRNA samples extracted from different stages of mouse preimplantation embryos and differentiating mESC to test the linearity of the transcript amplification of 8 candidate reference genes using real-time quantitative PCR (RT-qPCR). In addition we have compared the expression stability of the same candidate reference genes among the embryonic and ESC samples using a mathematical method implemented in the geNorm software. As a result, seven out of the eight candidate genes showed linear transcript amplification, and *Ppia* and *Hprt1* were ranked as the two most stably expressed genes in preimplantation mouse embryos and throughout different timepoints of mESC induced differentiation. This information is essential in order to perform accurate gene expression assays using RT-qPCR or validation of other expression assays using preimplantation embryos and differentiating ESC populations.

Introduction

To date, several gene expression studies have been published to gain insight of the timing of transcription patterns of oocytes and early developing mammalian embryos (reviewed in (Rodriguez-Zas et al., 2008; Bell et al., 2008). The most extended approach being the use expression arrays(Sharov et al., 2003; Hamatani et al., 2004; Zeng et al., 2004), and more recently mRNA-sequencing(Tang et al., 2010b, 2010a) , due to the possibility of parallel analysis of thousands of transcripts . However, small RNA quantities, such as those isolated from embryos, create a technical limitation to the use of this high throughput techniques. In addition to the mentioned high throughput expression analysis techniques, real-time quantitative PCR (RT-qPCR) has also been used for measurement transcript abundance in embryos (Gutierrez-Adan et al., 1997; Mamo et al., 2008). RT-qPCR delivers a very fine determination of the gene expression and it is an especially accurate and sensitive technique when the source of RNA is limited(Bustin et al., 2005). Furthermore, RT-qPCR is used to validate expression array and mRNA-seq results and consequently it is an essential method to confirm and understand the comparative roles of different transcripts in gene expression experiments.

To overcome the transcript abundance limitation imposed by the small samples, several preamplification methods have been developed to date (Ginsberg, 2005; Nygaard and Hovig, 2006; Noutsias et al., 2008). These methods can imply a total or a selective amplification of the transcripts in the analyzed samples. Nevertheless, regardless to the method used, linearity of amplification is an essential requirement the reaction must satisfy to faithfully maintain the transcriptomic representativity of the unamplified samples. However, transcript amplification efficiency has has been shown to depend on sequence characteristics, such as GC content and length of the poli-A tag (Duftner et al., 2008). Therefore, when mRNA amplification is used in gene expression studies, preceding experiments controlling for amplification bias should be performed for the transcripts of interest.

In addition, in order to compare transcript abundance between different samples, it is critical to consider experimental variations inherent to the quantification, such as the amount of starting material, RNA extraction, RNA quality and enzymatic efficiencies during retro-transcription(Bustin, 2002). To account for these possible sources of variability there are different normalization approaches, but the use of internal reference genes has been described as the most accurate way to normalize the transcript measurements(Huggett et al., 2005). Reference genes, also known as

housekeeping genes (Butte et al., 2001), should be constitutively and stably expressed among the tissues or cells under investigation. However, many studies use some reference genes without a proper validation of their stability of expression. Indeed, beta-actin (*Actb*), glyceraldeyde 3-phosphate (*Gapd*) or 18S RNA have been used for transcript normalization taking for granted their constitutive and stable expression in all type of tissues and cells (Thellin et al., 1999). Nonetheless, many studies have presented evidences that expression of these genes are not constant between different developmental stages and experimental conditions (Tricarico et al., 2002; Dheda et al., 2005). Normalization of the data using the wrong control gene can result in erroneous conclusions being drawn. In order to validate the presumed stable expression of a given control gene, it is required a previous knowledge of a reliable measure to normalize this gene in order to remove any nonspecific variation. To address this circular problem, (Vandesompele et al., 2002) developed a gene-stability measure (M) based on non-normalized expression levels of several candidate reference genes in given set of samples. However, recent reports describe transcription as a randomly oscillatory cellular process (Raj and van Oudenaarden, 2008) that occurs in bursts in each cell individually. Consequently gene expression experiments measure the average transcriptional activity of the cell population tested. Trying to address this intrinsic transcriptional variability, (Vandesompele et al., 2002) showed that instead of using one gene as reference in order to normalize expression values, the use of a composite normalization factor (product of the geometric average of expression 2 or more stable reference genes) compensates for the transcriptional fluctuations.

In this study we have addressed the amplification linearity and the reference gene selection by comparing expression stability of seven widely used housekeeping genes in embryos and ES cells, in both amplified and non-amplified samples. To do so, we have set up and validated an experimental procedure which involves: a) maximization of RNA isolation from small samples; b) efficient and representative retro-transcripton of the isolated RNA into cDNA; c) parallel preamplification of selected transcripts using TaqMan PreAmp Master Mix kit (TPAMMK); d) candidate reference gene expression stability analysis; and e) normalization of expression data. The described experimental procedure is essential in order to perform accurate RT-qPCR gene expression studies using small samples that have undergone transcript preamplification.

Results:

Pre-amplification uniformity

cDNA obtained from preimplantation embryos and differentiating mESC (table 1) were used to check the amplification uniformity of all the transcripts in order to exclude any bias . The specifications of the pre-amplification reaction kit defines a threshold range of ΔΔCT =±1.5 for an acceptable linearity of amplification deviation when performing relative quantification experiments between any two pair of genes.(Applied-Biosystems, 2005). Figure 1 shows the median and the 90% range of ΔΔCT between pre-amplified and non-amplified samples for every pair of genes. Of all the candidate reference genes, only *Actb* proved a deviation in the uniformity of pre-amplification reaction: when compared to *B2m, Gusb* and *Gapd* the 90% the ΔΔCt range was out of the ±1.5 threshold. Therefore, *Actb* remained excluded of the experiment to determine the most stably expressed reference gene for the mouse samples. The rest of candidate genes (*Hprt1, B2m, Gusb, Tbp, Ppia, Tfrc1 and Gapd*) performed within the range of "unbiased" amplification.

Expression stability

M value is a gene expression stability index the calculation of which is based on the principle that the expression ratio of two ideal internal control genes should be identical in all the samples. Therefore, variation of the expression ratios of two particular reference genes shows that one or both of the genes are not constantly expressed, with increasing deviation in the ratio corresponding to diminishing stability(Vandesompele et al., 2002).

GeNorm software was implemented originally to calculate M from non-normalized expression levels. For every control gene tested, the pair-wise variation of expression levels with all other control genes is calculated. After the first run, the program eliminates the worst scoring reference gene, and recalculates a new M value for the rest of genes after leaving the worst scoring one out. At the end, the result is the most stable pair of reference genes of the panel. Genes with the lowest M values have the most stable relative expression to all others (the ratio of expression is logarithmically transformed). Table 2A shows the reference genes ordered according to their M value. The ranking of expression stability of the candidate genes within the three experiments showed a variability that was reduced when the M-value was calculated for specific sample groups. Embryonic samples were grouped into cleavage stage embryos (2-cell and 8-cell embryo samples) and post cavitation embryos (cavitating morula and blastocysts), due to a different Ct range in their expression values. On the other hand, ESC samples showed stable and similar Ct-ranges of each

reference genes tested independently to their differentition status . When M is recalculated for all the reference genes within these three subgroups (cleavage stage embryos, post-cavitation embryos and ESC), M values were generally lower, and the rankings showed lower variability between the three replicates (table 2B) in comparison to the values calculated for the total group of samples.

Since M is a ratio-derived measure, it is based on the assumption that normalization by more than one reference gene is more accurate to calculate expression levels than using just the best performing one. In order to compute a normalization factor (NF) based on the expression levels of more than one reference genes, a geometric mean of individual normalization factors is used. The geometric mean is preferred over the arithmetic mean in order to control for outlying values and abundance differences among genes.

The number of stable genes used for geometric averaging of the normalization factor is result of a balance between practical considerations and accuracy. The more stable genes used to calculate the average normalization factor, the more accurate the gene expression experiment will be. However, since the starting amount of material is a limitation when working with embryos, the use of a large number of reference genes would be impractical. For that reason we decided to determine the two best reference genes for the tested samples. To do so, M was computed for all the possible pair-wise combinations of reference genes using the expression data sub-grouped depending on the type of samples (cleavage-stage embryos, the post-cavitation embryos and the ESC) in the three experiments. We then calculated an average value of M for each pair-wise combination of reference genes. As a result, the three most stably expressed pair of reference genes are the three different combinations of *Ppia*, *Hprt1* and *Tbp*, being *Ppia-Hprt1* the most stable one (table 3).

The expression patterns of all the candidate genes was calculated for the embryo samples by normalizing the qRT-PCR data with the normalization factor calculated by geNorm and based on the combination of *Ppia* and *Hprt1* (Figure 2). In early mouse development, *Ppia* and *Hprt1* appear as the most stably expressed genes from the tested ones: their range of expression within the tested developmental stages is stable when compared to the rest of genes (*B2m*, *Gusb*, *Tbp*, *Tfrc11* and *Gapd*). *Tbp* and *B2m* show a very similar expression pattern: they are highly expressed in the early cleavage stages but their expression after cavitation drops sharply. Gusb on the other hand, shows an irregular pattern of expression even among different samples of the same type. *Tfrc1* shows also an irregular pattern of expression, as well, with high expression during cleavage stage that diminishes until the blastocyst stage. Gapd shows also an irregular transcriptional profile during

early mouse development, with great variability among the embryonic stages tested. Interestingly, differentiating mouse ESC samples do not show the same degree of transcriptional variability found during embryonic development. Instead, mESC show a very stable expression patterns of all the candidate reference genes tested.

Discussion

Performing a RT-qPCR experiment in embryos is challenging specially due to the critical importance of RNA quality and quantity. Due to the limited amount of RNA contained in embryonic samples, we have optimized all the steps previous to RT-qPCR experiments:

a) RNA extraction and purification of embryonic samples was optimized using a specially designed extraction column for pico-scale RNA samples. Moreover, RNA was treated with DNAse while in the column to minimize loses. The quality of the RNA was optimal and no trace of gDNA was found on the samples.

b) Efficiency of retrotranscription was previously tested using three different commercial kits. As a result, the kit which yielded the highest concentration of cDNA was used with the experimental samples. Moreover, a mixture of oligo-dT primers and random nonamers was used to maximize the reaction and the representativity of the RNA retrotranscrived.

c) In order to increase the amount of starting material to perform the RT-qPCR reactions, we preamplified the cDNA. This reaction consists in a selective linear amplification (in parallel) of the amplicons of the genes of interest using the same Taqman primers to be used in the posterior RT-qPCR. Preamplification has been proven to be a faithful way to increase the starting cDNA availability without introducing a significant bias(Noutsias et al., 2008; Mengual et al., 2008). In order to verify that this step does not change the representativity of the RNAs of the genes of interest, we had to investigate the linearity of the amplification, and then we excluded the genes that did not amplify linearly. Actb was excluded from the candidate pool of genes due to the bias introduced by the amplification reaction.

d) Standard curves were used to quantify the efficiency of the reactions during RT-qPCR. Efficiency values were used properly modify the expression data results.

Control of the technical issues is essential, but in order to make sense out the gene expression data, it is crucial to find a stable internal standard gene in the cell types or tissues examined, independent of the differentiation or biological state of the cell. These internal control genes have been normally obtained from the literature and used in different experimental models. An ideal reference gene should be expressed at a stable level among different experimental conditions at all the embryonic stages. However, previous studies showed that several commonly used reference genes did not have a sable expression throughout preimplantational bovine embryo development(Robert et al., 2002). Therefore, it is a vital step to investigate the stability of expression of the possible reference genes. In this experiment we wanted to rate the stability of expression of seven commonly used reference genes during embryonic preimplantational development of human and mouse.

But the expression of reference genes, even the most stable ones can oscillate. Therefore, the safest approach is to use a geometric average of expression of several stable reference genes to compensate for those fluctuations. Vandesompele *et al*.(Vandesompele et al., 2002) suggested that genes showing stable expression patterns in relation to one another are good control genes. Nevertheless, to validate the presumed stable expression of a given control gene, it is required a previous knowledge of a reliable measure to normalize this gene in order to remove any nonspecific variation. To address this circular problem, the authors developed a gene-stability measure based on non-normalized expression levels[28]. Thus, by using this approach, the changing RNA content throughout development was taken into account.

We have used the mentioned method to compute the expression data obtained from the embryonic murine samples. After three replicates of the same experiment, we found that *Ppia* and *Hprt1*are the most stably expressed genes of the set of candidate housekeeping genes in the human and murine samples we have tested. These three genes are involved in clearly different cellular pathways: Ppia catalyzes the cis-trans isomerization of proline imidic peptide bonds in oligopeptides and accelerate the folding of proteins; Hprt1 acts as a catalyst in the reaction between guanine and phosphoribosyl pyrophosphate to form GMP which altogether discards any co-regulation as an explanation for their co-joined expression stability.

Remarkably, mESC showed a higher degree of transcriptional stability for all the candidate reference genes tested, suggesting that induction of differentiation of plurpotent stem cells is not a comparable scenario to the embryonic development. In such samples, all the reference genes tested showed an acceptable value of expression stability. However, Ppia and HPRT would be the

reference genes of election for a comparison of transcriptional activities between mESC and embryonic samples due to the stability between these two different type of samples.

Conclusion

We have found that Ppia and HPRT are a couple of stable expressing genes throughout murine embryonic preimplantational development and mESC in vitro differentiation.. This finding is methodologically relevant in order to perform gene reliable expression experiments using these types of samples.

Material and Methods

Samples used are detailed in table 1.

Embryos

Frozen mouse embryos were commercially obtained at 2 cell stage (B6C3F-1 x B6D2F-1, Embryotech, Wilmington, MA). 130 of them were thawed and cultured in sequential IVF media. Embryos were collected at the indicative time period after the thaw: 2 cell (12h), 8cell (24h), early blastocyst (55h), expanded blastocyst (72h). At each stage, three replicates of ten pooled embryos were obtained and introduced in 20 µl of RNA extraction buffer and stored them at -80ºC until further use.

Embryonic stem cells

A murine ESC (derived previously in our lab from B6C3F-1 x B6D2F-1 blastocysts) was routinely cultured on mouse embryonic fibroblasts (MEF, Chemicon, Temecula, CA) using standard media: KO-DMEM (Invitrogen, Carlsband, CA) supplemented with 20% Knockout Serum Replacement (Invitrogen, Carlsband, CA), 1X nonessential amino acids (Gibco, Carlsband, CA), 1X L-Glutamine (Invitrogen, Carlsband, CA), 1X Pen/Strep (Invitrogen, Carlsband, CA), 1X Mercaptoethanol (Gibco, Carlsband, CA) supplemented with 1ng/ml mouse LIF (Sigma, Saint Louis, MO).

In order to form embryoid bodies (EB's), ESC were grown for 4 days after passage in standard media and then incubated in 30 units/ml of collagenase (Worthington, Lakewood, NJ) until colonies were completely detached from the MEF. The colonies were washed two times with PBS and the cultured in differentiation media: KO-DMEM media supplemented with 10 % of Fetal Bovine Serum (Hyclone, Logan, UT), 1X nonessential amino acids (Gibco), 1X L-Glutamine, 1X Pen/Strep, 1X Mercaptoethanol. Ultra low attachment plates (Corning, Lowell, MA) were used to promote suspended culture of the EBs, that were collected at 0h, 24h, 48h and 144 h (day +6) and

introduced in 20 µl of RNA extraction buffer and stored them at -80ºC until further use.

RNA extraction.

Embryos. RNA was isolated from embryos using the Picopure RNA Isolation Kit (Arcturus, Mountain View, CA) according to the manufacturer's instructions. The kit was engineered to recover high quality RNA from pico-scale samples. In column DNase digestion was carried out using RNAse free DNase Set (Qiagen, Valencia CA). The total volume of the extraction was 20 µl. To check the quality of the RNA, we used the Agilent Bioanalyzer, Pico LabChip kit and confirmed that an extra set of samples had optimal rRNA ratios and clean run patterns (cleavage stage and blastocysts). The quality of the samples met the standards established by the method.

ESC. Total RNA from cultured cells and EB's was obtained using the RNeasy micro extraction kit (Qiagen, Valencia, CA). In column DNAse digestion was performed using RNAse free DNase Set (Qiagen, Valencia CA). RNA concentration and quality was measured using Nanodrop spectrophotometer (Nanodrop Inc., Wilmington, DE). The 280/260 ratio of the samples was always between 2 and 2.15.

Reverse transcription

The total RNA reverse transcription was performed using Sensiscript cDNA synthesis kit (Qiagen, Valencia, CA) optimizing manufacturer's instructions after checking the final yield of this and two more available kits(data not shown). A mixture of Oligo-dT primers (final concentration of 1µM) and random nonamers (at a final concentration of 10 µM) were used in the reaction. Negative controls were included. For embryo samples, total RNA was used. For ESC samples 50ng of RNA were used. After reverse transcription, a final 1 to 10 dilution was made from each cDNA sample (20 µl of total reaction volume to a final 200µl).

Preamplification

For all the samples, 10 µl of the diluted cDNA were used as a template for pre-amplification reactions. We used Taqman Preamp Master Mix (Applied Biosystems) and the Taqman primers following the manufacturer's instructions to perform the pre-amplification. The 14 rounds of amplification protocol was used. The final volume of the reaction was 40 µl. The result samples were diluted 1/5 up to 200 µl total volume.

For each sample the remaining 190 µl of diluted cDNA were used as a template to perform the RT-qPCR reactions in which no amplification was performed, thus from here non-amplified samples.

RT-qPCR

Taqman primer/probes were obtained commercially for the following genes: *b-actin (Actb)*, *glyceraldeyde 3-phosphate (Gapd*), *beta glucoronidase* (*Gusb*), *hypoxantine phosphoribosil-*

transferase I (Hprt1), *peptidylprolyl isomerase A (Ppia*), *TATA box binding protein* (*Tbp*) *and transferrin receptor 1 (Tfrc1)* (see table 2 for full gene name, accession number, function, alias, and indication that the primers span an intron).

Real time quantitative PCR was performed on a 7500 Fast Real Time PCR System (Applied Biosystems) using Taqman PCR Master Mix (Applied Biosystems). Mouse specific Taqman primer/probes can be seen in the table 1. We run replicates for each sample. Cycle conditions were: one cycle at 50°C for 2 min followed by 1 cycle at 95°C for 10 minutes followed by 40 cycles at 95°C for 15s and 60°C for 1 minute. Three replicates per samples were used for each RT-PCR reaction.

All primers were checked for amplification efficiency. First, non-amplified cDNA was obtained by pooling samples from embryos and ESC. A 5 point ¼ dilution series was made to be able to calculate the each primer efficiency.

Amplification Uniformity

A relative quantification study among the complete set of reference genes (each one of the 8 candidates against the rest) to determine ΔΔ Ct values between the preamplified cDNA samples and the non-amplified cDNA samples. The rationale behind this step is that if the reaction uniformly amplifies all the cDNA fragments of the tested reference genes, the ΔCt values will be the same using the original cDNA and the amplified cDNA when considering any particular pair of reference genes. Therefore, the closest the ΔΔCt value is to zero, the closes to perfect linearity is the amplification reaction of the two genes tested.

Expression analysis

Expression analyses were performed using geNorm software[21]. This application calculates a gene-stability measure (M) to determine the expression stability of all the genes tested on the basis of non-normalized expression levels. For each gene, average Ct values of unknown samples were transformed into the log of the starting quantities with the formula obtained from the standard curve, thereby taking into account the efficiency of the PCR reaction. Raw starting quantities were analyzed with geNorm to determine gene expression stability over the different developmental stages, which resulted in a gene expression stability measure *M* for each gene.

Bibliography

Applied-Biosystems (2005). Taqman PreAmp Master Mix KitProtocol. Available at: https://docs.appliedbiosystems.com/pebiodocs/04366127.pdf.

Bell, C. E., Calder, M. D., and Watson, A. J. (2008). Genomic RNA profiling and the programme controlling preimplantation mammalian development. Molecular human reproduction *14*, 691-701.

Bustin, S. A. (2002). Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. Journal of molecular endocrinology *29*, 23-39.

Bustin, S. A., Benes, V., Nolan, T., and Pfaffl, M. W. (2005). Quantitative real-time RT-PCR--a perspective. Journal of molecular endocrinology *34*, 597-601.

Butte, A. J., Dzau, V. J., and Glueck, S. B. (2001). Further defining housekeeping, or "maintenance," genes Focus on "A compendium of gene expression in normal human tissues". Physiological genomics *7*, 95-6.

Dheda, K., Huggett, J. F., Chang, J. S., Kim, L. U., Bustin, S. A., Johnson, M. A., Rook, G. A., and Zumla, A. (2005). The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. Analytical biochemistry *344*, 141-3.

Duftner, N., Larkins-Ford, J., Legendre, M., and Hofmann, H. A. (2008). Efficacy of RNA amplification is dependent on sequence characteristics: Implications for gene expression profiling using a cDNA microarray. Genomics *91*, 108-117.

Ginsberg, S. D. (2005). RNA amplification strategies for small sample populations. Methods *37*, 229-237.

Gutierrez-Adan, A., Behboodi, E., Murray, J. D., and Anderson, G. B. (1997). Early transcription of the SRY gene by bovine preimplantation embryos. Molecular reproduction and development *48*, 246-50.

Hamatani, T., Carter, M. G., Sharov, A. A., and Ko, M. S. (2004). Dynamics of global gene expression changes during mouse preimplantation development. Developmental cell *6*, 117-31.

Huggett, J., Dheda, K., Bustin, S., and Zumla, A. (2005). Real-time RT-PCR normalisation; strategies and considerations. Genes and immunity *6*, 279-84.

Mamo, S., Gal, A. B., Polgar, Z., and Dinnyes, A. (2008). Expression profiles of the pluripotency marker gene POU5F1 and validation of reference genes in rabbit oocytes and preimplantation stage embryos. BMC molecular biology *9*, 67.

Mengual, L., Burset, M., Marin-Aguilera, M., Ribal, M. J., and Alcaraz, A. (2008). Multiplex preamplification of specific cDNA targets prior to gene expression analysis by TaqMan Arrays. BMC research notes *1*, 21.

Noutsias, M., Rohde, M., Block, A., Klippert, K., Lettau, O., Blunert, K., Hummel, M., Kuhl, U., Lehmkuhl, H., Hetzer, R., et al. (2008). Preamplification techniques for real-time RT-PCR analyses of endomyocardial biopsies. BMC molecular biology *9*, 3.

Nygaard, V., and Hovig, E. (2006). Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. Nucl. Acids Res. *34*, 996-1014.

O'Donnell, K. A., Yu, D., Zeller, K. I., Kim, J. W., Racke, F., Thomas-Tikhonenko, A., and Dang, C. V. (2006). Activation of transferrin receptor 1 by c-Myc enhances cellular proliferation and tumorigenesis. Molecular and cellular biology *26*, 2373-86.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216-226.

Robert, C., McGraw, S., Massicotte, L., Pravetoni, M., Gandolfi, F., and Sirard, M. A. (2002). Quantification of housekeeping transcript levels during the development of bovine preimplantation embryos. Biology of reproduction *67*, 1465-72.

Rodriguez-Zas, S. L., Schellander, K., and Lewin, H. A. (2008). Biological interpretations of transcriptomic profiles in mammalian oocytes and embryos. Reproduction *135*, 129-139.

Sharov, A. A., Piao, Y., Matoba, R., Dudekula, D. B., Qian, Y., VanBuren, V., Falco, G., Martin, P. R., Stagg, C. A., Bassey, U. C., et al. (2003). Transcriptome Analysis of Mouse Stem Cells and Early Embryos. PLoS Biol *1*, E74.

Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., and Surani, M. A. (2010a). Tracing the Derivation of Embryonic Stem Cells from the Inner Cell Mass by Single-Cell RNA-Seq Analysis. Cell Stem Cell *6*, 468-478.

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V. I., Lao, K., and Surani, M. A. (2010b). RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc *5*, 516-535.

Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., and Heinen, E. (1999). Housekeeping genes as internal standards: use and limits. Journal of biotechnology *75*, 291-5.

Tricarico, C., Pinzani, P., Bianchi, S., Paglierani, M., Distante, V., Pazzagli, M., Bustin, S. A., and Orlando, C. (2002). Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. Analytical biochemistry *309*, 293-300.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology *3*.

Zeng, F., Baldwin, D. A., and Schultz, R. M. (2004). Transcript profiling during preimplantation mouse development. Developmental biology *272*, 483-96.

Specie	Sample name	Time	Replicates per experiment
Mouse	2-cell embryo	2h after thaw	3
	8-cell embryo	18h after thaw	3
	Cavitating morula	26h after thaw	3
	Expanded blastocyst	36h after thaw	3
	ESC 0	0h after induced differentiation	3
	ESC 24	24h after induced differentiation	3
	ESC 48	48h after induced differentiation	3
	ESC 144	144h after induced differentiation	3

Table1. Samples used to determine the amplification linearity of the 8 candidate reference genes and their subsequent expression stability. Each embryo samples contained 10 embryos of the specified type.

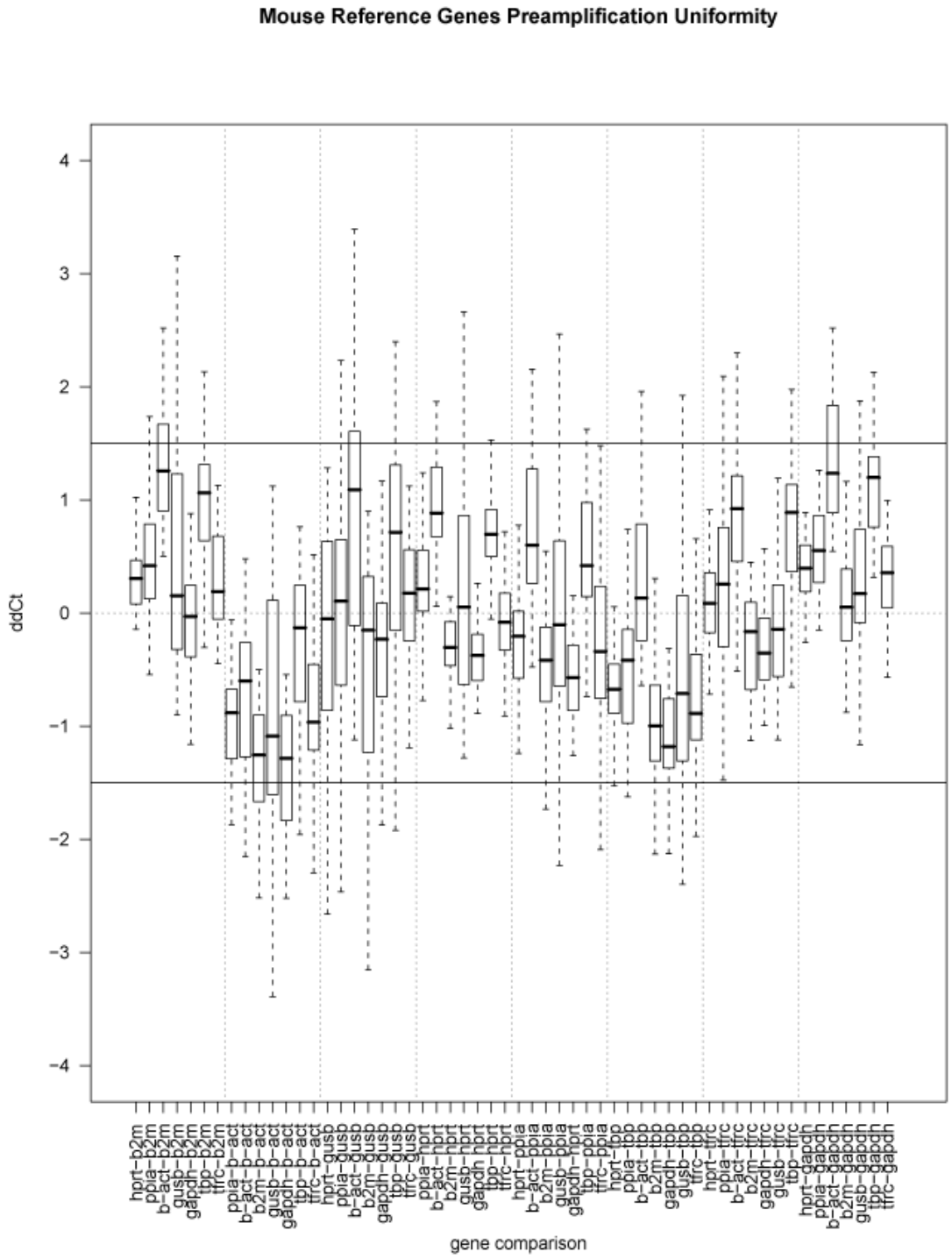


Figure 1. Relative quantification study among the candidate housekeeping genes. Box and whiskers plot of the ΔΔ Ct values between the preamplified cDNA samples and the non-amplified cDNA samples. The data is shown as medians (lines), 10th percentile to the 90th percentile (boxes) and ranges (whiskers) for all the samples tested.

A

Experiment 1		Experiment 2		Experiment 3	
PPIA+TBP	0.85	HPRT1+PPIA	0.79	TFRC1+PPIA	1.07
B2M	1.24	TFRC1	0.84	TBP	1.08
HPRT1	1.30	TBP	0.88	B2M	1.15
GUSB	1.42	GUSB	0.94	HPRT1	1.20
TFRC1	1.71	B2M	1.08	GUSB	1.66
GAPDH	2.41	GAPDH	1.25	GAPDH	1.91

B

Cleavage-stage						Blastocyst						ESC					
Eperiment 1		Experiment 2		Experiment3		Eperiment 1		Experiment 2		Experiment3		Eperiment 1		Experiment 2		Experiment3	
PPIA + HPRT1	0.37	GUSB + PPIA	0.35	GUSB + PPIA	0.29	GAPDH + GUSB	0.63	GAPDH + PPIA	0.51	PPIA + TBP	0.62	GAPDH + HPRT1	0.08	HPRT1 + GUSB	0.10	TBP + TFRC1	0.09
TFRC	0.60	HPRT1	0.54	HPRT1	0.48	TBP	0.84	TFRC	0.57	TFRC1	0.78	TBP	0.17	TFRC1	0.12	HPRT1	0.22
TBP	0.83	TFRC	0.74	TFRC1	0.74	PPIA	0.93	HPRT1	0.65	B2M	0.87	TFRC	0.21	PPIA	0.25	PPIA	0.26
B2M	0.89	TBP	0.87	TBP	0.98	B2M	1.08	TBP	0.71	GAPDH	0.95	GUSB	0.33	TBP	0.30	GUSB	0.35
GUSB	0.98	B2M	0.97	B2M	1.07	HPRT1	1.19	GUSB	0.78	HPRT1	1.05	B2M	0.47	GADPH	0.43	GAPDH	0.42
GAPDH	1.13	GAPDH	1.17	GAPDH	1.34	TFRC	1.38	B2M	0.89	GUSB	1.41	PPIA	0.62	B2M	0.53	B2M	0.57

Table 2. **Gene stability values (M) for the candidate reference genes**. A- M values calculated using all the samples for the three replicate experiments. There is a considerable degree of variability in the positions the candidate genes occupy between experiments, but *Ppia*, *Hprt1* and *Tbp* occupy the top positions of stability. B- Detailed M values calculated for each subtype of samples in each experiment.

	AverageM-value	SD
Hprt1-Ppia	0.93	±0.45
Ppia-Tbp	1.09	±0.51
Hprt1-Tbp	1.14	±0.49

Table 3. **Gene stability values for the combination of two reference genes**. M values were calculated for all the possible combinations of two genes within all the subgroups of samples (cleavage stage, blastocysts, ESC) for the three experiments. Average M values of the three most stable gene-combinations are shown.

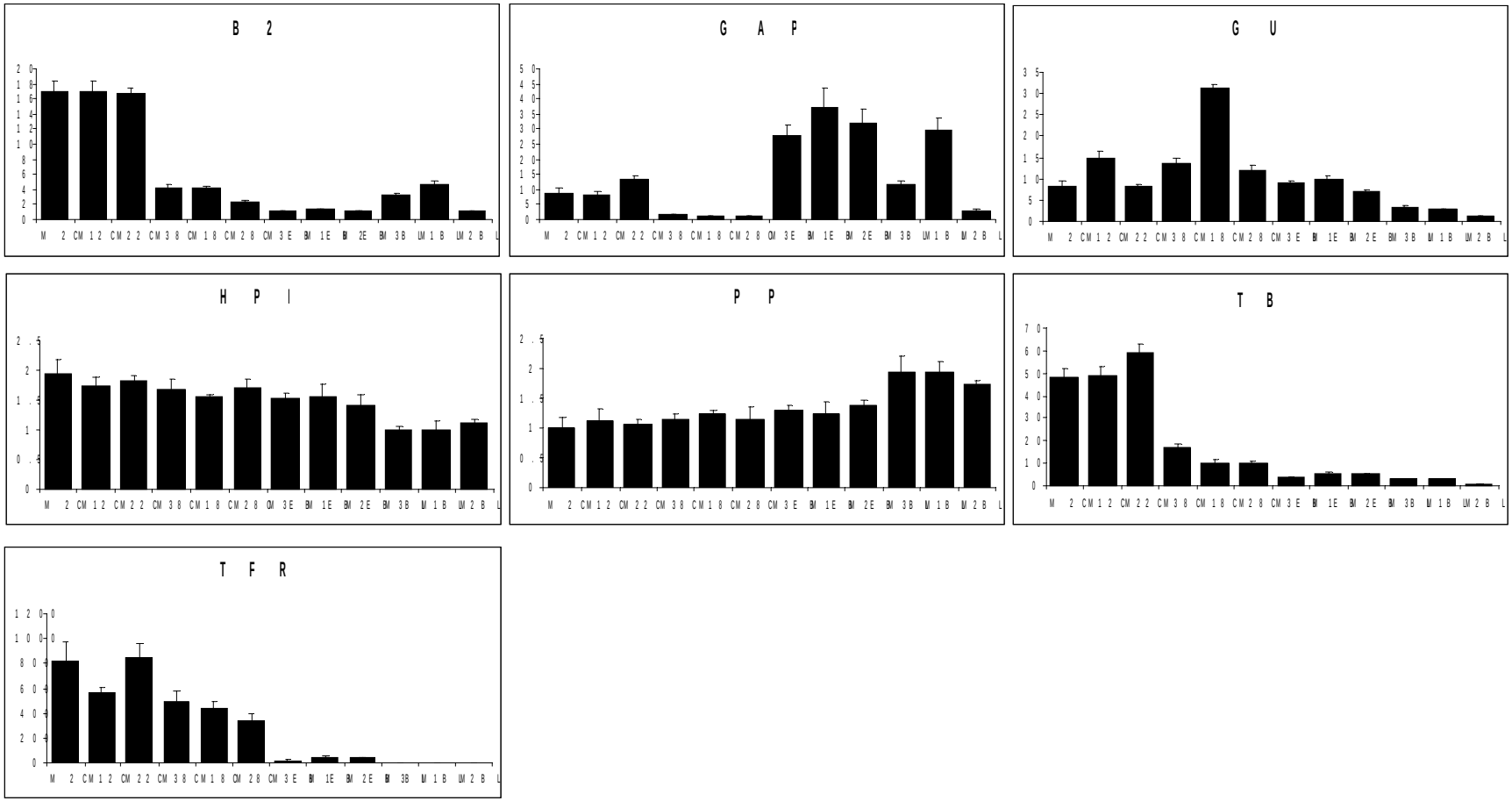


FIGURE 2 Relative expression of *B2m*, *Gapd*, *Gusb*, *Hprt1*, *Ppia*, *Tbp* and *Tfrc1* at different stages of murine embryonic development. X-axis: developmental stage (CL= cleavage stage, BL = blastocyst stage, M2C = 2-cell stage, M8C = 8-cell stage, MEB = cavitating morula, , MBL = expanded blastocyst). Y-axis: normalized relative expression. Data was normalized to the geometric mean of 2 stably expressed genes (*Hprt1*, *Ppia*) as determined by geNorm analysis. Error bars represent SEM.

4.2 Functional and transcriptional analysis cleavage-stage mouse blastomeres

Functional and transcriptional analysis cleavage-stage mouse blastomeres

Josep Pareja-Gomez[1*], Nikica Zaninovic[1], Jiying Hao[1], Oriol Vidal[2], Anna Veiga[3] and Zev Rosenwaks[1]

1. Ronald O. Perelman and Claudia Cohen Center for Reproductive Medicine, Weill Cornell Medical College, New York, New York, USA.
2. Biology Department, Girona University, Catalonia, Spain.
3. Center for Regenerative Medicine in Barcelona, Catalonia, Spain.

* Corresponding author

Formatted to be submitted to:
Reproduction (ISSN: 1470-1626)

Abstract

Mouse blastocyst formation from the totipotent zygote involves an ordered series of lineage specifications to give rise three type of cells: trophectoderm cells, primitive endoderm cells and epiblast cells. Different models have been suggested to explain embryonic cell specification but the molecular mechanism establishing the first differences between embryonic cells during preimplantational development remains unknown. Here we have molecularly characterized the mouse lineage specification: a) using simultaneous immuno-fluorescent detection of GATA6, CDX2, OCT4 and NANOG and b) performing single embryo RT-qPCR for relevant transcripts. To test whether any developmental potential difference exists among blastomeres at different cleavage stages, we disaggregated two-cell, four-cell and eight-cell mouse embryos to molecularly characterize them and to derive embryonic stem cells from them. The molecular differences found and the ESC derivation rates suggest that the second embryonic division triggers a molecular maturation that imposes a developmental potential restriction that may be different among blastomeres in the same embryo. However, before compaction, these molecular differences seem to be not definitively established and could be reversible. Thus, we suggest that lineage markers differential expression during post-compaction mouse embryonic development would be a downstream consequence of the molecular differences established during cleave stages.

Introduction

Murine embryo preimplantational development involves an ordered series of lineage specifications and axial asymmetries that gives rise to the blastocyst, a hollow structure that will attach to the uterine wall. Structurally, the blastocyst consists of an outside cell layer, the trophectoderm (TE), that accommodates an eccentrically located inner cell mass (ICM). Three types of cells can be found in the blastocyst: trophectoderm (TE) cells that will originate the future placenta; epiblast (EPI) cells in the ICM will give rise to the embryo itself; and primitive endoderm (PE) cells in the ICM that will form the future yolk sack. Therefore, the process by which the totipotent zygote gives rise to the three cell lineages existing in the blastocyst must imply a series of events that restrict the developmental potential of some of the embryonic cells(Ang and Constam, 2004; Beddington and Robertson, 1999; Rossant, 2004; Rossant and Tam, 2009).

To study lineages specification during preimplantational development, several molecular markers have been used. On one hand, the genes known to be required for specifying the pluripotent epiblastic cells are *Sox2*(Avilion et al., 2003; Nichols et al., 1998), *Oct4*(Nichols et al., 1998) and *Nanog*(Mitsui et al., 2003; Chambers et al., 2003). Remarkably, *Oct4* and *Nanog* have been shown to be expressed in every mouse blastomere from eight-cell stage until compaction(Dietrich and Hiiragi, 2007). On the other hand, Cdx2, a caudal related homeodomain protein, is an essential regulator of the mature TE. Previous reports show that *Cdx2* is expressed after the eight-cell stage to progressively become restricted and upregulated in the outside cells of the compacting embryo(Dietrich and Hiiragi, 2007; Ralston and Rossant, 2008). Experiments with *Cdx2* knock-out mice(Strumpf et al., 2005) suggested that *Cdx2* is not necessary for TE specification but it is essential for the maintenance of the molecular identity(Nishioka et al., 2008; Yagi et al., 2007) of TE cells. Conversely, *Gata6* is a PE specific zinc-finger transcription factor(Morrisey et al., 1998) known to be essential for normal embryo development which has been reported to be expressed as early as 8 cell stage in mouse embryo(Plusa et al., 2008) .

Therefore, the formation of the blastocyst cell lineages requires transforming an indeterminate expression pattern of key lineage regulators into a spatially restricted and regulated pattern, in addition to the changing cellular context of the developing blastomeres. The mechanism of morphogenesis and patterning of the mouse embryo prior to blastocyst formation have been addressed intensely, but a definitive model remains elusive.

Some authors have proposed a pre-patterning model, suggesting that lineage specification is pre-determined at the egg stage(Gardner, 2001; Zernicka-Goetz, 2002), inducing a certain cleavage pattern that establishes differences in the developmental potential as early as at 4-cell stage blastomeres(Piotrowska-Nitsche et al., 2005; Piotrowska-Nitsche and Zernicka-Goetz, 2005). However, abundant data contradicting this pre-patterning model have been published to date: unlike some non-mammalian species(Schier and Talbot, 2005; Weaver and Kimelman, 2004), the mouse oocyte has no clear polarity, and no internal or external determinants important for lineage specification are yet known (Yamanaka et al., 2006). Moreover, until the second division cleavage patterns are not stereotypic, and blastomere ablation or addition at this stage does not affect the development of a normal fetus(Johnson and McConnell, 2004; Ciemerych et al., 2000; Tarkowski, 1959; Tsunoda and McLaren, 1983; Zernicka-Goetz, 1998). In addition, chimeras can be made by combining two whole eight cell embryos or by combining subsets of blastomeres of two different cleavage (two to eight-cell) stages embryos(Tam and Rossant, 2003). However, in those cases, a biased contribution of blastomeres of different stages to the three different lineages is found(Spindle, 1982).

In opposition to this pre-patterning model, the regulative model advocates that after compaction takes place, successive divisions create topological differences between the blastomeres positioned outside and those remaining on the inside of the embryo. These differences would be based on the fact that inside-cells symmetrically contact their neighboring cells, whereas outside cells would do it asymmetrically(Johnson and Ziomek, 1981a, 1981b). This asymmetry would induce cell polarization followed by epithelialisation on the outside cell layer(Rossant and Tam, 2009; Yamanaka et al., 2006) . Moreover, after lineage tracing experiments, some authors have suggested that TE cells in the blastocyst derive from the outside cells of the 16-cell stage embryo, whilst the ICM cells will derive from those remaining on the inside(Yamanaka et al., 2006; Johnson and McConnell, 2004).

A third model, know as the hidden pre-formation model suggest that cleavage pattern influence the future allocation of cells at compaction and therefore, the lineage allocation in the blastocyst(Graham, 1971) . Although the two blastomeres in a two-cell cleavage stage embryo gives rise to EPI, PE and TE lineages, some authors claim that one cell tends to contribute more to the embryonic part of the blastocyst and the other contributes to the abembryonic part of the blastocyst(Gardner, 2001; Piotrowska-Nitsche et al., 2005; Fujimori et al., 2003). Since the molecular base of the developmental potential circuitry and the lineage determination is not yet

fully understood, this possible bias in the progeny of the cleavage stages blastomeres suggests that the developmental potential differences among sibling blastomeres in a given embryo, although reversible, are perhaps being established at this stage. This model is regarded as a synthesis between the two previous models because it underlines the importance of the highly regulative nature of the mouse embryos while stressing the influence of the cleavage to lineage determination.

Several experiments have been conducted independently to find out how the molecular markers of the three blastocystic become restricted to their characteristic cellular location during mouse embryo development(Rossant, 2004; Rossant and Tam, 2009; Ralston and Rossant, 2008; Strumpf et al., 2005; Plusa et al., 2008; Bischoff et al., 2008; Chazaud et al., 2006; Jedrusik et al., 2008). However, the mechanism(s) that initiate the molecular difference between blastomeres and the ultimate lineage specification of the preimplantation embryo remains unknown. To further investigate whether these differences are established prior to compaction, different molecular and functional standard methods can be used. On one hand, we have analyzed the transcriptional and the protein expression patterns in normal mouse embryo development to determine the co-expression of the lineage markers up to the formation of the blastocyst. On the other hand, to investigate the differences among individual blastomeres at different cleavage stages, we have used isolated blastomeres to derive embryonic stem cells (ESC), which is a methodology previously described(Chung et al., 2006; Delhaise et al., 1996; Lorthongpanich et al., 2008; Tesar, 2005; Wakayama et al., 2007; Wilton and Trounson, 1989). This would be a functional method to determine the developmental potential of blastomeres: theoretically, if no differences in developmental potential are to be found among sibling blastomeres in a cleavage stage embryo (up to eight-cell) similar derivation rates should be found when using blastomeres of different stages. Moreover, it has previously been reported that when a cleavage stage embryo is disaggregated (up to eight cells), blastomeres can develop into blastocysts-like structures(Lorthongpanich et al., 2008; Rossant, 1976). We have analyzed the transcriptional pattern and the protein expression these blastomere derived blastocysts, to compare them with the control blastocysts and to determine any difference in developmental potential among the sibling blastomeres they derive from.

Results

Molecular signature of normally developing embryos cultured *in vitro*

The two different immunofluorescent labeling experiments let us determine the presence of the transcription factors known to be differentially expressed in the three cell lineages of the blastocyst.

Thus, the first labeling combination simultaneously shows the allocation of NANOG, OCT4 and GATA6 (Figure 1A, movie 1) allowing us to differentiate the expression of the markers that will differentiate the EPI (OCT4 and NANOG) and PE (OCT4 and GATA6) cell populations of the ICM, and the second combination the allocation of CDX2, OCT4 and GATA6 (Figure 1B), that gives a global visualization of the three markers that characterize the three lineages in the blastocyst.

The labeling experiments at 2C stage do not detect any expression of the transcription factors investigated. OCT4, GATA6 and NANOG are first detected in 8C cell embryos, and are virtually expressed simultaneously in all the blastomeres of this stage embryos analyzed (figure 1A). This result coincides with previous independent findings(Dietrich and Hiiragi, 2007; Plusa et al., 2008). However, no expression of CDX2 detected at 8C stage, being first seen at 16-32C stage (figure 1B) which is also in concordance with previously published data(Ralston and Rossant, 2008; Strumpf et al., 2005; Jedrusik et al., 2008; Suwinska et al., 2008).

It is during the 16-32C stage that the two labeling experiments show that on one hand, CDX2 is co-expressed with OCT4 and GATA6 in virtually all the cells (figure 1A), and on the other hand, NANOG is equally co-express with OCT4 and GATA6 in nearly all the nuclei, with different intensities. These two independent findings lead us to assume that CDX2, NANOG, OCT4 and Gata6 proteins are co-expressed transiently in the transition between 16 to 32 cells of the embryo development.

The labeling of the expanded blastocyst shows that at this stage two clearly differentiated cell populations can be found in the ICM: a) OCT4+NANOG expressing cells and b) OCT4+GATA6 expressing cells. Remarkably, in most of the blastocysts, a proportion of ICM cells co-expresses GATA6, NANOG and OCT4, and therefore is not definitively committed to any of the two ICM lineages. CDX2 expression is reduced to the TE cells, but in some of them CDX2 is co-expressed with GATA6 and OCT4 during the early steps of the blastocyst formation. Only upon full expansion GATA6 and OCT4 expression is gradually reduced in TE (figure 2). Thus, in the expanded blastocyst the three lineages are established and characterized by a distinctive protein expression pattern: CDX2 is exclusively expressed in the TE cells, whereas in the ICM, PE cells co-express GATA6 and OCT4 and epiblast cells co-express NANOG and OCT4. However, some cells can still be found in the expanded blastocyst with expression dynamics that show not a definitive lineage commitment at this stage.

RT-qPCR experiment on single embryos allowed us to evaluate transcriptional dynamics in the developing embryo and also assess the correspondence between transcript abundance and protein detection experiments. The mean relative abundances of the *Oct4*, *Nanog*, *Sox2*, *Gata6*, and *Cdx2* transcripts throughout the four investigated stages are shown in Figure 3.

The 2C is the stage with lowest relative transcript abundance of all the stages tested. At this early developmental phase *Oct4* transcript was found to be the most abundant among the investigated. At eight-cell stage relative abundance of all tested transcripts increased except for *Cdx2*, being *Sox2* (13.3 times), *Nanog* (9 times), *Gata6* (7.7 times) the ones with bigger increase. However, it is during the 16-32C stage that the abundance of the transcripts tested is at its highest level. Remarkably, *Cdx2* expression is increased by 61 times when compared to the eight-cell stage, and both *Nanog* and *Sox2* transcripts abundance increased by 4.6 times, whereas the rest of the genes increase was more moderate. In fact, all the transcripts reached their highest level of expression during the compaction-cavitation of the embryo. At the blastocyst stage, the relative expression of all the genes tested dropped when compared to the previous tested stage, being *Cdx2* the transcript with the highest relative abundance at the peak of expansion.

Molecular signature and developmental potential of embryos derived from cleavage stage blastomeres cultured in vitro

Blastomeres obtained from disaggregating embryos at two-cell (2C), four-cell (4C) and eight cell (8C) stages cleaved at normal pace in sequential culture media to originate two-cell blastomere derived embryo (2CBL), four-cell blastomere-derived embryo (4CBL), and eight-cell blastomere-derived embryo (8CBL) respectively. Interestingly, the timing of cavitation was approximately the same as if the separation of blastomeres had not taken place. Thus, control embryos and blastomere-derived ones, started cavitating around the same time independently of the number of cells (38-40h after thaw).

Blastocysts rate from 2CBL embryos was lower but comparable to control embryos (table 1A). 2CBL blastocysts were about the half the size of the control blastocyst, and morphologically identical. However, blastocyst rates from 4CBL and 8CBL embryos were significantly lower than the control embryos (table 1A). Those 4CBL that reached the blastocyst stage morphologically resembled to the control blastocysts being about a quarter their size. However, 8CBL were

extremely small and if cavitation occurred, trophoblastic cellular structures of about one eighth the size of a control blastocyst were obtained (figure 4).

The labeling experiment showed that for 100% of the 2CBL blastocysts distribution of NANOG, OCT4, GATA6 and CDX2 proteins is comparable to control blastocysts (Fig 5A and 5B). Thus, the described cell populations in the ICM of control blastocysts can be found in the 2CBL embryos: NANOG+OCT4 and OCT4+GATA6 expressing cells. Remarkably, the ICM of these embryos is composed by a smaller number of cells when compared to the control blastocysts. CDX2 is only expressed in TE cells, although when compared to control embryos, a higher proportion of TE cells display an immature protein profile, co-expressing GATA6 and CDX2.

The labeling of the 4CBL derived blastocysts shows that 42% of the embryos (3 out of 7) express Nanog in the ICM: two of them siblings derived from the same original 4 cell embryo. Moreover, those 4CBL embryos expressing Nanog had a remarkably reduced ICM and the TE cell population was composed with high proportion of cells with an immature expression profile (GATA6+CDX2).

Only 11% of the 8CBL pseudo-blastocysts tested (1 out of 9) with the first labeling showed a positive labeling for Nanog. Interestingly, this one 8CBL embryo contained 3 NANOG+OCT4 positive cells and 5 OCT4+GATA6 cells surrounding them. However, most of the cells in the embryos labeled using the second combination of antibodies showed the molecular signature of the immature TE cells, expressing GATA6, CDX2 and OCT4, but some embryos showed some non-grouped cells expressing GATA6+OCT4 which is the characteristic pattern of the PE in the ICM (Figure 6B)

Another set of sibling blastocysts derived from a disaggregated embryos of each stage (2C, 4C, 8C) were analyzed by RT-qPCR using the same conditions described above for the normally developing embryos. Results for 2 x 2CBL, 4 x 4CBL and 6 x 8CBL sibling blastocysts was obtained and the average expression for 3 control blastocysts are shown in figure 6A. Remarkably, while relative expression of *Gata6* and *Oct4* in all blastomere-derived embryos is comparable to the control blastocysts group, *Cdx2* expression is higher on the 4CBL and 8CBL blastocyst than in the control and 2CBL embryo. In general, the transcripts levels of the blastomere-derived blastocysts shows that *Sox2* and *Nanog* transcript abundances are lower than the control blastocyst. In fact, *Sox2* transcript is undetected in one 4CBL embryo and five out of six 8CBL blastocysts. Contrary to the protein, Nanog mRNA is detected in all the embryos analyzed at generally lower levels than the

control embryos (with the exception of two embryos). Strikingly, when calculating ratios of expression between *Sox2* and *Cdx2* on one hand, and *Nanog* and *Cdx2* on the other, we noticed that the proportion of blastomere-derived embryos showing a transcriptional pattern similar to the control blastocysts are parallel to the rates of embryos with EPI molecular signature: 100% of the 2CBL embryos, 50% of the 4CBL embryos and 16.6% of the 8CBL embryos (figure 6B). Interestingly, the rest of 8CBL embryos showed an equal or higher relative expression of *Cdx2* to *Nanog*, but no *Sox2* expression. *Nanog* is transcribed but not expressed, suggesting a possible post-transcriptional modulation(Tay et al., 2008) . Moreover *Sox2* is the only gene of the pluripotency-inducing genes with an affected expression pattern in these embryos.

Derivation of ESC from blastomere-derived embryos and isolated blastomeres

We first calculated the blastocyst rate of the blastomeres cultured in vitro (Table 1), to determine how the disaggregation at different cleavage stages affects the capability of the blastomeres to form a blastocyst. 2CBL embryos have a lower but comparable blastocyst rate (78.2%) than the control group (88.2%), whereas both 4CBL and 8CBL embryos display a significantly lower blastocyst rate (42.5% and 38.7% respectively) than the control group.

The results of the ESC derivation experiment using 2CBL, 4CBL and 8CBL blastocysts are shown on Table1. ESC lines were derived from 2CBL and 4CBL blastocyst, but no line was obtained from 8CBL embryos. The ESC line per blastocyst rate for both 2CBL and 4CBL embryos (21.3% and 11.8% respectively) are significantly different from the control group (33.0%). Interestingly, two pairs of sibling 2CBL embryos were able to produce sibling ESC lines. ESC rate per original embryo (the one that gave rise to the sibling blastomere-derived blastocysts) is the same between control blastocysts and 2CBL blastocysts (33%). Contrary, ESC rate per original embryo for 4CBL blastocysts is lower than control embryos (20%).

In order to determine whether the derivation technique influences the outcome of the derivation process, blastomeres obtained from disaggregating embryos at 2C, 4C and 8C stages were laid on a feeder cell layer separately to derive ESC lines bypassing the blastocyst stage in vitro. After attachment, blastomeres started cleaving, some of them presenting a pseudo-cavitation process between day 1 and day 3 of the culture (figure 7A). After that, at day 4, the proliferative cell clumps started spreading on the MEF and by day 9, outgrowths were split on fresh MEF cultures for expansion and characterization.

ESC lines were obtained from 4CBL and 8CBL cultures but none from 2CBL or control embryos, which consisted in zona-free non disaggregated two-cell embryos (table 1B). Remarkably, this second method to derive ESC increases the ESC per blastomere rate for 4CBL and 8CBL when compared with the first method, and importantly, the rate of ESC lines per original embryo is higher for both 4CBL and 8CBL blastomeres than for the traditional method of derivation using a blastocyst.

The basic pluripotent characterization of the ESC lines were confirmed by expression of stem cell markers: Oct4, Nanog (figure 7B).

Discussion

Protein and transcript profile of normal developing mouse embryos

To gain insight into the molecular basis of preimplantation embryo development, we investigated the expression pattern of a selected transcripts and proteins which characterize the three lineages in the blastocyst. On one hand, the two immunolabeling combinations allowed us track the expression pattern of OCT4, NANOG, GATA6 and CDX2 throughout the embryo development until blastocyst stage, when their expression differentiates the three cell types: TE, EPI and PE. On the other hand, here we describe the relative transcript abundance on single embryonic samples of these relevant transcription factors, being the first, to our knowledge, to describe the transcriptional dynamics of these genes during mouse preimplantation development.

The molecular portrait obtained from the immunolabeling experiments reveals that at 8C stage virtually all the blastomeres co-express GATA6, NANOG and OCT4, which is in concordance with previously published data(Dietrich and Hiiragi, 2007; Plusa et al., 2008; Tesar, 2005; Niwa et al., 2005). However, no CDX2 protein is seen in the nuclei of the embryonic cells at 8C stage, coinciding with previous reports(Dietrich and Hiiragi, 2007; Strumpf et al., 2005). It is during early compaction (16 to 32 cells), that the two labeling experiments show co-expression of, on one hand, OCT4-GATA6-CDX2 and OCT4-GATA6-NANOG on the other hand, indirectly demonstrating that, at this stage, an overlapping expression of the four transcription factors occurs. Remarkably, from 16-32 cell stage on, the lineage determination process implies a selective restriction of the expression of the studied transcription factors in the distinct cell types EPI, PE and TE.

Our staining results show that after compaction, lineage determination of TE cells implies an increase the expression of CDX2, and a sudden stop of NANOG expression. Contrarily to NANOG, we found that OCT4 and GATA6 expression is slowly turned off until full expansion, when only CDX2 will be expressed in most these cells. Interestingly, GATA6 expression is still seen in some TE cells at full expansion. The lingering expression of this PE characteristic TF may imply a certain degree of immaturity in some of TE cells.

When it comes to the lineage determination of the ICM cells our results show that in order to keep the pluripotency molecular circuitry to originate the embryo, the future EPI cells stop expressing GATA6 and upregulate the expression of NANOG. Contrarily, in the future PE cells, GATA6 expression is increased and NANOG becomes undetected. Remarkably, those cells with the most intense expression of NANOG display no signal for GATA6 and the reverse observation is made for the brightest GATA6 expressing cells. However, even at full expansion, a few ICM cells can be detected co-expressing NANOG and GATA6 not as intensely as the ones that express these TFs exclusively. These particular cells seem to have not reached the molecular tipping point where they commit to either of the two lineages. Interestingly, OCT4 is consistently expressed in both ICM lineages.

Therefore, at compaction, blastomeres display an OCT4, NANOG, GATA6 and CDX2 expression pattern that for some reason, is not affected by the expression of the rest, and thus is mutually independent. However, a certain degree of variability in the intensity of expression can be observed (and has been also previously reported(Dietrich and Hiiragi, 2007)), establishing a molecular difference that may be determinant for the cell fate determination. Our results and previous lineage tracing studies(Dietrich and Hiiragi, 2007; Ralston and Rossant, 2008; Plusa et al., 2008), led us to assume that at some point between cavitation and expansion, differences in the molecular pathways that define lineage commitment among blastomeres are reflected in the differential expression of these four TFs. How these differences are established remains unknown. Actually, we can distinguish two mechanisms of expression restriction among these four transcription factors: on one hand, CDX2 and NANOG are generally expressed in all blastomeres at 16-32 cell stage to be restricted very rapidly during expansion to the blastomeres that will give rise to TE and EPI lineages respectively. In contrast, OCT4 and GATA6 expression is steady in all blastomeres until 16-32 cell stage and their expression becomes gradually intensified into their characteristic cellular lineage and also gradually turned off in those cellular lineages that won’t be expressing them at full

expansion of the blastocyst.

From a transcriptional point of view, 16-32C stage is when the transcript abundance of the genes investigated reaches its maximum because virtually all the cells transcribe them. Remarkably, SOX2, which is known to promote OCT4 transcription to keep the pluripotency molecular circuitry on(Masui et al., 2007), is the second most abundant transcript right after CDX2, underlining the "crossroad" at which cells are in the mentioned stage. The drop in the relative transcript abundance between cavitating and expanded blastocyst could be explained by the restriction of the expression of the analyzed TFs to their putative expressing cell type in the blastocyst. Thus, the characteristic TF of the most abundant cell type in the blastocyst, CDX2, is the most abundant transcript in the blastocyst.

The information on the co-expression of these determinant transcription factors during early stages of mouse preimplantation embryo shows that even if previous reports show proof of reciprocal inhibition between CDX2 and OCT4 (Niwa et al., 2005) on one side, and NANOG and GATA6(Singh et al., 2007) on the other, additional regulatory mechanisms must exist controlling their expression. This fact emphasizes the importance of finding the molecular mechanisms that underlie in the control of the expression of these transcription factors and therefore, control the lineage determination process.

The experiment assessing functionally and molecularly the differences among blastomeres of cleavage stage embryos up to eight cells gave us valuable information on the relative importance of the investigated proteins and transcripts and how they correlate or not with the developmental potential. Like in previous reports(Lorthongpanich et al., 2008; Wakayama et al., 2007), we found that the further ahead in the cleaving process the blastomeres are separated, the lower is their capability to form a blastocyst-like structure and the lower is the ESC derivation rate per blastomere. On one hand, the competence to form a blastocyst-like structure was considered to be correlated with the maintenance of the totipotency of the original zygote in that particular blastomere. On the other hand, the capacity to form ESC colonies is considered to be correlated with the epiblastic pluripotent molecular signature that is the one that sustains the formation of the embryo. It is clear that totipotency implies pluripotency but not the other way around. Therefore, the results obtained with the two derivation methods let us see that while 8C blastomeres had a very diminished capacity to form blastocyst-like structures and no ESC line was derived from those structures, we could derive ESC lines from 8C blastomeres cultured on the MEF bypassing the

blastocyst stage. On the other end, 2CBL embryos had a blastocyst rate and an ESC rate per blastocyst comparable to control embryos, while we could not derive any ESC from the blastomeres directly bypassing the formation of the blastocyst.

These results depict a very interesting picture of the restriction of the developmental potential during the three first divisions of the mouse embryos. After the second division the competence to form the blastocyst structure (totipotency or the capacity to form the three lineages in a blastocyst) is gradually lost by the blastomeres due to a probable gradual molecular restriction. However, the faculty to form a blastocoel does not imply that the three lineages are formed in the blastocyst-like structures. In fact, the molecular analysis of the blastomere-derived pseudo-blastocysts shows that only half of the 4CBL blastocysts and one eight of the 8CBL blastocysts were formed by cells showing the expression patterns of EPI, PE and TE lineages. Interestingly, in the rest of the embryos, the epiblast molecular signature was not present and their cells were characterized by what we had determined as trophectodermal developmental immaturity (especially in the 8CBL embryos): a high proportion of them expressed CDX2+GATA6+OCT4. Only a few cells in these pseudo-blastocysts expressed what is considered the PE molecular signature (GATA6+OCT4), with no organized distribution inside the embryonic structure. Thus, this inability to form the three blastocystic lineages implies a developmental potential restriction, which is less constraining than that of their sibling non-cavitating blastomeres, but a restriction that implies the loss of the zygotic totipotency nonetheless.

This idea of gradual restriction of developmental potential was complemented by the transcriptional profile of the blastomere-derived embryos. The results obtained suggested two remarkable facts: firstly, a correlation between the proportion of blastomere-derived embryos showing a positive labeling for the three lineages and proportion of blastomere-derived embryos keeping the SOX2 transcriptional level of control embryos, and secondly, a higher relative expression of CDX2 in the rest blastomere-derived embryos. The first observation adds up to the previously published proof of SOX2 being essential to keep the pluripotency molecular circuitry on (Rodda et al., 2005; Masui et al., 2007; Jaenisch and Young, 2008), and therefore the underlines the importance of this gene in the establishment of the epiblastic molecular identity of some blastomeres.

These observations are complemented by the results of direct derivation of ESC from the blastomeres, bypassing the formation of the blastocyst. Using this method, blastomeres divide on the feeder cell layer, to form a clump of cells that sometimes shows a hollow structure in the middle

(considered as cavitation attempts). Interestingly, 2C blastomeres did not form any ESC colony and only after 4C and 8C blastomeres did, with a diminishing ESC per blastomere rate correlating with the blastomere size. Therefore, a proportion of 4C and 8C blastomeres confirmed the capacity to derive ESC proving to have the competence to originate an epiblastic cell lineage. Since 2C blastomeres have proven to have that capacity too after culturing them until blastocyst stage (using the two-step method to derive ESC), we hypothesize that a certain degree of molecular maturation in the embryonic context is needed for blastomeres to be able to derive ESC when attached directly on the MEF. We suggest that this molecular maturation is promoted by the second cleaving in the embryo (cleaving on the feeder layer would not be equivalent), and thus, 4C blastomeres are the first to have the capacity to develop an epiblastic molecular identity when cultured directly on the feeder cells. The inability of control 2C embryos to give rise to a single ESC colony corroborates this hypothesis. In fact, only one author has reported direct derivation of ESC from 2C embryos blastomeres or zygotes (bypassing the blastocyst stage)(Tesar, 2005) and those results have not been reproduced yet.

Interestingly, when we then calculated the ESC rate per original embryo (the ones we originally disaggregated to obtain the blastomeres) we observed that the one-step method is generally more efficient deriving ESC from blastomeres than the two-steps method. This higher efficiency is reflected in higher derivation rates for the one-step method using 4C and 8C blastomeres than using the traditional method with control blastocysts. This finding suggests that in the two step method, epiblastic cells that originate the ESC are kept inside the blastocoelic cavity, with the trophectodermal cells as a barrier to overcome to establish an effective expansion on the feeder cells. Contrarily, the one-step method grants contact of the embryonic cells with the feeder cells at all times. Moreover, separating the blastomeres prior to compaction may imply maximizing the possibility of activation of the epiblastic molecular signature in those blastomeres that are undergoing a gradual restriction of their developmental potential. This would be in accordance to the observation that 4C blastomeres have a higher ESC derivation rate per original embryo than the 8C blastomeres.

Here we have tested both molecularly and functionally whether cleavage-stage blastomeres up to the third division showed any significant difference on their developmental potential when isolated. Overall, our findings indicate that 2C-embryo blastomeres are developmentally equivalent to the zygote (totipotent). Actually, only 2C embryo blastomeres have been shown develop to term and originate pups(Tarkowski, 1959). Interestingly, we found that around the time of the second division

the molecular mechanisms that will originate the three lineages in the blastocyst are triggered. Our findings suggest that turning on this molecular mechanism implies a progressive maturation process in the blastomeres that involves a gradual reduction of the ability to form the three blastocystic lineages at this point. Importantly, the first difference is established when some blastomeres lose the capacity to form the epiblast. However, these blastomeres can form the TE and the PE.

The protein labeling experiments allowed us to visualize the evolution of the localization of the markers of the established lineages. These transcription factors have shown to be discriminating for the three lineages at the blastocyst stage and it is a general consideration that their co-expression around compaction and cavitation implies that blastomeres are not yet committed to any lineage. However, our results led us to hypothesize that the coexpression of OCT4, NANOG, GATA6 and CDX2 may be a default expression pattern at this stage, and that the molecular mechanisms triggered at the second cleavage influence the progressive restriction of expression of these four transcription factors and cell sorting after compaction.

However, we consider that during the first three divisions the interaction between blastomeres may influence the molecular mechanism that controls the restriction of the developmental potential with a certain degree of variability. Therefore, we hypothesize that although molecular differences between blastomeres at any point before compaction may be dynamic they determine their molecular fate within a given embryo, and therefore, if disaggregated and deprived of the embryonic environment, blastomeres show differences in their developmental potential acquired until that precise moment. Since our experimental design does not take into consideration the relative position of the blastomeres, our results obviate that influence, but the dynamic progressive reduction in the developmental potential when blastomeres are separated at 4C and 8C stages would be in agreement with the hidden pre-formation model.

Material and Methods

Embryos

Mouse embryos were commercially obtained frozen at 2 cell stage (B6C3F-1 x B6D2F-1, Embryotech, Wilmington, MA). After thaw, embryos were cultured in phase I sequential IVF media(Gardner and Lane, 1997) and cultured at 37ºC under an atmosphere of 5% CO_2. After 48 hours of culture, embryos were transferred to phase II sequential IVF media(Gardner and Lane, 1997) and cultured under the same conditions.

Immunolabelling

Embryos were fixed at four different developmental stages: 2 cells (2C, 2h after thaw, 10 embryos), 8 cells (8C, 18h after thaw, 10 embryos), 16-32 cells (16-32C 30h after thaw, 10 embryos), and expanded blastocyst (BL, 60h after thaw, 10 embryos). Fixation was performed introducing the embryos in a paraformaldehyde 4% solution for 5 minutes. Afterwards, embryos were permeabilized and blocked using a Triton-X100 1% + bovine serum albumin (BSA) 0.1% solution in PBS buffer for 60 minutes. Samples were incubated for 1 hour in the primary antibodies solution at room temperature, and then washed 3 times in PBS and finally incubated in Alexa-Flour conjugated secondary antibodies (Molecular Probes, Carlsband, CA) for 60 minutes. Two combinations of primary antibodies were used. Labeling #1: OCT4 (1:50 rabbit-raised, Abcam, Cambridge, MA), CDX2 (1:200, mouse raised, Biogenex, San Ramon, CA), GATA6 (1:50, goat raised, R&D, Minneapolis, MN). Labelling #2 OCT4 (2,5 µg/ml, mouse raised; BD, San Jose, CA), NANOG (1:100, rabbit raised, Abcam), GATA6 (1:50, goat raised, R&D). 5 embryos per stage per labelling were analayzed. All imaging was performed using a Zeiss 510 META confocal microscope.

Real-Time quantitative PCR

The quantification of transcripts by real-time quantitative PCR was performed on single embryo samples at same four developmental stages used in immunolabeling: two-cell (2C), eight-cell (8C), 16 to 32 cells (16-32C), and expanded blastocyst (BL). Four embryos of each stage were analyzed for transcript abundance of *CDX2*, *GATA6*, *OCT4*, *NANOG*, *SOX2*.

RNA was isolated from embryos using the Picopure RNA Isolation Kit (Arcturus, Mountain View, CA) according to the manufacturer's instructions. In column DNase digestion was carried out using RNAse free DNase Set (Qiagen, Valencia CA). The total volume of the extraction was 10 µl. The total RNA reverse transcription was performed using Sensiscript cDNA synthesis kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. A mixture of Oligo-dT primers (final concentration of 1µM) and random nonamers (at a final concentration of 10 µM) were used in the reaction in a final volume of 20 µl. Negative controls were included. The resulting cDNA solution was used as a template for pre-amplification of selected amplicons. To do so, we used Taqman Preamp Master Mix (Applied Biosystems) and the Taqman primers (*Oct4*, *Nanog*, *Cdx2*, *Sox2*, *Gata6*, *Ppia*, *Hprt1*) following the manufacturer's instructions to perform the pre-amplification The thermal profile was: one cycle at 95ºC for 10 minutes, followed by 14 cycles at 95°C for 15s and 60°C for 4 minutes. The final volume of the reaction was 40 µl. The result samples were diluted 1/5 up to 200 µl total volume.

Real time quantitative PCR was performed on a 7500 Fast Real Time PCR System (Applied

Biosystems) using Taqman PCR Master Mix (Applied Biosystems) and the same commercial primers used in the pre-amplification reaction. We run replicates for each sample. Thermal profile was: one cycle at 50°C for 2 min followed by 1 cycle at 95°C for 10 minutes followed by 40 cycles at 95°C for 15s and 60°C for 1 minute.

Primers were checked for amplification efficiency in the linear range: we pre-amplified murine embryonic stem cell cDNA and made a dilution series. RT-qPCR results of this series let us calculate the efficiency of the reaction. The normalized values for each target transcript were analyzed using the 95% confidence intervals.

Previous analyses to check the preamplification faithfulness had been performed in a previous work (submitted). Hprt1 and *Ppia* were chosen as housekeeping genes after a thorough analysis (submitted) using geNorm software(Vandesompele et al., 2002). This application calculates a gene-stability measure (M) to determine the expression stability of all the genes tested on the basis of non-normalized expression levels. For each gene, Ct values of unknown samples were transformed into the log of the starting quantities with the formula obtained from the standard curve, thereby taking into account the efficiency of the PCR reaction. Raw starting quantities were analyzed with geNorm to determine gene expression stability over the different developmental stages, which resulted in a gene expression stability measure *M* for each gene.

Blastomere isolation

For those embryos object of blastomere separation zona pellucida was removed using Tyrode's acid solution 2h after thaw. Mouse embryos were disaggregated at two-cell, four-cell and eight cell stages gently pipeting the cells in a Ca^{2+}-Mg^{2+} free media until the embryo was completely disaggregated.

To obtain blastomere-derived embryos, the resulting blastomeres were cultured in sequential media in parallel conditions to the control embryos until full expansion (65h after thaw).

Sibling blastocysts derived from disaggregated embryos of each stage were immunolabeled to determine whether distribution of NANOG, OCT4, GATA6 and CDX2 was comparable to the normally developing embryos. Four embryos of each stage (2C, 4C and 8C) were disaggregated to produce 8 x 2CBL embryos, 14 x 4CBL and 18 x 8CBL (not all the 4CBL and 8CBL reached the blastocyst stage). Two sets of sibling blastomere derived embryos were labeled with each of the two different antibody combinations, following the previously described conditions.

Sibling blastomere-derived blastocyst from one embryo at each cleavage stage (2C, 4C, and 8C), along with three control blastocysts, were used to perform an RT-qPCR to analyze the transcript abundance of Oct4, NANOG, SOX2, CDX2 and GATA6 following the same conditions as

previously described.

ESC derivation from blastomeres

To derive the ESC lines from blastomeres, two techniques were used: A) blastomere-derived embryos were laid on a culture of murine embryonic fibroblasts (MEF), and cultured until outgrowth was observed. Standard culture media was used: KO-DMEM (Invitrogen, Carlsband, CA) supplemented with 20% Knockout Serum Replacement (Invitrogen), 1X nonessential amino acids (Gibco, Carlsband, CA), 1X L-Glutamine (Invitrogen), 1X Pen/Strep (Invitrogen), 1X Mercaptoethanol (Gibco) supplemented with 1ng/ml mouse LIF (Sigma, Saint Louis, MO). A control group of denudated 2-cell embryos was cultured until blastocyst and laid on MEF to compare derivation efficiency. B) Mouse embryos were disaggregated at two-cell, four-cell and eight cell stage. The resulting blastomeres were carefully attached to the MEF and cultured in ESC medium (described above) for 9 days.

After expansion, outgrowths were picked up using 2.5% trypsin/EDTA and put on fresh MEFs in ES medium. After 5 days, ESC-like colonies developed and were further passaged by disassociation using 0.25% trypsin/EDTA and transferred onto fresh MEFs. Mouse ESC colonies were passaged every 4-5 days. A control group of denudated 2 cell embryos was also cultured in the same conditions.

ESC pluripotency characterization

ESC lines derived from the three type of blastomeres (2C, 4C and 8C) where tested for expression of pluripotency markers OCT4 and NANOG. ESC cultures derived from blastomeres were cultured for 8 passages and a sample was fixed using the same immunolabeling protocol as described above for the embryonic samples. Primary antibody solution contained OCT4 (2,5 μg/ml BD), NANOG (1:100 Abcam). Detection solution contained Alexafluor conjugated secondary antibodies. Samples were visualized in a fluorescence microscope.

Statistical Analysis

Fisher exact probability test and Student t-test were performed using the data presented in the tables, and $P < 0.05$ was considered statistically significant

Bibliography

Ang, S. L., and Constam, D. B. (2004). A gene network establishing polarity in the early mouse embryo. Seminars in cell & developmental biology *15*, 555-61.

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & development *17*, 126-40.

Beddington, R. S., and Robertson, E. J. (1999). Axis development and early asymmetry in mammals. Cell *96*, 195-209.

Bischoff, M., Parfitt, D. E., and Zernicka-Goetz, M. (2008). Formation of the embryonic-abembryonic axis of the mouse blastocyst: relationships between orientation of early cleavage divisions and pattern of symmetric/asymmetric divisions. Development (Cambridge, England) *135*, 953-62.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-55.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. Developmental cell *10*, 615–624.

Chung, Y., Klimanskaya, I., Becker, S., Marh, J., Lu, S. J., Johnson, J., Meisner, L., and Lanza, R. (2006). Embryonic and extraembryonic stem cell lines derived from single mouse blastomeres. Nature *439*, 216-9.

Ciemerych, M. A., Mesnard, D., and Zernicka-Goetz, M. (2000). Animal and vegetal poles of the mouse egg predict the polarity of the embryonic axis, yet are nonessential for development. Development (Cambridge, England) *127*, 3467-74.

Delhaise, F., Bralion, V., Schuurbiers, N., and Dessy, F. (1996). Establishment of an embryonic stem cell line from 8-cell stage mouse embryos. European journal of morphology *34*, 237-43.

Dietrich, J. E., and Hiiragi, T. (2007). Stochastic patterning in the mouse pre-implantation embryo. Development (Cambridge, England) *134*, 4219-31.

Fujimori, T., Kurotaki, Y., Miyazaki, J., and Nabeshima, Y. (2003). Analysis of cell lineage in two- and four-cell mouse embryos. Development (Cambridge, England) *130*, 5113-22.

Gardner, D. K., and Lane, M. (1997). Culture and selection of viable blastocysts: a feasible proposition for human IVF? Human reproduction update *3*, 367-82.

Gardner, R. L. (2001). Specification of embryonic axes begins before cleavage in normal mouse development. Development (Cambridge, England) *128*, 839-47.

Graham, C. F. (1971). The design of the mouse blastocyst. Symposia of the Society for Experimental Biology *25*, 371-8.

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell *132*, 567-82.

Jedrusik, A., Parfitt, D. E., Guo, G., Skamagki, M., Grabarek, J. B., Johnson, M. H., Robson, P., and Zernicka-Goetz, M. (2008). Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. Genes & development *22*, 2692-706.

Johnson, M. H., and McConnell, J. M. (2004). Lineage allocation and cell polarity during mouse embryogenesis. Seminars in cell & developmental biology *15*, 583-97.

Johnson, M. H., and Ziomek, C. A. (1981a). Induction of polarity in mouse 8-cell blastomeres: specificity, geometry, and stability. The Journal of cell biology *91*, 303-8.

Johnson, M. H., and Ziomek, C. A. (1981b). The foundation of two distinct cell lineages within the mouse morula. Cell *24*, 71-80.

Lorthongpanich, C., Yang, S. H., Piotrowska-Nitsche, K., Parnpai, R., and Chan, A. W. (2008). Development of single mouse blastomeres into blastocysts, outgrowths and the establishment of embryonic stem cells. Reproduction (Cambridge, England) *135*, 805-13.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A. A., et al. (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nature cell biology *9*, 625-35.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-42.

Morrisey, E. E., Tang, Z., Sigrist, K., Lu, M. M., Jiang, F., Ip, H. S., and Parmacek, M. S. (1998). GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. Genes & development *12*, 3579-90.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-91.

Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K., and Sasaki, H. (2008). Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mechanisms of development *125*, 270-83.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917-29.

Piotrowska-Nitsche, K., Perea-Gomez, A., Haraguchi, S., and Zernicka-Goetz, M. (2005). Four-cell stage mouse blastomeres have different developmental properties. Development (Cambridge, England) *132*, 479-90.

Piotrowska-Nitsche, K., and Zernicka-Goetz, M. (2005). Spatial arrangement of individual 4-cell stage blastomeres and the order in which they are generated correlate with blastocyst pattern in the mouse embryo. Mechanisms of development *122*, 487-500.

Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A. K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development (Cambridge, England) *135*, 3081-91.

Ralston, A., and Rossant, J. (2008). Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. Developmental biology *313*, 614-29.

Rodda, D. J., Chew, J. L., Lim, L. H., Loh, Y. H., Wang, B., Ng, H. H., and Robson, P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. The Journal of biological chemistry *280*, 24731-7.

Rossant, J. (2004). Lineage development and polar asymmetries in the peri-implantation mouse blastocyst. Seminars in cell & developmental biology *15*, 573-81.

Rossant, J. (1976). Postimplantation development of blastomeres isolated from 4- and 8-cell mouse eggs. Journal of embryology and experimental morphology *36*, 283-90.

Rossant, J., and Tam, P. P. (2009). Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. Development (Cambridge, England) *136*, 701-13.

Schier, A. F., and Talbot, W. S. (2005). Molecular genetics of axis formation in zebrafish. Annual review of genetics *39*, 561-613.

Singh, A. M., Hamazaki, T., Hankowski, K. E., and Terada, N. (2007). A heterogeneous expression pattern for Nanog in embryonic stem cells. Stem cells (Dayton, Ohio) *25*, 2534-42.

Spindle, A. (1982). Cell allocation in preimplantation mouse chimeras. The Journal of experimental zoology *219*, 361-7.

Strumpf, D., Mao, C. A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development (Cambridge, England) *132*, 2093-102.

Suwinska, A., Czolowska, R., Ozdzenski, W., and Tarkowski, A. K. (2008). Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16- and 32-cell embryos. Developmental biology *322*, 133-44.

Tam, P. P., and Rossant, J. (2003). Mouse embryonic chimeras: tools for studying mammalian development. Development (Cambridge, England) *130*, 6155-63.

Tarkowski, A. K. (1959). Experiments on the development of isolated blastomers of mouse eggs. Nature *184*, 1286-7.

Tay, Y., Zhang, J., Thomson, A. M., Lim, B., and Rigoutsos, I. (2008). MicroRNAs to Nanog, Oct4

and Sox2 coding regions modulate embryonic stem cell differentiation. Nature *455*, 1124-8.

Tesar, P. J. (2005). Derivation of germ-line-competent embryonic stem cell lines from preblastocyst mouse embryos. Proceedings of the National Academy of Sciences of the United States of America *102*, 8239-44.

Tsunoda, Y., and McLaren, A. (1983). Effect of various procedures on the viability of mouse embryos containing half the normal number of blastomeres. Journal of reproduction and fertility *69*, 315-22.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology *3*.

Wakayama, S., Hikichi, T., Suetsugu, R., Sakaide, Y., Bui, H. T., Mizutani, E., and Wakayama, T. (2007). Efficient establishment of mouse embryonic stem cell lines from single blastomeres and polar bodies. Stem cells (Dayton, Ohio) *25*, 986-93.

Weaver, C., and Kimelman, D. (2004). Move it or lose it: axis specification in Xenopus. Development (Cambridge, England) *131*, 3491-9.

Wilton, L. J., and Trounson, A. O. (1989). Biopsy of preimplantation mouse embryos: development of micromanipulated embryos and proliferation of single blastomeres in vitro. Biology of reproduction *40*, 145-52.

Yagi, R., Kohn, M. J., Karavanova, I., Kaneko, K. J., Vullhorst, D., DePamphilis, M. L., and Buonanno, A. (2007). Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. Development (Cambridge, England) *134*, 3827-36.

Yamanaka, Y., Ralston, A., Stephenson, R. O., and Rossant, J. (2006). Cell and molecular regulation of the mouse blastocyst. Dev Dyn *235*, 2301-14.

Zernicka-Goetz, M. (1998). Fertile offspring derived from mammalian eggs lacking either animal or vegetal poles. Development (Cambridge, England) *125*, 4803-8.

Zernicka-Goetz, M. (2002). Patterning of the embryo: the first spatial decisions in the life of a mouse. Development (Cambridge, England) *129*, 815-29.

A

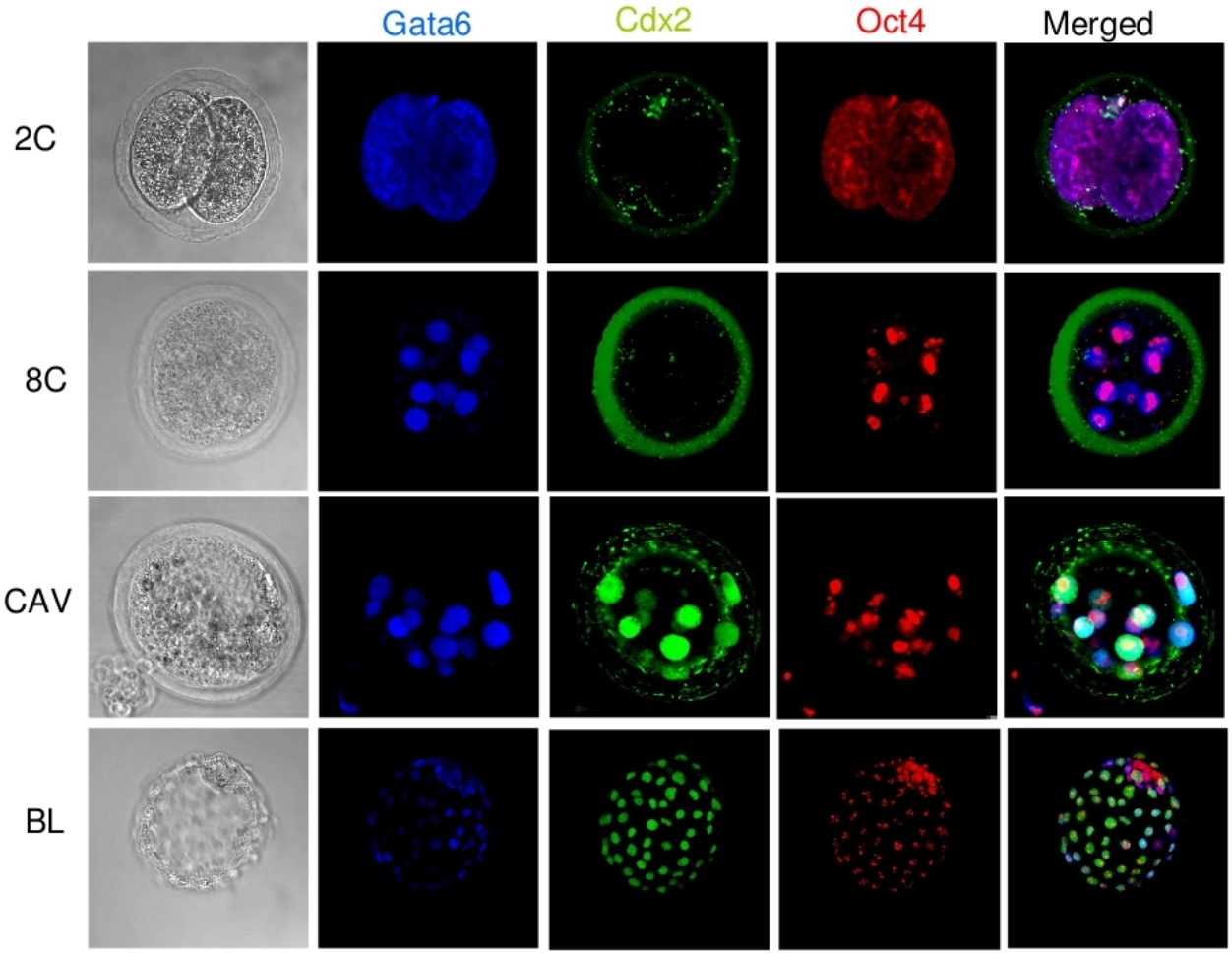


Figure 1. A. Results of the immunofluorescent experiments for GATA6, CDX2, OCT4. Confocal laser scanning images of *in vitro* cultured mouse embryos labeled with primary antibodies for GATA6, CDX2 and OCT4

B

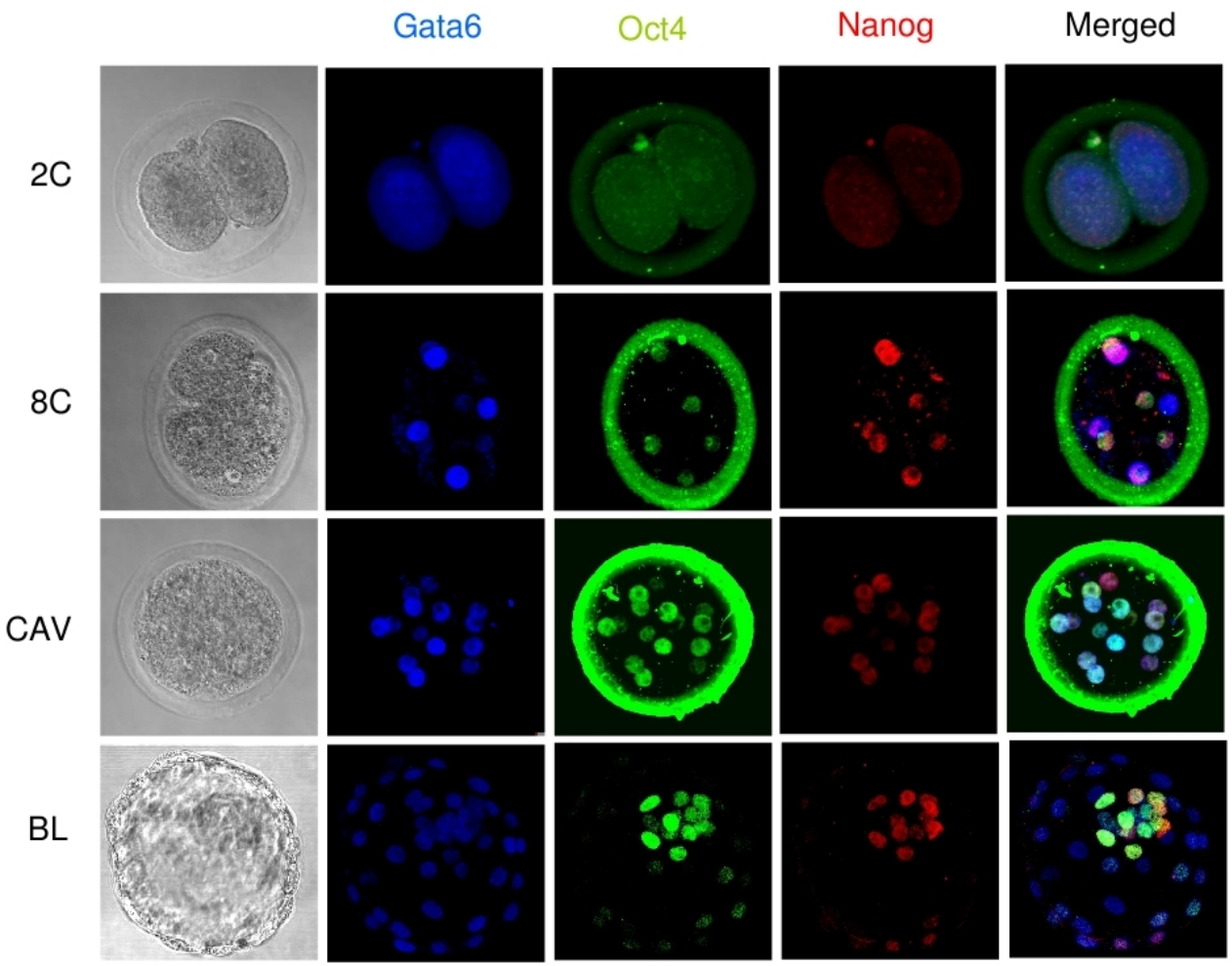


Figure 1. B. Results of the immunofluorescent experiments for GATA6, NANOG, OCT4. Confocal laser scanning images of *in vitro* cultured mouse embryos labeled with primary antibodies for GATA6, NANOG and OCT4. For both experiments, secondary detection was performed using ALEXA flour labeled secondary antibodies. Different stages of preimplantation embryo development were analyzed (2C: two cell; 8C: eight cell, CAV: compacting-cavitating, 16 to 32 cells; BL: blastocyst).

Figure 2. Expanding mouse blastocyst. GATA6 and OCT4 intensity of expression is diminishing in CDX2 expressing cells. A high proportion of the trophectodermal cells only express CDX2 Two clearly distinct cell populations can be observed in the growing ICM: OCT4+ GATA6- (EPI) and OCT4+GATA6+ cells (PE).

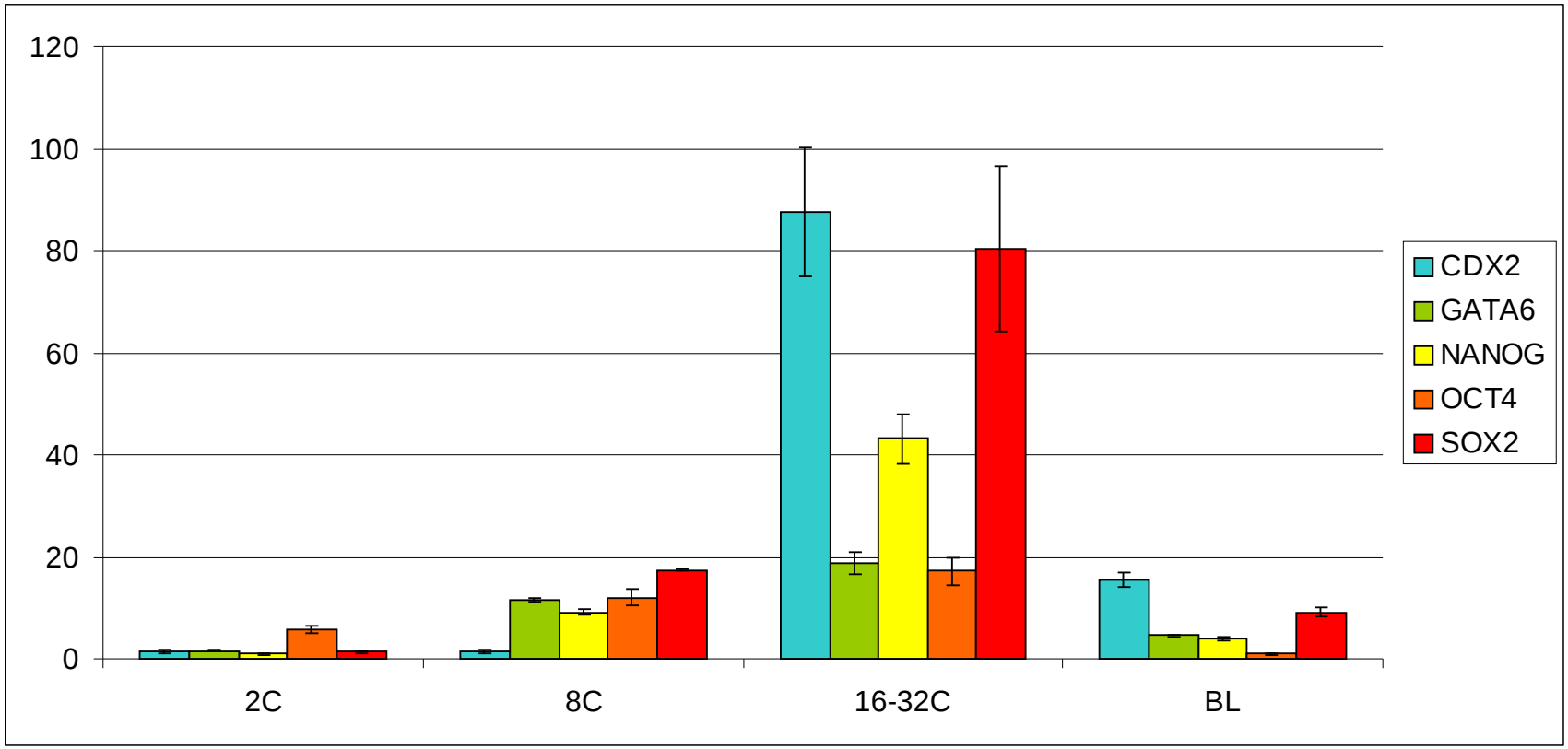


Figure 3. RT-qPCR results of normally developing mouse embryos. Developmental stage (2C = 2-cell stage, 8C = 8-cell stage, CAV= compacting-cavitating stage, BL = expanded blastocyst). Y-axis: mean relative expression. Data was normalized to the geometric mean of 2 stably expressed genes (HPRT1, PPIA) as determined by geNorm analysis. Median values of the 4 replicates are shown. Error bars represent standard deviation of the measurements.

A

Stage	Disaggregated Embryos		Blastocyst rate	ESC derivation			
	# Embryos	# Blastomeres		# ESC lines	ESC per blastomere rate	ESC per blastocyst rate	ESC per original embryo rate
Control	76	-	88.2 % (67)	22	-	33.0	33.0
2C	96	192	78.2% (150)	32	16.6	21.3*	33.3
4C	10	40	42.5% (17)*	2	5.0	11.8*	20*
8C	20	160	9.4% (15)*	0	0	0*	0*

B

Group / stage	# Embryos	# Blastomeres	# ESC lines	% ESC per Blastomere	% ESC per Embryo
Control	40	-	0	0	0
2C	40	80	0	0	0
4C	20	78	11	14.1	55
8C	20	158	8	5.1[§]	40

Table 1. **A**. **Experiment 1 results**. Disaggregated blastomeres were cultured in vitro to form a blastocyst and then at 72h approximately, laid onto MEF layer in order to derive ESC. **B**: **Experiment 2 results**. Disaggregated blastomeres were laid on MEF directly to derive ESC . *Statistically significant differences ($p<0.05$). § Two ESC lines were derived from sibling eight-cell embryo blastomeres.

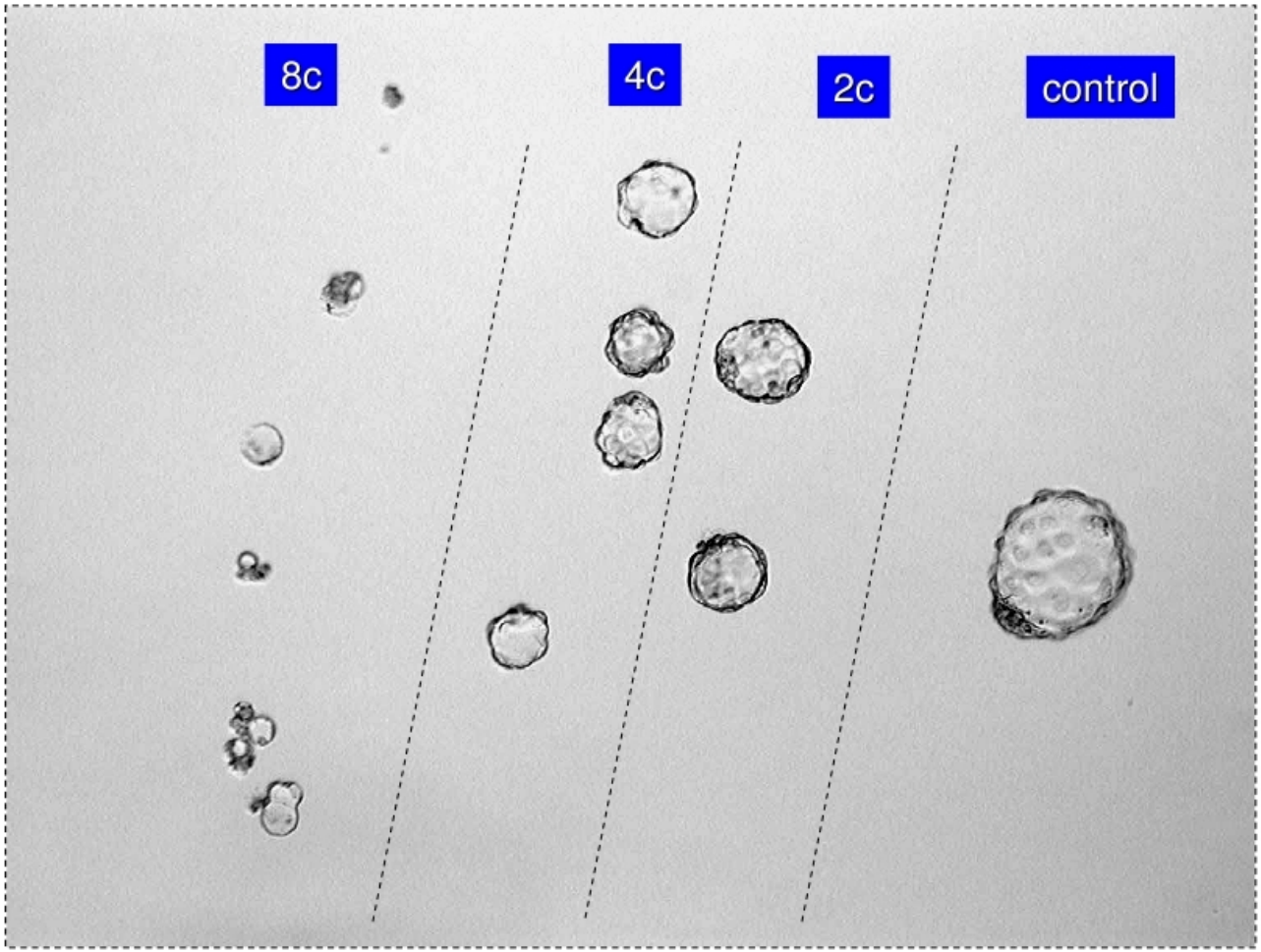


Figure 4. Blastomere derived embryo after 70-72 h of culture. Size of blastomere derived embryos is proportional to the original blastomere that originated them.

A

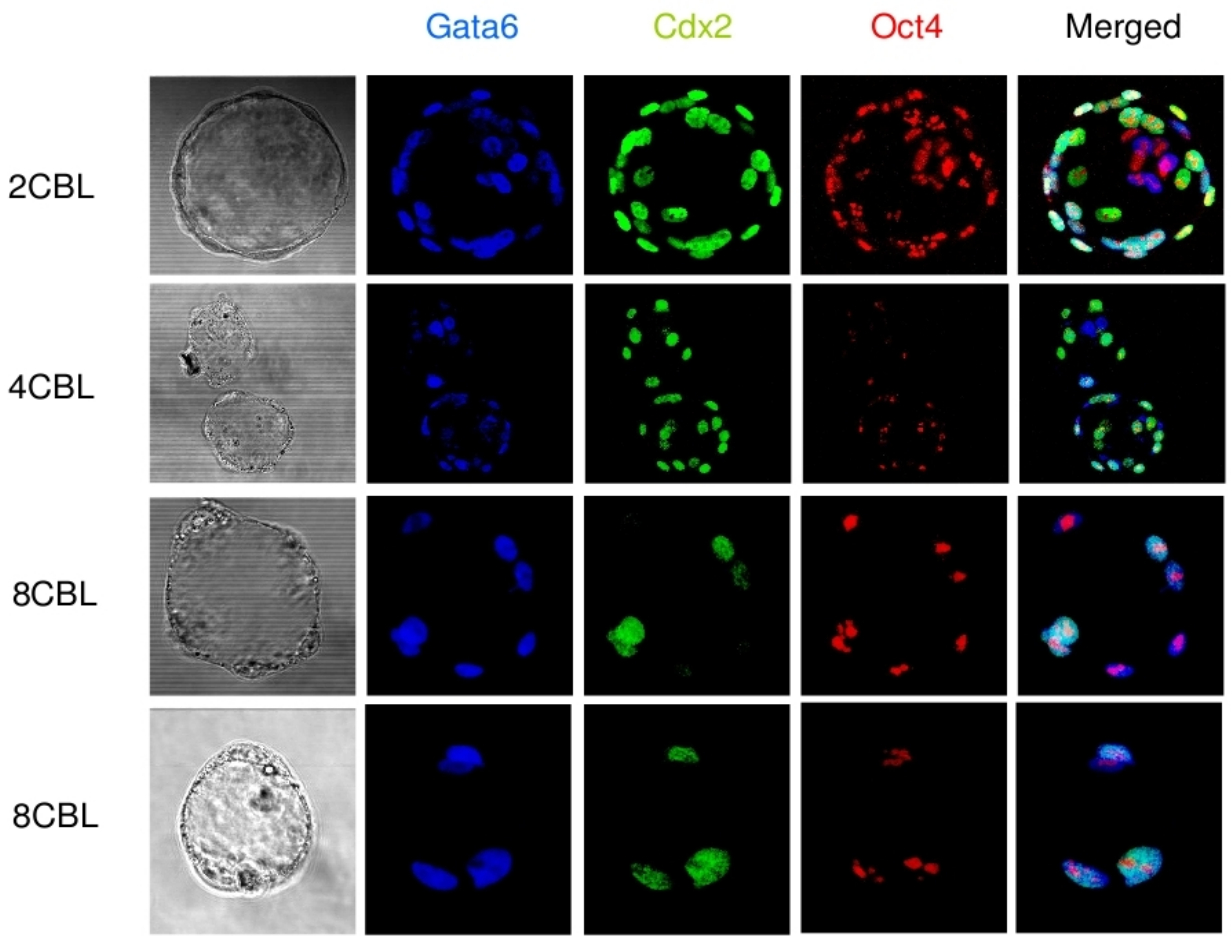


Figure 5. A. Results of the immunofluorescent experiments for GATA6, CDX2, OCT4. Confocal laser scanning images of *in vitro* cultured blastomere-derived embryos labeled with primary antibodies for GATA6, CDX2 and OCT4.

B

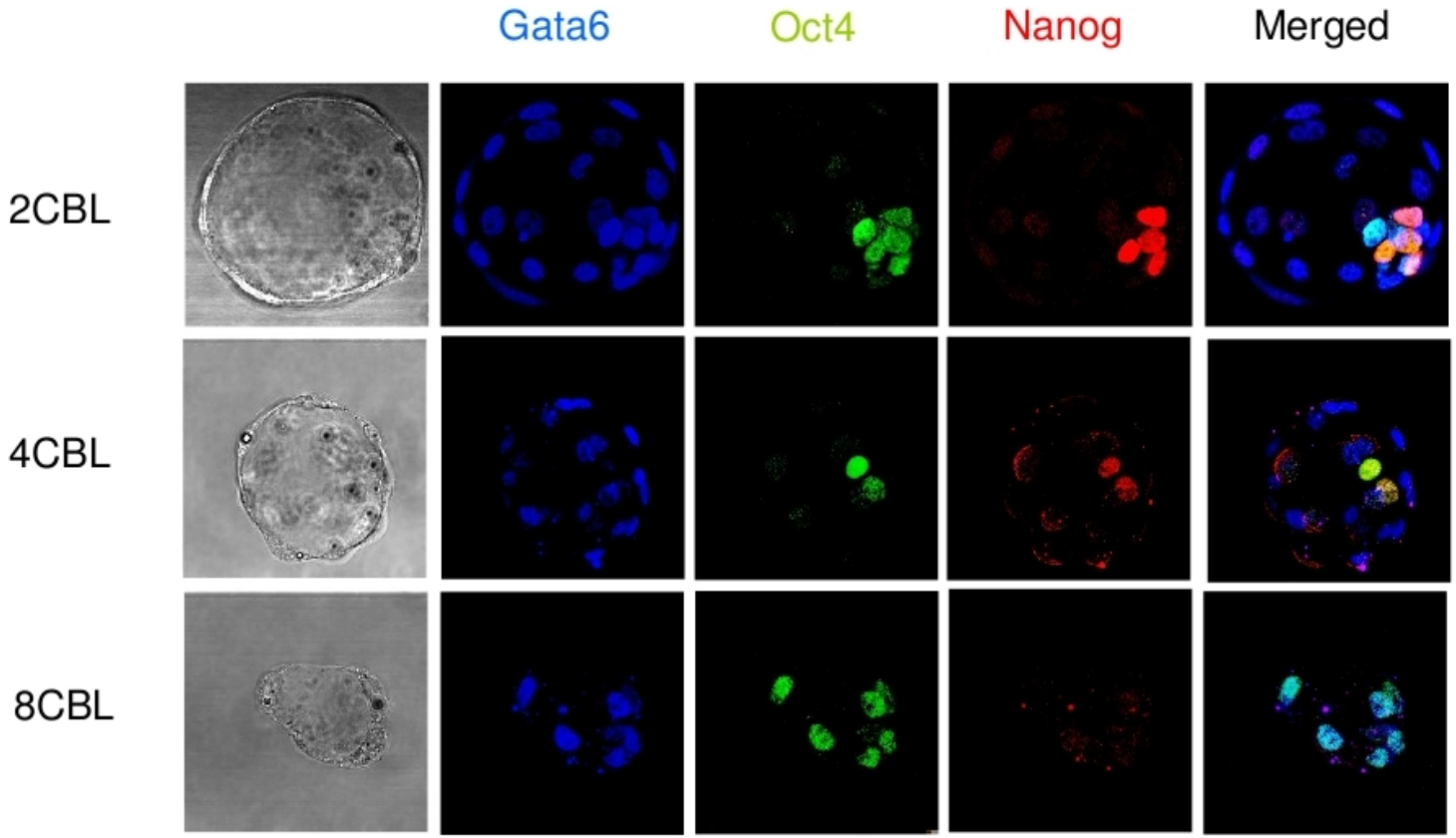


Figure 5.B. Results of the immunofluorescent experiments for GATA6, NANOG, OCT4. Confocal laser scanning images of *in vitro* cultured cultured blastomere-derived embryos labeled with primary antibodies for GATA6, NANOG and OCT4. For both experiments, secondary detection was performed using ALEXA flour labeled secondary antibodies. All the embryos were cultured for approximately 70-72 hours before fixation.

A

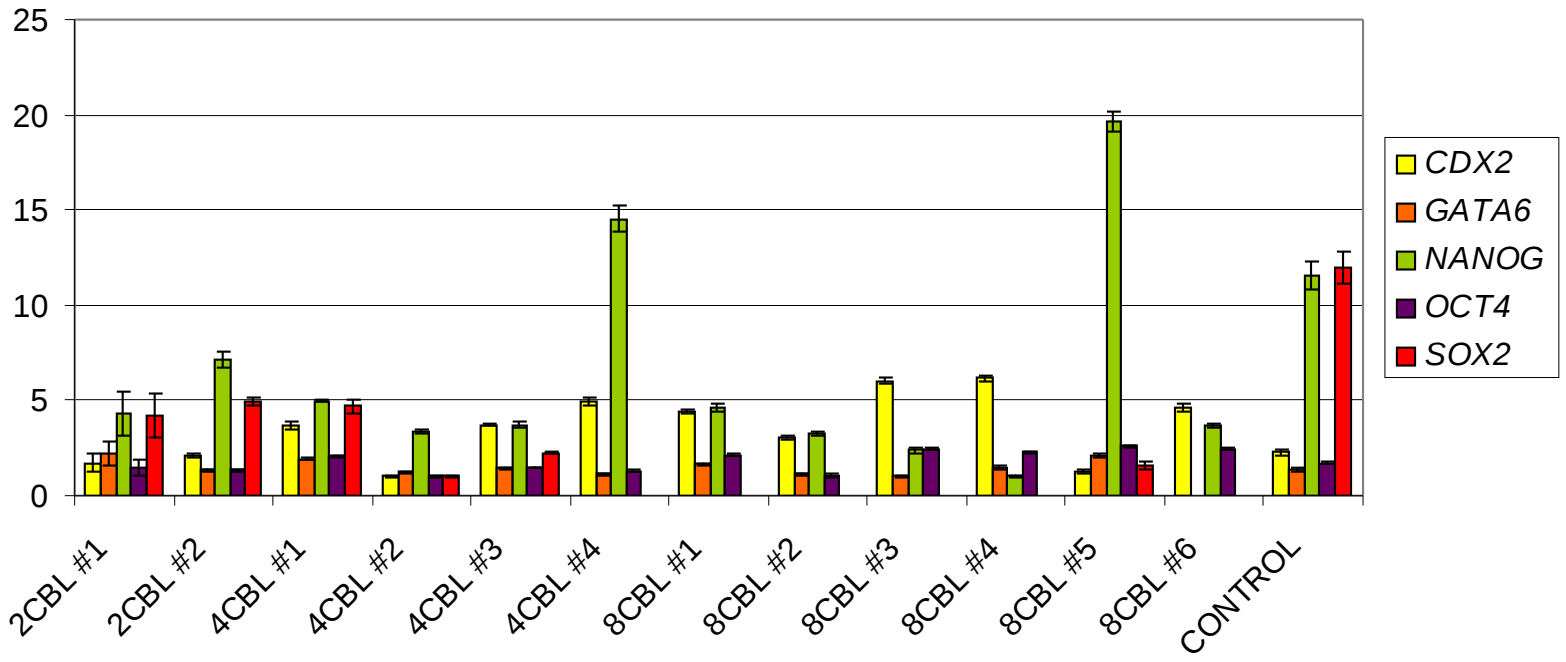


Figure 6 . **A**. RT-qPCR results of a sibling 2CBL, 4CBL and 8CBL blastocysts compared to a mean control blastocyst. X-axis: embryo type. Y-axis: normalized relative expression. Data was normalized to the geometric mean of 2 stably expressed genes (HPRT1, PPIA) as determined by geNorm analysis. Median values of the 4 replicates are shown. Error bars represent SEM.

B

	Nanog/Cdx2	*Sox2/Cdx2*
2CBL #1	2.51749367	2.4711
2CBL #2	3.37312213	2.33981694
4CBL #1	1.33742416	1.26482077
4CBL #2	3.33852209	1.02928330
4CBL #3	0.9933713	0.60686644
4CBL #4	2.95482795	0
8CBL #1	1.06142889	0
8CBL #2	1.06881986	0
8CBL #3	0.3944277	0
8CBL #4	0.16238625	0
8CBL #5	15.145541	1.24070622
8CBL #6	0.7945399	0
CONTROL	5.03357971	5.21566448

Figure 6 . **B**. Relative expression ratios between *NANOG* / *CDX2* and *SOX2* / *CDX2*. Only those blastomere-derived blastocyst with both ratios > 1 show a similar pattern to the control blastocyst. NANOG transcript is detected in all the 8CBL embryos, but SOX2 transcript is only found in one of them.

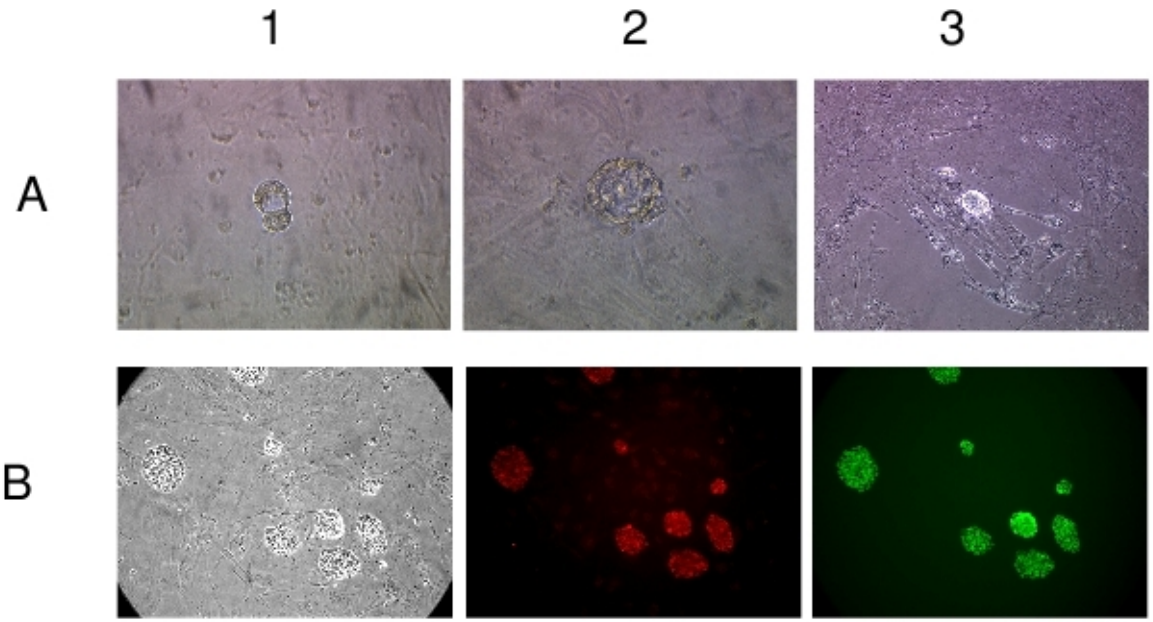


Figure 7. A. Evolution of a eight-cell blastomere on MEF. A1.24h. A2.72h. A3.216h **B**. Pluripotency characterization of the ESC. B1. DIC picture of the colonies. B2. NANOG positive staining. B3. OCT4 positive staining.

4.3 Human embryonic stem cells undergo LOH during long term culture in vitro

Human embryonic stem cells undergo LOH during long term culture in vitro

Josep Pareja[1], Nikica Zaninovic[1], Oriol Vidal[2], Anna Veiga[3] and Zev Rosenwaks[1]

1. Center for Reproductive Medicine and Infertility, Weill Cornell Medical College, New York
2. Departament de Biologia, Facultat de Ciències, Universitat de Girona, Catalonia, Spain.
3. Centre de Medicina Regenerativa de Barcelona, Catalonia, Spain.

Abstract:

Genomic stability of human embryonic stem cells (hESC) during long-term culture has been questioned by recent findings of aneuploidies and other subkaryotypic aberrations. Moreover, these genomic abnormalities have been found to correlate with an enhanced adaptability of hESC to standard culture conditions, characterized by some oncogenic-like features. Our aim was to investigate the genomic integrity of two hESC lines cultured in vitro at different temporal points of their passage history using both classical cytogenetic banding techniques and whole genome single nucleotide polymorphisms (WG-SNP) arrays. These type of arrays allow to detect amplification, deletions as well as homozygous segments in the genome. We also analyzed two pairs of sibling hESC lines at very early passages in order to investigate coincidences and divergences in the genomic events after derivation. Our findings show that both amplifications/deletions and loss of heterozygosity (LOH) occur during in vitro culture. Strikingly, all type of aberrations (specially LOH) were present from the earliest passages from all the cell-lines analyzed. Interestingly, the detected genomic events found comprise coding regions of relevant genes with roles in signal transduction, regulation of transcription, differentiation (polycomb machinery), and significantly in the DNA-damage response machinery. Our results reveal the previously unreported importance of LOH events in the hESC genome during extended in vitro culture. Significantly, the haploinsufficiency of genes involved in DNA repair may enhance adaptability of hESC to in vitro culture conditions by promoting new aberrations.

Key words: Stem cells, genomic stability, LOH.

Introduction

More than ten years after the first derivation event[1], human embryonic stem cells (hESC) still need to meet several safety requirements to be used in clinical protocols. Among other requisites, like the reproducibility of differentiation processes, genomic stability is essential for hESC to be used in cell therapies[2] . Assurance of genomic integrity would imply the elimination of the risk of undesired genetic and epigenetic modifications during extended *in vitro* culture[3,4]. Indeed, previous reports show that although pluripotent embryonic stem cells display a lower mutation frequency than somatic cells, they also exhibit a higher susceptibility to undergo structural chromosomal rearrangements[5]. In accordance to this finding, several papers describe recurrent numeric chromosomal abnormalities in hESC after extended in vitro culture, singularly the gain of chromosomes 12, 17q and X[6-9], as well as genomic instability in specific sub-chromosomal regions[10,11].

Chromosomal abnormalities acquired by hESC after extended in vitro culture are considered to be the reflection of the progressive adaptation of the cells to their culture conditions, enhancing their capacity to proliferate in such enviroment[6]. Moreover, the process by which karyotypically abnormal cells display an increased adaptability to culture conditions raised an obvious parallelism with the process of malignant transformation[6,11]. Indeed, other than the accrued growth rate, previous reports describe tumorigenic characteristics that chromosomally abnormal hESC display: enhanced cloning efficiencies after plating single cells[12], reduced capacity for apoptosis[13] and/or retention of undifferentiated identity in xenograf teratomas[14]. In spite of the correlation between these features and chromosomal abnormalities, adaptation could also involve other genetic and epigenetic changes beyond the resolution of standard cytogenetic techniques[12,15]. Loss of heterozygosity (LOH) is one of the possible subkaryotypic events occurring in these cells. LOH occurs when a genomic DNA region originally heterozygous becomes homozygous. Indeed, LOH has been observed to affect from small regions comprising dozens of base pairs to entire chromosomes[16,17]. LOH can be originated due two different mechanisms: induced by DNA double strand breaks[18-20] or produced by an error during chromosome segregation in mitosis or meiosis[21,22]. Recombination mechanisms of double strand break repair can produce LOH by homologous recombination between chromosomes or by non-homologous end joining[18-20]. Chromosomal segregation errors that produce LOH have been observed during gametogenesis (either in meiosis I and II) and during early embryonic cell divisions, by chromosomal deletion or mitotic non-disjunction and subsequent duplication[22,23]. As a result, of any of the chromosomal segregation

errors can induce a total o partial isodisomy which is a specific case of LOH called uniparental disomy (UPD) [23,24]. Moreover, LOH is one of the alterations commonly observed in sporadic tumors[25-30].The acquired haploinsufficiency in particular genomic regions has been proven to be an obligate step in tumorigenesis in several familiar cancer syndromes involving tumor suppressor genes[16,25,30-32]. In these cases, the loss of a non-pathological allele in genomic regions containing a pathological copy of a tumor-suppressor gene would lead the cells down the oncogenesis path. A different step into the same direction would be the case of a region with a mutant or a dominant allele in an oncogene where homozygosity could result in a super-oncogenic effect[33-36].

Pluripotent stem cells have been found to undergo spontaneous extensive LOH[37,38]. Previously reported experiments using reporter genes at mapped and unmapped loci in murine ESC describe a LOH rates ranging between 10^{-3} and 10^{-5} events per cell generation[37-42], which represents a 18-fold higher frequency than that occurring in somatic cells[39]. The majority of these LOH events in ESC were found to be originated through homologous recombination after double strand break and affected small genomic segments around the breaks [40]. Moreover, other authors showed that exposure to non-toxic concentrations of certain mutagens promotes LOH in murine ESC with varying rates depending on the genetic background of the mice strains[43,44]. Yet, little is known of the impact of LOH during extended in vitro culture of human ESC.

In this paper have used whole genome SNP chips to check the genomic integrity of two hESC cell lines derived in our lab at varying time points in the course of tissue culture. This approach provides a unique opportunity for the assessment of genomic fidelity during in vitro extension. Therefore, we analyzed two pairs of early and late cell samples separated by multiple tissue culture passages. We evaluated the big chromosomal aberrations, deletions, duplications but also the LOH events during the two time points. In addition, we have evaluated the original genomic profile of two pairs of hESC cell lines derived from sibling embryos to determine how different genetic backgrounds affect genomic stability in the newly derived sibling hESC lines. This information can be very valuable to appraise natural occurring genomic variants and those, if any, that are coincident among the different hESC populations.

Material and Methods

Cell culture

WMC1, WMC2, WMC4, WMC5, WMC6 and WMC7 are human embryonic stem cell lines derived

in our lab on mouse embryonic fibroblasts (MEF) feeder cells following described standard protocols[45]. The pairs WMC4 / WMC5 and WMC6 / WMC7 are two pairs of sibling cell lines (derived from embryos with the same parental origin).

These cell lines were routinely cultured on MEF (Chemicon, Temecula, CA) previously inactivated with Mytomicin-C. Culture medium consisted of KO-DMEM (Invitrogen, Carlsbad, CA) supplemented with 20% Knockout Serum Replacement (Invitrogen, Carlsbad, CA), 1X nonessential amino acids (Gibco, Carlsbad, CA), 1X L-Glutamine (Invitrogen, Carlsbad, CA), 1X Penicillin/Streptomycin (Invitrogen, Carlsbad, CA), 1X Mercaptoethanol (Gibco, Carlsbad, CA) and 4ng/ml of FGF-2 (Invitrogen, Carlsbad, CA). hESC were passaged every 4-6 days by incubation with 4mg/ml of collagenase IV (Worthington, Lakewood, NJ), for 60 minutes at 37ºC.

Samples for temporal assessment of genomic stability

WMC1 and WMC2 were tested for genomic stability during extended in vitro culture. Two samples corresponding to early and late passages of of WMC1 and WMC2 were genomically assessed (figure 1). During the timeline of passage history, subclonal lineages of both WMC1 and WMC2 (also called sublines) were started and cultured for several passages. Particularly, WMC1-sub globally underwent 32 splits, deriving from original passage 11 and being further cultured for 21 passages after the separation from the main clonal line (p11+21). WMC2-sub was obtained after 34 passages, deriving from passage 19 of the main line, and being further cultured separately for 15 more splits (p19+15). All the details can be seen in table 1 and an schematic view of the cell lineages is depicted in the figure 1.

Samples for genomic comparison between sibling cell lines

Early passages of the sibling lines were investigated for genomic stability: passage 7 for WMC4/WMC5, and passage 5 for the pair WMC6/WMC7 (table1).

Kariotyping

Between 20 and 30 metaphase spreads were karyotyped for each cell sample. Cells were treated with 0.1 lg/ml Colcemid (Invitrogen, Carlsbad, CA) for up to 4 hours, followed by dissociation with trypsin/versene. The cells were pelleted via centrifugation, resuspended in prewarmed 0.0375 M KCl hypotonic solution, and incubated for 10 minutes. Following a further centrifugation step, cells were resuspended in fixative (methanol:acetic acid 3:1). Metaphase spreads were prepared on glass microscope slides and G-banded by brief exposure to trypsin and stained with 4:1 Gurr's/Leishmann's.

Genomic DNA extraction

Genomic DNA (gDNA) extraction from cell samples was performed using DNeasy extraction kit (QIAGEN, Valencia, CA) following manufacturer's instructions. DNA concentration was quantified

using Nanodrop (Thermo Scientific, Willmington, DE). Genomic DNA quality was assessed by a 1% agarose gel electrophoresis with 1µg of material.

Samples whole-genome genotyping

Illumina HumanCNV-370 BeadChips (Illumina Inc., San Diego, CA) were used to genotype the different hESC lines samples. 750 ng of starting gDNA is required for each sample. Pre-amplification and hybridization process were performed following manufacturer's instruction which are based in the protocol described elsewhere[46]. Fluorescent signals read by BeadStation hardware (Illumina Inc., San Diego, CA) were imported into the BeadStudio software version 3.2 (Illumina Inc., San Diego, CA) and normalized. For each sample, the program's output are the genomic plots of two valuable measures obtained from each SNP: a) B allelic frequency (BAF), which is a measure of the two possible haplotyptes at each position and b) the the Log_2R ratio, where the expected SNP signal intensities are compared with those from HapMap canonycal genotype clusters[46](Log_2 [Intensity $_{subject}$/Intensity$_{expected}$]). The representation of these two measures into genomic profiles are the basis for detecting chromosomal aberrations using WG-SNP arrays. Their visualization side by side provides a simple image of numerical abnormalities and the biggest sub-chromosomal aberrations. Manifest used for 300 k version 2 BeadChips was HumanHap300v2_A. BeadStudio v3.2 was used to visualize the plots of B allele frequency and log R ratio against each chromosome.

Data quality

Call rates were in in the range of [0.9958-0.9982] for all the samples. To ensure reproducibility in the chips preparation and processing, concordances between replicate samples were calculated for two of the lines -WMC1late and WMC5 in two different chips- as the number of concordant pairs divided by the number of successfully genotyped pairs. Results show a concordance of 99.92% and 99.90% respectively.

Array data analysis

For amplification and deletions analysis, two programs were used for robustness of the analysis: SOMATICs[47] and PennCNV[48]. PennCNV version (april 2009) was used in Linux platform, using the default analysis pipeline. SOMATICS[47] is set of scripts for R (downloaded from http://www.r-project.org/). R version 2.9.2 for Linux was used to run the scripts. Data mining was performed using Perl scripting language.

Both programs provide reliable genotypes and can detect chromosomal aberrations at a high resolution from the output files obtained from the HumanCNV370-Duo arrays reading. PennCNV it is a widely used software package for Illumina Beadchip arrays that incorporates a Hidden Markov Model strategy in the detection of the alterations. Nevertheless, many samples present genomic

heterogeneity (within the sample) and that is a major limitation to achieve an accurate analysis of the aberrations. Previous reports account for the extension of numerical aberrations in ESC culture in a matter of a few passages[10,7]. SOMATICs confers a cellular score (cs) to each event (range [0-1]) discriminates fixed alterations (cs=1) and partial alterations (cs <1). In contrast to SOMATICs, PennCNV does not specify any information on the level of extension of the amplifications and deletions detected. Thus, we decided to combine both for comparison and robustness.

LOH Analysis

LOH detection implies an imbalance in the Log_2R ratio and both PennCNV and SOMATICs detect regions with allelic imbalance when one of the two copies of the region is deleted (hemizygosity). When the allelic imbalance is caused by the presence of two homozygous copies of the studied genomic region (LOH), SOMATICs and PennCNV do not perform well. For that reason, the LOH analysis was performed using two specific programs, LOH-Score and dChip, in order to compare their findings and ensure robustness of the results.

LOH Score plug-in for BeadStudio 3.2 was used to score all the SNP in the samples, using the default parameteres. Release 9 of dChip[49] (December 2009) was applied to normalized allele intensities exported from Illumina BeadStudio 3.2, along with genotype calls, as recommended. The SNP annotations were also included (physical positions provided by Illumina). The LOH analysis was performed using Hidden Markov Models for unpaired data, assuming a proportion of heterozygous SNPs of 35% for the Illumina HumanHap300v2 (determined from the normal samples). All other parameters were set to default values.

Software comparison

The way we implemented to quantify the coincidence in event finding between programs is to compute the event-overlap rate for each program. This rate for each sample analyzed by a given program is calculated dividing the number of coincident base-pairs (bp) found to be “altered” by both programs by the number of total bp found to be altered by the tested program.

Another way we used to compare the performance of each software is to check the coherence in the events found on two temporally different samples of the same cell line (with respect to passage history). The temporal event overlap of a cell sample is calculated dividing the number of coincident bp found to be altered in two temporally different passages of the same cell line by the number of total bp found to be altered in each of the passages, when both samples have been analyzed with the same program. This analysis was also applied to sibling cell lines.

Genomic analysis

In order to ensure robustness of the functional analysis, we chose those events that : a) are present in more than one sample of the same cell line (different passages), in siblings cell lines or in

different cell lines; and b) to be detected by the two programs used in each type of alterations, SOMATICs /PennCNV for deletions and amplifications, and LOHScore/dChip for LOH events. Functional analysis consisted in investigating coding regions comprised in the events using the latest release from ENSEMBL (version 56.37a from human assembly GRCh37).

Gene Ontology classification

Biological processes terms from Gene Ontology (GO) database for each gene comprised in the events found were investigated and classified in order to search for any bias in the representativity of these terms in the set of genes found. In order to do so, we first classified the all the ENSEMBL human genes according to their biological process as dictated by GO terms. In total, 11528 human genes had a "biological process" GO-term assigned, and frequency for each term was calculated for each term. Subsequently, chi-square value was calculated to assess the representativity of the most abundant GO terms associated with the set of genes comprised in the deletion/amplification events and the LOH events.

Gene parsing in Cancer databases

Genes in the altered genomic segments were further parsed onto two databases: the Cancer Gene Census[50] (CGC) database, and the COSMIC[51,52] database. CGC database contains genes which their genomic alteration have been proven to be causative of oncogenesis. COSMIC database lists genes that are known to be mutated in tumors.

Results

Cytogenetic analysis

WMC1 and WMC2 were karyotyped at the early, late and sub-passage time points stated in the Table1. Early and subline passages of both WMC1 and WMC2 do not present any detectable aberration. On the late passages, contrarily, WMC1 and WMC2 present one numerical chromosomal aberration each: 86% of WMC2 late metaphase extensions present a trisomy of chromosome 12 and 78% of WMC1 metaphase extensions exhibit a dicentric isochromosome 20, involving loss of most of 20p and duplication of 20q (for karyotype images, see figure 2). WMC4, WMC5, WMC6 and WMC7 present normal karyotypes at the analyzed early passages (table 1).

Array Analysis

-Genomic plots

Unexpectedly, WMC1-early plots of chromosome 20 (Figure 3A) show a Log_2R ratio= 0.101 for the 20p arm (positions from 11799 to 25645042) while the average Log_2R ratio for the rest of early cell samples (WMC2-early, WMC4, WMC5, WMC6 and WMC7) in this genomic segment is 0.0041±0.0032. This Log_2R value implies that an indefinite proportion of the cells in the WMC1-early sample contain an amplification of the 20p arm, which had not been seen in the karyotyping metaphase extensions. The Log_2R value for 20q arm is -0.092, which compared to -0.016±0.012 of the average value of the other cell lines, implies a higher degree of hemizygosity. Strikingly, BAF plot begins to show a separation in two clusters of the different SNPs, which is indicative of an undetermined degree of duplication of the 20q arm in the cellular population, which is in contradiction with the Log_2R value observed. These findings contrast with the BAF plot and the Log_2R value (-0.248) in WMC1-late for the same genomic region which is concordant with the observed karyotypic hemizygosity of the 20p arm (Figure 3B). In this same sample, duplication of the q arm is reflected by the evident split of the heterozygous state in the BAF plot into two clusters plus a Log_2R value of 0.117 (-0.016±0.012 being the average value for the rest of the cell lines). Interestingly, the change of tendency in the Log_2R values in both WMC1-early and WMC1-late samples is centered in the 20p11.21 band of the chromosome.

WMC2-early average Log_2R of 0.011 for the whole chromosome plus the BAF plot confirm the euploid cytogenetic karyotype (Figure 3C). For the WMC2-late sample (Figure 3D), the Log_2R value (0.167) of the whole chromosome 12, plus the clear split of the heterozygous state in two clusters observed in the BAF plot is in agreement with the trisomy 12 observed in the 86% of the metaphase extensions. However, the Log_2R value indicates that the trisomy is not extended in the whole cell population: using genomically homogeneous aberrant samples, Peiffer *et al*.[46] determined that Log_2R value for a trisomic chromosome should be +0.395.

Log_2R ratio and BAF plots confirm the euploid karyotypes of the rest of the lines analyzed: WMC4, WMC5, WMC6, and WMC7.

Deletions and Amplification Analysis

-Software comparison

Table 2 shows that the number of events detected by SOMATICs and PennCNV. No match between programs is seen in the number of deletions/amplifications found for any of the samples. The average length per event and the global length of all the events detected by SOMATICs is higher than PennCNV for all the samples. Remarkably, on the late passages of WMC1 and WMC2, PennCNV divides the known isodisomy 20 (WMC1-late) and the trisomy 12 (WMC2-late) in dozens of smaller different events, while SOMATICs segments them in 3 and 5 events respectively. SOMATICs globally detects more deletions than amplifications, while PennCNV finds the opposite results for the same samples (Table 3). However, the increased segmentation of the events found by PennCNV influences these differences in the global number of amplification and deletions found.

When we investigated the average event overlap ratio between programs respect the total aberration length we observed a 13,9% overlap for SOMATICs and a 34,73% for PennCNV. This percentages correspond to the length of the regions found coincidentally by both programs with respect to the total length of events found by each of them. The lower overlap ratio for SOMATICs is in accordance with the higher average total event length found by this program, attributable to the *a priori* higher sensitivity of this software. However, there is a wide range of overlapping ratios between the events found by SOMATICs and PennCNV for the different samples (table 2). In the case of the WMC1-late and WMC2-late passages, the existence of extended large numerical abnormalities accounts for the highest overlapping ratios between programs (99% for PennCNV). For the earlier passages, contrarily, the fewer and the less extended (in the cell population) genomic amplifications or deletions evidence the challenge in aberration-defining task that face the analysis software: the programs overlap ratios range is [0.1-22.04%] for SOMATICs and [0.1-41,40%] for PennCNV.

-Temporal and sibling overlap between samples

The temporal/sibling event-overlap ratios between samples (Table 4) differ depending on the software used, but results obtained from SOMATICs show a higher temporal overlap even if the total length of the events found by this program is larger than PennCNV. Therefore, we used the SOMATICs results to explore the coherence and the meaning of these temporal coincidences in the events found in the three different passages of WMC1 and WMC2 and in the sibling cell lines.

WMC1 and WMC2 independently present 7 common events for the three different temporal samples analyzed (different between WMC1 and WMC2, table 5). In accordance to the scheme of the clonal lineages (Figure 1), early samples do not present any unique fixed event, and only late and sub samples of WMC1 and WMC2 exhibit events with cs=1 that are specific to those samples due to accumulation of alterations after their common history ended. Accordingly, WMC1-early and WMC1-late exhibit fixed common events (15 of them) not present in WMC1-sub (table 6).

Results obtained from the analysis of the WMC4/5 pair show 16 common events and 6 of them for the WMC6/7 pair, all enlisted in the table 7. These results show a high degree of coincidence in the deletion and amplification events between cell lines that are different but share the parental origin.

Interestingly, the overall results show that 95% of the coincident events had a cs=1, and therefore, were completely extended in the cell populations. Of all the non-extended events (cs $<$1), a 98% did not overlap with any other event in the compared cell lines. In fact, only one coincident event between WMC2-early and WMC2-late had a cs<1 in both samples. Moreover, we did not observe any event that is partially extended in the early samples that becomes fully extended in the late samples.

-Gene-content analysis of the amplifications and deletions

We explored the gene content of all the events found to be coincident between samples (passage samples, sibling samples) and those detected in one sample simultaneously by SOMATICs and PennCNV. The gene content of all the events is enlisted in the table 8. A total of 57 % of the regions affected by deletions or amplifications contain coding sequences of known genes and/or miRNA, and a 58% totally or partially overlap a previously described CNV[53] event, implying a 42% of copy number variants in these cell lines at previously unreported genomic locations.

At least two coding regions involved in tumorigenesis are found in the altered regions of each set of cell lines analyzed (table 9). It is noteworthy that the coding regions for this type of genes found in the WMC1 and WMC2 cell lines are present in all the early samples in addition to the late or the sub samples. In these group of genes, two of them are genes encoding for proteins involved in xenobiotic metabolism (ARNT, FMO3), and four are genes encoding for components of the double-stranded DNA-break repair machinery (NBN, RFC4, UBE2I and PPP1R10). Detailed information

of these genes is contained in table 10). Remarkably, all the cell lines except WMC4 and WMC5 contain a deletion on the 3q26.1 region, that encodes the microRNA hsa-mir-720.

To determine if there is any bias in the biological processes represented by the genes comprised in the deletions and amplifications found, we investigated the Gene Ontology (GO) terms associated. A total of 1845 “biological process” different terms were present in the deletions/amplification events investigated. Table 11 lists the GO terms that are most significantly over-represented in the genes found. Cell signaling, transcriptional activity, cell cycle regulation and DNA repair are the biological processes with statistically significant over-representation in the set of genes affected by the genomic amplifications and deletions.

LOH Analysis

-Software comparison

The comparison of the LOH events found by LOH-Score and dChip (table 12) shows that the average total length of homozygous segments found by dChip (85,1Mb, 2,76% of the human genome) is higher than LOH score (31,4 Mb, 1.02% of the genome). A majority of the LOH segments found by LOH-score (average 86%) overlapped those found by dChip, while one third (average 32%) of the regions found by dChip were detected by LOH-Score.

-LOH-events overlap between samples

Coherence of LOH events detected in different passages and between sibling samples was investigated (table 13). For WMC1 and WMC2, both programs achieve a very high degree of event overlap [68-100%] between different passage samples, with one notable exception: WMC1-late overlap ratios with respect WMC1-early and WMC1-sub are lower than WMC2 [38.1-51.2%] due to the hemizygosity of the large 20p arm. Using these two programs, we lack of a measure of how extended these LOH regions are in the cell population. Nevertheless, agreeing with the time-line of clonal lineages of WMC1, only the sub and the late samples exhibit unique LOH-events. Specifically, for WMC1-late, we detect one “unique segment” where there is loss of heterozygosity: p-arm of chromosome 20. WMC2-late displays 7 unique LOH-events (4,7% of total), not present in the earlier passages.

Interestingly, the sibling cell lines exhibit total length values of LOH events higher than WMC1-early and WMC2-early samples. Moreover, the total LOH-event lengths for WMC4, WMC5, WMC6 and WMC7 are comparable to the total length of WMC1-late sample, which contains an hemizygous chromosome 20p arm. In addition, WMC4/5 and WMC6/7 display lower overlapping ratios between them than temporally different passages of WMC1 and WMC2: [34.9 – 47.3%] for WMC4/5 and [14.65-30.26%] for WMC6/7 (table 13). However, these overlapping figures are significantly higher than the average overlap found between any two non-related cell lines (4.93%).

-Gene-content analysis of LOH

We observed that 88% of the LOH-events detected contain either protein-coding or miRNA-coding regions (Supplementary Tables 14-17). We then investigated the biological processes in which those genes were involved using the GO terms in the same fashion as with the amplifications/deletions events. A total of 2759 biological processes terms were linked to the genes in the LOH events investigated. Table 18 lists the terms that are most significantly over-represented in the genes found. The first 4 positions are occupied by the same GO terms seen in the classification of the deletions/amplifications events: cell signaling, transcriptional activity, cell cycle regulation are over-represented in the set of genes in the LOH segments, as well as genes involved in apoptosis mechanisms.

We further investigated these LOH regions for genes known to be involved in tumorigenesis. From all LOH segments analyzed (287) nearly one third (29, 27%, listed on tables 14-17) contain a gene or a miRNA known to be involved in oncogenesis, and 8,01% contain a gene for which mutations are known to cause cancer (CGC database) (table 19). These latter genes were further examined for their molecular function (table 20). Remarkably among the genes investigated we found tumor suppressor genes (CBLC, BRCA1, BACH1, FANC, CDKN2C), oncogenes (CCNB1IP1, ETV4, CDK4, EPS15, ERBB2, BCL9, NOTCH2, CDC73). Interestingly, genes coding for essential components of the Polycomb machinery were also found in segments with haploinsufficiency (PHF19, SEMBT1, EZH1, CBX6, RNF2, RARA) as well as genes involved in chromosomal segregation during mitosis (KIF18A, TPR, HIST1H4L, CEP100) and hemimethylation during replication and p53 repressing activity (PCNA).

In addition, we observed that two homozygous regions in WMC6/7 sibling lines comprised the coding regions of 3 miRNA (let-7a-3, let-7b, let-7g) that are known to play an active role in

oncogenesis targeting proto-oncogene RAS transcripts.

Discussion

Here we have used standard cytogenetic banding and WG-SNP-CGH arrays to define the dynamics in the prevalence of numeric and subkaryotypic genomic aberrations of two human embryonic stem cell lines at three different time-points of their in vitro culture passaging history. We also evaluated the genomic integrity of two pairs of sibling hESC lines at their earliest passages searching for genomic coincidences and divergences in their genomes. These experiments allowed us to investigate chromosomal numerical aberrations, deletions and amplification of genomic segments, and importantly, the LOH events which previous reports on hESC genomic stability obviated.

Numeric aberrations

The karyotypes showed that the considered late passages of both cell lines had acquired one numerical chromosomal aberration each. WMC2, in particular, presented a trisomy 12, which has been previously described as a common alteration observed in hESC lines in culture6. WMC1 presented a dicentric isochromosome 20, involving loss of most of 20p and duplication of 20q. This specific aberration had not been described before. However, chromosome 20 had been implicated in chromosomal aberrations in hESC cultured in vitro[7,10]. Specially, the region 20p11.21, found in WMC2 to be the genomic spot where the WMC1 rearrangement occurred had been previously described recurrent hotspot of genomic aberrations in long-term cultured hESC lines.

Numeric aberrations of these two chromosomes (12 and 20) have led to direct comparison of aberrations in hESC to those found in testicular germ cell tumors (TGCT). Particularly, the gain of chromosome 12 (either in form of a trisomy or an isochromsome 12p) is so prevalent in TGCT that it can be used as a diagnostic marker of this form of malignancy[54]. This recurrent gain of chromosome 12 in TGCT and in extensively cultured hESC have been linked to the presence in the chromosome of several genes involved in self-renewal like NANOG, DPPA, GDF3, CCND2, the overexpression of which could provide cells with an advantage in culture[55,56]. On the other hand, numerical aberrations of chromosome 20 has also been reported to be involved in tumorigenesis: particularly, the alteration of 20q11.21 region is found in breast carcinomas[57,58], lung cancer[59], melanoma[60], hepatocellular carcinoma[61], bladder cancer[62] and early-stage cervical cancer[63]. All these data suggest that this region may contain genes whose amplification provides a proliferative

advantage, both for cancer progression and for hES cell growth[7,10].

Comparison of analysis software for WG-CGH arrays

Although WG-CGH arrays are widely used tools for genomic studies, there are multiple applications using different analytic approaches to detect genomic aberrations from the output files of the array readings. For that reason, our strategy was to combine the use of two methodologically different analysis programs for both deletions/amplifications analysis (PennCNV and SOMATICs) and for LOH detection (LOH-Score and dChip) in order to compare the detected events and to ensure a more consistent global analysis.

The main difference between PennCNV and SOMATICS is that the latter was designed to discriminate between genomic events present in all the cell population (cs =1) from those that are not fully extended in all the cells (cs $<$1). However, even if PennCNV does not provide that information, the fact that it detects events observed by SOMATICs with a cs=[0.25-1] proves that PennCNV finds events partially extended in the cell population. Nevertheless, using the same input files, we found a great differences in the event-overlapping ratios detected by SOMATICs and those detected by PennCNV [1-99%]. Thus, this great variation on the events found using the same information underscores the necessity for more robust analysis algorithms to study the chromosomal aberrations in genomically heterogeneous samples. To ensure a minimum strength in the results, we only considered for further genomic analysis the overlapping segments detected by both programs and those aberrations consistently by either SOMATICs or PennCNV in more than one temporal sample.

Contrarily, LOH-Score and dChip showed a higher degree of coincidence between programs than the analysis of CNVs. Since these two programs do not calculate the information on the extension of LOH-segments are, the default assumption made by the programs is that the haploinsufficiency stretches are present in the whole cell population. In this case, we only considered for subsequent genomic analyses those LOH-segments detected simultaneously by both programs.

Deletions, amplifications and LOH dynamics during hESC passaging history

The LOH segments detected by LOH-Score and dChip in addition to the deletions and amplifications detected using SOMATICs and PennCNV in the samples analyzed confirm the

hypothesis suggesting that hESC populations cultured *in vitro* present an heterogeneous genomic composition due to the random aberrations taking place in these highly proliferating cells.

A priori, a simple explanation for extension or the disappearance in the cell population of a new genomic aberration would be the differential selective advantage or disadvantage that it may confer to the cells. However, other important factors may influence this processes, such as the driver/passenger effect. This effect implies that any given aberrations that is advantageous (driver event) and becomes fixed may carry over other events present in the same cellular genome (passenger events) that get extended in the cell population as well. In addition, other external elements have been described to play an important role in these aberration extension dynamics too. Specifically, the drift effect imposed to the cellular population by the splitting process may influence decisively the population genomic dynamics. In other words, only the genomic variants present in the cells sub-population replated during the passaging process may be present in the global genome of the cell population. Moreover, previous reports suggest that manually passaged cell cultures retained normal karyotypes more efficiently than those passaged using other bulk methods (enzymatic and non enzymatic cell culture disaggregation techniques)[15,64,65]. These reports suggest that stress induced by the bulk methods may increase the underlying mutation rate, in addition to other selective pressures not exerted by the manual split (pure cell survival to the chemical or enzymatic agents). In addition, cell density in the culture dish may exercise too a selective pressure on cells. On one hand, high cell density conditions may produce a positive selection of those cells that can survive in lower concentration of nutrients. On the other hand, since maintenance of hESC self-renewal has been described to depend upon cellular signaling mediated by cell-cell contact, low cell-density plating may prime cells with self-renewal abilities that overcome this dependence[12]In our experiment, we passaged the hESC culture using enzymatic methods, but we were careful trying to avoid confluence of colonies in the culture dish by passaging every 4-5 days.

In the particular case of the detected deletions and amplifications in our analysis, the cellular score information given by SOMATICs applied to the samples obtained at different passages provides us with an idea of timeline of the alteration occurrence within a given cell line: the higher the cellular score, the more extended an event is at that specific time-point. Our analysis shows that during cell culture some deletions/amplifications became fixed (extended in the entire cell population) between the passages analyzed, implying that all the cells in the culture have acquired the aberration, while others disappeared from the cell population. In fact, all the extended aberrations in the early

samples were extended in the late samples too (specifically, 7 events for WMC1, and 7 events for WMC2), but the newly extended deletions/amplifications in the late or sub samples were not detected at all in the early samples, not even as partially extended events. Moreover, the great majority of non-extended events (cs<1) were non-coincident between passages. Due to sensibility threshold of SOMATICs (20% of cell population), it is difficult to state that the unique deletions/amplifications events found in the late and sub passages were not present in cell population at the early passage, but our results indicate a very rapid extension dynamics of deletions and amplifications in the cell population.

Interestingly, each of the two pairs of sibling cell lines also exhibited common deletions and amplifications: 16 for WMC4/5 and 6 for WMC6/7. From these coincidental 22 events, 18 were fully extended events in the cell populations, while only 4 displayed a cs<1 in at least one of the sibling cell lines comprising them. The fact that these 4 common events were not fully extended in at least one of the sibling cell populations may indicate a coincidental *de novo* appearance of these genomic aberrations. Hypothetically, the common and extended events in the early passages of WMC1 and WMC2 or in sibling cell lines could be either inherited from the parental germ-lines or originated between the derivation of the cell line and the temporal point of sample collection. In addition, for WMC1 and WMC2, we observed that late and sub passages exhibited unique extended events, but also common ones, not present in the early-passage samples underlining the possibility of both coincidental and differential aberrations acquired by different clonal lineages of the same cell line. In this case too, the sensibility threshold of the method detecting aberrations introduces the reasonable doubt of whether the unique or common aberration in the sub and late passages were already present in the early passage in a proportion of cells under the detection limits. However, evidences of coincident aberrations originated after the derivation of sibling cell lines may indicate that the simultaneous occurrence of deletions or amplifications in a same genomic locus after the cell lineages were separated is plausible and that this phenomenon may indicate the existence of genomic hotspots for the appearance of the aberrations that maybe dependent on the genetic background.

The LOH incidence found in all the cell lines analyzed reveals that this type of subkaryotypic abnormality is an important phenomena contributing to hESC lines instability during in vitro culture, which had not been quantified yet. Although we lack of a measure of how extended in the cell population these homozygous regions are, the consistency and reproducibility of the found events suggests that the great majority of the LOH events detected by both programs are extended

in the cell population. Interestingly, even though we found that new LOH events took place in culture between the early and late passages of the hESC analyzed,a great majority of these aberrations (between a 64-99% depending on the cell line) were already present in the earliest passages. In fact, the sibling cell line pairs, WMC4/WMC5 and WMC6/WMC7, that had been cultured for only 7 and 5 passages respectively, were the cell samples with the highest total length of LOH-segments. Interestingly, an average one third of these were common between sibling lines. A possible explanation for the early passages' common LOH-segments would be that they are inherited form the parental genomes. However, the high incidence of these events (ranging 1-2% of the genome in the hESC lines analyzed) and the fact that we detected new LOH events in a matter of a few passages in WMC1 and WMC2, may question the parental genomes origin of the shared acquired haploinsufficiency events in the WMC4/5 and WMC6/7 lines. A prospective study comparing these events in parental germlines and hESC lines after derivation would shed some light into this issue.

Gene-content analysis

Searching for the cause of the incidence of the deletions, amplifications and haploinsufficiency at the described genomic locations, we investigated the coding regions comprised in them. The fact that an 86% of the LOH stretches comprised totally or partially at least protein-coding and/or a miRNA-coding gene may give a hint. The gene-ontology analysis of these coding regions revealed that these LOH-segments were enriched in coding sequences of genes involved in signaling (especially, downstream of G-protein receptors), regulation of transcription, and importantly genes involved in cellular differentiation and proliferation, cell cycle regulation. When we further investigated the known biological function of the proteins and miRNA encoded by these genes we found confirmed oncogenes and tumor suppressor proteins (such as BCRA1, BACH1, CBLB, FANC, NOTCH2, CDC73) with the potential consequences of haploinsufficiency confers to these aberration events if the remaining allele is a pathogenic variant. Components of the polycomb machinery were also found in LOH stretches as well as a gene encoding for a member in the DNA hemimethylation complex. Remarkably, the LOH incidence in genes encoding for the polycomb machinery had not been described previously. The polycomb is formed by a group proteins that exert a transcriptional repression function that regulates lineage choices during development and differentiation. Recent studies have described how these proteins regulate cell fate decisions and how their deregulation potentially contributes to cancer[66,67] due to the deregulatory effect on tightly controlled balance between self-renewal and differentiation. In addition, induced hypomethylation

in murine ESC has been also linked to an elevated LOH rate[20]. Interestingly, PCNA is an essential component of the replication fork complex and it is directly involved in the hemimethylation of the nascent DNA chain during this process. Therefore, an impaired function of PCNA produced by an LOH event may induce an increment in LOH incidence itself.

We also observed a significant over-representation of genes involved biological processes that may influence the adaptability of the cells in culture, in the deletion/amplification events. Namely, signal transduction (genes encoding proteins of the downstream G-proteins coupled-receptors signaling pathway), regulators of transcription (specially transcription factors), kinases, genes involved in DNA damage response and DNA repair, as well as genes involved in the multicellular organism development and cell cycle. These findings draw a parallelism with the aberrations found in tumors[68]. In fact, each cell line analyzed had suffered at least two deletion/amplification events that included the coding region of a gene, the alteration of which is known to trigger oncogenic effects (table 3 and 4). Specifically, these particular events were present from the earliest passages analyzed and affected genes that are involved in xenobiotic metabolism and poignantly, genes involved in the maintenance of genomic stability.

miRNA coding regions were also affected by deletions/amplifications and LOH events. The recent realization of importance of miRNA deregulation during oncogenesis[69-72], underlines the observation that hsa-mir-720 was found to be deleted in four of the six cell lines analyzed. This miRNA is known to be downregulated in renal tumors[69]. A detailed analysis of the biological processes that this small RNA regulates will help to determine its implication on the adaptability of ESC lines in culture. Moreover, three members of the let-7 miRNA subfamily (let-7a-3, let-7b, let-7g) were found to be included in LOH segments in WMC7 cell line. Let-7 miRNAs are known to play an active role in oncogenesis targeting proto-oncogene RAS transcripts.

ESC derivation: testing the genomic stability

Here we have described that hESC accumulate new genomic aberrations between passages during extended in vitro culture. These events detected can be divided into copy number aberrations (deletions/amplifications) and acquired homozygosity. Interestingly, a majority of LOH events detected (65-99%) seem to occur previously to the first passage analyzed. Equally, if we only take into consideration those deletions and amplification events that are totally extended in the cell population, half of them are already present in the earliest passages. Lacking information of the

parental germline we hypothesize that a proportion of the detected aberrations at early passages may be have been inherited by the embryo from which they derive. However, the accumulation of new aberrations between passages, suggested that new aberrations may have occurred after derivation and previous to the first passage analyzed.

Interestingly, it had been previously reported that pluripotent embryonic cells are more prone to undergo LOH than somatic cells during extended in vitro culture[38,37,39,40]. *A priori*, LOH has a potentially a less disturbing effect than point mutations in the genome: actually, a number of homozygous segments have been found in WG-SNP array experiments exploring the genome in apparently healthy adult individuals[17]. These homozygous genomic segments are most probably product of UPD during embryogenesis, due to the homogeneous presence in different tissues tested. However, other WG-SNP arrays experiments matching tumors and normal samples of the same individuals have showed a high incidence and heterogeneity of acquired haploinsufficiency events in the cancerous tissue[73,74]. Therefore, the parallelism between the acquisition of LOH events during progressive adaptation of ESC to in vitro culture and the allelic imbalances observed during somatic cell transformation process is evident[16,75]. However, determining how haploinsufficiency is acquired by ESC during in vitro culture, and how the concomitant loss of heterozygosity contributes to hESC adaptability in addition to deletion/amplification events and aneuploidies is not obvious.

Since the ICM cells in a blastocyst give rise to all cell types in the body, some mutations occurring in them could have potential catastrophic consequences for the future organism. Hence, it is reasonable to expect that ICM cells should be more efficiently protected against genomic instability than somatic cells. Indeed, ESC show a lower mutation frequency that somatic cells[5]. However, mouse and primates embryonic pluripotent cells lack a G1 checkpoint, which is the step in the cell cycle when the repair the damaged DNA takes place in somatic cells[76]. Therefore, embryonic cells are proposed to undergo apoptosis when DNA damage is detected, instead of promoting DNA-repair. Indeed, it has been speculated that embryonic cells and stem cells in general rely heavily on the cellular DNA-damage response machinery to induce apoptosis, formed by ATM-Chk2-p53[77-80]. Thus, defects in base-excision repair, mismatch repair, homologous recombination or the replication fork complex impair stem cell function promoting accumulation of genomic aberrations[81,82]. Even if they may share the same mechanisms of DNA-damage response, there is a clear difference between ICM cells and ESC: when ICM cells are plated on the culture dish, these cells are enforced to maintain in vitro the same pluripotent state that otherwise would last a few hours during *in vivo* normal embryo development. Moreover, they are subjected to a selective pressure to survive and

proliferate in a totally different environment, to which they have to become adapted. In this new conditions, ESC are under stress for survival. Moreover, the use of undefined media (serum-replacement products) may have unknown effects in ESC, which are known to be extremely sensitive toxic elements. In this environment, double strand breaks may be product of the action of genotoxic agents or (chemical or physical) as well as, the replicative stress these cells are in. Therefore, some of the LOH-events may be product of the repair process by homologous recombination. Needless to say, that under these conditions there is a great selection pressure for loss-of-function mutations, or epigenetic silencing of factors within the activated DNA-damage repair (DDR) machinery, and clones of cells with such DDR defects likely emerge and may progress: the cells survive at the expense of enhanced genomic instability.

Indeed, the earliest passages of the hESC lines tested already contained aberrations that comprised genes encoding for members of the machinery involved in keeping the genomic stability and known to be involved in tumorigenesis: NBN (deletion), RFC4 (deletion), UBE2I (deletion), PPP1R10 (deletion), BCRA1 (LOH), FANC (LOH), BACH1 (LOH), CBLB (LOH), NOTCH2 (LOH), CDC73 (LOH). The temporal timeline and the dynamics of the appearance and extension of these events at these early passages is unknown, but some of them are coincidental between different cell lines (table 9 and table 19). Bypassing the DNA-repair machinery, these cells promote that the rest of aleatory genomic aberrations that may affect characteristics susceptible to be altered in order to enhance survival in such conditions will be favorably selected. In a similar way, the aberrations affecting genes encoding for proteins involved in the xenobiotic-metabolism response (ARNT and FMO3) may have been selected due to the possible increase in the chance of genomic aberrations they probably induce due to the lack of activation of DNA protection mechanism in front of toxic agents.

Nevertheless, more than ten years after the first human derivation experiments, efficiencies of hESC derivation continue to be stubbornly low[83]. This fact reflects the inability of a big number of ICMs to pass the survival test that is imposed by the standard ESC derivation conditions. Strikingly, we derived two pairs of sibling hESC lines from only two available embryos (with the same parental origin) in each case. This high frequency of sibling hESC lines has been previously reported by other groups[83]. This divergence in the derivation efficiency from morphologically similar embryos emphasizes the probable existence of intrinsic genetic factors in every embryo that may condition their capacity to be derived into an cell line.

The observation of this data and our results leads us to hypothesize that under the influence of a certain genetic background, the degree of homozygosity may be a factor favoring the embryo derivability. In other words, in the case of ICM cells derived into ESC line, haploinsufficiency in certain key loci with the right allele, either inherited from parental genomes or product of LOH, may promote self-renewal potential ability in order to survive the derivation process. In other words, if different alleles of specific locus have different effects on the "adaptability" of ICM cells to the specific ESC derivation conditions, homozygosity of the most favorable allele for the survival of the cells in such environment would help the derivation process. Therefore, in embryos containing a favorable allele at heterozygous locus, the selection pressure would prime LOH events at those genomic locations and a concomitant higher capacity of being derived into a stem cell line and survive in the specific culture conditions.

This hypothesis would be backed by two facts. The first, the reported increased frequency of germline homozygosity at certain loci in individuals with high incidence of carcinomas compared with ancestry-matched controls[33]. Interestingly, a correlation has been found between and increased acquired LOH events in spontaneous tumors at the same "homozygosity hotspots" genomic locations, when originally the germline was heterozygous[33]. The second, the variable differences in ESC derivation rates in different mouse depending on the strains. All inbred mouse strains are characterized by an average 98.6% homozygosity rate in their genomes[84]. However, some strains have notoriously been reported as non-permissive to derivation of hESC[85]. Contrarily, 129 mouse strain has the highest derivability rate (near 100%[86]) and, like most of inbred strains[87]a high tumor incidence as well. Specifically, an average 8% of male individuals from 129 strain suffer from spontaneous TGCT [88-90] which had been previously reported to undergo strikingly similar chromosomal aberrations as hESC in long-term culture[6]. Interestingly, the substitution in 129 strain mice of chromosome 18 for the same chromosome of another strain (MOLF) has been recently reported to promote a decrease in the TGCT incidence that was concomitant to a drastic reduction of the mESC derivation rate[91].Therefore, a high incidence of homozygosity in specific genomic loci in this mice strain correlates with both a high derivation rate and a high incidence of spontaneous germ cell tumors.

Concluding remarks

Here we have showed that hESC culture in vitro in standard conditions promotes the accumulation of genomic alterations involving deletions, amplifications and acquired haploinsufficiency. The

derivation process itself could expose the embryonic ICM cells to a strong selective pressures to become adapted to the new culture conditions. And in spite of the fact that the cellular passaging routine leads to a genomic drift effect every time a group of cells is passaged to continue the culture, we have observed a great deal of coincidence in the type of genes affected by the most successful aberrations (those that get fixed in the entire cell population). This forces would prime those cells that promote aberrations affecting DNA-damage repair machinery, because the concomitant genomic instability would make adaptability easier for those cells containing them.

Bibliography

1. Thomson, J.A. et al. Embryonic stem cell lines derived from human blastocysts. *Science (New York, N.Y* **282**, 1145-7 (1998).
2. Daley, G.Q. & Scadden, D.T. Prospects for Stem Cell-Based Therapy. *Cell* **132**, 544-548 (2008).
3. Deb, K.D. & Sarda, K. Human embryonic stem cells: preclinical perspectives. *J Transl Med* **6**, (2008).
4. Carpenter, M.K., Frey-Vasconcells, J. & Rao, M.S. Developing safe therapies from human pluripotent stem cells. *Nat. Biotechnol* **27**, 606-613 (2009).
5. Cervantes, R.B., Stringer, J.R., Shao, C., Tischfield, J.A. & Stambrook, P.J. Embryonic stem cells and somatic cells differ in mutation frequency and type. *Proceedings of the National Academy of Sciences* **99**, 3586 (2002).
6. Baker, D.E.C. et al. Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. *Nat Biotechnol* **25**, 207-215 (2007).
7. Spits, C. et al. Recurrent chromosomal abnormalities in human embryonic stem cells. *Nat Biotechnol* **26**, 1361-1363 (2008).
8. Hanson, C. & Caisander, G. Human embryonic stem cells and chromosome stability. *Apmis* **113**, 751-5 (2005).
9. Imreh, M. et al. In vitro culture conditions favoring selection of chromosomal abnormalities in human ES cells. *J. Cell. Biochem.* **99**, 508-516 (2006).
10. Lefort, N. et al. Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. *Nat Biotechnol* **26**, 1364-1366 (2008).
11. Werbowetski-Ogilvie, T.E. et al. Characterization of human embryonic stem cells with features of neoplastic progression. *Nat Biotechnol* **27**, 91-97 (2009).
12. Enver, T. et al. Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. *Hum. Mol. Genet* **14**, 3129-3140 (2005).
13. Herszfeld, D. et al. CD30 is a survival factor and a biomarker for transformed human pluripotent stem cells. *Nat. Biotechnol* **24**, 351-357 (2006).
14. Andrews, P.W. et al. Embryonic stem (ES) cells and embryonal carcinoma (EC) cells: opposite sides of the same coin. *Biochem. Soc. Trans* **33**, 1526-1530 (2005).
15. Maitra, A. et al. Genomic alterations in cultured human embryonic stem cells. *Nat Genet* **37**, 1099-1103 (2005).
16. Tuna, M., Knuutila, S. & Mills, G.B. Uniparental disomy in cancer. *Trends in Molecular Medicine* **15**, 120-128 (2009).
17. Simon-Sanchez, J. et al. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Human molecular genetics* **16**, 1-14 (2007).
18. Shrivastav, M., De Haro, L.P. & Nickoloff, J.A. Regulation of DNA double-strand break repair pathway choice. *Cell Res* **18**, 134-147 (2008).
19. Helleday, T. Pathways for mitotic homologous recombination in mammalian cells. *Mutat. Res* **532**, 103-115 (2003).
20. Helleday, T., Lo, J., van Gent, D.C. & Engelward, B.P. DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA Repair (Amst.)* **6**, 923-935 (2007).
21. Robinson, W.P. et al. Somatic segregation errors predominantly contribute to the gain or loss of a paternal chromosome leading to uniparental disomy for chromosome 15. *Clin. Genet* **57**, 349-358 (2000).
22. Robinson, W.P. Mechanisms leading to uniparental disomy and their clinical consequences. *Bioessays* **22**, 452-459 (2000).
23. Engel, E. A fascination with chromosome rescue in uniparental disomy: Mendelian recessive outlaws and imprinting copyrights infringements. *Eur J Hum Genet* **14**, 1158-1169 (2006).

24. Engel, E. Uniparental disomy revisited: the first twelve years. *Am. J. Med. Genet* **46**, 670-674 (1993).
25. Smilenov, L.B. Tumor development: haploinsufficiency and local network assembly. *Cancer Lett* **240**, 17-28 (2006).
26. Luo, G. et al. Cancer predisposition caused by elevated mitotic recombination in Bloom mice. *Nat. Genet* **26**, 424-429 (2000).
27. Karanjawala, Z.E. et al. Complete maternal isodisomy of chromosome 8 in an individual with an early-onset ileal carcinoid tumor. *Am. J. Med. Genet* **93**, 207-210 (2000).
28. Murthy, S.K., DiFrancesco, L.M., Ogilvie, R.T. & Demetrick, D.J. Loss of Heterozygosity Associated with Uniparental Disomy in Breast Carcinoma. *Mod Pathol* **15**, 1241-1250 (2002).
29. Yip, L. et al. Loss of heterozygosity of selected tumor suppressor genes in parathyroid carcinoma. *Surgery* **144**, 949-955; discussion 954-955 (2008).
30. Santarosa, M. & Ashworth, A. Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way. *Biochim. Biophys. Acta* **1654**, 105-122 (2004).
31. Henry, I. et al. Uniparental paternal disomy in a genetic cancer-predisposing syndrome. *Nature* **351**, 665-667 (1991).
32. Bignell, G.R. et al. Signatures of mutation and selection in the cancer genome. *Nature* **463**, 893-898 (2010).
33. Assie, G., LaFramboise, T., Platzer, P. & Eng, C. Frequency of germline genomic homozygosity associated with cancer cases. *Jama* **299**, 1437 (2008).
34. Soh, J. et al. Oncogene mutations, copy number gains and mutant allele specific imbalance (MASI) frequently occur together in tumor cells. *PLoS ONE* **4**, e7464 (2009).
35. Tiu, R.V. et al. New lesions detected by single nucleotide polymorphism array-based chromosomal analysis have important clinical impact in acute myeloid leukemia. *J. Clin. Oncol* **27**, 5219-5226 (2009).
36. Gondek, L.P., Dunbar, A.J., Szpurka, H., McDevitt, M.A. & Maciejewski, J.P. SNP array karyotyping allows for the detection of uniparental disomy and cryptic chromosomal abnormalities in MDS/MPD-U and MPD. *PLoS ONE* **2**, e1225 (2007).
37. Mortensen, R.M., Conner, D.A., Chao, S., Geisterfer-Lowrance, A.A. & Seidman, J.G. Production of homozygous mutant ES cells with a single targeting construct. *Mol. Cell. Biol* **12**, 2391-2395 (1992).
38. Lefebvre, L., Dionne, N., Karaskova, J., Squire, J.A. & Nagy, A. Selection for transgene homozygosity in embryonic stem cells results in extensive loss of heterozygosity. *Nat. Genet* **27**, 257-258 (2001).
39. Van Sloun, P.P. et al. Determination of spontaneous loss of heterozygosity mutations in Aprt heterozygous mice. *Nucleic Acids Res* **26**, 4888-4894 (1998).
40. Stark, J.M. & Jasin, M. Extensive loss of heterozygosity is suppressed during homologous repair of chromosomal breaks. *Molecular and Cellular Biology* **23**, 733 (2003).
41. Chan, M.F. et al. Reduced rates of gene loss, gene silencing, and gene mutation in Dnmt1-deficient embryonic stem cells. *Mol. Cell. Biol* **21**, 7587-7600 (2001).
42. Yusa, K. et al. Genome-wide phenotype analysis in ES cells by regulated disruption of Bloom's syndrome gene. *Nature* **429**, 896-899 (2004).
43. Bielas, J.H., Venkatesan, R.N. & Loeb, L.A. LOH-proficient embryonic stem cells: a model of cancer progenitor cells? *Trends in Genetics* **23**, 154–157 (2007).
44. Donahue, S.L., Lin, Q., Cao, S. & Ruley, H.E. Carcinogens induce genome-wide loss of heterozygosity in normal stem cells without persistent chromosomal instability. *Proceedings of the National Academy of Sciences* **103**, 11642 (2006).
45. Cowan, C.A. et al. Derivation of embryonic stem-cell lines from human blastocysts. *N. Engl. J. Med* **350**, 1353-1356 (2004).
46. Peiffer, D.A. High-resolution genomic profiling of chromosomal aberrations using Infinium

whole-genome genotyping. *Genome Research* **16**, 1136-1148 (2006).
47. Assié, G. et al. SNP arrays in heterogeneous tissue: highly accurate collection of both germline and somatic genetic information from unpaired single tumor samples. *Am. J. Hum. Genet* **82**, 903-915 (2008).
48. Wang, K. et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Research* **17**, 1665-1674 (2007).
49. Lin, M. et al. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* **20**, 1233-1240 (2004).
50. Futreal, P.A. et al. A census of human cancer genes. *Nat Rev Cancer* **4**, 177-183 (2004).
51. Forbes, S.A. et al. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer. *Nucl. Acids Res.* **38**, D652-657 (2010).
52. Forbes, S.A. et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet* **Chapter 10**, Unit 10.11 (2008).
53. Conrad, D.F. et al. Origins and functional impact of copy number variation in the human genome. *Nature* (2009).doi:10.1038/nature08516
54. Reuter, V.E. Origins and molecular biology of testicular germ cell tumors. *Mod. Pathol* **18 Suppl 2**, S51-60 (2005).
55. Korkola, J.E. et al. Down-regulation of stem cell genes, including those in a 200-kb gene cluster at 12p13.31, is associated with in vivo differentiation of human male germ cell tumors. *Cancer Res* **66**, 820-827 (2006).
56. Skotheim, R.I. et al. Novel genomic aberrations in testicular germ cell tumors by array-CGH, and associated gene expression changes. *Cell. Oncol* **28**, 315-326 (2006).
57. Guan, X.Y. et al. Hybrid selection of transcribed sequences from microdissected DNA: isolation of genes within amplified region at 20q11-q13.2 in breast cancer. *Cancer Res* **56**, 3446-3450 (1996).
58. Tanner, M.M. et al. Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. *Cancer Res* **56**, 3441-3445 (1996).
59. Tonon, G. et al. High-resolution genomic profiles of human lung cancer. *Proc. Natl. Acad. Sci. U.S.A* **102**, 9625-9630 (2005).
60. Koynova, D.K. et al. Gene-specific fluorescence in-situ hybridization analysis on tissue microarray to refine the region of chromosome 20q amplification in melanoma. *Melanoma Res* **17**, 37-41 (2007).
61. Midorikawa, Y. et al. Molecular karyotyping of human hepatocellular carcinoma using single-nucleotide polymorphism arrays. *Oncogene* **25**, 5581-5590 (2006).
62. Hurst, C.D. et al. High-resolution analysis of genomic copy number alterations in bladder cancer by microarray-based comparative genomic hybridization. *Oncogene* **23**, 2250-2263 (2004).
63. Scotto, L. et al. Identification of copy number gain and overexpressed genes on chromosome arm 20q by an integrative genomic approach in cervical cancer: potential role in progression. *Genes Chromosomes Cancer* **47**, 755-765 (2008).
64. Brimble, S.N. et al. Karyotypic stability, genotyping, differentiation, feeder-free maintenance, and gene expression sampling in three human embryonic stem cell lines derived prior to August 9, 2001. *Stem Cells Dev* **13**, 585-597 (2004).
65. Mitalipova, M.M. et al. Preserving the genetic integrity of human embryonic stem cells. *Nat. Biotechnol* **23**, 19-20 (2005).
66. Simon, J.A. & Kingston, R.E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol* **10**, 697-708 (2009).
67. Bracken, A.P. & Helin, K. Polycomb group proteins: navigators of lineage pathways led astray

in cancer. *Nat. Rev. Cancer* **9**, 773-784 (2009).
68. Furney, S.J., Higgins, D.G., Ouzounis, C.A. & López-Bigas, N. Structural and functional properties of genes involved in human cancer. *BMC Genomics* **7**, 3 (2006).
69. Yi, Z., Fu, Y., Zhao, S., Zhang, X. & Ma, C. Differential expression of miRNA patterns in renal cell carcinoma and nontumorous tissues. *J Cancer Res Clin Oncol* (2009).doi:10.1007/s00432-009-0726-x
70. McCarthy, N. Cancer: Small losses, big gains with microRNAs. *Nat Rev Genet* **11**, 8-8 (2009).
71. Afanasyeva, E.A., Hotz-Wagenblatt, A., Glatting, K.H. & Westermann, F. New miRNAs cloned from neuroblastoma. *BMC genomics* **9**, 52 (2008).
72. Cummins, J.M. et al. The colorectal microRNAome. *Proc. Natl. Acad. Sci. U.S.A* **103**, 3687-3692 (2006).
73. Fujii, H., Marsh, C., Cairns, P., Sidransky, D. & Gabrielson, E. Genetic divergence in the clonal evolution of breast cancer. *Cancer Res* **56**, 1493-1497 (1996).
74. Wang, Z.C. et al. Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. *Cancer Res* **64**, 64-71 (2004).
75. Chari, R. et al. Integrating the multiple dimensions of genomic and epigenomic landscapes of cancer. *Cancer Metastasis Rev* **29**, 73-93 (2010).
76. Hong, Y., Cervantes, R.B., Tichy, E., Tischfield, J.A. & Stambrook, P.J. Protecting genomic integrity in somatic cells and embryonic stem cells. *Mutat. Res* **614**, 48-55 (2007).
77. Jackson, S.P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071-1078 (2009).
78. Branzei, D. & Foiani, M. Maintaining genome stability at the replication fork. *Nat. Rev. Mol. Cell Biol* **11**, 208-219 (2010).
79. Riches, L.C., Lynch, A.M. & Gooderham, N.J. Early events in the mammalian response to DNA double-strand breaks. *Mutagenesis* **23**, 331-339 (2008).
80. Bartkova, J. et al. DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* **434**, 864-870 (2005).
81. Park, Y. & Gerson, S.L. DNA repair defects in stem cell function and aging. *Annu. Rev. Med* **56**, 495-508 (2005).
82. Rossi, D.J. et al. Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. *Nature* **447**, 725-729 (2007).
83. Chen, A.E. et al. Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell stem cell* **4**, 103-6 (2009).
84. Beck, J.A. et al. Genealogies of mouse inbred strains. *Nat. Genet* **24**, 23-25 (2000).
85. McWhir, J. et al. Selective ablation of differentiated cells permits isolation of embryonic stem cell lines from murine embryos with a non-permissive genetic background. *Nat Genet* **14**, 223-226 (1996).
86. Brook, F.A. & Gardner, R.L. The origin and efficient derivation of embryonic stem cells in the mouse. *Proc. Natl. Acad. Sci. U.S.A* **94**, 5709-5712 (1997).
87. Naf, D., Krupke, D.M., Sundberg, J.P., Eppig, J.T. & Bult, C.J. The Mouse Tumor Biology Database: A Public Resource for Cancer Genetics and Pathology of the Mouse. *Cancer Res* **62**, 1235-1240 (2002).
88. STEVENS, L.C. & HUMMEL, K.P. A description of spontaneous congenital testicular teratomas in strain 129 mice. *J. Natl. Cancer Inst* **18**, 719-747 (1957).
89. Stevens, L.C. Spontaneous and experimentally induced testicular teratomas in mice. *Cell Differ* **15**, 69-74 (1984).
90. Stevens, L.C. & Little, C.C. Spontaneous Testicular Teratomas in an Inbred Strain of Mice. *Proc. Natl. Acad. Sci. U.S.A* **40**, 1080-1087 (1954).
91. Anderson, P.D., Nelson, V.R., Tesar, P.J. & Nadeau, J.H. Genetic factors on mouse chromosome

18 affecting susceptibility to testicular germ cell tumors and permissiveness to embryonic stem cell derivation. *Cancer Res* **69**, 9112-9117 (2009).
92. Lee, S.H., Sterling, H., Burlingame, A. & McCormick, F. Tpr directly binds to Mad1 and Mad2 and is important for the Mad1-Mad2-mediated mitotic spindle checkpoint. *Genes Dev* **22**, 2926-2931 (2008).
93. Walczak, C.E., Cai, S. & Khodjakov, A. Mechanisms of chromosome behaviour during mitosis. *Nat Rev Mol Cell Biol* (2010).doi:10.1038/nrm2832
94. Menendez, D., Inga, A. & Resnick, M.A. The expanding universe of p53 targets. *Nat. Rev. Cancer* **9**, 724-737 (2009).
95. Ohm, J.E. et al. A stem cell–like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet* **39**, 237-242 (2007).
96. Sanada, M. et al. Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. *Nature* **460**, 904-908 (2009).
97. Huen, M.S., Sy, S.M. & Chen, J. BRCA1 and its toolbox for the maintenance of genome integrity. *Nat Rev Mol Cell Biol* (2009).doi:10.1038/nrm2831

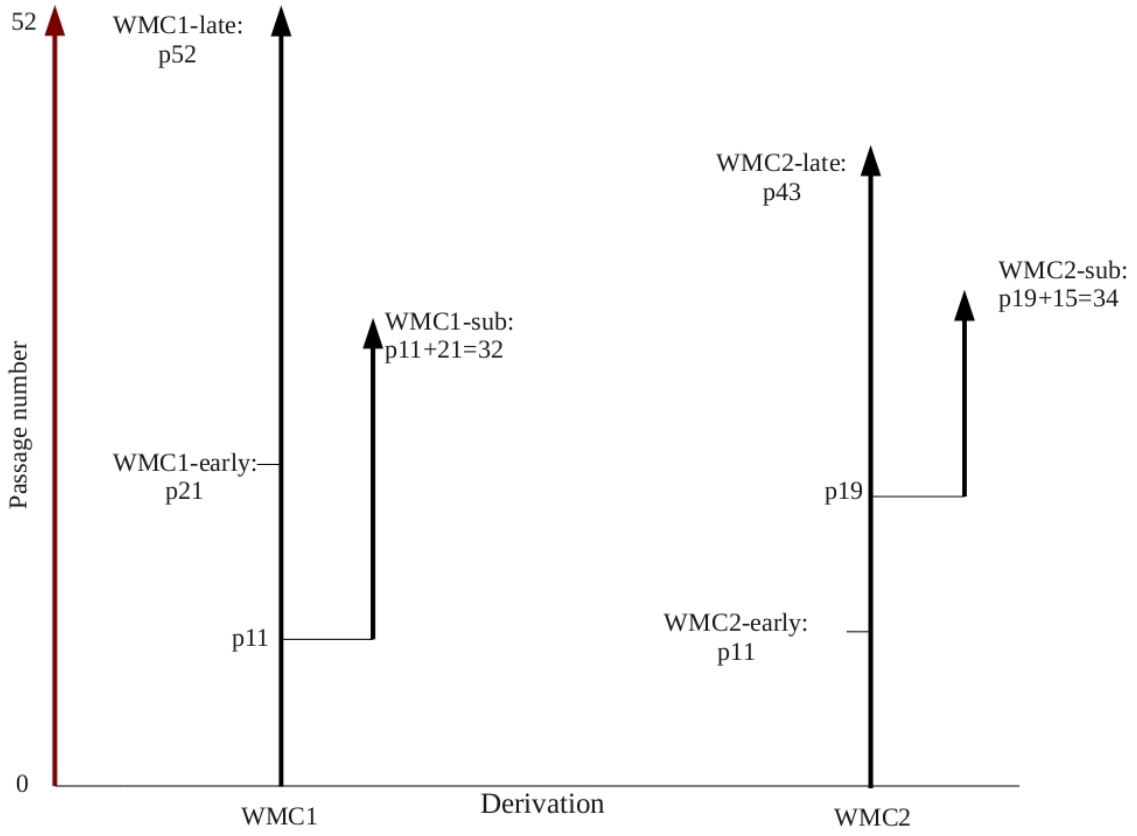


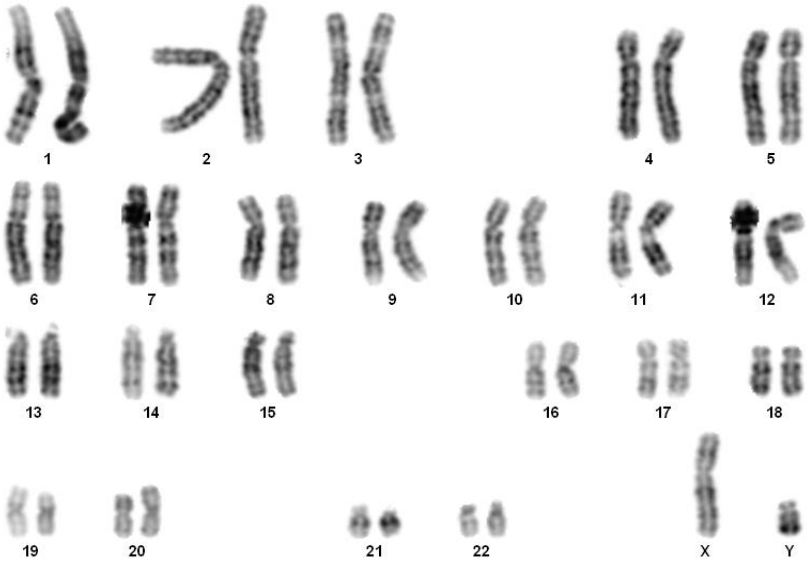
Figure 1. Schematic view of the different cell samples analyzed in the context of the clonal lineages from which they derive. WMC1 early is obtained at passage 21 from the main clonal lineage, and late sample is obtained at passage 52. WMC1-sub, is originated by an new lineage initiated from passage 11 from the main one, which was independently cultured for further 21 passages, being 32 the total number of splits that have originated this sample. As for the WMC2 cell line, WMC2-sub sample is 34 passages old and originated from a subline split oft he main clonal lineage at passage 19. In contrast to WMC1-sub, the WMC2-sub clonal lineage was initiated temporally later than the WMC2-early (passage 11) sample.

Cell line	Time point	Passage number	Karyotype	Extensions
WMC1	Early	21	46 XY	22/22
	Sub-passage	11+21 (32)	46 XY	23/24
	Late	52	46 XY idic(20)(p11)	22/28
WMC2	Early	11	46 XY	20/20
	Sub-passage	19+15 (34)	46 XY	22/22
	Late	43	47 XY,+12	26/30
WMC4	Early	7	46 XY	23/23
WMC5	Early	7	46 XY	22/23
WMC6	Early	5	46 XY	20/20
WMC7	Early	5	46 XY	20/20

Table 1. Description of the samples analyzed and result of the cytogenetic analysis. Three samples of the WMC1 and WMC2 cell lines were analyzed. We intentionally chose a WMC1-sub sample that had been separated from the main clonal lineage previously to the early passage and a WMC2-sub that was separated from the main lineage between early and late passages to contrast findings. All the samples had a normal karyotype except for WMC1-late and WMC2-late samples. For WMC1-late, 76% of the extensions presented a chromosome 20 isodisomy, while 86% of the WMC2-late extensions exhibited a chromosome 12 trisomy.

WMC1-late: 46,XY,idic(20)(p11)

A



WMC2-late : 47,XY,+12

B

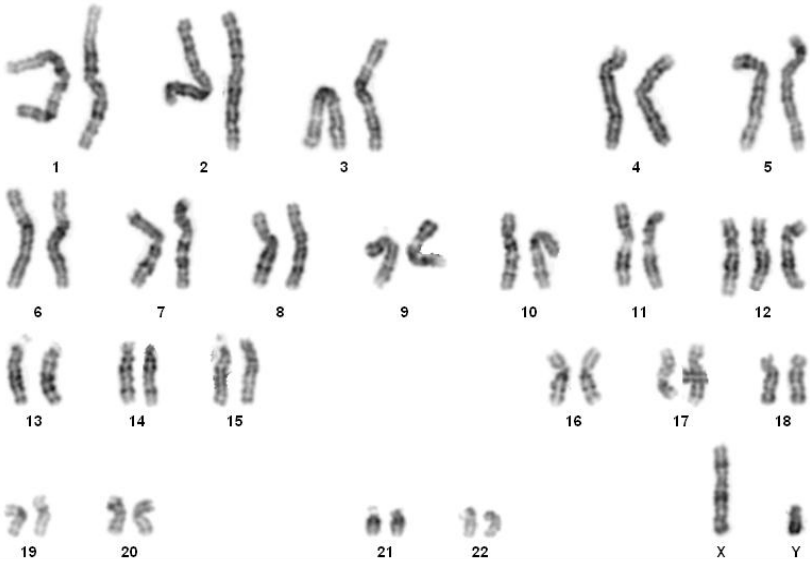


Figure 2. Karyotypic of analyses of the WMC1-late and WMC2-late samples. The ideograms show the isodisomy 20 for WMC1-late and the trisomy 12 for WMC2-late.

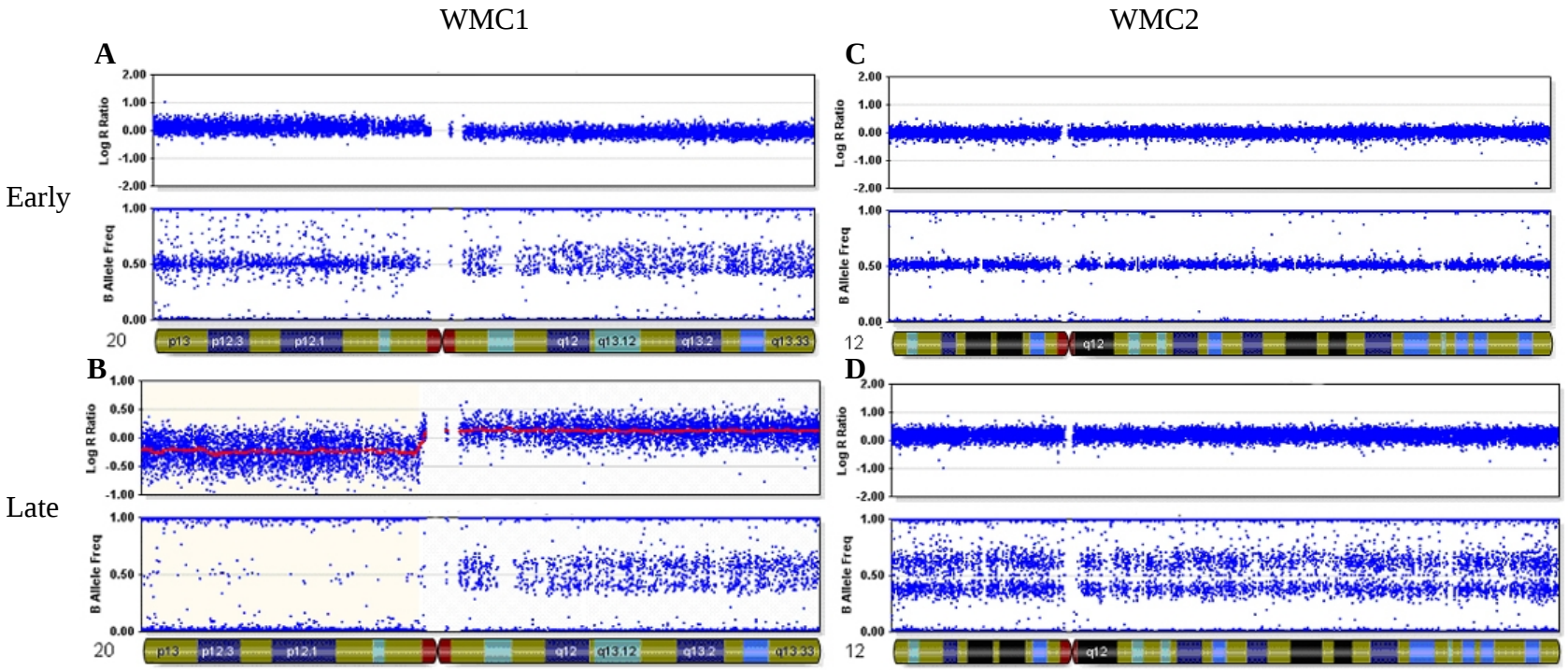


Figure 3. Genomic plots of WMC1 and WMC2 lines, at early and late passages. (A) WMC-1 early passage average Log_2R ratio of the 20p arm (11799-25645042) is 0.101, which is above the average euploid value (0.0041±0.032), implying an amplification until the 20p11.21 band in a undefined proportion of the cell population. Average Log_2R ratio value for 20q, -0.092 is under the value of the other cell lines (-0.016±0.012), indicating an undetermined degree of hemizygosity. However, BAF plot at 20q begins to show a separation in two clusters of the different SNPs, which is a contradictory to the Log_2R value. This BAF split is originated by the heterozygous SNPs due to the change in the proportion of intensities: one allele will tend to be in a 1-to-3 ratio and the others in a 2-to-3 ratio. (B) In the WMC1-late hemizygosity of the p-arm (11799-25645042) is reflected in the Log_2R value of -0.2476, and in lack of heterozygous SNPs in the BAF plot . Duplication of the q arm is visualized in an increase in the Log_2R ratio +0.1173, plus the evident split of the heterozygous state into two clusters in the BAF plot. (C) Chromosome 12 plot for WMC2-early sample, shows the average euploid plots for any chromosome. Log_2R value is 0.011 and distribution of heterozygous and homozygous SNPs in the BAF plot does not indicate any aberration. (D) Plot of the trisomic 12 chromosome on WMC2-late sample. Log_2R ratio is +0.1667 and there is a clear split in the heterozygous state in two differentiated clusters. The Log_2R value indicates that trisomy is not extended in whole cell population: Peiffer *et al.*[1] suggest a Log_2R ratio increase to +0.395 for an homogeneous trisomy in the entire cell population.

	SOMATICS					PENNCNV					Overlap SOMATICS	Overlap PennCNV
	# Events	Min. Len. (Kb)	Max. Len. (Kb)	Av. Len. (Kb)	Total Len. (Kb)	# Events	Min. Len. (Kb)	Max. Len. (Kb)	Av. Len. (Kb)	Total Len. (Kb)		
WMC1-early	49	0.010	3273.929	259.800	12730.26	41	0.806	128.169	44.140	1809.91	3.03%	21.32%
WMC1-sub	20	0.010	12623.111	259.800	14648.63	14	0.806	155.230	23.750	332.58	0.10%	0.10%
WMC1-late	32	0.010	32867.780	19335.348	61931.17	131	0.002	1277.660	145.179	19018.48	30.62%	99.70%
WMC2-early	23	0.040	1043.471	96.782	2225.99	15	0.214	170.220	31.195	467.93	8.70%	41.40%
WmC2-sub	21	0.040	1043.471	123.811	2252.9	14	0.214	163.600	28.760	402.75	1.84%	10.30%
WMC2-late	43	0.040	65392.440	3240.256	139331.01	308	0.214	1843.240	245.930	78520.99	56.05%	99.45%
WMC4	69	0.060	505.492	94.041	6488.86	77	0.177	324.131	31.490	2425.07	5.96%	15.96%
WMC5	74	0.177	592.071	81.411	6024.47	55	0.214	245.000	39.388	2166.34	0.83%	2.30%
WMC6	53	0.040	747.220	98.779	5235.32	81	0.214	304.621	35.354	2863.7	22.04%	40.30%
WMC7	38	0.04	747.220	98.78	4118.55	75	0.218	263.357	32.326	2424.51	9.69%	16.45%

Table 2. Comparison between the events found by SOMATICs and PennCN for all the samples. All lengths in Kb. The overlap percentages are referred to the total length of aberrations found in each program.

Sample	SOMATICs		PennCNV	
	AMPL	DELS	AMPL	DELS
WMC1-early	10	47	38	3
WMC1-sub	3	42	5	9
WMC1-late	5	32	51	80
WMC2-early	3	25	3	12
WMC2-sub	3	23	4	10
WMC2-late	10	44	292	16
WMC4	55	29	66	11
WMC5	51	31	48	7
WMC6	41	15	61	20
WMC7	20	44	52	23

Table 3. Comparison of total amplification and deletion events detected by SOMATICs and PennCNV for each sample.

	SOMATICS		PennCNV		SOMATICS		PennCNV		SOMATICS		PennCNV	
	WMC1-early	WMC1-late	WMC1-early	WMC1-late	WMC1-early	WMC1-sub	WMC1-early	WMC1-sub	WMC1-sub	WMC1-late	WMC1-sub	WMC1-late
# Events	49	32	41	131	49	20	41	14	20	32	14	131
# Coincident Events	42	27	31	19	10	8	1	1	8	11	1	1
Total Event length (Kb)	12730.26	61931.17	1809.91	19018.48	12730.26	14648.63	1809.91	332.58	14648.63	61931.17	14648.63	19018.48
Overlapping length (Kb)	7959.08		1266.35		288.25		1		12480.08		3.19	
Overlapping rate %	62.52	12.85	69.96	6.65	2.22	1.93	0.04	0.02	85.19	20.15	0.95	0.01

	SOMATICS		PennCNV		SOMATICS		PennCNV		SOMATICS		PennCNV	
	WMC2-early	WMC2-late	WMC2-early	WMC2-late	WMC2-early	WMC2-sub	WMC2-early	WMC2-sub	WMC2-sub	WMC2-late	WMC2-sub	WMC2-late
# Events	23	43	15	308	23	21	15	14	21	43	14	308
# Coincident Events	15	19	7	7	16	16	11	11	13	18	7	7
Total Event length (Kb)	2225.99	139331.01	467.93	78520.99	2225.99	2252.9	467.93	402.75	2252.9	139331.01	402.75	78520.99
Overlapping length (Kb)	1269.61		121.54		306.34		339.65		228.07		128.74	
Overlapping rate %	57.03	0.91	25.97	0.15	13.76	13.6	72.58	84.33	10.12	0.16	31.96	0.16

	WMC6	WMC7	WMC6	WMC7
# Events	53	38	81	75
# Coincident Events	7	6	9	9
Total Event length (Kb)	5235.32	4118.55	2863.7	2424.51
Overlapping length (Kb)	55.99		177.39	
Overlapping rate %	1.06	1.35	6.2	7.32

	WMC4	WMC5	WMC4	WMC5
# Events	69	74	77	55
# Coincident Events	17	18	4	4
Total Event length (Kb)	6488.86	6024.47	2425.07	2166.34
Overlapping length (Kb)	817.81		224.28	
Overlapping rate %	12.6	13.57	10.07	11.28

Table 4. Coincident CNV events between related cell samples. **A)** General comparison of the coincident alterations between temporally different samples of the same cell line (WMC1 and WMC2, early, sub and late passages), and the sibling cell lines (WMC4 / WMC5 and WMC6/WMC7). SOMATICs overlapping rate are generally higher than PennCNV's, specially taking the total event length into account.

Cell sample	Event #	Chr	Start postiion	End Position	Score	CNV
WMC1	1	chr2	159667833	159668238	1	deletion
WMC1	2	chr3	164004033	164101579	1	deletion
WMC1	3	chr3	191220845	191221392	1	deletion
WMC1	4	chr4	19130834	19131054	1	deletion
WMC1	5	chr4	68905652	69515619	1	deletion
WMC1	6	chr7	81279747	81280187	1	deletion
WMC1	7	chr10	62386724	62386725	1	deletion
WMC2	1	chr3	164046708	164123191	1	deletion
WMC2	2	chr4	173226303	173226559	1	deletion
WMC2	3	chr5	10327112	10327399	1	deletion
WMC2	4	chr8	586317	588391	1	deletion
WMC2	5	chr8	38932203	39697120	1	amplification
WMC2	6	chr8	281394	281400	1	deletion
WMC2	7	chr9	89564857	89564863	1	deletion

Table 5 WCM1 and WMC2 coincident deletion/amplification events in the three temporal samples analyzed using SOMATICs. No coincident event with a cs <1 is found in this list.

	GO id	Biological process	# of genes	$\chi 2$
1	GO:0007165	signal transduction	165	24,81
2	GO:0006355	regulation of transcription, DNA-dependent	130	18,26
3	GO:0007186	G-protein coupled receptor protein signaling pathway	100	29,49
4	GO:0045449	regulation of transcription	100	14,42
5	GO:0007275	multicellular organismal development	79	9,18
6	GO:0008152	metabolic process	64	6,96
7	GO:0006810	transport	61	10,63
8	GO:0006811	ion transport	58	0,18
9	GO:0006508	proteolysis	55	1,1
10	GO:0006468	protein amino acid phosphorylation	52	5,01
16	GO:0007049	cell cycle	38	4,93
19	GO:0030154	cell differentiation	35	10,12
56	GO:0006281	DNA repair	15	5,22
57	GO:0006334	nucleosome assembly	15	4,89
64	GO:0006974	response to DNA damage stimulus	14	8,23

Table 6 WMC1-early and WMC-late exhibit 15 fixed common events not present in WMC1-sub. These events have appeared in the cell population after WMC-1 sub sample was separated from the main lineage and previously to the WMC1-early sample passage. These events imply a quick fixation of the aberration in a matter of 10 passages.

Event #	Chr	Sibling line 1	Start site 1	End site 2	Score	CNV	Sibling line2	Start site 2	End Site 2	Score	CNV
1	chr1	WMC4	150828032	150850302	1	deletion	WMC5	150828032	150850302	1	deletion
2	chr1	WMC4	246811393	246862475	1	deletion	WMC5	246811393	246862475	1	deletion
3	chr2	WMC4	146583025	146592386	1	deletion	WMC5	146583025	146592386	1	deletion
4	chr4	WMC4	19130834	19131054	1	deletion	WMC5	19130834	19131054	1	deletion
5	chr4	WMC4	134352228	134352498	1	deletion	WMC5	134352228	134352498	1	deletion
6	chr6	WMC4	32610165	32611857	1	deletion	WMC5	32610165	32611857	1	deletion
7	**chr6**	**WMC4**	**133965998**	**134143187**	**0.29**	**amplification**	**WMC5**	**133976522**	**134143187**	**0.25**	**amplification**
8	chr6	WMC4	31405690	31406526	1	deletion	WMC5	31405690	31406526	1	deletion
9	chr6	WMC4	32061446	32063551	1	deletion	WMC5	32061446	32063551	1	deletion
10	chr6	WMC4	32094298	32096483	1	deletion	WMC5	32094298	32096483	1	deletion
11	chr6	WMC4	32563460	32563711	1	deletion	WMC5	32563460	32563711	1	deletion
12	chr6	WMC4	65404871	65406119	1	deletion	WMC5	65404871	65406119	1	deletion
13	chr8	WMC4	584761	588391	1	deletion	WMC5	584761	588391	1	deletion
14	**chr8**	**WMC4**	**90945088**	**91271065**	**0.43**	**amplification**	**WMC5**	**90964181**	**91202854**	**0.33**	**amplification**
15	chr9	WMC4	17900043	17901466	1	deletion	WMC5	17900043	17901466	1	deletion
16	**chr10**	**WMC4**	**46507836**	**47013328**	**0.5**	**amplification**	**WMC5**	**46765637**	**47067933**	**0.2**	**amplification**
1	chr1	WMC6	66974753	66974854	1	deletion	WMC7	66974753	66974854	1	deletion
2	chr2	WMC6	44320603	44322423	1	deletion	WMC7	44320603	44322423	1	deletion
3	chr2	WMC6	159667833	159669697	1	deletion	WMC7	159667833	159669697	1	deletion
4	**chr6**	**WMC6**	**32560168**	**32611857**	**0.85**	**deletion**	WMC7	**32266463**	**32650112**	**0.89**	**deletion**
5	**chr6**	**WMC6**	**32611466**	**32611857**	**1**	**deletion**	WMC7	**32266463**	**32650112**	**0.89**	**deletion**
6	chr6	WMC6	30571147	30571187	1	deletion	WMC7	30571147	30571187	1	deletion
7	chr6	WMC6	32703606	32704014	1	deletion	WMC7	32703606	32704014	1	deletion

Table 7 Common CNV events found between the two pairs of sibling cell lines. For some events (bold), the cellular score is < 1, implying a common but independent occurrence in both cell lines, suggesting the possibility of hotspots where genomic deletions/amplifications occur. However, for most of them, the cellular score is 1, and therefore discerning the parentally inherited aberrations from those that have occurred after derivation is not possible without the parental genomes information.

Table 8 (part 1)

Event	Chr	Samples	Type	program	Band	Start site	End Site	Length (kb)	Genes
1	1	WMC6 WMC7	Del	somatics	1p31.3	66974753	66974854	0.1	
2	1	WMC5 WMC6	Ampl	somatics	1p31.1	76024923	76045901	20.98	SLC44A5
3	1	WMC2-sub WMC2-late	Del	somatics	1p31.1	80778892	80831841	52.95	
4	1	WMC2-early WMC2-sub WMC2-late	Del	penncnv	1p31.1	80793066	80831841	38.78	
5	1	WMC7 WMC5	Ampl	somatics	1p21.3	96791322	96879593	88.27	AL138801.1 AC092393.3
6	1	WMC1-early WMC1-late	Del	somatics	1p13.3	108535014	108538635	3.62	AL390036.2
7	1	WMC4 WMC7	Ampl	penncnv	1p11.2	120866130	120902013	35.88	AL357493.2 AL357493.4
8	1	WMC4 WMC7	Ampl	penncnv	1p11.2	120990641	120991757	1.12	
9	1	WMC2-early WMC2-late WMC4 WMC5	Del	somatics penncnv	1q21.3	150828032	150850302	22.27	ARNT
10	1	WMC4 WMC5	Del	somatics	1q44	246811393	246862475	51.08	AL591848.1 C1orf71
11	1	WMC4 WMC5	Del	penncnv	1q44	246828735	246852068	23.33	C1orf71
12	2	WMC2-early WMC2-sub WMC2-late	Del	penncnv somatics	2p22.3	34556561	34580068	23.51	
13	2	WMC2-late WMC7 WMC2-early	Del	somatics	2p22.3	34558104	34570457	12.35	
14	2	WMC1-early WMC1-late WMC6 WMC7	Del	somatics penncnv	2p21	44320603	44322423	1.82	
15	2	WMC2-early WMC2-late	Ampl	somatics	2p11.2	88906246	89949717	1043.47	RPIA EIF2AK3 IGKV2-4 IGKC IGKV2-14 IGKV1-13 IGKV4-1 IGKCI GKV1-5
16	2	WMC1-early WMC1-late WMC4	Del	somatics penncnv	2q14.3	123197294	123198771	1.48	
17	2	WMC4 WMC5	Del	somatics	2q22.3	146583025	146592386	9.36	AC079248.1
18	2	WMC4 WMC5 WMC6	Del	somatics	2q22.3	146589604	146592386	2.78	
19	2	WMC1-early WMC1-sub WMC1-late WMC6 WMC7	Del	somatics	2q24.2	159667833	159668238	0.41	DAPL1
20	3	WMC4 WMC6	Del	penncnv	3p26.1	6626929	6629060	2.13	AC069277.1
21	3	WMC2-early WMC2-sub WMC2-late WMC4 WMC5 WMC6	Del	penncnv	3p12.1	86992144	86992358	0.21	VGLL3
22	3	WMC6 WMC4 WMC7	Del	penncnv	3q22.1	133194645	133195707	1.06	
23	3	WMC2-early WMC2-sub WMC2-late	Del	somatics	3q23	142028284	142029696	1.41	XRN1
24	3	WMC1-late WMC2-early WMC2-late WMC7 WMC1-early WMC1-sub WMC2-sub	Del	somatics penncnv	3q26.1	164037547	164085280	47.73	hsa-mir-720

Table 8 (part 2)

29	6	WMC7	loh dchip	6p21.32	32413348	32905854	492,51	TAP2	
30	6	WMC7	loh dchip	6p21.1	42862503	44146800	1284,3	VEGFA POLH XPO5 TTBK1 SRF PTK7 CUL7 PPP2R5D	
31	6	WMC7	loh dchip	6p12.1	54644135	55304546	660,41		
32	6	WMC6	loh dchip	6q14.3	86193380	86335410	142,03	SNX14	
33	6	WMC6	loh dchip	6q14.3	86557675	86763559	205,88		
34	6	WMC6	loh dchip	6q22.31	122330405	122502832	172,43		
35	7	WMC6 WMC7	loh dchip	7q31.32	122299725	122461500	161,78	RNF133	
36	8	WMC6 WMC7	loh dchip	8p23.2	4867155	5117620	250,47		
37	8	WMC7	loh dchip	8p23.1	7575048	8019212	444,16		hsa-mir-548i-3
38	8	WMC6 WMC7	loh dchip	8p23.1	9249850	9449218	199,37	TNKS	
39	8	WMC6	loh dchip	8p23.1	9743802	9855341	111,54		hsa-mir-124-1
40	8	WMC6 WMC7	loh dchip	8q11.1	46994719	47702006	707,29		
41	8	WMC7	loh dchip	8q12.1	61083149	61372013	288,86		
42	9	WMC6	loh dchip	9q31.1	105153812	106169221	1015,41		
43	9	WMC7	loh dchip	9q33.1 9q33.2	121840185	123972838	2132,65	RAB14 CEP110 TRAF1 PHF19 PSMD5 FBXW2 CDK5RAP2	hsa-mir-147
44	9	WMC6	loh dchip	9q33.3	129047502	129167958	120,46		
45	10	WMC6	loh dchip	10p15.3	59083	491550	432,47	ZMYND11	
46	10	WMC7	loh dchip	10q24.1	98861595	99116239	254,64	FRAT2 FRAT1 ARHGAP19	
47	10	WMC6	loh dchip	10q25.1	107637696	107729018	91,32		
48	11	WMC7	loh dchip	11p14.3	25686584	25939791	253,21		
49	11	WMC6	loh dchip	11p14.2	27033553	27066308	32,76		
50	11	WMC6	loh dchip	11p14.1	28121394	29498108	1376,71	KIF18A	
51	11	WMC6	loh dchip	11p13 11p12	36101652	39957376	3855,72	TRAF6 RAG1	
52	11	WMC6	loh dchip	11q21	93387383	93480352	92,97		hsa-mir-1304
53	11	WMC6	loh dchip	11q21	93668535	93851696	183,16		
54	11	WMC6	loh dchip	11q22.1	101164194	101499261	335,07		
55	12	WMC6	loh dchip	12q21.32	85500368	85825624	325,26		
56	12	WMC6	loh dchip	12q23.3	105596041	105753835	157,79		
57	12	WMC6 WMC7	loh dchip	12q24.11	109094097	109294030	199,93	CORO1C	hsa-mir-619
58	12	WMC6	loh dchip	12q24.12 12q24.13	110379655	110955473	575,82	RAD9B ANAPC7 GIT2	

Table 8 (part 3)

49	6	WMC4 WMC5 WMC6 WMC7	Del	somatics	6p21.32	32563460	32563711	0.25	HLA-DRB1
50	6	WMC1-sub WMC5 WMC6 WMC7	Del	somatics	6p21.32	32587485	32587521	0.04	
51	6	WMC4 WMC5 WMC6 WMC7	Del	somatics	6p21.32	32610165	32611857	1.69	HLA-DQA1
52	6	WMC6 WMC4 WMC7 WMC5	Del	somatics	6p21.32	32611466	32611857	0.39	HLA-DQA1
53	6	WMC1-early WMC4 WMC7	Del	somatics	6p21.32	32611857	32650112	38.26	AL662789.4 AL662789.3 HLA-DQB1 AL662789.2 HLA-DQA1
54	6	WMC6 WMC4 WMC7	Ampl	penncnv	6p21.32	32643872	32656281	12.41	
55	6	WMC1-early WMC1-late WMC4 WMC7	Del	somatics	6p21.32	32643946	32647375	3.43	
56	6	WMC6 WMC7	Del	somatics	6p21.32	32703606	32704014	0.41	
57	6	WMC4 WMC5	Del	somatics	6q12	65404871	65406119	1.25	EYS
58	6	WMC2-early WMC2-sub WMC2-late	Del	penncnv	6q12	67075448	67084250	8.8	
59	6	WMC4 WMC5	Ampl	somatics	6q23.2	133976522	134143187	166.67	AL078586.2 AL137011.1 AL078586.1
60	7	WMC1-early WMC1-late	Del	somatics	7p12.1	51562900	51563915	1.02	
61	7	WMC1-early WMC1-late	Del	somatics	7q21.11	81279747	81280187	0.44	
62	7	WMC1-sub WMC1-late WMC1-early	Del	somatics	7q21.11	81279971	81280187	0.22	
63	7	WMC1-early WMC1-late WMC2-late	Del	somatics	7q31.33	125836509	125836837	0.33	
64	8	WMC2-early WMC2-sub WMC2-late	Del	somatics	8p23.3	281394	281400	0.01	
65	8	WMC2-late WMC4 WMC5	Del	somatics	8p23.3	584761	588391	3.63	
66	8	WMC2-early WMC2-sub WMC2-late WMC4 WMC5	Del	somatics	8p23.3	586317	588391	2.07	
67	8	WMC2-early WMC2-sub WMC2-late WMC6 WMC7	Del	somatics penncnv	8p11.23	39379684	39457081	77.4	ADAM18 ADAM3A ADAM3A
68	8	WMC2-early WMC2-sub WMC2-late	Del	penncnv	8p11.1	43782832	43789367	6.54	
69	8	WMC4 WMC6	Ampl	penncnv	8q21.11	77260467	77308136	47.67	
70	8	WMC4 WMC5;WMC7	Ampl	somatics	8q21.3	90964181	91202854	238.67	NBN DECR1 CALB1 AC123779.1

Table 8 (part 4)

23	7	WMC2	late	loh dchip	7q11.21	64051842	64407931	356,09		
24	7	WMC2	late sub early	loh dchip	7q11.21	64407696	64407931	0,24		
25	7	WMC2	late sub early	loh dchip	7q11.21	64436574	65189366	752,79		
26	7	WMC2	late sub early	loh dchip	7q21.13	90493976	90611296	117,32	PFTK1'	
27	7	WMC2	late	loh dchip	7q31.2	116622099	116716301	94,2		
28	7	WMC2	late sub early	loh dchip	7q31.33	124154815	124387741	232,93		
29	7	WMC2	late sub early	loh dchip	7q31.33	126068490	126331677	263,19		
30	8	WMC2	late sub early	loh dchip	8q11.21	49767144	49932007	164,86	SNAI2#	
31	8	WMC2	late sub early	loh dchip	8q13.1	66510382	67322987	812,61	DNAJC5B TRIM55'	
32	8	WMC2	late sub early	loh dchip	8q23.2	110947338	111044375	97,04	KCNV1'	
33	10	WMC2	late sub early	loh dchip	10p11.21	37919016	37983337	64,32		
34	10	WMC2	late sub early	loh dchip	10p11.21 10p11.1	38369843	38930301	560,46		
35	10	WMC2	late sub early	loh dchip	10p11.1	38998574	39141132	142,56		
36	10	WMC2	late sub early	loh dchip	10q25.1	109985274	110058401	73,13		
37	11	WMC2	late sub early	loh dchip	11p11.12 11p11.11	50950526	51447829	497,3		
38	11	WMC2	late sub early	loh dchip	11q11	54840010	55633891	793,88	TRIM48	
39	11	WMC2	late sub early	loh dchip	11q13.5	74916133	75669579	753,45	UVRAG' DGAT2 MOGAT2 MAP6 SERPINH1#	hsa-mir-326
40	12	WMC2	late sub early	loh dchip	12p12.3	18335076	18655407	320,33	PIK3C2G	
41	12	WMC2	late sub early	loh dchip	12q11	36144166	36431552	287,39		
42	13	WMC2	late	loh dchip	13q22.3	77608029	77767282	159,25	MYCBP2'	
43	15	WMC2	late sub early	loh dchip	15q23 15q24.1	70338766	71108832	770,07		hsa-mir-629
44	17	WMC2	late sub early	loh dchip	17q21.2 17q21.31	37702888	38581147	878,26	TOP2A' RARA*'# CDC6 RAPGEFL1 NR1D1' THRA MED24 IKZF3' GRB7 ERBB2*'# PPP1R1B	
45	17	WMC2	late sub early	loh dchip	17q21.32	42854104	43994702	1140,6	MAP3K14' PLCD3' GJC1 ADAM11	

100	15	WMC5 WMC6	Ampl	somatics	15q11.2	19875658	20090262	214.6	LOC729722 U6
101	15	WMC4 WMC6	Ampl	penncnv	15q13.3	29058472	29063737	5.27	AC055876.5
102	16	WMC1-early WMC1-late WMC2-late WMC4	Del	somatics	16p13.3	1363159	1363876	0.72	UBE2I
103	18	WMC4 WMC6	Ampl	penncnv	18p11.31	6273038	6306027	32.99	L3MBTL4
104	18	WMC2-late WMC4	Del	somatics	18q12.3	36514913	36519446	4.53	
105	18	WMC2-late WMC5	Del	penncnv	18q12.3	36518186	36519446	1.26	
106	19	WMC1-early WMC1-late	Del	somatics	19q13.33	56832974	56840616	7.64	ZSCAN5A
107	22	WMC2-late WMC5	Del	somatics	22q13.33	49531259	49562479	31.22	

Table 8. Deletions and amplifications events detected by SOMATICs and/or PennCNV simultaneously in different cell samples. Description, including the chromosomal location, total length (Kb) and the genes and miRNA comprised in them.

Chr	Cell line	S	Type	Detected by	Locus	Genes	miRNA
1	WMC2 WMC4 WMC5	early late	Del	somatics penncnv	1q21.3	**ARNT**	
1	WMC6		Ampl	somatics penncnv	1q24.3	**FMO3**	
2	WMC2	early late	Ampl	somatics	2p11.2	**EIF2AK3**	
3	WMC1 WMC2 WMC6 WMC7	early sub late	Del	somatics penncnv	3q26.1		**hsa-mir-720**
3	WMC6		Ampl	somatics penncnv	3q27.2	**RFC4** **EIF4A2**	hsa-mir-1248
6	WMC6 WMC7		Del	somatics	6p21.33	**PPP1R10**	
8	WMC2	early sub late	Del	somatics penncnv	8p11.23	**ADAM18**	
8	WMC4 WMC5		Ampl	somatics	8q21.3	**NBN**	
16	WMC1 WMC2	WMC1 (early/late) WMC2 late	Del	somatics	16p13.3	**UBE2I**	

Table 9. Amplification and deletion events found to comprise a gene known to be involved in carcinogenesis (according to COSMIC and/or CGC databases). One exception is the deletion found in the 3q26.1 region, that was included in the table due to its coincidental occurrence in four different cell lines. This region comprises the hsa-mir-720 miRNA coding region, which is known to be downregulated in renal carcinoma.

Gene	Database	Description
ARNT	CGC	Aryl hydrocarbon nuclear translocator, which form parts of the signalling route activated by a large group of environmental pollutants, promoting the transcriptional activation of several xenobiotic metabolizing enzymes. Increasing evidence suggests that deregulation of ARNT may lead to deregulation of cell-cell contact thereby inducing unbalanced proliferation, dedifferentiation and enhanced motility[74].
FMO3	CGC	Flavin-containing monooxygenase 3. Is one of the major hepatic metabolic enzymes that catalyze the NADPH-dependent attachment of molecular oxygen to endogenous and foreign chemicals containing nucleophilic N, S and P heteroatoms[75].
RFC4	CGC	Part of the large multi-subunit protein complex called BASC, for BRCAl-associated genome surveillance complex, consisting of various repair proteins including the Mre11, Rad50, NBN and others. Essential for preserving genome stability.
PPP1R10	CGC	Also known as PNUTS (PP1 nuclear targeting subunit) regulates the phosphorylation p53 and MDM2, which is an essential step to regulate the apoptotic activity of these two proteins[76].
ADAM18	CGC	A-Disintegrin and Metalloprotease-18 is a multifunctional, membrane-bound cell surface glycoprotein, that has numerous functions in cell growth, differentiation, and motility. Over
NBN	CGC	Nibrin, also known as "Mediator of damage checkpoint protein 1" (NBS1). Forms part of the Mre11/Rad50/Nbs1 complex that contributes to the preservation of genome stability by participating in DNA double-strand break repair, cell cycle check point control, and telomere maintenance [80,61] . This complex is part of a large multi-subunit protein complex called BASC, for BRCAl-associated genome surveillance complex.
UBE2I	CGC	Also known as UBC9, is an essential sumoylating enzyme required in the the post-replication repair pathway (PRR) in order to maintaining genome integrity after double-strand breaks[77].

Table 10. Detailed description of the molecular function of the genes comprised in deletion and amplification events found in CGC/COSMIC databases.

	GO id	Biological process	# of genes	$\chi 2$
1	GO:0007165	signal transduction	165	24,81
2	GO:0006355	regulation of transcription, DNA-dependent	130	18,26
3	GO:0007186	G-protein coupled receptor protein signaling pathway	100	29,49
4	GO:0045449	regulation of transcription	100	14,42
5	GO:0007275	multicellular organismal development	79	9,18
6	GO:0008152	metabolic process	64	6,96
7	GO:0006810	transport	61	10,63
8	GO:0006811	ion transport	58	0,18
9	GO:0006508	proteolysis	55	1,1
10	GO:0006468	protein amino acid phosphorylation	52	5,01
16	GO:0007049	cell cycle	38	4,93
19	GO:0030154	cell differentiation	35	10,12
56	GO:0006281	DNA repair	15	5,22
57	GO:0006334	nucleosome assembly	15	4,89
64	GO:0006974	response to DNA damage stimulus	14	8,23

Table 11. Selected GO annotations of genes comprised in the CNV events found in all the hESC lines analyzed. $\chi 2$ indicates the over-representation of the GO term in this group of genes.

	LOH Score					dChip						
	# Events	Min. Len. (Kb)	Max. Len. (Kb)	Av. Len. (Kb)	Total Len. (Kb)	# Events	Min. Len. (Kb)	Max. Len. (Kb)	Av. Len. (Kb)	Total Len. (Kb)	Overlap LOH %	Overlap dChip %
WMC1-late	19	0	26244	2147,25	40797,67	65	54,52	28147,33	1319,41	85761,51	98,94	48,12
WMC1-early	19	0	8474,71	793,65	15079,42	42	88,63	9316,14	1044,86	43884,17	91,14	33,22
WMC1-sub	24	12,33	8484,18	807,47	19379,25	41	88,63	9288,43	1085,13	44490,13	86,7	41,06
WMC2-late	52	0	3660,17	417,7	21720,62	116	120,89	5745,8	785,55	91124,36	69,14	18,58
WMC2-early	47	10,43	1963,19	438,49	20608,84	112	57,01	5547,02	804,8	90137,84	83,27	19,95
WMC2-sub	47	10,43	1963,19	438,49	20608,84	113	55,29	5554,11	800,37	90441,9	83,27	19,89
WMC4	66	8,08	5378,57	577,57	38119,68	128	7,53	4997,36	725,42	92853,59	77,54	33,48
WMC5	53	15,6	5378,57	565,8	29987,2	123	129,56	4997,36	690,05	84876,49	75,63	40,09
WMC6	70	0	14713,46	920,51	64435,84	138	76,5	15571,57	888,63	122630,9	72,58	30,55
WMC7	60	0	8514,61	734,55	44073,3	138	76,5	9117,51	760,96	105012,82	74,62	26,58

Table 12. Comparison between the events found by LOH-Score and dChip. All lengths in Kb. The overlap percentages are referred to the total length of aberrations found in each program.

	LOH-Score		dChip		LOH-Score		dChip		LOH-Score		dChip	
	WMC1-early	WMC1-late	WMC1-early	WMC1-late	WMC1-early	WMC1-sub	WMC1-early	WMC1-sub	WMC1-sub	WMC1-late	WMC1-sub	WMC1-late
# Events	25	25	42	65	25	29	42	41	29	25	41	65
# Coincident Events	25	25	42	42	22	20	35	35	20	22	41	41
Total Event length (Kb)	15995.58	41713.83	43884.17	85761.51	15995.58	21070.54	43884.17	44490.13	21070.54	41713.83	44490.13	85761.51
Overlapping length (Kb)	15995.58		43686.2		15698.84		33919.81		15698.84		43902.14	
Overlapping rate %	100	38.34	99.55	50.93	98.14	74.5	90.94	89.7	74.5	37.63	98.67	51.19

	LOH-Score		dChip		LOH-Score		dChip		LOH-Score		dChip	
	WMC2-early	WMC2-late	WMC2-early	WMC2-late	WMC2-early	WMC2-sub	WMC2-early	WMC2-sub	WMC2-sub	WMC2-late	WMC2-sub	WMC2-late
# Events	57	66	112	116	57	57	112	113	57	66	113	116
# Coincident Events	51	51	102	102	57	57	112	112	51	51	102	102
Total Event length (Kb)	21602,69	24487.58	90137.84	91124.36	21602,69	21602.69	90137.84	904419.03	21602.69	24487.58	904419.03	91124.36
Overlapping length (Kb)	16793.86		81276.22		21602.69		90058.19		16793.86		81205.38	
Overlapping rate %	77.73	68.58	90.16	89.19	100	100	99.91	99.57	68.58	77.73	89.78	89.11

	WMC6	WMC7	WMC6	WMC7
# Events	76	70	138	138
# Coincident Events	19	20	49	49
Total Event length (Kb)	65013.59	45355.15	122630.9	105012.82
Overlapping length (Kb)	9525.85		317828.87	
Overlapping rate %	14.65	21.01	25.91	30.26

	WMC4	WMC5	WMC4	WMC5
# Events	82	60	128	123
# Coincident Events	20	20	43	43
Total Event length (Kb)	39081.15	30237.83	92853.59	84876.49
Overlapping length (Kb)	14300.7		32402.54	
Overlapping rate %	36.59	47.29	34.89	38.17

Table 13. Coincident LOH events between related cell samples. General comparison of the coincident allelic imbalance segments between temporally different samples of the same cell line (WMC1 and WMC2, early, sub and late passages), and the sibling cell lines (WMC4 / WMC5 and WMC6/WMC7).

Table 14 (part 1)

Event	Chr	Cell line	S	Prg.	Locus	Start site	End site	Size (Kb)	Genes	miRNA
1	1	WMC1	late sub early	loh dchip	1p31.1	73517431	73549261	31,83		
2	1	WMC1	sub	loh dchip	1p31.1	74012802	74084554	71,75		
3	2	WMC1	sub	loh dchip	2q32.1 2q32.2	189085754	189264326	178,57		hsa-mir-561
4	3	WMC1	late early sub	loh dchip	3p12.2	82226341	83334750	1108,41		
5	3	WMC1	sub	loh dchip	3p11.1	89598926	89758408	159,48		
6	3	WMC1	late early sub	loh dchip	3q11.2	94994003	95129086	135,08		
7	4	WMC1	late early sub	loh dchip	4q13.3	71217945	71390897	172,95		
8	4	WMC1	late early sub	loh dchip	4q31.3	151380792	151519206	138,41		
9	4	WMC1	late early sub	loh dchip	4q31.3	151667186	151821568	154,38		
10	5	WMC1	late early sub	loh dchip	5p15.33	73747	399636	325,89	PDCD6 AHRR' SDHA PLEKHG4B	
11	6	WMC1	late early sub	loh dchip	6p22.1	27619350	28338696	719,35	HIST1H4L* HIST1H2BO	
12	6	WMC1	late early sub	loh dchip	6q24.3	146082544	146163830	81,29	FBXO30	
13	7	WMC1	late early sub	loh dchip	7q31.1	110476229	110604870	128,64		
14	8	WMC1	sub	loh dchip	8q11.22	50415911	51681799	1265,89		
15	8	WMC1	late early sub	loh dchip	8q11.22	50437208	51643947	1206,74		
16	11	WMC1	sub	loh dchip	11p11.12 11p11.11	50109294	51447829	1338,54		
17	11	WMC1	late early sub	loh dchip	11p11.12	50136414	50343409	207		
18	11	WMC1	late early sub	loh dchip	11p11.12 11p11.11	50950526	51447829	497,3		
19	12	WMC1	late early sub	loh dchip	12q11 12q12	36144166	36633905	489,74		
20	12	WMC1	sub	loh dchip	12q21.32	86719096	87335820	616,72		
21	12	WMC1	late early sub	loh dchip	12q21.32	86731475	87290698	559,22		
22	14	WMC1	late early	loh dchip	14q24.2	70748956	70748956	0		
23	16	WMC1	sub	loh dchip	16q22.1	65609548	66582496	972,95		
24	16	WMC1	late early sub	loh dchip	16q22.1	66459571	66582496	122,93		
25	19	WMC1	late early sub	loh dchip	19p12-19q12	24228244	32702956	8474,71	TSHZ3' C19orf2' CCNE1 C19orf12 PLEKHF1 POP4 UQCRFS1 ZNF254	

26	20	WMC1	late	loh dchip	20p13-20p11.1	11799	26255797	26244	PAX1' FOXA2'# THBD# C20orf191 RIN2 RBBP9 SNX5 RRBP1# OTOR KIF16B' JAG1'# PAK7' PLCB4' PLCB1' BMP2 MCM8' CDS2 PCNA'# RASSF2 RNF24 CDC25B CENPB HSPA12B ADAM33' UBOX5' PTPRA IDH3B SNRPB STK35 SIRPA SIRPB2 FKBP1A C20orf46 CSNK2A1 TRIB3' SIRPD'	hsa-mir-1292 hsa-mir-103-2 hsa-mir-103-2-as hsa-mir-663
27	20	WMC1	sub	loh dchip	20q11.22	33282831	33762042	479,21	TP53INP2 NCOA6' ACSS2 PROCR	hsa-mir-499

Table 14. LOH events in WMC1 samples (early, late, sub) simultaneously detected by LOH-Score and dChip description, including the chromosomal location, total length (Kb) and the genes and miRNA comprised in them.

Table 15 (part 1)

Event	Chr	Cell line	S	Prg.	Locus	Start site	End site	Size (Kb)	Genes	miRNA
1	1	WMC1	late sub early	loh dchip	1p31.1	73517431	73549261	31,83		
2	1	WMC1	sub	loh dchip	1p31.1	74012802	74084554	71,75		
3	2	WMC1	sub	loh dchip	2q32.1 2q32.2	189085754	189264326	178,57		hsa-mir-561
4	3	WMC1	late early sub	loh dchip	3p12.2	82226341	83334750	1108,41		
5	3	WMC1	sub	loh dchip	3p11.1	89598926	89758408	159,48		
6	3	WMC1	late early sub	loh dchip	3q11.2	94994003	95129086	135,08		
7	4	WMC1	late early sub	loh dchip	4q13.3	71217945	71390897	172,95		
8	4	WMC1	late early sub	loh dchip	4q31.3	151380792	151519206	138,41		
9	4	WMC1	late early sub	loh dchip	4q31.3	151667186	151821568	154,38		
10	5	WMC1	late early sub	loh dchip	5p15.33	73747	399636	325,89	PDCD6 AHRR' SDHA PLEKHG4B	
11	6	WMC1	late early sub	loh dchip	6p22.1	27619350	28338696	719,35	HIST1H4L* HIST1H2BO	
12	6	WMC1	late early sub	loh dchip	6q24.3	146082544	146163830	81,29	FBXO30	
13	7	WMC1	late early sub	loh dchip	7q31.1	110476229	110604870	128,64		
14	8	WMC1	sub	loh dchip	8q11.22	50415911	51681799	1265,89		
15	8	WMC1	late early sub	loh dchip	8q11.22	50437208	51643947	1206,74		
16	11	WMC1	sub	loh dchip	11p11.12 11p11.11	50109294	51447829	1338,54		
17	11	WMC1	late early sub	loh dchip	11p11.12	50136414	50343409	207		
18	11	WMC1	late early sub	loh dchip	11p11.12 11p11.11	50950526	51447829	497,3		
19	12	WMC1	late early sub	loh dchip	12q11 12q12	36144166	36633905	489,74		
20	12	WMC1	sub	loh dchip	12q21.32	86719096	87335820	616,72		
21	12	WMC1	late early sub	loh dchip	12q21.32	86731475	87290698	559,22		
22	14	WMC1	late early	loh dchip	14q24.2	70748956	70748956	0		
23	16	WMC1	sub	loh dchip	16q22.1	65609548	66582496	972,95		
24	16	WMC1	late early sub	loh dchip	16q22.1	66459571	66582496	122,93		
25	19	WMC1	late early sub	loh dchip	19p12-19q12	24228244	32702956	8474,71	TSHZ3' C19orf2' CCNE1 C19orf12 PLEKHF1 POP4 UQCRFS1 ZNF254	

23	7	WMC2	late	loh dchip	7q11.21	64051842	64407931	356,09		
24	7	WMC2	late sub early	loh dchip	7q11.21	64407696	64407931	0,24		
25	7	WMC2	late sub early	loh dchip	7q11.21	64436574	65189366	752,79		
26	7	WMC2	late sub early	loh dchip	7q21.13	90493976	90611296	117,32	PFTK1'	
27	7	WMC2	late	loh dchip	7q31.2	116622099	116716301	94,2		
28	7	WMC2	late sub early	loh dchip	7q31.33	124154815	124387741	232,93		
29	7	WMC2	late sub early	loh dchip	7q31.33	126068490	126331677	263,19		
30	8	WMC2	late sub early	loh dchip	8q11.21	49767144	49932007	164,86	SNAI2#	
31	8	WMC2	late sub early	loh dchip	8q13.1	66510382	67322987	812,61	DNAJC5B TRIM55'	
32	8	WMC2	late sub early	loh dchip	8q23.2	110947338	111044375	97,04	KCNV1'	
33	10	WMC2	late sub early	loh dchip	10p11.21	37919016	37983337	64,32		
34	10	WMC2	late sub early	loh dchip	10p11.21 10p11.1	38369843	38930301	560,46		
35	10	WMC2	late sub early	loh dchip	10p11.1	38998574	39141132	142,56		
36	10	WMC2	late sub early	loh dchip	10q25.1	109985274	110058401	73,13		
37	11	WMC2	late sub early	loh dchip	11p11.12 11p11.11	50950526	51447829	497,3		
38	11	WMC2	late sub early	loh dchip	11q11	54840010	55633891	793,88	TRIM48	
39	11	WMC2	late sub early	loh dchip	11q13.5	74916133	75669579	753,45	UVRAG' DGAT2 MOGAT2 MAP6 SERPINH1#	hsa-mir-326
40	12	WMC2	late sub early	loh dchip	12p12.3	18335076	18655407	320,33	PIK3C2G	
41	12	WMC2	late sub early	loh dchip	12q11	36144166	36431552	287,39		
42	13	WMC2	late	loh dchip	13q22.3	77608029	77767282	159,25	MYCBP2'	
43	15	WMC2	late sub early	loh dchip	15q23 15q24.1	70338766	71108832	770,07		hsa-mir-629
44	17	WMC2	late sub early	loh dchip	17q21.2 17q21.31	37702888	38581147	878,26	TOP2A' RARA*'# CDC6 RAPGEFL1 NR1D1' THRA MED24 IKZF3' GRB7 ERBB2*'# PPP1R1B	
45	17	WMC2	late sub early	loh dchip	17q21.32	42854104	43994702	1140,6	MAP3K14' PLCD3' GJC1 ADAM11	

Table 15. LOH events in WMC2 samples (early, late, sub) simultaneously detected by LOH-Score and dChip description, including the chromosomal location, total length (Kb) and the genes and miRNA comprised in them

.

Table 16 (part 1)

Event	Chr	Cell line	Prg	Locus	Start site	End site	Size (Kb)	Genes	miRNA
1	5	WMC4 WMC5	dchip loh	5p12-5q11.1	45229226	50187722	4958.5		
2	12	WMC4 WMC5	dchip loh	12q21.31	78840229	79424479	584.25	SYT1	
3	1	WMC5	loh dchip	1p33 1p32.3	51039062	52109627	1070.57	EPS15 RNF11 CDKN2C FAF1	
4	1	WMC5	loh dchip	1p22.2 1p22.1	91867666	92604157	736.49	BRDT CDC7	
5	1	WMC5	loh dchip	1p12	120220367	120569493	349.13	NOTCH2 ADAM30	
6	1	WMC4	loh dchip	1q31.3	194886125	195089923	203.8		
7	2	WMC5	loh dchip	2p12	81865855	82740073	874.22		
8	2	WMC4 WMC5	loh dchip	2p12	82635251	82740073	104.82		
9	2	WMC4	loh dchip	2p11.2	86059993	86672920	612.93	JMJD1A	
10	2	WMC4	loh dchip	2q11.1 2q11.2	94691119	96038893	1347.77		
11	2	WMC4	loh dchip	2q14.1 2q14.2	117904368	118778863	874.5	DDX18	
12	2	WMC4	loh dchip	2q14.2	120011170	120215230	204.06		
13	2	WMC4	loh dchip	2q24.1	155515297	155758262	242.97		
14	2	WMC4	loh dchip	2q32.2	189612919	190432386	819.47		hsa-mir-1245
15	2	WMC4	loh dchip	2q33.2	203543422	203934576	391.15		
16	2	WMC4	loh dchip	2q37.3	242135194	242262119	126.93	HDLBP	
17	3	WMC4	loh dchip	3p22.2 3p22.1	39068330	39308314	239.98	CX3CR1 AXUD1	
18	3	WMC5	loh dchip	3p22.1	41574530	41656237	81.71	ULK4	
19	3	WMC4	loh dchip	3p11.1	90088422	90576572	488.15		
20	3	WMC4	loh dchip	3q13.11	105562571	105605364	42.79	CBLB	
21	3	WMC4	loh dchip	3q13.13	112433230	112733898	300.67		
22	3	WMC4	loh dchip	3q13.2	112918647	113026985	108.34		
23	3	WMC4 WMC5	loh dchip	3q22.3	137691478	137793675	102.2		
24	4	WMC4	loh dchip	4p15.31	21033532	21733158	699.63		
25	4	WMC5	loh dchip	4p15.1	31140595	31310370	169.78	PCDH7	
26	4	WMC5	loh dchip	4p15.1	31329557	31375805	46.25		
27	4	WMC4	loh dchip	4q13.1	64532226	64795145	262.92		
28	4	WMC4	loh dchip	4q28.1	125253668	125328534	74.87		
29	4	WMC5	loh dchip	4q28.1	125713954	126396093	682.14		
30	4	WMC5	loh dchip	4q28.1	127907884	128357306	449.42		

Table 16 (part 2)

31	4	WMC4	loh dchip	4q28.3	133114901	134071884	956,98	PCDH10	
32	4	WMC5	loh dchip	4q31.21	143359589	143844712	485,12	INPP4B	
33	4	WMC4	loh dchip	4q32.1	155401217	155669079	267,86		
34	5	WMC4 WMC5	loh dchip	5p13.2	37209687	37651745	442,06	WDR70 NUP155	
35	5	WMC4 WMC5	loh dchip	5p13.1	39711367	40429362	718		
36	5	WMC4 WMC5	loh dchip	5q12.1	59782677	59844012	61,34		
37	5	WMC5	loh dchip	5q14.3	87632431	87699366	66,94		
38	5	WMC4	loh dchip	5q21.1	101336908	101847880	510,97		
39	5	WMC5	loh dchip	5q32	143678804	144295321	616,52		
40	6	WMC5	loh dchip	6p22.1	29566456	29716673	150,22		
41	6	WMC4 WMC5	loh dchip	6p21.1 6p12.3	45082289	45363272	280,98	RUNX2	hsa-mir-586
42	6	WMC4 WMC5	loh dchip	6p12.3	45203141	45248739	45,6		
43	6	WMC5	loh dchip	6q22.33	128957292	129292114	334,82		
44	6	WMC4	loh dchip	6q26	162892230	163168579	276,35		
45	7	WMC4 WMC5	loh dchip	7p14.1	38836969	38852569	15,6		
46	7	WMC5	loh dchip	7q11.22	68725586	69074601	349,02		
47	7	WMC5	loh dchip	7q31.32	122090557	123112695	1022,14	RNF133	
48	8	WMC4 WMC5	loh dchip	8p11.1 8q11.1	43225875	47654762	4428,89		
49	8	WMC4	loh dchip	8q21.3	89713547	89949472	235,93		
50	8	WMC5	loh dchip	8q23.1	110026075	110256866	230,79	TRHR NUDCD1	
51	9	WMC5	loh dchip	9q33.2	122626558	122922855	296,3		
52	10	WMC4	loh dchip	10p14	9477937	9638948	161,01		
53	10	WMC5	loh dchip	10p12.1	25353824	25434343	80,52		
54	10	WMC4 WMC5	loh dchip	10p11.21	35514705	35704462	189,76		
55	10	WMC5	loh dchip	10q24.2	100706226	100814477	108,25		
56	10	WMC4	loh dchip	10q25.3	117132364	117187445	55,08		
57	10	WMC4	loh dchip	10q25.3	117293434	117496769	203,34		
58	11	WMC4 WMC5	loh dchip	11p11.2	47764350	48461172	696,82	NUP160 PTPRJ	
59	11	WMC4 WMC5	loh dchip	11q12.3	61426522	62195395	768,87	INCENP RAB3IL1 FEN1	hsa-mir-611 hsa-mir-1908
60	11	WMC4	loh dchip	11q14.3	88456186	88752298	296,11		

Event	Chr	Cell line	Prg	Locus	Start site	End site	Size (Kb)	Genes	miRNA
61	11	WMC4	loh dchip	11q14.3	88975724	89014498	38.77		
62	11	WMC4	loh dchip	11q22.1	97023895	97149034	125.14		
63	12	WMC4	loh dchip	12q11 12q12	36144166	38022789	1878.62		
64	12	WMC4	loh dchip	12q12	38682027	38974348	292.32		
65	12	WMC5	loh dchip	12q14.1	57985964	58378810	392.85	CYP27B1 CDK4 TSPAN31 CENTG1 DTX3 PIP4K2C	hsa-mir-26a-2
66	12	WMC4 WMC5	loh dchip	12q21.31	78764959	79121975	357.02		
67	12	WMC4	loh dchip	12q24.31	122309880	122387562	77.68	PSMD9	
68	13	WMC4	loh dchip	13q21.1	54532312	54807631	275.32		
69	13	WMC4	loh dchip	13q21.1	55035424	55956435	921.01		
70	13	WMC5	loh dchip	13q31.1	80513537	80521044	7.51		
71	13	WMC4	loh dchip	13q31.1	86261101	86269179	8.08		
72	17	WMC4 WMC5	loh dchip	17q11.2	24350947	24734553	383.61		
73	17	WMC4	loh dchip	17q21.32	43280889	43374600	93.71	MAP3K14	
74	18	WMC5	loh dchip	18q22.1	61761692	61992270	230.58		
75	19	WMC4	loh dchip	19q13.12	42468279	42962003	493.72	ERF GSK3A POU2F2	
76	19	WMC4	loh dchip	19q13.13	43011044	43117876	106.83		
77	21	WMC4 WMC5	loh dchip	21q22.11	30576319	30716253	139.93	BACH1	
78	22	WMC4	loh dchip	22q12.2 22q12.3	30062461	30540763	478.3	MTMR3 NF2	
79	22	WMC4	loh dchip	22q13.1 22q13.2	39244895	39357765	112.87	CBX6 APOBEC3A	

Table 16. LOH events in WMC4/WMC5 simultaneously detected by LOH-Score and dChip description, including the chromosomal location, total length (Kb) and the genes and miRNA comprised in them.

Table 17 (part 1)

Event	Chr	Cell line	Prg	Locus	Start site	End site	Size (Kb)	Genes	miRNA
1	1	WMC7	loh dchip	1p34.3	34523523	35852095	1328,57	ZMYM4 SFPQ GJB4	hsa-mir-552 hsa-mir-552
2	1	WMC6	loh dchip	1p21.3	96335934	96752202	416,27		
3	1	WMC7	loh dchip	1p21.2	100365241	100517386	152,15		
4	1	WMC6	loh dchip	1q31.1-1q32.1	184217757	198931212	14713,46	PTPRC NEK7 ASPM CFH CDC73 GLRX2 RGS2 UCHL5 RGS13 RGS1 PLA2G4A FAM5C RGS18 PTGS2 PDC TPR RNF2	hsa-mir-1278 hsa-mir-181b-1 hsa-mir-181a-1
5	2	WMC6 WMC7	loh dchip	2p16.1	56756213	59884084	3127,87	FANCL VRK2	
6	2	WMC7	loh dchip	2p14	69807891	69961575	153,68	AAK1	
7	2	WMC6 WMC7	loh dchip	2q12.3 2q13	108495294	108780535	285,24		
8	2	WMC6 WMC7	loh dchip	2q31.2	179359199	179421846	62,65	TTN PLEKHA3	
9	3	WMC6 WMC7	loh dchip	3p21.1	52217942	53909211	1691,27	IL17RB PRKCD SFMBT1 NEK4 WDR82 TLR9	hsa-let-7g hsa-mir-135a-1
10	3	WMC6	loh dchip	3p12.2	83508170	83547371	39,2		
11	3	WMC6	loh dchip	3p12.1	84711551	85845797	1134,25	CADM2	
12	3	WMC7	loh dchip	3q11.2-3q13.11	98146934	106661546	8514,61	CBLB SENP7 TFG	
13	3	WMC6	loh dchip	3q23	142995203	143391883	396,68	SLC9A9	
14	3	WMC6	loh dchip	3q26.1	164123191	164764164	640,97		
15	4	WMC6 WMC7	loh dchip	4p16.1	9218604	10044442	825,84		hsa-mir-548i-2
16	4	WMC7	loh dchip	4p13	42067762	44330950	2263,19		
17	4	WMC6	loh dchip	4q24	106482899	106592309	109,41		
18	4	WMC7	loh dchip	4q33	170498679	171083128	584,45	NEK1	
19	5	WMC7	loh dchip	5p15.2	12554892	13150705	595,81		
20	5	WMC6	loh dchip	5p11 5q11.1	46019437	49844804	3825,37		
21	5	WMC6	loh dchip	5q11.2	51768871	51980713	211,84		
22	5	WMC7	loh dchip	5q11.2	54711979	54740120	28,14		
23	5	WMC7	loh dchip	5q12.1	60365381	60462017	96,64		
24	5	WMC7	loh dchip	5q15	96960495	96985884	25,39		
25	5	WMC7	loh dchip	5q22.2	112338295	112802324	464,03	TSSK1B	
26	5	WMC6	loh dchip	5q23.1	120344934	120502430	157,5		
27	5	WMC6	loh dchip	5q31.1	133462813	133941452	478,64	SKP1 PPP2CA CDKL3 UBE2B SAR1B	
28	6	WMC6	loh dchip	6p22.2	24208169	24662741	454,57	GPLD1 TTRAP	

Table 17 (part 2)

29	6	WMC7	loh dchip	6p21.32	32413348	32905854	492,51	TAP2	
30	6	WMC7	loh dchip	6p21.1	42862503	44146800	1284,3	VEGFA POLH XPO5 TTBK1 SRF PTK7 CUL7 PPP2R5D	
31	6	WMC7	loh dchip	6p12.1	54644135	55304546	660,41		
32	6	WMC6	loh dchip	6q14.3	86193380	86335410	142,03	SNX14	
33	6	WMC6	loh dchip	6q14.3	86557675	86763559	205,88		
34	6	WMC6	loh dchip	6q22.31	122330405	122502832	172,43		
35	7	WMC6 WMC7	loh dchip	7q31.32	122299725	122461500	161,78	RNF133	
36	8	WMC6 WMC7	loh dchip	8p23.2	4867155	5117620	250,47		
37	8	WMC7	loh dchip	8p23.1	7575048	8019212	444,16		hsa-mir-548i-3
38	8	WMC6 WMC7	loh dchip	8p23.1	9249850	9449218	199,37	TNKS	
39	8	WMC6	loh dchip	8p23.1	9743802	9855341	111,54		hsa-mir-124-1
40	8	WMC6 WMC7	loh dchip	8q11.1	46994719	47702006	707,29		
41	8	WMC7	loh dchip	8q12.1	61083149	61372013	288,86		
42	9	WMC6	loh dchip	9q31.1	105153812	106169221	1015,41		
43	9	WMC7	loh dchip	9q33.1 9q33.2	121840185	123972838	2132,65	RAB14 CEP110 TRAF1 PHF19 PSMD5 FBXW2 CDK5RAP2	hsa-mir-147
44	9	WMC6	loh dchip	9q33.3	129047502	129167958	120,46		
45	10	WMC6	loh dchip	10p15.3	59083	491550	432,47	ZMYND11	
46	10	WMC7	loh dchip	10q24.1	98861595	99116239	254,64	FRAT2 FRAT1 ARHGAP19	
47	10	WMC6	loh dchip	10q25.1	107637696	107729018	91,32		
48	11	WMC7	loh dchip	11p14.3	25686584	25939791	253,21		
49	11	WMC6	loh dchip	11p14.2	27033553	27066308	32,76		
50	11	WMC6	loh dchip	11p14.1	28121394	29498108	1376,71	KIF18A	
51	11	WMC6	loh dchip	11p13 11p12	36101652	39957376	3855,72	TRAF6 RAG1	
52	11	WMC6	loh dchip	11q21	93387383	93480352	92,97		hsa-mir-1304
53	11	WMC6	loh dchip	11q21	93668535	93851696	183,16		
54	11	WMC6	loh dchip	11q22.1	101164194	101499261	335,07		
55	12	WMC6	loh dchip	12q21.32	85500368	85825624	325,26		
56	12	WMC6	loh dchip	12q23.3	105596041	105753835	157,79		
57	12	WMC6 WMC7	loh dchip	12q24.11	109094097	109294030	199,93	CORO1C	hsa-mir-619
58	12	WMC6	loh dchip	12q24.12 12q24.13	110379655	110955473	575,82	RAD9B ANAPC7 GIT2	

Table 17 (part 3)

Event	Chr	Cell line	Prg	Locus	Start site	End site	Size (Kb)	Genes	miRNA
59	13	WMC6	loh dchip	13q13.3	34712496	34750860	38.36		
60	13	WMC7	loh dchip	13q21.1	55757454	56968400	1210.95		
61	13	WMC6	loh dchip	13q21.1 13q21.2	57299434	57698033	398.6		
62	13	WMC7	loh dchip	13q21.32	64597274	64716983	119.71		
63	13	WMC6 WMC7	loh dchip	13q32.1	95429593	95923396	493.8	ABCC4	
64	14	WMC7	loh dchip	14q11.2	20113123	21160945	1047.82	NP APEX1 PARP2 CCNB1IP1	hsa-mir-1201
65	14	WMC7	loh dchip	14q11.2	22665561	22715558	50		
66	14	WMC6	loh dchip	14q32.32 14q32.33	102848843	103178496	329.65		
67	15	WMC7	loh dchip	15q22.31	62915894	62927573	11.68		
68	15	WMC6	loh dchip	15q23	68999891	69162978	163.09		
69	16	WMC6	loh dchip	16p11.2	31750349	32112362	362.01		
70	16	WMC6	loh dchip	16p11.2 16p11.1	34285183	35141900	856.72		
71	17	WMC6 WMC7	loh dchip	17q11.2	25270484	25273428	2.94		
72	17	WMC6	loh dchip	17q21.31 17q21.32	37981755	42751086	4769.33	FZD2 ITGA2B ASB16 HDAC5 DUSP3 ETV4 DHX8 ARL4D BRCA1 RND2 G6PC BECN1 WNK4 EZH1 PTRF STAT3 STAT5A STAT5B RAB5C HSPB9 NKIRAS2 DNAJC7 CNP ACLY KLHL10 FKBP10 JUP KRT14 SMARCE1 CCR7 TOP2A RARA CDC6 RAPGEFL1 NR1D1 THRA MED24 IKZF3 TNFRSF13B	hsa-mir-2117
73	18	WMC6 WMC7	loh dchip	18q11.1	16790610	17043299	252.69		
74	18	WMC7	loh dchip	18q12.3	36257764	36278746	20.98		
75	18	WMC7	loh dchip	18q12.3	36318755	36327870	9.12		
76	18	WMC6	loh dchip	18q22.1	60043551	60441471	397.92	TNFRSF11A	
77	19	WMC7	loh dchip	19p13.3	3399694	3979783	580.09	DAPK3 ITGB1BP3 MATK PIP5K1C FZR1 NFIC	hsa-mir-637
78	20	WMC6 WMC7	loh dchip	20p13	1995966	3966579	1970.61	RNF24 CDC25B CENPB HSPA12B ADAM33 UBOX5 PTPRA IDH3B SNRPB STK35	hsa-mir-1292 hsa-mir-103-2
79	20	WMC6 WMC7	loh dchip	20p13	3606931	3826038	219.11	CDC25B CENPB HSPA12B ADAM33	
80	20	WMC7	loh dchip	20p12.1	13802827	13832010	29.18		

Event	Chr	Cell line	Prg	Locus	Start site	End site	Size (Kb)	Genes	miRNA
80	20	WMC7	loh dchip	20p12.1	13802827	13832010	29.18		
81	22	WMC6	loh dchip	22q12.1	27036220	27201570	165.35		
82	22	WMC7	loh dchip	22q12.2	28096966	28436340	339.37	PITPNB	
83	22	WMC6	loh dchip	22q13.31	44911445	46811549	1900.1	SMC1B PPARA GTSE1	hsa-mir-1249 hsa-let-7a-3 hsa-let-7b
84	22	WMC6	loh dchip	22q13.31 22q13.32	46981959	47079687	97.73		
85	22	WMC6	loh dchip	22q13.32	47476820	47541323	64.5		

Table 17. LOH events in WMC6/WMC7 simultaneously detected by LOH-Score and dChip description, including the chromosomal location, total length (Kb) and the genes and miRNA comprised in them.

	GO id	Biological process	# of genes	$\chi 2$
1	GO:0007165	signal transduction	258	29.7
2	GO:0007186	G-protein coupled receptor protein signaling pathway	184	18.96
3	GO:0006355	regulation of transcription, DNA-dependent	157	57.69
4	GO:0045449	regulation of transcription	123	42.62
5	GO:0007155	cell adhesion	100	0.75
6	GO:0007275	multicellular organismal development	98	28.36
7	GO:0050896	response to stimulus	91	2.21
8	GO:0006810	transport	90	16.75
9	GO:0007608	sensory perception of smell	78	8.50E-004
10	GO:0006468	protein amino acid phosphorylation	76	8.47
21	GO:0030154	cell differentiation	49	18.07
22	GO:0007049	cell cycle	45	16.4
23	GO:0006915	apoptosis	45	10.12
25	GO:0008283	cell proliferation	42	2.85

Table 18. Biological processes GO terms over-represented in the group of genes included in the LOH-segments found in the hESC lines. $\chi 2$ indicates the over-representation of the GO term in this group of genes.

Cell line	Chr.	Sample	Band	Start site	End site	Length (Kb)	Genes	miRNA
WMC1	6	late early sub	6p22.1	27619350	28338696	719,35	HIST1H4L	
WMC1	20	late	20p13-20p11.1	11799	26255797	26244	FOXA2 PCNA	hsa-mir-1292 hsa-mir-103-2 hsa-mir-663
WMC2	1	late sub early	1q21.1	144838594	145463869	625,28	PDE4DIP	
WMC2	1	late sub early	1q21.1 1q21.2	146788907	149174579	2385,67	BCL9	
WMC2	3	late sub early	3q22.3	139204565	139832022	627,46	RBP1	
WMC2	17	late sub early	17q21.2 17q21.31	37702888	38581147	878,26	RARA ERBB2	
WMC4	3		3q13.11	105562571	105605364	42,79	CBLB	
WMC4	22		22q13.1 22q13.2	39244895	39357765	112,87	CBX6	
WMC4 WMC5	21		21q22.11	30576319	30716253	139,93	BACH1	
WMC5	1		1p33 1p32.3	51039062	52109627	1070,57	EPS15 CDKN2C	
WMC5	1		1p12	120220367	120569493	349,13	NOTCH2	
WMC5	12		12q14.1	57985964	58378810	392,85	CDK4	hsa-mir-26a-2
WMC6	1		1q31.1-1q32.1	184217757	198931212	14713,46	CDC73 TPR RNF2	hsa-mir-1278 hsa-mir-181b-1 hsa-mir-181a-1
WMC6	11		11p14.1	28121394	29498108	1376,71	KIF18A	
WMC6	17		17q21.31 17q21.32	37981755	42751086	4769,33	ETV4 BRCA1 EZH1 STAT3 RARA	hsa-mir-2117
WMC6	22		22q13.31	44911445	46811549	1900,1		hsa-mir-1249 hsa-let-7a-3 hsa let-7b
WMC6 WMC7	2		2p16.1	56756213	59884084	3127,87	FANCL	
WMC6 WMC7	3		3p21.1	52217942	53909211	1691,27	SFMBT1	hsa-let-7g hsa-mir-135a-1
WMC7	3		3q11.2-3q13.11	98146934	106661546	8514,61	CBLB TFG	
WMC7	9		9q33.1 9q33.2	121840185	123972838	2132,65	CEP110 PHF19	hsa-mir-147
WMC7	14		14q11.2	20113123	21160945	1047,82	CCNB1IP1	hsa-mir-1201

Table 19. LOH events that include tumorigenesis promoting genes. Genes were found in CGC/COSMIC Databases

Gene	Description
TPR	Component of the nuclear pore complex (NPC), as a novel Mad2-interacting protein, essential for the mitotic spindle checkpoint [92] .
KIF18A	Kinesin 8. Contributes to the suppression of the amplitude of kinetochore oscillations by increasing the frequency by which kinetochores change direction. It regulates the dynamics of the plus ends of kinetochore microtubules[93].
PCNA	DNMT1 is a maintenance methyltransferase that prefers hemimethylated templates and is recruited to actively replicating DNA through an association with PCNA, the replication fork clamp[4,5]. It is a responsive element for p53[94].
CCNB1IP1	Cyclin B1 interacting protein 1
ETV4	ETV4, a member of ETV transcription factors family, can alter the expression of proteins involved in a range of pathways including stem cell development, cell senescence, proliferation, migration, apoptosis and tumorigenesis.
CDK4	Cyclin kinase 4, key regulator of G1 progression and the G1/S transition ($p16^{INK4A}$-Cdk4-pRb pathway).
EPS15	Epidermal growth factor receptor pathway substrate 15
ERBB2	Cell surface transmembrane epidermal growth factor receptor, tyrosine kinase activity.
BCL9	BCL9 perform its oncogenic effect enhancing beta-catenin–mediated transcriptional activity and Wnt signaling pathway.
RARA	The PRC1 Polycomb group complex interacts with PLZF/RARA to mediate leukemic transformation[3]
PHF19	member of a polycomb-like complex, that Stimulates H3K27 trimethylation and recruits PRC2[2].
SEMBT1	Member of the complex required to recruit the polycomb complex to a certain region of the genome[2] .
EZH1	Member of the human polycomb repressive complex 2 and 4 (PRC2 and PRC4). Essential for methyltransferase activity[2]. PRC4 is active in ESC and Cancer cells[95].
CBX6	Member of the Polycomb Repressive Complex1(PCR1)[2].
RNF2	Member of the Polycomb Repressive Complex 1(PRC1). PRC1 recruitment results in the RNF1 and RNF2-mediated ubiquitylation of histone H2A on lysine 119, which is thought to be important for transcriptional repression[2].
CBLB	Together with its close homologue, C-CBL, CBLB is thought to be involved in the negative modulation of tyrosine kinase signalling, primarily through their E3 ubiquitin ligase activity[96]
BRCA1	DNA repair function[91]. Mutated in several tumors, particularly gynecologyc tumors.
BACH1	Helicase-like protein,interacts directly with BRCA1 and contributes to its DNA repair function[97].

FANC	It has an early role in the processing of crosslinked DNA, which may subsequently require the participation of HR repair factors (for example, BRCA1, BRCA2, PAlB2 and BACH1) to repair the break[97].
CDKN2C	Frequent LOH seen in pituitary adenomas. Acts downstream of GATA3 and restrains tumorigenesis (breast cancer)[3].
NOTCH2	NOTCH controls key steps of development, cell growth and differentiation.
CDC73	Cell division cycle 73, Paf1/RNA polymerase II complex component,

Table 20. Biological function of carcinogenesis-promoting genes comprised in the LOH-events found in the different hESC samples analyzed.

4.4 Genomic analysis of euploid human embryonic stem cell lines derived from PGS-diagnosed aneuploid embryos: no trace of self-correction detected.

Genomic analysis of euploid human embryonic stem cell lines derived from PGS-diagnosed aneuploid embryos: no trace of self-correction detected.

Josep Pareja-Gómez[1], Nikica Zaninovic[1], Oriol Vidal[2] , Anna Veiga[3] and Zev Rosenwaks[1]

1. Center for Reproductive Medicine and Infertility (CRMI), Weill-Cornell Medical College, New York, NY
2. Departament de Biologia, Universitat de Girona, Girona
3. Centre de Medicina Regenerativa de Barcelona (CMRB), Barcelona.

Formatted to be submitted to:
Human Reproduction (ISSN: 1460-2350)

Abstract

Aneuploidies are known to be incompatible with healthy human embryonic development. Prenatal genetic screening (PGS) using fluorescent in situ hybridization (FISH) aims to detect any chromosomal numerical aberration (of a selected group of chromosomes) in single blastomeres biopsied from day-3 embryos during in-vitro fertilization (IVF) cycles. Embryos diagnosed with aneuploidies have become a possible source to derive human embryonic stem cells (hESC) from. Interestingly, recent reports account for 18 out of 20 hESC lines derived from aneuploid embryos are karyotypically normal, which is in contradiction with the PGS diagnosis rationale. Some authors have suggested a mechanism of self-correction of during hESC derivation to explain the euploidization of the hESC. Here we describe the derivation of two euploid hESC lines from a total of eleven PGS-diagnosed embryos. After analyzing the genomes of both hESC lines using microsatellite analysis and whole-genome genotyping arrays, we found no proof of the hypothesized self-correction mechanisms. However, no direct evidences of a PGS misdiagnosis can be stated for these type of embryos due to the reported high degree of euploidy/aneuploidy mosaicism reported at these early stages of human embryogenesis.

Introduction

Human embryonic stem cells (hESC) can be derived from human embryos at different stages, but most commonly from the inner cell mass (ICM) of blastocyst-stage embryo(Thomson et al., 1998; Reubinoff et al., 2000). So far, all the hESC lines derived until the present date have been originated from a human embryo generated after in vitro fertilization (IVF). These embryos are donated by their progenitors, who are couples that normally have accomplished their parental project after a fertility treatment. In addition, embryos diagnosed as abnormal after preimplantation genetic diagnosis (PGD) are susceptible of donation for hESC derivation. Indeed, several groups have reported the derivation of hESC lines from embryos affected of different monogenic diseases(K.D. Sermon et al., 2009; Frumkin et al.)after PGD diagnosis. Similarly, preimplantational genetic screening (PGS), a type of PGD focused in the detection of chromosomal aberrations(Y.[1] Verlinsky et al., 2005; Santiago Munné and [1], 2002; Gianaroli et al., 1999; Y. Verlinsky et al., 1995), has become a source of embryos donated for research. In order to perform PGS three approaches can be used: biopsy of the second polar body, biopsy of a blastomere from a day-3 cleavage stage embryo, or biopsy of trophectodermal cells of a day 5 embryo. After polar body or cell fixation on glass slides, the presence of up to 12 chromosomes is investigated using loci-specific or centromeric probes by fluorescent in situ hybridization (FISH). The chromosomes most frequently targeted by the probes used are the ones described to be affected by numerical aberrations in spontaneous miscarriages(Celep et al., 2006; Menasha et al., 2005; Lebedev et al., 2004). In fact, PGS was originally conceived to reduce the spontaneous miscarriage rate per embryo transferred(Y.[1] Verlinsky et al., 2005; Santiago Munné and [1], 2002; Gianaroli et al., 1999; Y. Verlinsky et al., 1995). Nowadays, PGS has been increasingly used in IVF cases world-wide(Joyce Harper et al., 2008). Reported indications for PGS, other than recurrent or spontaneous miscarriages, are advanced maternal age, severe male infertility and recurrent implantation failure and previously aneuploidy-affected pregnancies(Joyce Harper et al., 2008), which expands enormously the number of embryos which could potentially undergo PGS. In fact, some authors have suggested that PGS should be a default procedure in all IVF cycles(Y.[1] Verlinsky et al., 2005).

However, despite the coherent scientific rationale behind PGS and some favorable reports matching indicated cases with controls(Santiago Munné et al., 2003; Gianaroli et al., 1999), recent randomized prospective clinical trials failed to observe a significantly increase in the birth rate after

transfer of chromosomally normal embryos selected after PGS(C. Staessen et al., 2008; Catherine Staessen et al., 2004; Twisk et al., 2008; Bart C.J.M. Fauser, 2008; Joyce Harper et al., 2010; Schoolcraft et al., 2009a). Moreover, several reports describe the derivation of euploid hESC lines from aneuploidy-diagnosed embryos(Lavon et al., 2008; Peura et al., 2008; Santiago Munné et al., 2005). New concerns have been raised recently about the chromosomal stability of embryos during early human preimplantation development suggesting an elevated aneuploidy mosaicism rate(Vanneste et al., 2009a).

In this study we used PGS-diagnosed aneuploid embryos to stablish their derivability into hESC, their pluripotency and differentiation capabilities and their genomic status right after derivation. A genomic analysis was performed on the stem cells lines in an attempt to elucidate the origin (parentally inherited or embryonic de novo appearance) of the diagnosed embryonic chromosomal aberration.

Material and Methods

Eleven embryos with different parental origins diagnosed by preimplantation genetic screening (PGS) as chromosomally abnormal were used for this study. Institutional review board approval for this study was obtained, and written consent was provided by each couple for the derivation of human embryonic stem cell lines (hESC) (Weill Cornell Medical College IRB # 0502007737).

Preimplantation Genetic Screening (PGS)

Preimplantation Genetic Screening was carried out by the PGD-team at the Center for Reproductive Medicine and Infertility of Weill Cornell Medical College. Briefly, on day 3 after intracytoplasmatic sperm injection (ICSI), each embryo had one cell biopsied (acid Tyrode's method). Fixation was performed as described previously(Xu et al., 1998). If the fixation was not successful, a second blastomere was biopsied. Cells were analyzed for 9 chromosomes (13,14, 15,16,18,21,22,X,Y) using fluorescence *in situ* hybridization (FISH) and diagnosed using a scoring function previously described(Xu et al., 1998).

Embryos were then cultured with the standard sequential culture medium until blastocyst stage (day 5). Embryos were then either used fresh for derivation or were frozen using a standard freezing protocol described elsewhere(A Trounson and Mohr, 1983).

ESC derivation

In order to derive hESC, the donated embryos were cultured until day 6. Stem cell derivation from these embryos was carried out on mouse embryonic fibroblast (MEF) feeder cell layer adapting

standard protocol described elsewhere(Cowan et al., 2004). Briefly, mechanical ICM isolation was performed disrupting the trophectoderm taking great care not to affect the ICM cells and physically attaching them to the feeder cell layer using a glass needle-pipette. Initial outgrowths were observed after 24h. Medium was renovated every three days extracting half of the volume and replacing it with freshly prepared medium.

Human ESC culture medium consisted of KO-DMEM (Invitrogen, Carlsbad, CA) supplemented with 20% Knockout Serum Replacement (Invitrogen, Carlsbad, CA), 1X nonessential aminoacids (Gibco, Carlsbad, CA), 1X L-Glutamine (Invitrogen), 1X Penicillin/Streptomycin (Invitrogen), 1X Mercaptoethanol (Gibco) and 4ng/ml of FGF-2 (Invitrogen).

After the lines were established, cells were passaged every 4-6 days by incubation with 4mg/ml of collagenase IV (Worthington, Lakewood, NJ), for 60 minutes at 37ºC.

Immunostaining

hESC were washed three times with phosphate-buffered saline (PBS) buffer and fixed onto the culture dish using a 4% para-formaldehyde solution. Permeabilization and Blocking was performed simultaneously using a Triton-X100 1% plus a bovine serum albumin 0.1% solution in PBS buffer. After permeabilizing/blocking, overnight incubation at 4ºC with primary IgG antibodies against human OCT4 (BD, Franklin Lakes, NJ), NANOG (AbCam, Cambridge, Ma), SOX2 (AbCam) was performed. Samples were washed 3 times using PBS buffer and incubated in the appropriate Alexa-Fluor conjugated secondary antibody (Molecular Probes, Carlsband, CA) solution following the manufacturer's indications. Nuclear counter-staining was performed using 1µg/ml of Hoechst 33258 (Sigma, Saint Louis, MO).

Karyotyping

Between 20 and 30 metaphase spreads were karyotyped for each cell sample. Cells were treated with 0.1 lg/ml Colcemid (Invitrogen) for up to 4 hours, followed by dissociation with trypsin/versene. The cells were pelleted via centrifugation, resuspended in prewarmed 0.0375 M KCl hypotonic solution, and incubated for 10 minutes. Following a further centrifugation step, cells were resuspended in fixative (methanol:acetic acid 3:1). Metaphase spreads were prepared on glass microscope slides and G-banded by brief exposure to trypsin and stained with 4:1 Gurr’s/Leishmann’s.

Genomic DNA extraction

Genomic DNA (gDNA) extraction from cell samples was performed using DNeasy extraction kit (QIAGEN following manufacturer’s instructions. DNA concentration was quantified using Nanodrop (Thermo Scientific, Willmington, DE). Genomic DNA quality was assessed by a 1%

agarose gel electrophoresis with 1µg of material.

Microsatellite analysis

Primer information was obtained from the UNISTS database(supplementary table 1). Forward primers were tagged with 6-FAM. HI-FI PCR master mix (Invitrogen) was used to perform the PCRs in a total volume of 15 µl (0.8 µl of forward primer 10uM, 0.8 µl reverse primer 10uM, 1µl gDNA 50 µM, 12.4 µl of master mix). Reaction mixtures were amplified in low-profile 96-well plates with an thermal cycler (Applied Biosystems) using the following protocol: 95°C for 10min, followed by 30 cycles of 95°C for 1min, 55°C for 30s, 72°C for 1min, and a final extension of 72°C for 20 min. 1 µl of the amplification products were combined with 20 µl deionized formamide, including 1 µl internal lane standard LIZ-500 (Applied Biosystems), heat-denatured at 95°C for 2 min, snap-cooled on ice, and subjected to electrophoresis on an ABI PRISM 3100 Genetic Analyzer using POP4 and default conditions.

Samples whole-genome genotyping

Illumina HumanCNV-370 BeadChips (Illumina Inc., San Diego, CA) were used to genotype the different hESC lines samples. 750 ng of starting gDNA was required for each sample. Pre-amplification and hybridization process were performed following manufacturer's instruction which are based in the protocol described elsewhere(Peiffer, 2006). Fluorescent signals read by BeadStation hardware (Illumina Inc., San Diego, CA) were imported into the BeadStudio software version 3.2 (Illumina Inc., San Diego, CA) and normalized.

WG-SNP array data quality

Array call rates were > 0.9958 for all the samples. To ensure reproducibility in the chips preparation and processing, concordances between replicate samples were calculated for one line (WMC3) in two different chips as the number of concordant pairs divided by the number of successfully genotyped pairs. Results showed a concordance of 99.90%.

Genomic analysis

dChip (version December 2009)(Lin et al., 2004) was applied to normalized allele intensities exported from Illumina BeadStudio 3.2, along with genotype calls, as recommended. The SNP annotations were also included (physical positions provided by Illumina). The homozygosity analysis was performed using Hidden Markov Models for unpaired data, assuming a proportion of heterozygous SNPs of 35% for the Illumina HumanHap300v2 (determined from the normal samples). All other parameters were set to default values.

Results

ESC derivation from aneuploid embryos

From the 11 hatched blastocysts seeded on the feeder cell layers, 3 of them (27,7%) had a visible ICM and healthy trophectoderm. Five embryos (54,5%) were scored as medium quality (grade B) displaying a small but perceptible ICM, while the rest of the embryos (27,7%) showed no visible ICM . Twenty-four hours after seeding, 8 embryos (72.2%) had attached to the feeder cells, but only two of the initial embryonic attachments resulted in cell lines: embryo 3 and embryo 7, that were embryos with an acceptable quality (4BB and 4CB respectively) and visible ICM (See table 1).

Characterization of the hESC

To determine whether these lines were composed of hESC, phenotypic analysis of undifferentiated colonies was performed using cells from passage 10 of their culture history. Undifferentiated colonies from both lines display protein expression of the tested the pluripotency markers: Oct4, Nanog, Sox2 (Figure1). Both cell lines passed the pluripotency test in vivo, since both of them were able to form teratomas in SCID mice with presence of tissues from the three germ layers (Figure 2).

Genomic status of the hESC

Karyotypic status. Once established that both lines were composed of phenotypically normal human embryonic stem cells, we considered their genomic status. Karyotypes of the cell lines performed at passage 10 showed that both cell lines were euploid in all the cellular extensions analyzed: 46 XY for WMC2 and 46 XX for WMC3 (figure 3).

Microsatelite analysis. Since both cell lines, WMC2 and WMC3, were derived from monosomic embryos (monosomy 15 and monosomy 8+monosomy21 respectively), we further investigated the homozygosity in their genomes. This step is directed to detect whether the aneuploidies had been “corrected” by a duplication of the monosomic chromosomes at some point between embryonic biopsy and the stem cell derivation. Analysis of 10 microsatellite markers in each of the affected chromosomes (chr 15 for WMC2 and chr 8 and 21 for WMC3) are shown for both lines on table 2.

These markers were chosen to be homogeneously distributed throughout the chromosome length. The results show that chromosomes 8, 15 and 21 are heterozygous in both cell lines.

WG-SNP analysis. To further asses the genomic status of both cell lines, extending it to the rest of the chromosomes, a WG-SNP array was used. The analysis of the array-readings confirm showed a general distribution of heterozygous SNPs along the chromosomes that were initially reported as monosomic in the embryos the cell lines derive from. Moreover, the rest of the chromosomes also showed a general distribution of heterozygous and homozygous SNPs, confirming the euploid karyotypes found and their heterozygosity (figure 4), discarding any type of chromosomal duplication or any other major chromosomal aberration.

Discussion

The embryos used here for hESC derivation had been diagnosed with at least one numerical chromosomal aberration using FISH in a standard 9 probes PGS protocol applied to one blastomere. These embryos are regarded as non-suitable for replacement in the motherly uterus, and therefore, they are normally discarded or used for research. We attempted to derive hESC lines from 11 aneuploid embryos in order to study their genomic status, and how it affected the cell culture phenotype. As a result, two euploid fully pluripotent hESC lines were derived from monosomic-diagnosed embryos. Normal undifferentiated phenotype and normal pluripotent capacities were found for these two cell lines. These results are concordant with previous reports of chromosomally normal cell lines derived from aneuploid embryos(Lavon et al., 2008; Peura et al., 2008; Ilic et al., 2010). To explain this apparently conundrum, some authors have suggested chromosomal self-correction mechanisms in aneuploid embryos during hESC derivation that would explain the euploidy of these cells(Santiago Munné et al., 2005).

Following the PGS rationale, the presence of an aneuploid blastomere in an embryo implies that the rest may be aneuploid as well, which makes the embryo non-suitable for replacement in the patient's uterus. Generalized aneuploidy in an embryo could be either inherited from any of the two gametes (due to errors in meiosis I or meiosis II) or could be originated by global mitotic errors during embryonic divisions after fertilization(Yury Verlinsky, Strelchenko, et al., 2009). Since the two analyzed euploid hESC lines derive from monosomic embryos, proof of uniparental disomy (UPD) of chromosome 15 in WMC2 and chromosomes 8 and 21 in WMC3 may suggest a parentaly inherited aneuploidy. To test this possibility, we analyzed 10 microsatellites distributed throughout

the monosomic chromosomes from both cell lines. The general distribution of heterozygosity in the monosomy-diagnosed chromosomes ruled out the possibility of a gamete-derived self-corrected monosomy.

Therefore, after discarding meiotic aberration hypothesis, the diagnosed aneuploidy in these embryos may have arisen from mitotic errors during cell division after fertilization. In this scenario, PGS rationale assumes that a indefinite proportion of the blastomeres may be aneuploid due to the multiplicative effects of chaotic chromosomal segregation (it affects the two arising cells). The self-correction hypothesis during hESC derivation suggests the appearance of mosaic embryonic cell population that includes some aneuploid and some euploid cells(Santiago Munné et al., 2005). The proposed euploidization mechanisms are chromosomal isoduplication for monosomies and anaphase-lag correction(Kalousek et al., 1991), non-disjunction correction(Tarín et al., 1992) or chromosomal abolition(Los et al., 1998) for trisomies(Santiago Munné et al., 2005). The same authors proposed that the derivation process would enforce the survival of the euploidized cells in due to selection pressure(Lavon et al., 2008; Santiago Munné et al., 2005). Thus, if correction of aneuploidy takes place in any of the blastomeres as described, 100% of monosomy correction events and 33,3 % of trisomy correction events would have produced UPD. Trying to find any proof of these hypothesized self-correction of mitotic errors during hESC derivation, we performed a WG-SNP genotyping analysis in order to search for any trace of chromosomal homozygosity in the whole chromosomal set. However, no evidence of UPD was found for any chromosome in any of the two cell lines analyzed. Thus, euploidy of the hESC showed no evidence of self-correction mechanisms.

Euploid hESC lines from aneuploid embryos: not necessarily a misdiagnosis during PGS

PGS relies on chromosomes being identified by hybridization of fluorescently-labeled probes that can be detected using a fluorescence microscope. Indeed, faithfulness of the diagnosis can be compromised by several technical factors like fixation techniques, artifacts like overlapping signals or and loss of material during sample processing(Wells et al., 2008; Ruangvutilert et al., 2000). As a result, the efficiency of the technique is variable and depends on different factors, ranging 70% and 95% for a single cell(Emiliani et al., 2004; Michiels et al., 2006; S Munné et al., 1994; M. C. Magli et al., 2001; DeUgarte et al., 2008). In consonance, reports on FISH data using blastomeres depict a systematic and significant bias of two to threefold excess of monosomy compared to trisomy(DeUgarte et al., 2008; M. Sandalinas et al., 2001; Gosden, 2007). Therefore, a higher error

rate is anticipated for monosomy as hybridization can fail for many reasons, while an artefactual third signal is less likely to happen(DeUgarte et al., 2008). Indeed, FISH analysis of two blastomeres of the same day-3 embryo showed a 74.5% of discordance in the diagnostic between them(Coulam et al., 2007). In addition, probe-rehybridization analysis in blastomere samples found false positive rates ranging 17-25.6% in single hybridization protocols(Michiels et al., 2006; DeUgarte et al., 2008). Hence, some authors are suggesting a second and a default round of rehybridization to confirm diagnosis of monosomies and non-informative blastomere samples(Uher et al., 2009), but others did not found any significant differences in the efficiency of the two-blastomere analysis procedure(Michiels et al., 2006). The publication of all these reports has encouraged a debate over the accuracy of FISH applied to PGS. Indeed, from a clinical perspective, 10 randomized controlled trials of PGS-FISH have been published detailing the analysis of blastomeres from day-3 cleavage stage human embryos(Catherine Staessen et al., 2004; D. K. Gardner et al., 2004; Mastenbroek et al., 2007; Blockeel et al., 2008; Hardarson et al., 2008; Mersereau et al., 2008; Debrock et al., 2010; Meyer et al., 2009; Schoolcraft et al., 2009a; C. Staessen et al., 2008), and none of them has shown an improvement in delivery rates, with some showing a significant decrease in delivery rates after PGS.

Notwithstanding the possible flaws of the use of FISH to evaluate the genomic status of a single cell, a recent report by Vanneste and colleagues, found a high degree of chromosomal instability in day-3 / day-4 human embryos, implying a high proportion of aneuploid/euploid blastomere mosaicism in these embryos(Vanneste et al., 2009b). This study accounts for the use WG-CGH-array to analyze individually all the blastomeres from 23 embryos donated by 9 young (under 35) fertile couples, which found that 21 of them (91%) contained at least one chromosomally aberrant blastomere while only half of the embryos (47%) contained a minimum of one euploid blastomere. However, the robustness of these results has yet to be confirmed by other similar experiments by other laboratories.

Like FISH, the application of WG-genotyping arrays to genetically screen single human blastomeres is technically challenged by the small amount of cellular material used. Particularly, WG-genotyping requires the amplification of the single-cell DNA to produce the yield necessary to perform the array-hybridization step. To do so, whole genome amplification (WGA) by genome fragmentation into PCR-amplifiable units using universal adaptors (Omniplex™ technology) (Langmore, 2002; Barker, 2004) or WGA by multiple displacement amplification (MDA) using the Phi29 polymerase(Blanco et al., 1989; Dean et al., 2002) have been implemented. Indeed, several

groups have reported single cell genomic DNA amplification to perform WG-SNP or WG-CGH arrays genotyping procedures(Spits et al., 2006; Pan et al., 2008). However, single-cell WGA methods are notoriously susceptible to strong amplification bias such as the failure of amplification of one of the two alleles (allele dropout) and excess amplification of one allele or unequal amplification of the two alleles (preferential amplification)(Spits et al., 2006; Berthier-Schaad et al., 2007; Iwamoto et al.; Park, 2005; Talseth-Palmer et al., 2008; Peiffer, 2006). In their recent report, Vanneste et al. claim that only 12 % of the blastomeres failed to reach the required DNA yield after MDA amplification, but an additional 34% of the blastomeres were excluded of the analysis after array-hybridization quality control. Overall, a 46% of blastomeres (almost 1 out of 2) had undergone an incomplete or biased DNA amplification. Remarkably, in non-amplified samples, array hybridization call rates under 95% are considered not acceptable for analysis whereas all the single blastomere DNA-hybridization experiments reported by Vanneste et al. yielded call rates under 86%. Thus, the CGH/SNP analysis algorithms were adapted for these single-cell samples with low call rates. Therefore, these “adapted algorithms” need to be rigorously revised using karyotypically undefined random samples amplified from different initial DNA concentrations (reaching the single cell DNA concentration) and be matched to the correspondent non-amplified genomic-DNA samples. Therefore, although the use of WG-CGH or WG-SNP genotyping arrays as a valid technique to substitute FISH in the PGS routine has been widely anticipated(Wells et al., 2008; Coskun and Alsmadi, 2007) and recently applied(Treff et al., 2010), a critical assessment of its accuracy would be required before applying this expensive technique to clinical protocols.

Evidence of aneuploidies, mosaicism during early embryonic development

Recurrent evidence of chromosomal aneuploidies during embryogenesis has been found thanks to the genomic analysis of early trimester miscarriages wastage. Like in prenatal genetic diagnosis protocols, miscarriage aneuploidy analysis is routinely performed using standard cytogenetic techniques (karyotype or FISH analysis), after in vitro culture the embryonic cell samples(Lebedev et al., 2004; Menasha et al., 2005). In some cases, when embryonic wastage fails to expand in vitro, WG-CGH protocols have been used to determine the genomic status of these samples, thanks to the sufficient DNA yield that can be obtained from them(Dória et al., 2009; Fritz et al., 2001) . Thus, in both type of protocols, the availability of cellular material in this cases makes the diagnosis far less susceptible of error and biases than the single-cell analysis during PGS.

Results of these analyses have shown that a very high proportion of the first trimester miscarriages

(50-70%) are correlated with aneuploidies(Lebedev et al., 2004; Dória et al., 2009). Statistics on the frequency of miscarriage during the first twelve weeks of pregnancy are elusive because an undefined proportion of embryonic implantation events remain unnoticed but some authors have suggested that only about 30% of conceptions reach the second trimester of gestation(Macklon et al., 2002). The combination of these two statistics depicts an scenario suggesting a determinant contribution of chromosomal aneuploidies to the lack of viability of the embryonic development after the first trimester of gestation. However, certain trisomies are known to be compatible with to-term pregnancies, namely trisomies 8, 9, 13 18 and 21. However, for these trisomies, very high lethality rates are observed either perinatally or during the first year after birth (with the exception of trisomy 21)(Fineman et al., 1979; ANNERÉN and SEDIN, 1981; Taylor, 1968; Magenis et al., 1968; Bornstein et al., 2009; Warburton et al., 2004). Interestingly, abundant cases of trisomic/euploid mosaicism affecting miscarriage wastage material(Wells et al., 2008; Dória et al., 2009; Celep et al., 2006) and also newborn babies (Fineman et al., 1975; Delatycki and R. J. M. Gardner, 1997; G. N. Wilson et al., 1985; Tucker et al., 2007; Bornstein et al., 2009)have also been reported.

In contrast to the chromosomal aberrations found in spontaneous abortions, very limited evidences of originally embryonic chromosomal aberrations have been detected through hESC derivation. In fact, in addition to our results, eighteen euploid hESC have been reported to be derived from chromosomally abnormal embryos diagnosed after PGS-FISH(Lavon et al., 2008; Peura et al., 2008; Ilic et al., 2010). Interestingly, one of these reports accounts for two hESC lines with a trisomic karyotype at passage 8, that had been derived from aneuploid embryos(Peura et al., 2008). However no concordance was found between trisomic chromosomes in these hESC lines and the abnormal chromosomes of the embryo they derive from. In addition to this finding, an other report claims the derivation of dozens of aneuploid cell lines from aneuploid embryos, with a direct correspondence between the aberration diagnosed in the embryos and the hESC lines(Yury Verlinsky, N. H. Zech, et al., 2009). However, no other group has replicated these findings and no other direct proof of correlation between diagnosed aneuploidy in embryos and aneuploidy in hESC derived from them has been published. However, if this results were to be confirmed, it would imply a major drawback in the efficiency and reliability of any type of PGS technique, because no correlation between a possible detected aneuploidy in a single blastomere and healthy embryonic development potential of that embryo would be possible to make. Nevertheless, in such scenario it would seem reasonable to suggest the existence of a selection pressure directed to ensure the survival and progression of the euploid cells during embryogenesis. This euploid blastomere

selection process during embryogenesis would be backed by the healthy newborn statistics in the general population (abnormal karyotypes rates after prenatal testing ranging between 0.85– 0.92%) (Gjerris et al., 2008; Bonduelle et al., 2002) and the fact that practically all hESC lines derived were primarily euploid (previous to extended in vitro culture).

Nonetheless, some factors involved in either the success of the to-term embryo development or the hESC derivation are unknown or can not be controlled. As a reflection of this, both the best delivery rates per transfer (12,8%(Ata and Urman, 2010)) and the best hESC derivation rates (54%, unpublished data) remain under certain thresholds after some years of perfecting their procedures. Indeed, despite being implemented to improve delivery rates for different etiologies of infertility, figures have challenged the efficiency of PGS using FISH to better the success rate of embryonic development in form of a higher take-home-baby statistics(Catherine Staessen et al., 2004; Mastenbroek et al., 2007; Blockeel et al., 2008; Hardarson et al., 2008; Mersereau et al., 2008; Debrock et al., 2010; C. Staessen et al., 2008; Schoolcraft et al., 2009b). In addition, we and others shown that some aneuploidy-diagnosed embryos after FISH-PGS originated euploid hESC lines showing no evidence of any chromosomal aberration. Whether this incoherence has been caused by a PGS misdiagnosis or is caused by the genomic mosaicism of human preimplantational embryos is an important scientific issue to be addressed to improve IVF pregnancy rates.

Bibliography

ANNERÉN, G., and SEDIN, G. (1981). TRISOMY 9 SYNDROME. Acta Paediatrica *70*, 125-128.

Ata, B., and Urman, B. (2010). ART register data on delivery rates. Hum. Reprod *25*, 805-806; author reply 806-807.

Barker, D. L. (2004). Two Methods of Whole-Genome Amplification Enable Accurate Genotyping Across a 2320-SNP Linkage Panel. Genome Research *14*, 901-907.

Berthier-Schaad, Y., Kao, W. H. L., Coresh, J., Zhang, L., Ingersoll, R. G., Stephens, R., and Smith, M. W. (2007). Reliability of high-throughput genotyping of whole genome amplified DNA in SNP genotyping studies. Electrophoresis *28*, 2812-2817.

Blanco, L., Bernad, A., Lázaro, J. M., Martín, G., Garmendia, C., and Salas, M. (1989). Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. J. Biol. Chem *264*, 8935-8940.

Blockeel, C., Schutyser, V., De Vos, A., Verpoest, W., De Vos, M., Staessen, C., Haentjens, P., Van der Elst, J., and Devroey, P. (2008). Prospectively randomized controlled trial of PGS in IVF/ICSI patients with poor implantation. Reprod. Biomed. Online *17*, 848-854.

Bonduelle, M., Van Assche, E., Joris, H., Keymolen, K., Devroey, P., Van Steirteghem, A., and Liebaers, I. (2002). Prenatal testing in ICSI pregnancies: incidence of chromosomal anomalies in 1586 karyotypes and relation to sperm parameters. Hum. Reprod. *17*, 2600-2614.

Bornstein, E., Lenchner, E., Donnenfeld, A., Kapp, S., Keeler, S. M., and Divon, M. Y. (2009). Comparison of modes of ascertainment for mosaic vs complete trisomy 21. American Journal of Obstetrics and Gynecology *200*, 440.e1-440.e5.

Celep, F., Karagüzel, A., Özeren, M., and Bozkaya, H. (2006). The frequency of chromosomal abnormalities in patients with reproductive failure. European Journal of Obstetrics & Gynecology and Reproductive Biology *127*, 106-109.

Coskun, S., and Alsmadi, O. (2007). Whole genome amplification from a single cell: a new era for preimplantation genetic diagnosis. Prenat. Diagn. *27*, 297-302.

Coulam, C. B., Jeyendran, R. S., Fiddler, M., and Pergament, E. (2007). Discordance among blastomeres renders preimplantation genetic diagnosis for aneuploidy ineffective. J. Assist. Reprod. Genet *24*, 37-41.

Cowan, C. A., Klimanskaya, I., McMahon, J., Atienza, J., Witmyer, J., Zucker, J. P., Wang, S., Morton, C. C., McMahon, A. P., Powers, D., et al. (2004). Derivation of embryonic stem-cell lines from human blastocysts. N. Engl. J. Med *350*, 1353-1356.

Dean, F. B., Hosono, S., Fang, L., Wu, X., Faruqi, A. F., Bray-Ward, P., Sun, Z., Zong, Q., Du, Y., Du, J., et al. (2002). Comprehensive human genome amplification using multiple displacement amplification. Proc. Natl. Acad. Sci. U.S.A *99*, 5261-5266.

Debrock, S., Melotte, C., Spiessens, C., Peeraer, K., Vanneste, E., Meeuwis, L., Meuleman, C.,

Frijns, J., Vermeesch, J. R., and D'Hooghe, T. M. (2010). Preimplantation genetic screening for aneuploidy of embryos after in vitro fertilization in women aged at least 35 years: a prospective randomized trial. Fertil. Steril *93*, 364-373.

Delatycki, M., and Gardner, R. J. M. (1997). Three cases of trisomy 13 mosaicism and a review of the literature. Clinical Genetics *51*, 403-407.

DeUgarte, C. M., Li, M., Surrey, M., Danzer, H., Hill, D., and DeCherney, A. H. (2008). Accuracy of FISH analysis in predicting chromosomal status in patients undergoing preimplantation genetic diagnosis. Fertility and Sterility *90*, 1049-1054.

Dória, S., Carvalho, F., Ramalho, C., Lima, V., Francisco, T., Machado, A. P., Brandão, O., Sousa, M., Matias, A., and Barros, A. (2009). An efficient protocol for the detection of chromosomal abnormalities in spontaneous miscarriages or foetal deaths. European Journal of Obstetrics & Gynecology and Reproductive Biology *147*, 144-150.

Emiliani, S., Gonzalez-Merino, E., Englert, Y., and Abramowicz, M. (2004). Comparison of the validity of preimplantation genetic diagnosis for embryo chromosomal anomalies by fluorescence in situ hybridization on one or two blastomeres. Genet. Test *8*, 69-72.

Fauser, B. C. (2008). Preimplantation genetic screening: the end of an affair? Hum. Reprod. *23*, 2622-2625.

Fineman, R. M., Ablow, R. C., Breg, W. R., Wing, S. D., RoseSteven, J. S., Rothman, L. G., and Warpinski, J. (1979). Complete and partial trisomy of different segments of chromosome 8: case reports and review. Clinical Genetics *16*, 390-398.

Fineman, R. M., Ablow, R. C., Howard, R. O., Albright, J., and Breg, W. R. (1975). Trisomy 8 Mosaicism Syndrome. Pediatrics *56*, 762-767.

Fritz, B., Hallermann, C., Olert, J., Fuchs, B., Bruns, M., Aslan, M., Schmidt, S., Coerdt, W., Müntefering, H., and Rehder, H. (2001). Cytogenetic analyses of culture failures by comparative genomic hybridisation (CGH)-Re-evaluation of chromosome aberration rates in early spontaneous abortions. Eur. J. Hum. Genet *9*, 539-547.

Frumkin, T., Malcov, M., Telias, M., Gold, V., Schwartz, T., Azem, F., Amit, A., Yaron, Y., and Ben-Yosef, D. Human embryonic stem cells carrying mutations for severe genetic disorders. In Vitro Cellular & Developmental Biology - Animal. Available at: http://dx.doi.org/10.1007/s11626-010-9275-5 [Accessed April 15, 2010].

Gardner, D. K., Surrey, E., Minjarez, D., Leitz, A., Stevens, J., and Schoolcraft, W. B. (2004). Single blastocyst transfer: a prospective randomized trial. Fertil. Steril *81*, 551-555.

Gianaroli, L., Magli, M. C., Ferraretti, A. P., and Munné, S. (1999). Preimplantation diagnosis for aneuploidies in patients undergoing in vitro fertilization with a poor prognosis: identification of the categories for which it should be proposed. Fertil. Steril *72*, 837-844.

Gjerris, A., Loft, A., Pinborg, A., Christiansen, M., and Tabor, A. (2008). Prenatal testing among women pregnant after assisted reproductive techniques in Denmark 1995-2000: a national cohort study. Hum. Reprod. *23*, 1545-1552.

Gosden, R. (2007). Genetic test may lead to waste of healthy embryos. Nature *446*, 372.

Hardarson, T., Hanson, C., Lundin, K., Hillensjö, T., Nilsson, L., Stevic, J., Reismer, E., Borg, K., Wikland, M., and Bergh, C. (2008). Preimplantation genetic screening in women of advanced maternal age caused a decrease in clinical pregnancy rate: a randomized controlled trial. Hum. Reprod *23*, 2806-2812.

Harper, J., Coonen, E., De Rycke, M., Fiorentino, F., Geraedts, J., Goossens, V., Harton, G., Moutou, C., Pehlivan Budak, T., Renwick, P., et al. (2010). What next for preimplantation genetic screening (PGS)? A position statement from the ESHRE PGD Consortium steering committee. Hum. Reprod *25*, 821-823.

Harper, J., Sermon, K., Geraedts, J., Vesela, K., Harton, G., Thornhill, A., Pehlivan, T., Fiorentino, F., SenGupta, S., de Die-Smulders, C., et al. (2008). What next for preimplantation genetic screening? Hum. Reprod *23*, 478-480.

Ilic, D., Caceres, E., Lu, S., Julian, P., Foulk, R., and Krtolica, A. (2010). Effect of karyotype on successful human embryonic stem cell derivation. Stem Cells Dev *19*, 39-46.

Iwamoto, K., Bundo, M., Ueda, J., Nakano, Y., Ukai, W., Hashimoto, E., Saito, T., and Kato, T. Detection of Chromosomal Structural Alterations in Single Cells by SNP Arrays: A Systematic Survey of Amplification Bias and Optimized Workflow. PLoS ONE *2*.

Kalousek, D. K., Howard-Peebles, P. N., Olson, S. B., Barrett, I. J., Dorfmann, A., Black, S. H., Schulman, J. D., and Wilson, R. D. (1991). Confirmation of CVS mosaicism in term placentae and high frequency of intrauterine growth retardation association with confined placental mosaicism. Prenat. Diagn *11*, 743-750.

Langmore, J. P. (2002). Rubicon Genomics, Inc. Pharmacogenomics *3*, 557-560.

Lavon, N., Narwani, K., Golan-Lev, T., Buehler, N., Hill, D., and Benvenisty, N. (2008). Derivation of euploid human embryonic stem cells from aneuploid embryos. Stem Cells *26*, 1874-1882.

Lebedev, I. N., Ostroverkhova, N. V., Nikitina, T. V., Sukhanova, N. N., and Nazarenko, S. A. (2004). Features of chromosomal abnormalities in spontaneous abortion cell culture failures detected by interphase FISH analysis. Eur J Hum Genet *12*, 513-520.

Lin, M., Wei, L., Sellers, W. R., Lieberfarb, M., Wong, W. H., and Li, C. (2004). dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics *20*, 1233-1240.

Los, F. J., van Opstal, D., van den Berg, C., Braat, A. P., Verhoef, S., Wesby-van Swaay, E., van den Ouweland, A. M., and Halley, D. J. (1998). Uniparental disomy with and without confined placental mosaicism: a model for trisomic zygote rescue. Prenat. Diagn *18*, 659-668.

Macklon, N. S., Geraedts, J. P. M., and Fauser, B. C. J. M. (2002). Conception to ongoing pregnancy: the 'black box' of early pregnancy loss. Hum. Reprod. Update *8*, 333-343.

Magenis, R. E., Hecht, F., and Milham, S. (1968). Trisomy 13 (D1) syndrome: studies on parental

age, sex ratio, and survival. J. Pediatr *73*, 222-228.

Magli, M. C., Sandalinas, M., Escudero, T., Morrison, L., Ferraretti, A. P., Gianaroli, L., and Munné, S. (2001). Double locus analysis of chromosome 21 for preimplantation genetic diagnosis of aneuploidy. Prenat. Diagn *21*, 1080-1085.

Mastenbroek, S., Twisk, M., van Echten-Arends, J., Sikkema-Raddatz, B., Korevaar, J. C., Verhoeve, H. R., Vogel, N. E. A., Arts, E. G. J. M., de Vries, J. W. A., Bossuyt, P. M., et al. (2007). In vitro fertilization with preimplantation genetic screening. N. Engl. J. Med *357*, 9-17.

Menasha, J., Levy, B., Hirschhorn, K., and Kardon, N. B. (2005). Incidence and spectrum of chromosome abnormalities in spontaneous abortions: new insights from a 12-year study. Genet. Med *7*, 251-263.

Mersereau, J. E., Pergament, E., Zhang, X., and Milad, M. P. (2008). Preimplantation genetic screening to improve in vitro fertilization pregnancy rates: a prospective randomized controlled trial. Fertil. Steril *90*, 1287-1289.

Meyer, L. R., Klipstein, S., Hazlett, W. D., Nasta, T., Mangan, P., and Karande, V. C. (2009). A prospective randomized controlled trial of preimplantation genetic screening in the "good prognosis" patient. Fertil. Steril *91*, 1731-1738.

Michiels, A., Van Assche, E., Liebaers, I., Van Steirteghem, A., and Staessen, C. (2006). The analysis of one or two blastomeres for PGD using fluorescence in-situ hybridization. Hum. Reprod. *21*, 2396-2402.

Munné, S., Weier, H. U., Grifo, J., and Cohen, J. (1994). Chromosome mosaicism in human embryos. Biology of Reproduction *51*, 373-379.

Munné, S., Sandalinas, M., Escudero, T., Velilla, E., Walmsley, R., Sadowy, S., Cohen, J., and Sable, D. (2003). Improved implantation after preimplantation genetic diagnosis of aneuploidy. Reprod. Biomed. Online *7*, 91-97.

Munné, S., Velilla, E., Colls, P., Garcia Bermudez, M., Vemuri, M. C., Steuerwald, N., Garrisi, J., and Cohen, J. (2005). Self-correction of chromosomally abnormal embryos in culture and implications for stem cell production. Fertil. Steril *84*, 1328-1334.

Pan, X., Urban, A. E., Palejev, D., Schulz, V., Grubert, F., Hu, Y., Snyder, M., and Weissman, S. M. (2008). A procedure for highly specific, sensitive, and unbiased whole-genome amplification. Proceedings of the National Academy of Sciences *105*, 15499-15504.

Park, J. W. (2005). Comparing Whole-Genome Amplification Methods and Sources of Biological Samples for Single-Nucleotide Polymorphism Genotyping. Clinical Chemistry *51*, 1520-1523.

Peiffer, D. A. (2006). High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. Genome Research *16*, 1136-1148.

Peura, T., Bosman, A., Chami, O., Jansen, R. P. S., Texlova, K., and Stojanov, T. (2008).

Karyotypically normal and abnormal human embryonic stem cell lines derived from PGD-analyzed embryos. Cloning Stem Cells *10*, 203-216.

Reubinoff, B. E., Pera, M. F., Fong, C., Trounson, A., and Bongso, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nat Biotech *18*, 399-404.

Ruangvutilert, P., Delhanty, J. D., Serhal, P., Simopoulou, M., Rodeck, C. H., and Harper, J. C. (2000). FISH analysis on day 5 post-insemination of human arrested and blastocyst stage embryos. Prenat. Diagn *20*, 552-560.

Sandalinas, M., Sadowy, S., Alikani, M., Calderon, G., Cohen, J., and Munne, S. (2001). Developmental ability of chromosomally abnormal human embryos to develop to the blastocyst stage. Hum. Reprod. *16*, 1954-1958.

Santiago Munné, and [1] (2002). Preimplantation genetic diagnosis of numerical and structural chromosome abnormalities. Reproductive BioMedicine Online *4*, 183-196.

Schoolcraft, W. B., Katz-Jaffe, M. G., Stevens, J., Rawlins, M., and Munne, S. (2009a). Preimplantation aneuploidy testing for infertile patients of advanced maternal age: a randomized prospective trial. Fertil. Steril *92*, 157-162.

Schoolcraft, W. B., Katz-Jaffe, M. G., Stevens, J., Rawlins, M., and Munne, S. (2009b). Preimplantation aneuploidy testing for infertile patients of advanced maternal age: a randomized prospective trial. Fertil. Steril *92*, 157-162.

Sermon, K., Simon, C., Braude, P., Viville, S., Borstlap, J., and Veiga, A. (2009). Creation of a registry for human embryonic stem cells carrying an inherited defect: joint collaboration between ESHRE and hESCreg. Hum. Reprod. *24*, 1556-1560.

Spits, C., Le Caignec, C., De Rycke, M., Van Haute, L., Van Steirteghem, A., Liebaers, I., and Sermon, K. (2006). Optimization and evaluation of single-cell whole-genome multiple displacement amplification. Hum. Mutat *27*, 496-503.

Staessen, C., Verpoest, W., Donoso, P., Haentjens, P., Van der Elst, J., Liebaers, I., and Devroey, P. (2008). Preimplantation genetic screening does not improve delivery rate in women under the age of 36 following single-embryo transfer. Hum. Reprod. *23*, 2818-2825.

Staessen, C., Platteau, P., Van Assche, E., Michiels, A., Tournaye, H., Camus, M., Devroey, P., Liebaers, I., and Van Steirteghem, A. (2004). Comparison of blastocyst transfer with or without preimplantation genetic diagnosis for aneuploidy screening in couples with advanced maternal age: a prospective randomized controlled trial. Hum. Reprod. *19*, 2849-2858.

Talseth-Palmer, B., Bowden, N., Hill, A., Meldrum, C., and Scott, R. (2008). Whole genome amplification and its impact on CGH array profiles. BMC Research Notes *1*, 56.

Tarín, J. J., Conaghan, J., Winston, R. M., and Handyside, A. H. (1992). Human embryo biopsy on the 2nd day after insemination for preimplantation diagnosis: removal of a quarter of embryo retards cleavage. Fertil. Steril *58*, 970-976.

Taylor, A. I. (1968). Autosomal trisomy syndromes: a detailed study of 27 cases of Edwards' syndrome and 27 cases of Patau's syndrome. J. Med. Genet *5*, 227-252.

Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., and Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. Science (New York, N.Y *282*, 1145-7.

Treff, N. R., Su, J., Tao, X., Levy, B., and Scott, R. T. (2010). Accurate single cell 24 chromosome aneuploidy screening using whole genome amplification and single nucleotide polymorphism microarrays. Fertil Steril. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20188357 [Accessed April 27, 2010].

Trounson, A., and Mohr, L. (1983). Human pregnancy following cryopreservation, thawing and transfer of an eight-cell embryo. Nature *305*, 707-709.

Tucker, M. E., Garringer, H. J., and Weaver, D. D. (2007). Phenotypic spectrum of mosaic trisomy 18: Two new patients, a literature review, and counseling issues. American Journal of Medical Genetics Part A *143A*, 505-517.

Twisk, M., Mastenbroek, S., Hoek, A., Heineman, M., van der Veen, F., Bossuyt, P. M., Repping, S., and Korevaar, J. C. (2008). No beneficial effect of preimplantation genetic screening in women of advanced maternal age with a high risk for embryonic aneuploidy. Hum. Reprod. *23*, 2813-2817.

Uher, P., Baborova, P., Kralickova, M., Zech, M. H., Verlinsky, Y., and Zech, N. H. (2009). Non-informative results and monosomies in PGD: the importance of a third round of re-hybridization. Reprod. Biomed. Online *19*, 539-546.

Vanneste, E., Voet, T., Le Caignec, C., Ampe, M., Konings, P., Melotte, C., Debrock, S., Amyere, M., Vikkula, M., Schuit, F., et al. (2009a). Chromosome instability is common in human cleavage-stage embryos. Nat Med *15*, 577-583.

Vanneste, E., Voet, T., Le Caignec, C., Ampe, M., Konings, P., Melotte, C., Debrock, S., Amyere, M., Vikkula, M., Schuit, F., et al. (2009b). Chromosome instability is common in human cleavage-stage embryos. Nat. Med *15*, 577-583.

Verlinsky, Y., Cieslak, J., Freidine, M., Ivakhnenko, V., Wolf, G., Kovalinskaya, L., White, M., Lifchez, A., Kaplan, B., Moise, J., et al. (1995). Diagnosing and preventing inherited disease: Pregnancies following pre-conception diagnosis of common aneuploidies by fluorescent in-situ hybridization. Hum. Reprod. *10*, 1923-1927.

Verlinsky, Y., Tur-Kaspa, I., Cieslak, J., Bernal, A., Morris, R., Taranissi, M., Kaplan, B., and Kuliev, A. (2005). Preimplantation testing for chromosomal disorders improves reproductive outcome of poor-prognosis patients. Reproductive BioMedicine Online *11*, 219-225.

Verlinsky, Y., Strelchenko, N., Kukharenko, V., Zech, N. H., Shkumatov, A., Zlatopolsky, Z., and Kuliev, A. (2009a). Impact of meiotic and mitotic non-disjunction on generation of human embryonic stem cell lines. Reprod. Biomed. Online *18*, 120-126.

Verlinsky, Y., Zech, N. H., Strelchenko, N., Kukharenko, V., Shkumatov, A., Zlatopolsky, Z., and

Kuliev, A. (2009b). Correlation between preimplantation genetic diagnosis for chromosomal aneuploidies and the efficiency of establishing human ES cell lines. Stem Cell Res *2*, 78-82.

Warburton, D., Dallaire, L., Thangavelu, M., Ross, L., Levin, B., and Kline, J. (2004). Trisomy Recurrence: A Reconsideration Based on North American Data. The American Journal of Human Genetics *75*, 376-385.

Wells, D., Alfarawati, S., and Fragouli, E. (2008). Use of comprehensive chromosomal screening for embryo assessment: microarrays and CGH. Mol. Hum. Reprod. *14*, 703-710.

Wilson, G. N., Raj, A., Baker, D., Opitz, J. M., and Reynolds, J. F. (1985). The phenotypic and cytogenetic spectrum of partial trisomy 9. American Journal of Medical Genetics *20*, 277-282.

Xu, K., Huang, T., Liu, T., Shi, Z., and Rosenwaks, Z. (1998). Improving the Fixation Method for Preimplantation Genetic Diagnosis by Fluorescent in Situ Hybridization. Journal of Assisted Reproduction and Genetics *15*, 570-574.Bibliography

Embryo	PGS diagnosis	Fresh / Frozen	Visible ICM	Initial attachment	Cell line
1	Trisomy 15	Frozen	No	Yes	NO
2	Monsomy 13	Frozen	Yes	No	NO
3	Monosomy 15	Frozen	Yes	Yes	WMC2
4	Trisomy 17	Fresh	No	No	NO
5	Trisomy 15	Frozen	Yes	Yes	NO
6	Monosomy 18	Fresh	Yes	Yes	NO
7	Monosomy 8 + Monosomy 21	Fresh	Yes	Yes	WMC3
8	Trisomy 22	Frozen	No	No	NO
9	Monosomy21	Frozen	Yes	Yes	NO
10	Mosomy 16+Trisomy 22	Fresh	Yes	Yes	NO
11	Trisomy 22	Frozen	Yes	Yes	NO

Table 1. Eleven embryos were seeded to derive hESC. Characteristics of the embryos donated after PGS-aneuploidy diagnosis, and their outcome.

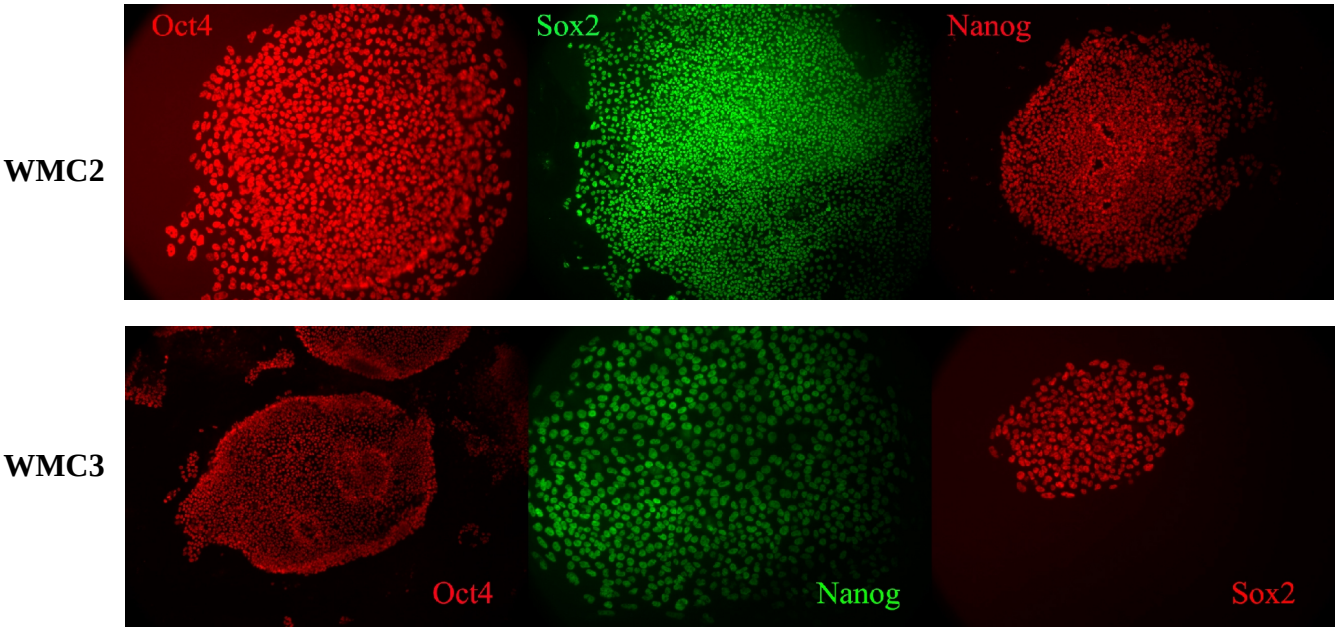


Figure 1. Immunofluroescent staining of pluripotency markers (SOX2, NANOG, and OCT4) in undifferentiated colonies of WMC2 and WMC3 cell lines.

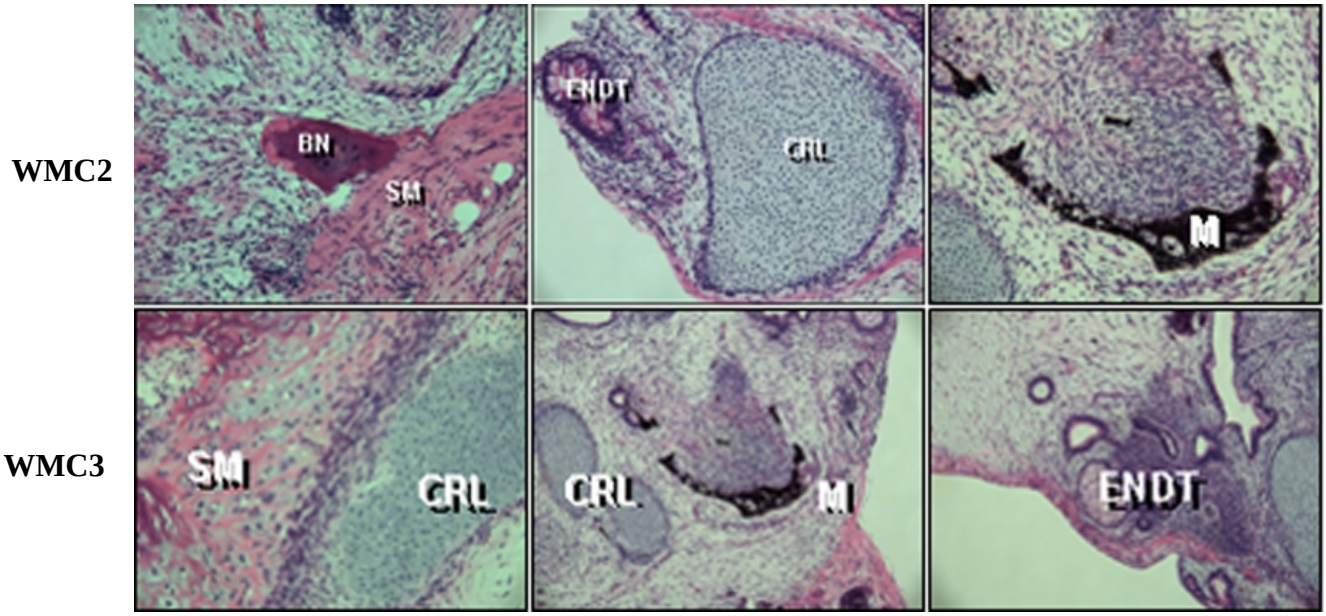
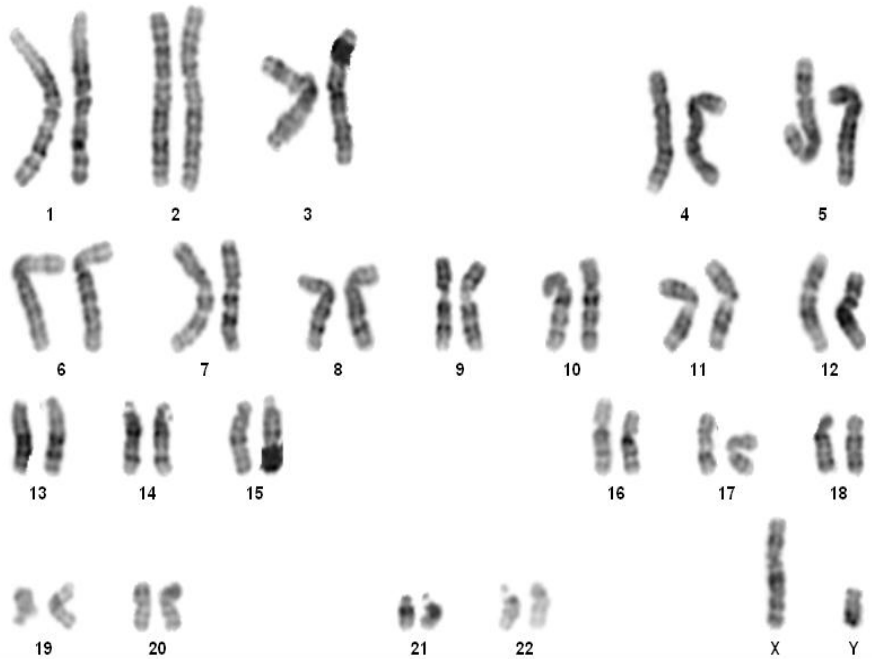


Figure 2. In vivo differentiation of WMC2 and WMC3 by teratoma formation after injection of one million cells (approximately) subcutaneously in SCID mice. Tissues found during histological analysis of the teratomas: BN, bone (mesoderm); SM, smooth muscle (endoderm); ENDT, endothelium (endoderm); CRL, cartilage (mesoderm); M, melanocites (ectoderm).

A WMC2



B WMC3

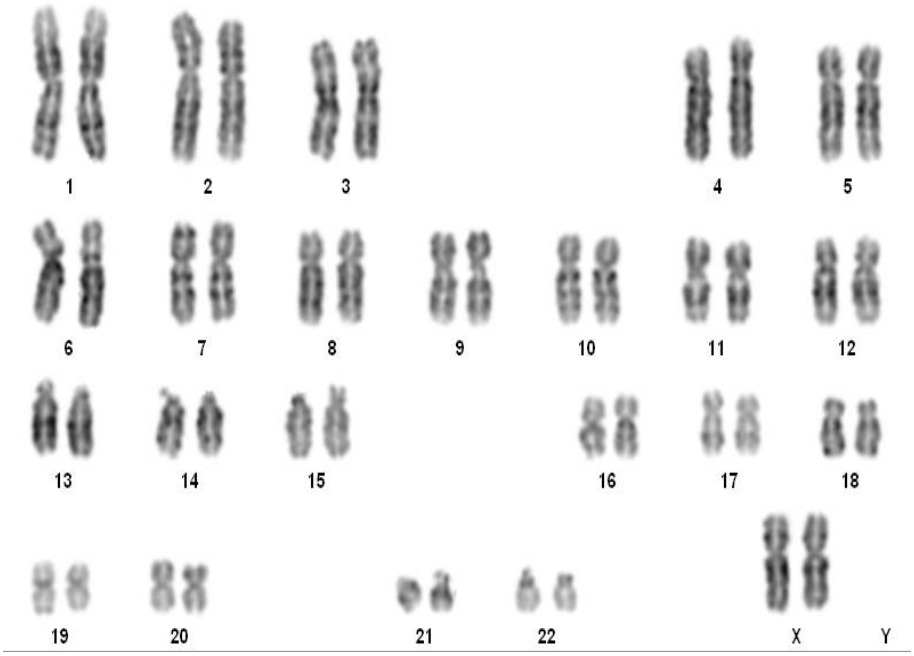


Figure 3. Karyotype analysis of WMC2 and WMC3 cell lines at passage 10. Both cell lines were euploid at this point of their passage history. WMC2 is a male line (46 XY) and WMC3 is a female line (46 XX).

A

Microsatellite	WMC2		WMC1 (control)	
Chromosome 15				
D15S1021	145	133	145	130
D15S1002	110	108	123	108
D15S1040	206	206	196	194
D15S196	352	348	352	352
D15S538	290	285	293	285
D15S818	151	155	151	151
D15S1045	217	205	225	217
D15S130	187	187	187	NA
D15S985	193	189	193	189
D15S642	197	197	209	207

B

Microsatellite	WMC3		WMC1 (control)	
Chromosome 8				
D8S504	200	192	200	200
D8S1731	222	218	232	216
D8S1725	158	158	158	156
D8S1121	113	113	119	119
D8S1816	143	139	143	139
D8S570	159	155	161	159
D8S88	92	88	94	78
D8S1768	203	201	203	199
D8S1793	110	108	140	108
D8S1926	156	156	156	156
Chromosome 21				
D21S1993	126	126	129	126
D21S1431	171	165	174	171
D21S1256	111	111	121	113
D21S1884	182	182	190	186
D21S269	255	241	249	249
D21S262	NA	NA	148	146
D21S1252	250	238	240	232
D21S168	113	106	117	113
D21S1260	210	204	210	208
D21S1912	196	192	200	196

Table 2. Microsatellite analysis results (in bp). Different length in the alleles imply heterozygosity of the specific microsatellite. **A.** WMC2 sample was investigated for chromosome 15. **B.** WMC3 sample was investigated for chromosome 8 and 21. None of the chromosomes tested had the same parental origin. A control cell line (WMC1) was used as a positive control and for comparative purposes.

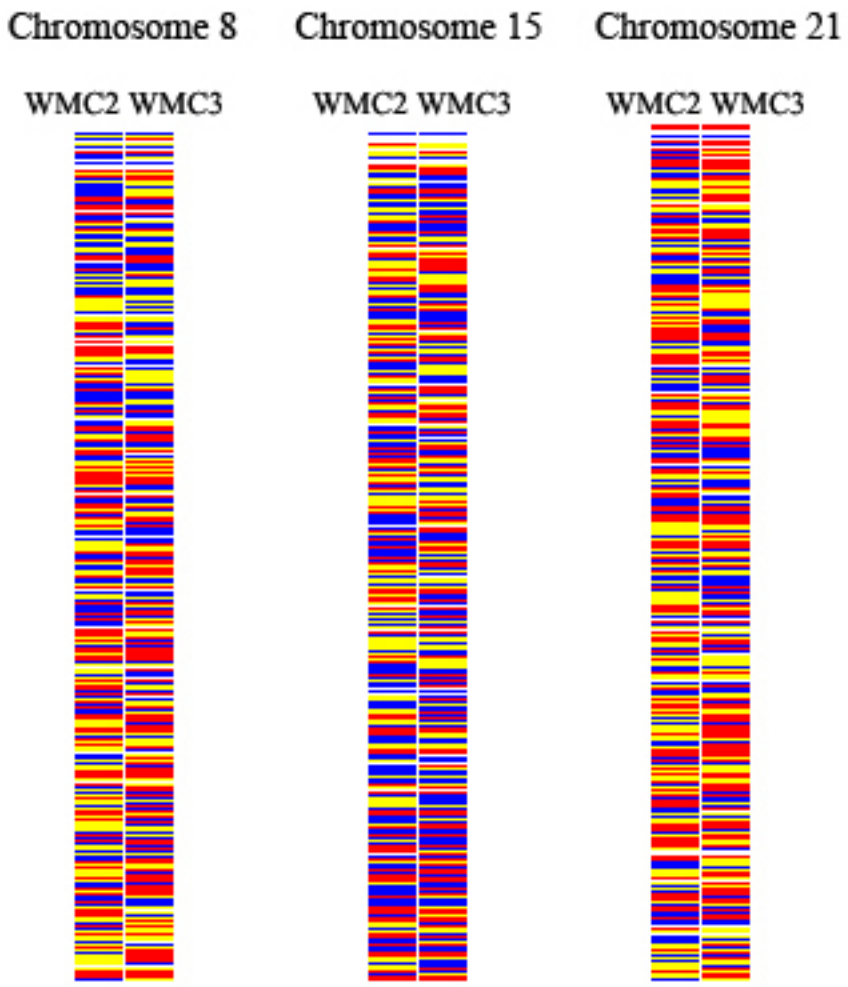


Figure 4. Inferred calls from the SNP Genotyping Array of chromosomes 8, 15 and 21 for WMC2 and WMC3. Red: AA, blue, BB, and yellow AB. Heterozygous locus (yellow) are distributed homogeneously in these three chromosomes as well as the rest of autosomes analyzed. Graphic generated with D-Chip

chr	marker	primer	phys.loc	gen.loc	Forward	Reverse	prod. size	GenBank
8	D8S504	AFM197xg5	2126744	0	ACTGGGTCACGAGGGA	CATGCCCATTTTCCAG	193-203 (bp)	Z23474
8	D8S1731	AFMa311wd1	18868342	26.31	CCAAGCAAATCATGGAAATC	AGCAAACTTATCCCACAAGG	217-241 (bp)	Z52884
8	D8S1725	AFMa226zh5	29719227	43.74	ACACCTACCAAGGACTGCTG	CGGGCTGGGACAATTTT	158-164 (bp)	Z52612
8	D8S1121	ATA22A07	42104050	55.41	TCACTCCATCAGTGGGTCTT	CTTTTTTGCTTCAGGAACCA	100-124 (bp)	G08653
8	D8S1816	AFM021te1	64618883	68.63	TGCACCCTTAAAAAGCATCG	ACTTGCGAACATGGGATCAC	143-147 (bp)	Z50883
8	D8S570	Mfd311	NA	84.04	CTTCTTCCTGGACTTTGCC	GTTCCTACTTGGAGCTTGGA	153-167 (bp)	9242
8	D8S88	Mfd45A	102824398	96.32	TCCAGCAGAGAAAGGGTTAT	GGCAAAGAGAACTCATCAGA	76-86 (bp)	X54564
8	D8S1768	AFMb311yh9	119472191	112.64	AGGTTTCCACATCCCA	AATCATTTCTGTATTACCTATGGT	174-210 (bp)	Z53488
8	D8S1793	AFMc014xf5	143576866	131.99	TGAGCCGAGTTCTTACCAC	AACAAGTCCAGCTTGATGAG	113-147 (bp)	Z54002
8	D8S1926	8QTEL25	NA	166.08	GGGCTTATTAACTTATGAGCAC	GAGTTTTACCTATCTCATTGC	154 (bp)	
15	D15S1021	AFMb344wc5	20653714	4.84	CCTGGCAGGTGGAAGT	ACAAAGATTAACCTCTATGTTTTCG	126-154 (bp)	Z53764
15	D15S1002	AFMb066yh9	24054000	15.05	GTATCCCAAGGCCATACCCT	CTCTTGCTAGAGACAGCAGG	105-129 (bp)	Z53249
15	D15S1040	AFM360te5	29906513	29.95	TGGGAGGCTGAGTCAC	AAAGCCAAATGTAGAGGAAT	181-211 (bp)	Z51533
15	D15S196	UT1532	45849579	47.33	GACCTGTAGCTGAAGGG	GGCACATTCATTTATAAAG	356 (bp)	L16404
15	D15S538	UT7521	60356070	64.09	GTCCCAGCTACTCAGGCGG	GGGGATCTTGAACTGGAAGG	291 (bp)	
15	D15S818	GATA85D02	73870220	79.03	TGTGCATCCTCTATGTCCCT	GCTAAGATGGCGCCATTG	150-170 (bp)	G07910
15	D15S1045	AFMa054wh9	93391203	94.83	AGCAGGCAGGCTAATC	CCTCATCTTCACATGGC	176-224 (bp)	Z51707
15	D15S130	AFM072yb11	98995604	108.21	CTGGTTGTAGAATCCATACCC	AGGCCTCCATGAACTAAACT	191 (bp)	
15	D15S985	AFMa283xh5	103726275	123.46	GCTGAATGACACTGGTGTGC	CTGACAAATCACAACTGAAGACTGG	186-194 (bp)	Z52762
15	D15S642	GATA27A03	107905620	133.61	CAGTTACCCAGGAAGCTGAA	AGATGCCGCCTGTACTAATG	195-218 (bp)	G07902
21	D21S1993	ACT3D07	12398911	3.74	CCGAGATTGCACCATTACTC	TTCTTTTTTCCTGGAAACTCA	117-136 (bp)	G08065
21	D21S1431	ACT2E10	13417760	7.04	ACAAAGCACCTAGCAAGCAC	AGCATAACCATGGGAGTGAA	168-177 (bp)	G08064
21	D21S1256	AFM284xe9	16268197	12.9	AGAAGTAAAAAGCCTATGGTCC	ATATCCACAGTTCTTAGATGGC	125 (bp)	Z24038
21	D21S1884	AFMa116yg1	19487827	20.81	TGATGCAAAAAATTATTGATAAC	GATGTGAATACAGGCAGATG	185-191 (bp)	Z52115
21	D21S269	AFM263xf5	24860105	27.24	AAAAAGTCTCCCATTATACAATAG	CCCTTTGCTTTACAAATCT	235-255 (bp)	Z17183
21	D21S262	AFM198tc5	30674015	35.68	TCTATGAGACAGGGCCAC	ATATTCCGTGTTGATTGTTGTT	142-152 (bp)	Z16848
21	D21S1252	AFM261zg1	34684426	42.96	TCTGTCTTTGTCTCACTATCTG	GCAATGCTCTGTGGCT	231-251 (bp)	Z23854
21	D21S168	D21S168	37799608	49.9	ATGCAATGTTATGTAGGCTG	CGGCATCACAGTCTGATAAA	104-118 (bp)	X52289
21	D21S1260	AFMa152wd1	39650496	56.33	TCCAAGGGGTTCATCC	CCCAAGGCACTGTTCC	200-214 (bp)	Z24672
21	D21S1912	AFM071xa1	42368244	65.59	CCCTCATACAGATTTAAAACACAC	GAGCCCACCCTGGTAAC	173-205 (bp)	Z50922

Supplementary table 1. Microsatellite markers description. Ten microsatellites were chosen to be homogeneously distributed in the three monosomy-diagnosed chromosomes: 8, 15 and 21. Physical location from the p-telomere of the chromosome in bases. Gene location was estimated sex-averaged genetic location from the p-telomere in centiMorgans.

5. Discussion

Pluripotent stem cells are unique due to their developmental potential and the possibility to study them is the key step to understand human development. These cells are characterized by their ability to originate all the cellular lineages within an adult organism. However, the characterization of these cells as pluripotent depends partially on our ability to assess the developmental potential of such samples.

The study of pluripotent stem cells represents a technical challenge. In mammals, these cells can be studied in their physiological cellular context -within the mammalian embryo- or once they have been isolated and kept in vitro to maintain their developmental potential indefinitely. Within embryonic milieu, pluripotent cells represent a dynamic fraction of the total cell number. Moreover, their physiological existence is constrained to early stages of embryonic development. Therefore, it was isolation and in vitro culture of the different types of mammalian pluripotent stem cells, and singularly ESC, that enabled the molecular characterization of the pluripotent state.

In the four articles included in this thesis we have addressed two different aspects of the molecular characterization of mammalian pluripotent stem cells: gene and protein expression on one side, and genomic stability on the other.

5.1 How gene and protein expression define pluripotency

Deciphering patterns of expressed genes in samples like single embryos and ESC can, on one side, lead to the identification of genes the expression of which is relevant to understand pluripotent state, an on the other side provide with an insight into the complex regulatory networks between these genes and the proteins they encode.

Expression assays start with the total mRNA extraction from the cell samples. Recent reports account for massive variations in the number of mRNA molecules present in single cells of the same cellular population[1]. These variations seem to occur due to mRNAs being synthesized in short but intense bursts of transcription beginning when the gene transitions from an inactive to an active state and ending when they transition back to the inactive state. These transitions appear to be intrinsically random and not due to global, extrinsic factors such as the levels of transcriptional activators[1]. Thus, expression assays imply an average evaluation of the transcription levels of the whole cellular population from which total mRNA is extracted, either whole embryos or hESC

samples. In single-embryo samples, the relative transcript abundance of certain lineage-specific genes will give an idea of cellular representativity of those lineages in the embryo analyzed.

There are three techniques that are currently used to evaluate gene expression: a) RT-PCR; b) gene-expression arrays and c) RNA-seq using ultrasequencing technology. Basically, gene expression arrays and RNA-seq allow the parallel analysis of thousands of transcripts while RT-PCR provides the simultaneous measurement of a limited number of transcripts in many different samples. Technically, expression arrays and RNA-seq have strict requirements about the starting RNA quantity to provide accurate results (100-500ng). Contrarily, RT-PCR is especially suitable when only a small number of cells are available. In addition, RT-PCR is frequently used to validate the results of the other two gene expression assays. Therefore, for any expression experiment, optimization of RT-PCR settings is an essential step to obtain accurate expression data or to validate it. Using this premise, we investigated the best conditions to analyze the transcript abundance using RT-PCR in different samples containing pluripotent cells: preimplantation embryos at different stages and stem cells. To do so we determined the most stably expressed gene among several widely used housekeeping genes. In addition to this objective, and due to the scarcity of the cellular material when using embryos (pre-implantation stages) we incorporated to the experiment the evaluation of the faithfulness of the transcript preamplification performed with a commercial method (Preamp from Applied Biosystems). Moreover, we pursued to optimize the conditions to proceed from the sample collection to the RT-PCR reaction set-up.

We first validated the preamplification method, finding that this step introduced no bias in the representativity of the majority of transcripts tested (except for GAPD and ACTB). Thus, this result confirmed this method as a way to obtain enough cDNA to perform transcript quantification reactions within the range to be accurately detected by RT-PCR machine. As for the rest of the candidate reference genes, the geometrical mean of their expression values showed that the most stably expressed in murine embryonic samples (from 2-cell embryos to blastocyst stage) and ESC samples are PPIA and HPRT1 . Interestingly, other reports investigating candidate reference genes found that HPRT1 and PPIA scored high in stability in non-amplified murine samples[2] but not in bovine samples[3].

Overall, these results confirm the importance of optimizing the settings of many commonly used procedures in the lab, especially when the cellular samples are scarce and precious as human the embryos. Our findings also highlight the common misuse of some so-called housekeeping genes

to normalize expression data without a previous assessment of their expression stability in the samples tested.

Since pluripotent state is, in part, a reflection of a certain mRNA/protein expression pattern, in the second paper included in this dissertation we attempted to characterize molecularly mouse embryos from 2-cell stage to blastocyst. Once this characterization was obtained, the following step would be to investigate whether there is any developmental potential difference among blastomeres at different cleavage stages (up to 8-cell). Expression assays using RT-PCR was one of the techniques used to approach this question, and thus, the information on stably expressed reference genes in mouse embryos was essential to accurately analyze the results. However, our molecular characterization was not limited to the transcript analysis. We used immunocytochemical staining to allocate the presence of the proteins known to be pluripotency and lineage commitment markers.

Our results show are in concordance with other recent reports on the molecular characterization of early mouse embryogenesis[4-9]. Indeed, we observed that only after embryonic genome activation (around 2C stage in mouse) that embryos start expressing proteins involved in pluripotency maintenance, namely OCT4 and NANOG, co-expressed with GATA6 (involved in PE determination) and with CDX2 after 8C stage (involved in TE determination). Despite the absence of these proteins in 2C embryos, transcripts of these four genes were present at that stage, indicating a probable maternal origin.

In addition, we found that during a variable period of time, blastomeres co-express TF that will become exclusively expressed in different cell populations after lineage commitments that will conform the expanded blastocyst. Interestingly, these findings contrast previous claims of reciprocal inhibition between some transcription factors that have been reported to control cell fate, singularly OCT4 vs CDX2[10] and GATA6 vs NANOG[11].

To test isolated blastomere totipotency and/or pluripotency during the three first cleavage divisions, we applied the described transcript and protein molecular analysis as well as two functional assays to test pluripotency. Isolated blastomeres from 2C, 4C and 8C cleavage-embryos cultured in vitro were capable of forming mini-blastocysts (their size was proportional to the volume of the blastomere they are originated from). Some of the mini-blastocysts obtained from 2C and 4C blastomeres presented transcript and protein signatures comparable to control blastocysts and ESC could be derived from them. However, no ESC line was derived from mini-blastocysts

obtained from a blastomere of an 8C embryo and their molecular characterization failed to show expression of pluripotency markers (TE and PE molecular signatures were detected). Contrarily, when blastomeres from 2C, 4C an 8C cleavage-stage embryos were isolated and seeded directly on feeder cells, ESC were derived from those obtained from 4C and 8C embryos, but none from 2C embryos. This data suggests that some 2C blastomeres previous to genome activation may be insensible to paracrine signaling that induces the pluripotent state in blastomeres isolated from 4C and 8C embryos when cultured on feeder cells.

Altogether, these results show that during the three first cleavage divisions mouse blastomeres can show full developmental potential giving rise to pluripotent, TE an PE lineages We postulate that molecular maturation occurs in a time-dependent and synchronously in all blastomeres of an embryo. Thus, 2C cleavage stage blastomeres show are less poised to acquire an pluripotent molecular signature (loosing their ability to form TE) than 4C blastomeres and 8C blastomeres. On the other hand, despite a proportion of 8C blastomeres showed to be able to originate ESC, we observed that these cells did not generate any ICM when developing into a mini-blastocyst. Since early embryogenesis is a time-dependent process, 8C blastomeres most probably lack the time to reach a minimum of cell number prior to cavitation that would provide the cell-cell contacts that seem to be decisive to epiblast determination. However, a proportion of 8C blastomere show pluripotent capacities when cell-contact and paracrine signaling is provided by feeder cells.

Lineage commitment molecular basis has been a hot topic during the recent years, and the debate is still candent, mainly using the mouse as model despite some experiments have been performed using human embryos[12,13]. Our results using the mouse model were obtained not taking into account the blastomere position with respect to the animal or vegetal poles of the embryo. Thus, it is not possible to state whether our data corroborates the hypothesis postulated by Dietrich and Hiiragi[14] that proposed a model by which early embryonic mouse patterning is product of stochastic molecular processes that occur in some blastomeres and are independent of the position of the blastomere with respect to the cleaving axes. Other authors suggest that the so-called animal and vegetal poles (determined by the cleavage axes and the polar bodies extrusion) influence the cell fate of the blastomeres[7]. Nevertheless, imbalances between determinant factors in a context of stochastic gene expression has been reported to be a recurrent scenario during differentiation[15]. This, accompanied by epigenetic programming events that impose a permissive (or non-permissive) environment for cell fate, might predispose a cell towards a particular lineage. It has been proposed that a particular TF expression may oscillate between a minimum and a maximum

level, and positive or negative interference of such transcriptional noise introduces a developmental bias in individual blastomeres. Indeed, we showed that expression of the key TF (CDX2, OCT4, GATA6) remain mosaic in individual blastomeres until the blastocyst stage in mouse embryos. It is possible then, that such stochastic expression would need very little to poise certain cells to a more stable identity due to natural feedback loops favoring it. Thus, this biological decision could, like many others, be a continuum from stochastic to biased, thus accounting for the great difficulty in determining the exact mechanism.

Moreover, recent reports show that de novo DNA methylation is initiated very early in embryonic development (around compaction)[16]. Upon blastocyst formation, global DNA methylation levels are markedly different between the embryonic and extraembryonic lineages; the TE is relatively hypomethylated compared with the ICM. Similar to DNA methylation, several histone modifications, also exhibit an asymmetry between the ICM and TE and might therefore have a role in lineage allocation and/or lineage commitment. Lineage fixation in mouse coincides with methylation of the gene *Elf5* in the ICM, which establishes a tight epigenetic boundary between the embryonic and the trophoblastic lineage (reviewed in Hemberger 2009[16]). Future experimental work should be directed to elucidate the interplay between transcription factors and epigenetic features that define pluripotency, and how these evolve towards lineage commitment during embryogenesis.

5.2 Genomic stability in pluripotent cells

The use of pluripotent cells, and particularly hESC, to provide differentiated cells for regenerative medicine will require the continuous maintenance of the undifferentiated stem cells for long periods in culture. However, chromosomal stability during extended passaging has been shown to be an issue of capital importance to ensure the safety of these future hypothetical cell therapies. In one of the chapters included in this dissertation we investigated the genomic stability of hESC during extended *in vitro* culture in addition to the comparison of the genomic features of newly derived sibling stem cell lines. Moreover, since hESC are derived from a few cells in the ICM from preimplantation embryos, we also investigated how diagnosis of aneuploidies in human embryos by cytogenetic techniques correlates with the genomic status of the embryos derived from them.

To assess the genomic stability of hESC we used WG-SNP arrays, adapting its utility to the particular characteristics of the populations of this type of cells. hESC derivation and culture are two processes that imply a great selection pressure for the cells. First, the process by which the

pluripotent cells in the ICM are explanted in vitro primes those cells that can adapt to the new situation over those that don't. Secondly, extended culture entails the same “selection of the fittest” circumstances. ESC containing genomic aberrations have been reported to have enhanced cell culture adaptability[17]. Thus, like in the case of tumor cells populations, hESC cultures were expected not to be a genomically homogeneous populations. As a consequence, WG-SNP array analysis had to be able to calculate the extension (in the cell population) of the detected genomic events.

Indeed, after analysis of different time-points of the hESC passaging history, we observed that in addition to numerical abnormalities, during extended in vitro culture amplification and deletion of subkaryotypic genomic segments occur and become extended in the entire hESC population. In addition, and to our knowledge, we are the first to describe how LOH events accumulate as well (after a cell selection process) in the hESC population genomes during extended in vitro culture. Strikingly, these two types of events (amplifications/deletions and LOH) are present in the earliest passages analyzed of all the cell lines, including the two couple of sibling lines, suggesting that selection of cells containing these events may also occur during the derivation process and during the earliest passages. In both cases, the detected aberrations comprised genes encoded members of the cellular machinery involved in maintaining the genomic stability the alteration (by mutation or genomic rearrangement) of which is known to trigger tumorigenesis (including some relevant ones like BRCA1, FANC, BACH1 and CBLB). In addition, cells containing subkaryotypic aberrations that compromise the cellular machinery involved in genomic stability would show a selective advantage over those that don't due to their ability “to adapt” by introducing new aberrations.

Our findings, recently corroborated by another report[18], suggest that genomic stability at subkaryotypic level is a key feature of hESC culture that has been underestimated to the date. If hESC or other pluripotent cells are to be used in clinical protocols in the future, the origin and implications of the described genomic events need to be investigated. Many laboratories check periodically their hESC lines for karyotypic aberrations not taking into account those occurring at a subchromosomal level, which we suggest seem that are primarily responsible for the ESC culture adaptability. Indeed, a profound review of each of the elements involved in the culture conditions needs to be assessed for their contribution to the occurrence of these aberrant genomic events in order to prevent them. However, the specific characteristics of ESC cell cycle progression (the short G1-phase, bypassing DNA-damage detection checkpoint), and the in vitro culture conditions may

originate the fixation of the spontaneously occurring genomic aberrations[19]. Under physiological conditions, within the embryo, pluripotent state is transient and cells undergoing any genomic rearrangement have been reported to undergo apoptosis to compensate for the lack of DNA-detection machinery[20]. However, ESC derivation, which implies the indefinite perpetuation of pluripotent state in vitro, entails as well the perpetuation relaxed DNA-damage detection activity during the cell cycle. If any spontaneous genomic aberration affects any loci involved in DNA-damage repair or apoptosis pathways, the slippery slope these cells may be falling down to seems inevitable. Recurrent aneuploidies of certain chromosomes (trisomy 12, trisomy 20q) seen in later passages may be one of the consequences. Indeed, many culture-adapted ESC lines have been shown refractory to differentiation protocols, stubbornly maintaining undifferentiated phenotypes, making them close to cancer stem cells[17].

The high efficiency deriving hESC from sibling embryos reported by us and others suggests that some genetic backgrounds may be more permissive to ESC derivation than others[21]. In fact, the parallel situation has been described in mice, with a high variation in the range of derivation efficiency between strains[22]. Therefore, further studies linking genetic variability and derivability should be conducted to identify the variants of the loci that may influence the success of the pluripotent cell derivation. In addition, it would be of sheer interest to compare the genomic events in the parental genomes of the embryos used to derive hESC and the genomic features of such lines in order to evaluate how the derivation process affects the genomic stability.

Surprisingly, two euploid cell lines were derived from embryos that had been diagnosed with monosomies (chr 15 in the embryo that originated WMC2 and chr 8 and chr 21 in the embryo that originated WMC3) after PGS analysis. Remarkably, other laboratories have accounted for the same result using embryos diagnosed as chromosomally abnormal (containing trisomies and/or monosomies) after PGS procedures. One of the papers reporting similar results hypothesized about a possible self-correction of the monosomies or trisomies by chromosomal duplication (for monosomies) and chromosomal abolition (for trisomies), disregarding any explanation linked to PGS misdiagnosis. However, our results using microsatellite markers showed that no chromosomal duplication of the originally monosomic chromosomes had taken place in the cell lines obtained, precluding the self-correction hypothesis.

A recent report describes the high rate of chromosomal instability found in the blastomeres of embryos from 9 young fertile couples that were analyzed using single-cell WG-SNP and WG-

CGH genomic arrays combined[23]. Taking these results into account, we used a WG-SNP array to investigate the euploidy of all the chromosomes of both hESC lines to rule out any other chromosomal aberration that may have been self-corrected after the suggested chromosomal abnormalities produced during early cleavage stage cell divisions. Again, no trace of self-correction was detected. Thus both cell lines were characterized by an euploid heterozygous chromosomal set, suggesting a possible misdiagnosis by the PGS procedure.

The mentioned report on chromosomal instability during early embryogenesis (only 47% of the tested embryos contained euploid blastomeres) need to be corroborated by further experimental data. In case these results are confirmed, it would imply a high proportion of euploid/aneuploid mosaicism in human embryos. Since most hESC lines have been reported to show an euploid karyotype during the early passaging history[24], the derivation process would entail a selection and survival of chromosomally normal blastomeres to be turned into hESC. This could be considered to be in conflict to the conclusions of the previous article, in which we postulated that certain aneuploidies could be selected due to the increased cell culture adaptability and doubling efficiency of the cells containing them. Indeed, contrarily to the general perception, recent findings show that not all aneuploidies contribute to an increased clonal efficiency of cell culture propagation[25]. In fact, many aneuploidies impose a lower cellular propagation efficiency, and thus may be erased in cell populations containing euploid cells. Aneuploid blastomeres may be unable to pass the derivability test due to the cellular stress induced by the in vitro culture conditions plus the aberrant chromosomal set. Therefore, we propose that euploid pluripotent cells with genetic and genomic variants enhancing adaptability to culture conditions would be the ones with a higher chances to be derived into ESC.

6. Conclusions

- When analysing the stability of seven reference genes in different stages of mouse preimplantation development and mESC differentiated in vitro we have shown that PPIA and HPRT1 were the most stably transcribed .In addition, the preamplification strategy used showed to faithfully amplify the different transcripts tested, suggesting that it is a useful step to increase original transcript abundance without introducing any bias to their representativity in the sample.

-The use of immunofluorescence (for protein expression detection) and RT-PCR (for mRNA detection) allowed us to molecularly characterize the formation of the trophectoderm and the ICM during mouse preimplantation embryo development. We concluded that until morula stage blastomeres co-express transcription factors (CDX, OCT4, NANOG, GATA6) that will become mutually exclusive after cavitation when the trophectoderm and the ICM are formed.

- Functional tests on isolated blastomeres from two-cell, four-cell and eight-cell mouse embryos show that despite molecular differences between blastomeres at any point before compaction may be dynamic and not irreversible, when deprived of the embryonic environment, blastomeres in the same embryo show differences in their developmental potential acquired.

-Human embryonic stem cell were observed to acquire subkaryotypic aberrations in vitro right after being derived and throughout their passage history. Common deletions, amplifications and significantly segments of loss of heterozygosity were observed between different timepoints of passaging history of different hESC lines. In addition, sibling cell lines show a high degree of coincidence in genomic aberrations from the earliest passages. These coincident aberration affect coding regions of genes involved in DNA-damage mechanisms, differentiation processes, apoptosis and cell signalling, which suggests a similitude to carcinogenic genomic aberrations.

-Two euploid human embryonic stem cell lines were derived from aneuploid embryos diagnosed after preimplantation genetic diagnosis (PGD). The genomic analysis of the hESC line showed no trace of any self-correction mechanism to recover from monosomy or trisomy in detected in the embryo. FISH failure and as a consequence misdiagnosis at the time of PGD account for the normality of the cell line.

7. Bibliography

1. Palermo, G.D. et al. Fertilization and pregnancy outcome with intracytoplasmic sperm injection for azoospermic men. *Hum. Reprod* **14**, 741-748 (1999).
2. Palermo, G., Munné, S. & Cohen, J. The human zygote inherits its mitotic potential from the male gamete. *Hum. Reprod* **9**, 1220-1225 (1994).
3. Veeck, L.L. & Zaninović, N. *An atlas of human blastocysts*. (Informa Health Care: 2003).
4. Herbert, M., Wolstenholme, J., Murdoch, A.P. & Butler, T.J. Mitotic activity during preimplantation development of human embryos. *J. Reprod. Fertil* **103**, 209-214 (1995).
5. Adjaye, J. et al. Primary Differentiation in the Human Blastocyst: Comparative Molecular Portraits of Inner Cell Mass and Trophectoderm Cells. *Stem Cells* **23**, 1514-1525 (2005).
6. Larsen, W.J., Sherman, L.S., Potter, S.S. & Scott, W.J. *Human embryology*. (Elsevier Health Sciences: 2001).
7. Surani, M.A., Hayashi, K. & Hajkova, P. Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747-762 (2007).
8. Gan, Q., Yoshida, T., McDonald, O.G. & Owens, G.K. Concise review: epigenetic mechanisms contribute to pluripotency and cell lineage determination of embryonic stem cells. *Stem Cells* **25**, 2-9 (2007).
9. Eggan, K. et al. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl. Acad. Sci. U.S.A* **98**, 6209-6214 (2001).
10. Nagy, A. et al. Embryonic stem cells alone are able to support fetal development in the mouse. *Development* **110**, 815-821 (1990).
11. Veiga, A., Calderon, G., Barri, P.N. & Coroleu, B. Pregnancy after the replacement of a frozen-thawed embryo with less than 50% intact blastomeres. *Hum. Reprod* **2**, 321-323 (1987).
12. Barker, N., van de Wetering, M. & Clevers, H. The intestinal stem cell. *Genes Dev* **22**, 1856-1864 (2008).
13. Willadsen, S.M. The viability of early cleavage stages containing half the normal number of blastomeres in the sheep. *J. Reprod. Fertil* **59**, 357-362 (1980).
14. Moore, N.W., Adams, C.E. & Rowson, L.E. Developmental potential of single blastomeres of the rabbit egg. *J. Reprod. Fertil* **17**, 527-531 (1968).
15. Tao, T. & Niemann, H. Cellular characterization of blastocysts derived from rabbit 4-, 8- and 16-cell embryos and isolated blastomeres cultured in vitro. *Hum. Reprod.* **15**, 881-889 (2000).
16. Willadsen, S. & Polge, C. Attempts to produce monozygotic quadruplets in cattle by blastomere separation. *Vet Rec.* **108**, 211-213 (1981).
17. Johnson, W., Loskutoff, N., Plante, Y. & Betteridge, K. Production of four identical calves by

the separation of blastomeres from an in vitro derived four-cell embryo. *Vet Rec.* **137**, 15-16 (1995).

18. Saito, S. & Niemann, H. Effects of extracellular matrices and growth factors on the development of isolated porcine blastomeres. *Biol. Reprod* **44**, 927-936 (1991).
19. Allen, W.R. & Pashen, R.L. Production of monozygotic (identical) horse twins by embryo micromanipulation. *J. Reprod. Fertil* **71**, 607-613 (1984).
20. Tsonuda Y, Usuai, Y, Sugie T Production of monozygotic twins following transfer of separated half embryos in the goat. *Jpn J Zootech Sci* **55**, 643-647 (1984).
21. Gardner, R.L. Contributions of blastocyst micromanipulation to the study of mammalian development. *Bioessays* **20**, 168-180 (1998).
22. Diwan, S.B. & Stevens, L.C. Development of teratomas from the ectoderm of mouse egg cylinders. *J. Natl. Cancer Inst* **57**, 937-942 (1976).
23. de Sousa Lopes, S.M.C., Hayashi, K. & Surani, M.A. Proximal visceral endoderm and extraembryonic ectoderm regulate the formation of primordial germ cell precursors. *BMC Dev. Biol* **7**, 140 (2007).
24. Lawson, K.A., Meneses, J.J. & Pedersen, R.A. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891-911 (1991).
25. Tam, P.P. & Zhou, S.X. The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev. Biol* **178**, 124-132 (1996).
26. Hayashi, K., de Sousa Lopes, S.M.C. & Surani, M.A. Germ cell specification in mice. *Science* **316**, 394-396 (2007).
27. Blackshear, P. et al. Extragonadal teratocarcinoma in chimeric mice. *Vet. Pathol* **36**, 457-460 (1999).
28. Stevens, L.C. & Little, C.C. Spontaneous Testicular Teratomas in an Inbred Strain of Mice. *Proc. Natl. Acad. Sci. U.S.A* **40**, 1080-1087 (1954).
29. Anderson, P.D., Nelson, V.R., Tesar, P.J. & Nadeau, J.H. Genetic factors on mouse chromosome 18 affecting susceptibility to testicular germ cell tumors and permissiveness to embryonic stem cell derivation. *Cancer Res* **69**, 9112-9117 (2009).
30. KLEINSMITH, L.J. & PIERCE, G.B. MULTIPOTENTIALITY OF SINGLE EMBRYONAL CARCINOMA CELLS. *Cancer Res* **24**, 1544-1551 (1964).
31. Kahan, B.W. & Ephrussi, B. Developmental potentialities of clonal in vitro cultures of mouse testicular teratoma. *J. Natl. Cancer Inst* **44**, 1015-1036 (1970).
32. Gachelin, G., Kemler, R., Kelly, F. & Jacob, F. PCC4, a new cell surface antigen common to

multipotential embryonal carcinoma cells, spermatozoa, and mouse early embryos. *Dev. Biol* **57**, 199-209 (1977).

33. Solter, D. & Knowles, B.B. Monoclonal antibody defining a stage-specific mouse embryonic antigen (SSEA-1). *Proc. Natl. Acad. Sci. U.S.A* **75**, 5565-5569 (1978).
34. Martin, G.R. Teratocarcinomas and mammalian embryogenesis. *Science* **209**, 768-776 (1980).
35. Brinster, R.L. The effect of cells transferred into the mouse blastocyst on subsequent development. *J. Exp. Med* **140**, 1049-1056 (1974).
36. Atkin, N.B., Baker, M.C., Robinson, R. & Gaze, S.E. Chromosome studies on 14 near-diploid carcinomas of the ovary. *Eur J Cancer* **10**, 144-146 (1974).
37. Hogan, B., Fellous, M., Avner, P. & Jacob, F. Isolation of a human teratoma cell line which expresses F9 antigen. *Nature* **270**, 515-518 (1977).
38. Andrews, P.W., Goodfellow, P.N., Shevinsky, L.H., Bronson, D.L. & Knowles, B.B. Cell-surface antigens of a clonal human embryonal carcinoma cell line: morphological and antigenic differentiation in culture. *Int. J. Cancer* **29**, 523-531 (1982).
39. Evans, M.J. & Kaufman, M.H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156 (1981).
40. Martin, G.R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U.S.A* **78**, 7634-7638 (1981).
41. Smith, A.G. et al. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-690 (1988).
42. Williams, R.L. et al. Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684-687 (1988).
43. Ying, Q.L., Nichols, J., Chambers, I. & Smith, A. BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-292 (2003).
44. Thomson, J.A. et al. Embryonic stem cell lines derived from human blastocysts. *Science (New York, N.Y* **282**, 1145-7 (1998).
45. Amit, M. et al. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev. Biol* **227**, 271-278 (2000).
46. Beattie, G.M. et al. Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. *Stem Cells* **23**, 489-495 (2005).
47. James, D., Levine, A.J., Besser, D. & Hemmati-Brivanlou, A. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132**, 1273-1282 (2005).

48. Vallier, L., Alexander, M. & Pedersen, R.A. Activin/Nodal and FGF pathways cooperate to maintain pluripotency of human embryonic stem cells. *J. Cell. Sci* **118**, 4495-4509 (2005).
49. Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A.H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nat. Med* **10**, 55-63 (2004).
50. Bendall, S.C. et al. IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. *Nature* **448**, 1015-1021 (2007).
51. Wang, L. et al. Self-renewal of human embryonic stem cells requires insulin-like growth factor-1 receptor and ERBB2 receptor signaling. *Blood* **110**, 4111-4119 (2007).
52. Klimanskaya, I., Chung, Y., Becker, S., Lu, S.J. & Lanza, R. Human embryonic stem cell lines derived from single blastomeres. *Nature* **444**, 512 (2006).
53. Chung, Y. et al. Embryonic and extraembryonic stem cell lines derived from single mouse blastomeres. *Nature* **439**, 216-9 (2006).
54. Tesar, P.J. Derivation of germ-line-competent embryonic stem cell lines from preblastocyst mouse embryos. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8239-44 (2005).
55. Thomson, J.A. et al. Isolation of a primate embryonic stem cell line. *Proc. Natl. Acad. Sci. U.S.A* **92**, 7844-7848 (1995).
56. Buehr, M. et al. Capture of authentic embryonic stem cells from rat blastocysts. *Cell* **135**, 1287-1298 (2008).
57. Vaags, A.K. et al. Derivation and Characterization of Canine Embryonic Stem Cell Lines with In Vitro and In Vivo Differentiation Potential. *Stem Cells* **27**, 329-340 (2009).
58. Wang, S. et al. Generation and characterization of rabbit embryonic stem cells. *Stem Cells* **25**, 481-489 (2007).
59. Intawicha, P. et al. Characterization of embryonic stem cell lines derived from New Zealand white rabbit embryos. *Cloning Stem Cells* **11**, 27-38 (2009).
60. Martins-Taylor, K. & Xu, R. Determinants of pluripotency: From avian, rodents, to primates. *Journal of Cellular Biochemistry* **109**, 16-25 (2010).
61. Muñoz, M. et al. Constraints to Progress in Embryonic Stem Cells from Domestic Species. *Stem Cell Reviews and Reports* **5**, 6-9 (2009).
62. Keefer, C., Pant, D., Blomberg, L. & Talbot, N. Challenges and prospects for the establishment of embryonic stem cell lines of domesticated ungulates. *Animal Reproduction Science* **98**, 147-168 (2007).
63. Talbot, N.C. & Blomberg, L.A. The pursuit of ES cell lines of domesticated ungulates. *Stem*

Cell Rev **4**, 235-254 (2008).

64. Tecirlioglu, R.T. & Trounson, A.O. Embryonic stem cells in companion animals (horses, dogs and cats): present status and future prospects. *Reprod. Fertil. Dev* **19**, 740-747 (2007).
65. Brons, I.G.M. et al. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191-195 (2007).
66. Tesar, P.J. et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196-199 (2007).
67. STEVENS, L.C. The biology of teratomas including evidence indicating their origin form primordial germ cells. *Annee Biol* **1**, 585-610 (1962).
68. Resnick, J.L., Bixler, L.S., Cheng, L. & Donovan, P.J. Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**, 550-551 (1992).
69. Matsui, Y., Zsebo, K. & Hogan, B.L. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841-847 (1992).
70. Labosky, P.A., Barlow, D.P. & Hogan, B.L. Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. *Development* **120**, 3197-3204 (1994).
71. Stewart, C.L., Gadi, I. & Bhatt, H. Stem cells from primordial germ cells can reenter the germ line. *Dev. Biol* **161**, 626-628 (1994).
72. Tada, M., Tada, T., Lefebvre, L., Barton, S.C. & Surani, M.A. Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO J* **16**, 6510-6520 (1997).
73. Shovlin, T.C., Durcova-Hills, G., Surani, A. & McLaren, A. Heterogeneity in imprinted methylation patterns of pluripotent embryonic germ cells derived from pre-migratory mouse germ cells. *Dev. Biol* **313**, 674-681 (2008).
74. Shamblott, M.J. et al. Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc. Natl. Acad. Sci. U.S.A* **95**, 13726-13731 (1998).
75. Turnpenny, L. et al. Derivation of human embryonic germ cells: an alternative source of pluripotent stem cells. *Stem Cells* **21**, 598-609 (2003).
76. Bryder, D., Rossi, D.J. & Weissman, I.L. Hematopoietic stem cells: the paradigmatic tissue-specific stem cell. *Am. J. Pathol* **169**, 338-346 (2006).
77. Orkin, S.H. & Zon, L.I. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. *Cell* **132**, 631-644 (2008).
78. Watt, F.M. The stem cell compartment in human interfollicular epidermis. *J. Dermatol. Sci* **28**, 173-180 (2002).

79. Morrison, S.J. & Spradling, A.C. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* **132**, 598-611 (2008).
80. Joseph, N.M. & Morrison, S.J. Toward an understanding of the physiological function of Mammalian stem cells. *Dev. Cell* **9**, 173-183 (2005).
81. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567-82 (2008).
82. Oswald, J. et al. Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol* **10**, 475-478 (2000).
83. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501-502 (2000).
84. Santos, F., Hendrich, B., Reik, W. & Dean, W. Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol* **241**, 172-182 (2002).
85. Pedersen, R.A., Wu, K. & Bałakier, H. Origin of the inner cell mass in mouse embryos: cell lineage analysis by microinjection. *Dev. Biol* **117**, 581-595 (1986).
86. Johnson, M.H. & Ziomek, C.A. Induction of polarity in mouse 8-cell blastomeres: specificity, geometry, and stability. *The Journal of cell biology* **91**, 303-8 (1981).
87. Johnson, M.H. & Ziomek, C.A. The foundation of two distinct cell lineages within the mouse morula. *Cell* **24**, 71-80 (1981).
88. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-91 (1998).
89. Schöler, H.R., Ruppert, S., Suzuki, N., Chowdhury, K. & Gruss, P. New type of POU domain in germ line-specific protein Oct-4. *Nature* **344**, 435-439 (1990).
90. Avilion, A.A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development* **17**, 126-40 (2003).
91. Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-55 (2003).
92. Mitsui, K. et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-42 (2003).
93. Zhang, J. et al. Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nat. Cell Biol* **8**, 1114-1123 (2006).
94. Elling, U., Klasen, C., Eisenberger, T., Anlag, K. & Treier, M. Murine inner cell mass-derived lineages depend on Sall4 function. *Proc. Natl. Acad. Sci. U.S.A* **103**, 16319-16324 (2006).
95. Yagi, R. et al. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development (Cambridge, England)* **134**, 3827-36

(2007).

96. Nishioka, N. et al. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech. Dev* **125**, 270-283 (2008).
97. Strumpf, D. et al. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development (Cambridge, England)* **132**, 2093-102 (2005).
98. Ralston, A. & Rossant, J. Cdx2 acts downstream of cell polarization to cell-autonomously promote trophectoderm fate in the early mouse embryo. *Developmental biology* **313**, 614-29 (2008).
99. Jedrusik, A. et al. Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. *Genes & development* **22**, 2692-706 (2008).
100. Russ, A.P. et al. Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95-99 (2000).
101. Suwinska, A., Czolowska, R., Ozdzenski, W. & Tarkowski, A.K. Blastomeres of the mouse embryo lose totipotency after the fifth cleavage division: expression of Cdx2 and Oct4 and developmental potential of inner and outer blastomeres of 16- and 32-cell embryos. *Developmental biology* **322**, 133-44 (2008).
102. Niwa, H. et al. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917-29 (2005).
103. Morrisey, E.E. et al. GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo. *Genes & development* **12**, 3579-90 (1998).
104. Koutsourakis, M., Langeveld, A., Patient, R., Beddington, R. & Grosveld, F. The transcription factor GATA6 is essential for early extraembryonic development. *Development* **126**, 723-732 (1999).
105. Chazaud, C., Yamanaka, Y., Pawson, T. & Rossant, J. Early lineage segregation between epiblast and primitive endoderm in mouse blastocysts through the Grb2-MAPK pathway. *Developmental cell* **10**, 615-24 (2006).
106. Yang, D., Cai, K.Q., Roland, I.H., Smith, E.R. & Xu, X. Disabled-2 Is an Epithelial Surface Positioning Gene. *Journal of Biological Chemistry* **282**, 13114-13122 (2007).
107. Gerbe, F., Cox, B., Rossant, J. & Chazaud, C. Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. *Dev. Biol* **313**, 594-602 (2008).
108. Loh, Y.H. et al. The Oct4 and Nanog transcription network regulates pluripotency in mouse

embryonic stem cells. *Nat Genet* **38**, 431-40 (2006).

109. Boyer, L.A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-56 (2005).
110. Avilion, A.A. et al. Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes & development* **17**, 126-40 (2003).
111. Chambers, I. et al. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643-55 (2003).
112. Nichols, J. et al. Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-91 (1998).
113. Mitsui, K. et al. The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631-42 (2003).
114. Boyer, L.A. et al. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-56 (2005).
115. Yuan, H., Corbi, N., Basilico, C. & Dailey, L. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes Dev* **9**, 2635-2645 (1995).
116. Nishimoto, M., Fukushima, A., Okuda, A. & Muramatsu, M. The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol. Cell. Biol* **19**, 5453-5465 (1999).
117. Wang, J. et al. A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364-8 (2006).
118. Becker, K.A. et al. Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase. *J. Cell. Physiol* **209**, 883-893 (2006).
119. Wobus, A.M. & Boheler, K.R. Embryonic stem cells: prospects for developmental biology and cell therapy. *Physiol. Rev* **85**, 635-678 (2005).
120. Savatier, P., Lapillonne, H., van Grunsven, L.A., Rudkin, B.B. & Samarut, J. Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. *Oncogene* **12**, 309-322 (1996).
121. Stead, E. et al. Pluripotent cell division cycles are driven by ectopic Cdk2, cyclin A/E and E2F activities. *Oncogene* **21**, 8320-8333 (2002).
122. Mantel, C. et al. Checkpoint-apoptosis uncoupling in human and mouse embryonic stem cells: a source of karyotpic instability. *Blood* **109**, 4518-4527 (2007).
123. White, J. et al. Developmental activation of the Rb-E2F pathway and establishment of cell cycle-regulated cyclin-dependent kinase activity during embryonic stem cell differentiation.

Mol. Biol. Cell **16**, 2018-2027 (2005).

124. Zhang, X. et al. A role for NANOG in G1 to S transition in human embryonic stem cells through direct binding of CDK6 and CDC25A. *J. Cell Biol* **184**, 67-82 (2009).
125. Boyer, L.A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-53 (2006).
126. Lee, T.I. et al. Control of developmental regulators by polycomb in human embryonic stem cells. *Cell* **125**, 301-13 (2006).
127. Azuara, V. et al. Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol* **8**, 532-538 (2006).
128. Bernstein, B. et al. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. *Cell* **125**, 315-326 (2006).
129. Mikkelsen, T.S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60 (2007).
130. Farthing, C.R. et al. Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet* **4**, e1000116 (2008).
131. Fouse, S.D. et al. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**, 160-169 (2008).
132. Meissner, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-770 (2008).
133. Yang, J. et al. Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A* **105**, 19756-19761 (2008).
134. Boyer, L.A. et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-53 (2006).
135. Hattori, N. et al. Epigenetic Control of Mouse Oct-4 Gene Expression in Embryonic Stem Cells and Trophoblast Stem Cells. *J. Biol. Chem.* **279**, 17063-17069 (2004).
136. Hattori, N. et al. Epigenetic regulation of Nanog gene in embryonic stem and trophoblast stem cells. *Genes Cells* **12**, 387-396 (2007).
137. Feldman, N. et al. G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat. Cell Biol* **8**, 188-194 (2006).
138. Epsztejn-Litman, S. et al. De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat. Struct. Mol. Biol* **15**, 1176-1183 (2008).
139. Li, J. et al. Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. *Mol. Cell. Biol* **27**, 8748-8759 (2007).

140. Waddington, C.H. *Organisers & genes*. (The University Press: 1940).
141. Graf, T. & Enver, T. Forcing cells to change lineages. *Nature* **462**, 587-594 (2009).
142. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676 (2006).
143. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872 (2007).
144. Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat. Biotechnol* **25**, 1177-1181 (2007).
145. Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T. & Yamanaka, S. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* **322**, 949-953 (2008).
146. Hochedlinger, K. & Plath, K. Epigenetic reprogramming and induced pluripotency. *Development* **136**, 509-523 (2009).
147. Mikkelsen, T.S. et al. Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49-55 (2008).
148. Shi, Y. et al. A combined chemical and genetic approach for the generation of induced pluripotent stem cells. *Cell Stem Cell* **2**, 525-528 (2008).
149. Huangfu, D. et al. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat. Biotechnol* **26**, 1269-1275 (2008).
150. Chang, H.H., Hemberg, M., Barahona, M., Ingber, D.E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544-547 (2008).
151. MacArthur, B.D., Ma'ayan, A. & Lemischka, I.R. Systems biology of stem cell fate and cellular reprogramming. *Nat Rev Mol Cell Biol* **10**, 672-681 (2009).
152. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-1234 (2007).
153. Chen, A.E. et al. Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell stem cell* **4**, 103-6 (2009).
154. Borstlap, J. et al. First evaluation of the European hESCreg. *Nat. Biotechnol* **26**, 859-860 (2008).
155. Reubinoff, B.E., Pera, M.F., Fong, C., Trounson, A. & Bongso, A. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. *Nat Biotech* **18**, 399-404 (2000).
156. Zhang, X. et al. Derivation of human embryonic stem cells from developing and arrested embryos. *Stem Cells* **24**, 2669-76 (2006).
157. Mitalipova, M. et al. Human embryonic stem cell lines derived from discarded embryos. *Stem*

Cells **21**, 521-526 (2003).

158. Chen, H. et al. The derivation of two additional human embryonic stem cell lines from day 3 embryos with low morphological scores. *Hum. Reprod* **20**, 2201-2206 (2005).
159. Sjögren, A. et al. Human blastocysts for the development of embryonic stem cells. *Reprod. Biomed. Online* **9**, 326-329 (2004).
160. Lavon, N. et al. Derivation of euploid human embryonic stem cells from aneuploid embryos. *Stem Cells* **26**, 1874-1882 (2008).
161. Peura, T. et al. Karyotypically normal and abnormal human embryonic stem cell lines derived from PGD-analyzed embryos. *Cloning Stem Cells* **10**, 203-216 (2008).
162. Munné, S. et al. Self-correction of chromosomally abnormal embryos in culture and implications for stem cell production. *Fertil. Steril* **84**, 1328-1334 (2005).
163. Mateizel, I. et al. Derivation of human embryonic stem cell lines from embryos obtained after IVF and after PGD for monogenic disorders. *Hum. Reprod* **21**, 503-511 (2006).
164. Plachot, M. Cytogenetic analysis of oocytes and embryos. *Ann. Acad. Med. Singap* **21**, 538-544 (1992).
165. Suss-Toby, E., Gerecht-Nir, S., Amit, M., Manor, D. & Itskovitz-Eldor, J. Derivation of a diploid human embryonic stem cell line from a mononuclear zygote. *Hum. Reprod* **19**, 670-675 (2004).
166. Levron, J., Munné, S., Willadsen, S., Rosenwaks, Z. & Cohen, J. Male and female genomes associated in a single pronucleus in human zygotes. *Biol. Reprod* **52**, 653-657 (1995).
167. Amit, M. & Itskovitz-Eldor, J. Sources, derivation, and culture of human embryonic stem cells. *Semin. Reprod. Med* **24**, 298-303 (2006).
168. Kim, H.S. et al. Methods for derivation of human embryonic stem cells. *Stem Cells* **23**, 1228-1233 (2005).
169. Solter, D. & Knowles, B.B. Immunosurgery of mouse blastocyst. *Proc. Natl. Acad. Sci. U.S.A* **72**, 5099-5102 (1975).
170. Heins, N. et al. Derivation, Characterization, and Differentiation of Human Embryonic Stem Cells. *Stem Cells* **22**, 367-376 (2004).
171. Ström, S. et al. Mechanical isolation of the inner cell mass is effective in derivation of new human embryonic stem cell lines. *Hum. Reprod* **22**, 3051-3058 (2007).
172. Turetsky, T. et al. Laser-assisted derivation of human embryonic stem cell lines from IVF embryos after preimplantation genetic diagnosis. *Hum. Reprod* **23**, 46-53 (2008).
173. Cortes, J.L. et al. Whole-blastocyst culture followed by laser drilling technology enhances the efficiency of inner cell mass isolation and embryonic stem cell derivation from good- and

poor-quality mouse embryos: new insights for derivation of human embryonic stem cell lines. *Stem Cells Dev* **17**, 255-267 (2008).

174. Klimanskaya, I., Chung, Y., Becker, S., Lu, S.J. & Lanza, R. Human embryonic stem cell lines derived from single blastomeres. *Nature* **444**, 481-5 (2006).
175. Chung, Y. et al. Human embryonic stem cell lines generated without embryo destruction. *Cell Stem Cell* **2**, 113-117 (2008).
176. Geens, M. et al. Human embryonic stem cell lines derived from single blastomeres of two 4-cell stage embryos. *Hum. Reprod.* dep262 (2009).doi:10.1093/humrep/dep262
177. Strelchenko, N., Verlinsky, O., Kukharenko, V. & Verlinsky, Y. Morula-derived human embryonic stem cells. *Reproductive biomedicine online* **9**, 623-9 (2004).
178. Tesar, P.J. Derivation of germ-line-competent embryonic stem cell lines from preblastocyst mouse embryos. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 8239-44 (2005).
179. Chin, A.C.P. et al. Identification of proteins from feeder conditioned medium that support human embryonic stem cells. *J. Biotechnol* **130**, 320-328 (2007).
180. Yu, J. & Thomson, J.A. Pluripotent stem cell lines. *Genes Dev* **22**, 1987-1997 (2008).
181. Hovatta, O. et al. A culture system using human foreskin fibroblasts as feeder cells allows production of human embryonic stem cells. *Hum. Reprod* **18**, 1404-1409 (2003).
182. Stojkovic, P. et al. An autogeneic feeder cell system that efficiently supports growth of undifferentiated human embryonic stem cells. *Stem Cells* **23**, 306-314 (2005).
183. Wang, Q. et al. Derivation and growing human embryonic stem cells on feeders derived from themselves. *Stem Cells* **23**, 1221-1227 (2005).
184. Choo, A., Ngo, A.S., Ding, V., Oh, S. & Kiang, L.S. Autogeneic feeders for the culture of undifferentiated human embryonic stem cells in feeder and feeder-free conditions. *Methods Cell Biol* **86**, 15-28 (2008).
185. Chavez, S.L., Meneses, J.J., Nguyen, H.N., Kim, S.K. & Pera, R.A.R. Characterization of six new human embryonic stem cell lines (HSF7, -8, -9, -10, -12, and -13) derived under minimal-animal component conditions. *Stem Cells Dev* **17**, 535-546 (2008).
186. Zhan, X., Hill, C., Brayton, C.F. & Shamblott, M.J. Cells derived from human umbilical cord blood support the long-term growth of undifferentiated human embryonic stem cells. *Cloning Stem Cells* **10**, 513-522 (2008).
187. Lee, J.B. et al. Establishment and maintenance of human embryonic stem cell lines on human feeder cells derived from uterine endometrium under serum-free condition. *Biol. Reprod* **72**, 42-49 (2005).

188. Ji, L. et al. Self-renewal and pluripotency is maintained in human embryonic stem cells by co-culture with human fetal liver stromal cells expressing hypoxia inducible factor 1alpha. *J. Cell. Physiol* **221**, 54-66 (2009).
189. Kleinman, H.K. et al. Isolation and characterization of type IV procollagen, laminin, and heparan sulfate proteoglycan from the EHS sarcoma. *Biochemistry* **21**, 6188-6193 (1982).
190. Xu, C. et al. Feeder-free growth of undifferentiated human embryonic stem cells. *Nat. Biotechnol* **19**, 971-974 (2001).
191. Li, Y., Powell, S., Brunette, E., Lebkowski, J. & Mandalam, R. Expansion of human embryonic stem cells in defined serum-free medium devoid of animal-derived products. *Biotechnol. Bioeng* **91**, 688-698 (2005).
192. Ludwig, T.E. et al. Feeder-independent culture of human embryonic stem cells. *Nat. Methods* **3**, 637-646 (2006).
193. Braam, S.R. et al. Recombinant vitronectin is a functionally defined substrate that supports human embryonic stem cell self-renewal via alphavbeta5 integrin. *Stem Cells* **26**, 2257-2265 (2008).
194. Amit, M., Shariki, C., Margulets, V. & Itskovitz-Eldor, J. Feeder Layer- and Serum-Free Culture of Human Embryonic Stem Cells. *Biol Reprod* **70**, 837-845 (2004).
195. Ellerström, C. et al. Derivation of a xeno-free human embryonic stem cell line. *Stem Cells* **24**, 2170-2176 (2006).
196. Ludwig, T.E. et al. Derivation of human embryonic stem cells in defined conditions. *Nat Biotechnol* **24**, 185-7 (2006).
197. Lu, J., Hou, R., Booth, C.J., Yang, S.H. & Snyder, M. Defined culture conditions of human embryonic stem cells. *Proc Natl Acad Sci U S A* **103**, 5688-93 (2006).
198. Yao, S. et al. Long-term self-renewal and directed differentiation of human embryonic stem cells in chemically defined conditions. *Proc Natl Acad Sci U S A* **103**, 6907-12 (2006).
199. Xu, C. et al. Basic fibroblast growth factor supports undifferentiated human embryonic stem cell growth without conditioned medium. *Stem Cells* **23**, 315-323 (2005).
200. Xu, R. et al. Basic FGF and suppression of BMP signaling sustain undifferentiated proliferation of human ES cells. *Nat. Methods* **2**, 185-190 (2005).
201. Greber, B. et al. Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell Stem Cell* **6**, 215-226 (2010).
202. Dvorak, P. et al. Expression and potential role of fibroblast growth factor 2 and its receptors in human embryonic stem cells. *Stem Cells* **23**, 1200-1211 (2005).
203. Kang, H.B. et al. Basic fibroblast growth factor activates ERK and induces c-fos in human

embryonic stem cell line MizhES1. *Stem Cells Dev* **14**, 395-401 (2005).

204. Li, J. et al. MEK/ERK signaling contributes to the maintenance of human embryonic stem cell self-renewal. *Differentiation* **75**, 299-307 (2007).
205. James, D., Levine, A.J., Besser, D. & Hemmati-Brivanlou, A. TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132**, 1273-1282 (2005).
206. Wong, R.C. et al. Gap junctions modulate apoptosis and colony growth of human embryonic stem cells maintained in a serum-free system. *Biochem Biophys Res Commun* **344**, 181-8 (2006).
207. Baker, D.E. et al. Adaptation to culture of human embryonic stem cells and oncogenesis in vivo. *Nat Biotechnol* **25**, 207-15 (2007).
208. Longo, L., Bygrave, A., Grosveld, F.G. & Pandolfi, P.P. The chromosome make-up of mouse embryonic stem cells is predictive of somatic and germ cell chimaerism. *Transgenic Res* **6**, 321-328 (1997).
209. Draper, J.S. et al. Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat Biotechnol* **22**, 53-4 (2004).
210. Inzunza, J. et al. Comparative genomic hybridization and karyotyping of human embryonic stem cells reveals the occurrence of an isodicentric X chromosome after long-term cultivation. *Molecular human reproduction* **10**, 461 (2004).
211. Atkin, N.B. & Baker, M.C. Specific chromosome change, i(12p), in testicular tumours? *Lancet* **2**, 1349 (1982).
212. Skotheim, R.I. et al. New insights into testicular germ cell tumorigenesis from gene expression profiling. *Cancer Res* **62**, 2359-2364 (2002).
213. Korkola, J.E. et al. Down-regulation of stem cell genes, including those in a 200-kb gene cluster at 12p13.31, is associated with in vivo differentiation of human male germ cell tumors. *Cancer Res* **66**, 820-827 (2006).
214. Mitalipova, M.M. et al. Preserving the genetic integrity of human embryonic stem cells. *Nat. Biotechnol* **23**, 19-20 (2005).
215. Närvä, E. et al. High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. *Nat. Biotechnol* **28**, 371-377 (2010).
216. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS biology* **4**, e309 (2006).
217. Mamo, S., Gal, A.B., Bodo, S. & Dinnyes, A. Quantitative evaluation and selection of

reference genes in mouse oocytes and embryos cultured in vivo and in vitro. *BMC developmental biology* **7**, 14 (2007).

218. Goossens, K. et al. Selection of reference genes for quantitative real-time PCR in bovine preimplantation embryos. *BMC developmental biology* **5**, 27 (2005).
219. Rossant, J. & Tam, P.P. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development (Cambridge, England)* **136**, 701-13 (2009).
220. Gerbe, F., Cox, B., Rossant, J. & Chazaud, C. Dynamic expression of Lrp2 pathway members reveals progressive epithelial differentiation of primitive endoderm in mouse blastocyst. *Developmental biology* **313**, 594-602 (2008).
221. Strumpf, D. et al. Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development (Cambridge, England)* **132**, 2093-102 (2005).
222. Bischoff, M., Parfitt, D.E. & Zernicka-Goetz, M. Formation of the embryonic-abembryonic axis of the mouse blastocyst: relationships between orientation of early cleavage divisions and pattern of symmetric/asymmetric divisions. *Development (Cambridge, England)* **135**, 953-62 (2008).
223. Jedrusik, A. et al. Role of Cdx2 and cell polarity in cell allocation and specification of trophectoderm and inner cell mass in the mouse embryo. *Genes & development* **22**, 2692-706 (2008).
224. Zernicka-Goetz, M., Morris, S.A. & Bruce, A.W. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat Rev Genet* **10**, 467-477 (2009).
225. Niwa, H. et al. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917-29 (2005).
226. Singh, A.M., Hamazaki, T., Hankowski, K.E. & Terada, N. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem cells (Dayton, Ohio)* **25**, 2534-42 (2007).
227. Cauffman, G., De Rycke, M., Sermon, K., Liebaers, I. & Van de Velde, H. Markers that define stemness in ESC are unable to identify the totipotent cells in human preimplantation embryos. *Hum. Reprod* **24**, 63-70 (2009).
228. Cauffman, G., Van de Velde, H., Liebaers, I. & Van Steirteghem, A. Oct-4 mRNA and protein expression during human preimplantation development. *Mol. Hum. Reprod* **11**, 173-181 (2005).
229. Dietrich, J.E. & Hiiragi, T. Stochastic patterning in the mouse pre-implantation embryo. *Development (Cambridge, England)* **134**, 4219-31 (2007).
230. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its

consequences. *Cell* **135**, 216-226 (2008).

231. Hemberger, M., Dean, W. & Reik, W. Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat. Rev. Mol. Cell Biol* **10**, 526-537 (2009).
232. Hong, Y., Cervantes, R.B., Tichy, E., Tischfield, J.A. & Stambrook, P.J. Protecting genomic integrity in somatic cells and embryonic stem cells. *Mutat. Res* **614**, 48-55 (2007).
233. Vanneste, E. et al. Chromosome instability is common in human cleavage-stage embryos. *Nat Med* **15**, 577-583 (2009).
234. Holland, A.J. & Cleveland, D.W. Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol* **10**, 478-487 (2009).