1. Introduction

1.1 hESCs: a General Introduction

1.1.1 Human embryonic stem cells. hESCs have been defined as the only known-to-date non transformed, non genetically modified human cells to exhibit an undifferentiated state as well as a stable and extensive proliferative and differentiation capacity, while retaining a normal karyotype throughout extensive culture (Hoffman and Carpenter, 2005). These cells have been derived for the first time by Jamie Thomson in 1998 at the University of Winscosin, by plating an isolated Inner Cell Mass (ICM) from a human blastocyst on inactivated Mouse Embryonic Fibroblasts (MEFs) (Thomson *et al.*, 1998) (**Fig. 1-1**). hES cells retain a distinct and unique subset of markers, from surface glycolipids and glycoproteins to transcription factors, distinct transcriptome profiles as well as enzyme subsets (Pera *et al.*, 2000). They also have the ability to differentiate in all the three germ layers – ectoderm, mesoderm and endoderm – probably forming every tissue and structure of an adult organism (Smith, 2001; Keller, 2005). Notably, these cells show only some of the features of their murine counterpart, mainly through a markedly different self renewal and pluripotency signalling pathways (Pera *et al.*, 2000). This may be due to the physiological differences that exist between murine and primate embryos during their development (Kaufman, 1992; Luckett, 1975), as monkey ES cells show a remarkable similarity to the human ones in terms of features and marker expression.

Typically, they seem to share the same Stage-Specific Embryonic Antigen (SSEA) as well as glycoprotein subset expression profile (Solter and Knowles, 1978; Thomson *et al.*, 1998; Pera *et al.*, 2000).

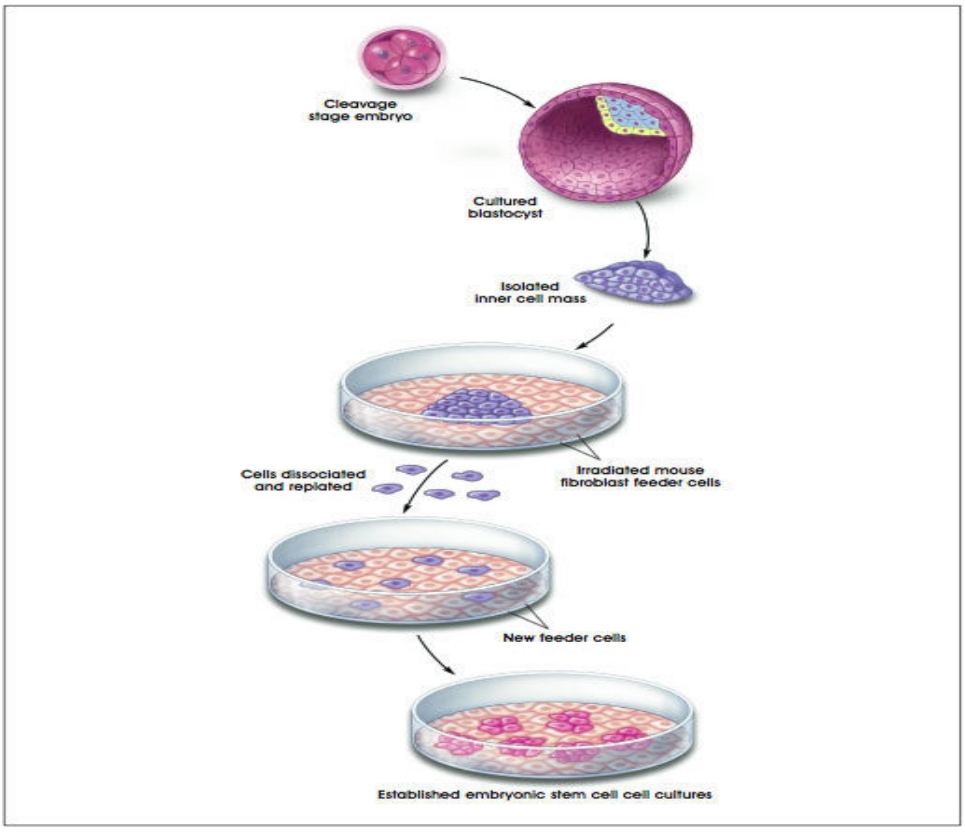

Figure 1-1. Derivation of hESC lines.

hESCs are isolated from the ICM of a cultured blastocyst, after removing the zona pellucida and the trophoblast. Colonies are then cultured on either irradiated or MMC treated MEFs for several passages until the establishment and the characterisation of the new stem cell line is carried out.
Key: hESCs = human Embryonic Stem Cells. ICM = Inner Cell Mass. MMC = MitoMycin C. MEF = Mouse Embryonic Fibroblasts.
Source: © Terese Wislow, Caitlin Duckwall, http://stemcells.nih.gov/info/scireport/appendixC.asp.
.

1.1.2 hES cell lines. Since the first derivation in 1998, many other scientists and laboratories have attempted to establish new stem cell lines, deriving more than two hundred new ones in the last few years (Brimble *et al.*, 2004). All these lines, however, show extensive divergences in terms of behaviour in culture, and of genetic and epigenetic signature

(Allegrucci and Young, 2007). In addition, scientists recently have succeeded in isolating hES cells from the morula stage embryo (Strelchenko *et al.*, 2004) as well as from later stage embryos (7-8 days) (Stojkovic *et al.*, 2004), likely adding further to the already existing great variability between lines. In fact, many are the variables involved in the process of derivation such as different genetic background and quality of the blastocysts employed, stress induced adaptation to culture and risk of selection of a subpopulation of cells. In addition, obvious and inevitable ethical issues, due to the manipulation of a human embryo, make the situation entirely different compared to the one researchers had to face for mES cells. For example, human ES lines are outbred, and all the embryos used to derive hES cells have come from *In Vitro* Fertilisation (IVF) clinics showing different ethnic and genetic backgrounds as well as quality (Abeyta *et al.*, 2004; Cai *et al.*, 2006). Strikingly, many differences can be seen between lines concerning growth rate, genetic and epigenetic stability in long term culture (Cowan *et al.*, 2004; Lee *et al.*, 2005; Allegrucci *et al.*, 2007), that could be due more to culture conditions, such as the presence of serum or specific growth factors, than to an actual difference in the genome (Skottman *et al.*, 2005; Allegrucci and Young, 2007). Many studies conducted using microarrays have also proved that variability in gene expression exists between independently derived stem cell lines (Abeyta *et al.*, 2004; Allegrucci and Young, 2007). These discrepancies are not likely to be dependent on the gender either (Abeyta *et al.*, 2004) and divergent genes do not cluster on any specific chromosome (Brandenberger *et al.*, 2004). Moreover, even passage number and enzymatic dissociation greatly contributes to hESC instability. Long-term culture can cause chromosomal abnormalities (especially gain of chromosomes carrying important self renewal or anti-apoptotic genes, like chromosome 12, 17 or X) or induce the cells to accumulate less visible

mutations because of the selective pressure resulting from the disruption of highly expressed adhesion molecules and gap junctions (Xu *et al.*, 2001; Caisander *et al.*, 2006). Furthermore, their epigenome and especially DNA methylation on Cytosine-phospho-Guanosine (CpG) islands can suffer from profound effects as a result of culture conditions (Allegrucci *et al.*, 2004; Steele *et al.*, 2005).

1.1.3 hESC features. Although many lines are now available, yet little is known about the signalling pathways that support pluripotency and self renewal in these cells. The picture that comes out from what has been investigated so far illustrates, surprisingly, a completely different condition from the mouse: Leukaemia Inhibitory Factor (LIF), the InterLeukin-6 (IL-6) family-related molecule that is necessary to support mES cells and maintain them in an undifferentiated state (Smith *et al.*, 1988), does not support hES cell. Even Bone Morphogenetic Protein 4 (BMP4), that in mouse embryonic stem (mES) cells acts together with LIF to promote pluripotency and avoid the differentiation towards neuroectoderm (Ying *et al.*, 2003), shows an entirely diverse function in hES cells. In hESCs, this molecule can induce differentiation to trophoblast (Xu *et al.*, 2002) while a family-related molecule, BMP2, induces extraembryonic endoderm (Pera *et al.*, 2004) through the activation of intracellular Sma/Mothers-Against-Decapentaplegic homologues (SMADs)-1/5/8. However, other growth factors and related proteins, including the Transforming Growth Factor-β (TGF-β) family (Amit *et al.*, 2004), Activin A (Beattie *et al.*, 2005) and Nodal (Vallier *et al.*, 2004) have been found to play a key role in the maintenance of hES cells by acting through Anaplastic Limphoma Kinase (ALK) receptors to activate SMADs-2/3 (Shi and Massague, 2003). In addition, a possible role in promoting self renewal has been suggested for WNT and related

ligands, acting through the canonical pathway, that inhibits the Glycogen Synthase Kinase-3β (GSK-3β) mediated differentiation (Sato *et al.*, 2004). However, other experiments proved that it was not the case, as inhibition of GSK-3β with lithium chloride resulted in differentiation of hES cells (Avery *et al.*, 2006). Roles in retaining hES cell pluripotency have been also suggested for Sphingosine-1-phosphate (S1P) and Platelet Derived Growth Factor (PDGF), involved in the WNT/GSK-3β pathway, when added to serum free and feeder free culture (Pebay *et al.*, 2005). A further role has been proposed for the Phosphatidylinositol-3 Kinase (PI3K) enzyme family that is activated by neurotrophins through TRK receptors (Pyle *et al.*, 2006), although PI3K seems to influence cell growth more than pluripotency. Finally, the Src-related family of Tyrosine Kinases, appears to be involved in self renewal. In particular cYes, a downstream target of LIF, probably stimulates proliferation (Anneren *et al.*, 2004) (**Fig. 1-2**).

Key transcription factors have been found that are now well established markers of both hES and mES cells. In both species, Oct3/4/OCT3/4, a member of the POU family (Hay *et al.*, 2004; Matin *et al.*, 2004), Sox2/SOX2, a Sex determining region of the Y chromosome (Sry) related transcription factor (Wei *et al.*, 2005), and Nanog/NANOG, a NK2 family homeobox protein (Hyslop *et al.*, 2005), together are capable of maintaining pluripotency and self renewal in the absence of feeders and growth factors. Conversely, hES cells have been found to express high levels of the cell surface antigens, SSEA-3 and SSEA-4, as well as of the TRA-1-60 and TRA-1-81 glycoproteins, whereas mES cells express SSEA-1 but neither SSEA-3 or SSEA-4 or the two TRA-glycoproteins (Reubinoff *et al.*, 2000; Thomson *et al.*, 1998).

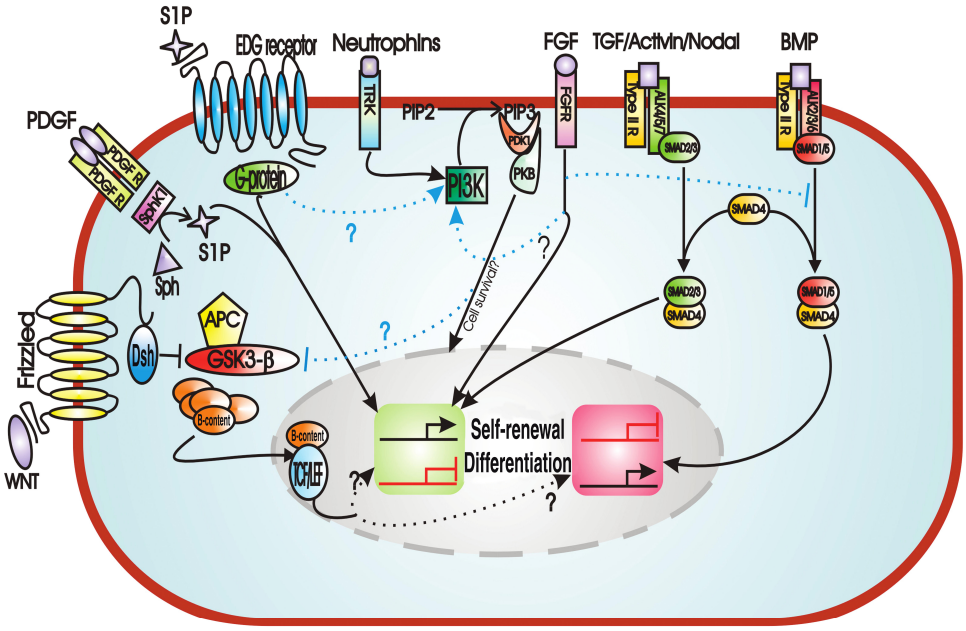

Figure 1-2. hESC signalling pathways.

Several pathways are thought to maintain and sustain self renewal and pluripotency in hESCs. TGF/Activin/Nodal can trigger the activation of SMAD-2/3 that, through the association with co-SMAD-4, results in the translocation to the nucleus and thus in self renewal gene expression. On the contrary, BMP related proteins can cause the activation of SMAD-1/5/8 that can recruit co-SMAD-4, translocate to the nucleus and promote differentiation. Furthermore, FGF is likely to act through its receptor, FGFR, to promote cell survival and sustain cell growth. In addition, neurotrophins and related receptors seem to contribute to self renewal through the recruitment of intracellular PI3K and consequent activation of PKB, probably sustaining cell survival. S1P and its receptor, EDG, can trigger the activation of G-protein dependent cascades in the cell that, together with intracellular S1P, generated by Sph PDGF mediated phosphorilation by SphK1, can promote self renewal. On the other hand, the role of WNT still remains not confirmed and confusing: this molecule seems to act through its canonical β-catenin dependent pathway, through the inhibition of GSK-3β and the nuclear translocation of the β-catenin/TCF/LEF complex; however, it is still to be confirmed whether this can trigger either differentiation or self renewal in hESCs. Cross-talk between pathways is highly probable.
Key: hESCs = human Embryonic Stem Cells. TGF = Transforming Growth Factor. SMAD = Sma/Mothers-Against-Decapentaplegic homologue. BMP = Bone Morphogenetic Protein. FGF = Fibroblast Growth Factor. FGFR = Fibroblast Growth Factor Receptor. PI3K = Phosphatidylinositol-3 Kinase. PKB = Protein Kinase B. S1P = Sphingosine-1-phosphate. EDG = Endothelial Differentiation Gene. PDGF = Platelet Derived Growth Factor. SphK1 = Sphingosine Kinase 1. GSK-3β = Glycogen Synthase Kinase 3-β. TCF = T Cell Factor. LEF = Lymphoid Enhancing Factor. Black arrows = Already proven pathways. Black dotted lines = Possible gene subset activations. Blue dotted lines = Possible interactions.
Source: Avery *et al.*, 2006.

1.1.4 hESC culture. Much effort is being made to optimise culture methods for hES cells, often by using growth factors to maintain pluripotency. Efforts have been directed to the use of defined, xeno-free media, avoiding the use of feeder cells and serum. hES cell have been derived and then maintained in long term culture using MEFs as a feeder layer in serum containing media (Thomson *et al.*, 1998; Reubinoff *et al.*, 2000). Later, MEF conditioned media were used in a feeder free system (Xu *et al.*, 2001). As scientists pursued development of defined media, other laboratories published the use of Knockout Serum Replacement (KSR) instead of whole serum. KSR has been supplemented with bFGF (Amit *et al.*, 2000), with a combination of TGF-β1, LIF and basic Fibroblast Growth Factor (bFGF) (Amit *et al.*, 2004), with bFGF and Noggin or with high concentrations of bFGF (Xu *et al.*, 2005). The urge for animal (xenogeneic) protein-free and easy to handle culture methods has boosted the search for feeder free substrates to culture hES cells. As ExtraCellular Matrix (ECM) is a very important factor to maintain pluripotency (Xu *et al.*, 2001), feeder layers have been substituted with tumor derived matrices like Matrigel (Wang et al., 2005: Xu *et al.*, 2005b), with humanised substrates such as Human Embryonic Fibroblasts (HEFs) Richards *et al.*, 2002; Hovatta *et al.*, 2003), or with purified human fibronectin/laminin (Amit *et al.*, 2004) to maintain unaltered embryonic features.

Another crucial aspect of stem cell culture is the passage method: both mechanical and enzymatic (usually collagenase IV or trypsin is used) passaging have been successfully employed. However, both methods present advantages and disadvantages: mechanical passage is time consuming and unsuitable for an even cell density, mostly because of the variability in colonies' size and their attachment in specific areas of the surface, whereas enzymatic passage is likely to result in culture disrupted

to single cell suspensions, so adding to the selective pressure that can generate anueploidies in culture (Brimble *et al.*, 2004; Hoffman and Carpenter, 2005). This is more of a problem when using trypsin rather than collagenase IV.

1.1.5 hES cell applications. The effort that scientists have been carrying out during the past 10 years to reach optimal and xeno-free culture condition are not aimless. Since the establishment of mES cells in culture in the early '80s, and then after the derivation of the first hES cell line in 1998, entirely new approaches to studying mammalian development and exploiting these cells for regenerative therapies have been created (Smith, 2001). In fact, hES cells not only represent an exciting, innovative and unique tool to get a deeper insight on early human development, but they could also be precious and powerful instruments to achieve wound regeneration, cardiac (Klug *et al.*, 1996) and brain repair, as well as a cure for some chronic and degenerative diseases, such as diabetes (Soria *et al.*, 2000) and Parkinson's disease (Svendsen and Smith, 1999), to cite only a few. Thus, stem cells clearly represent a very promising instrument: indeed, they could be employed in modelling human diseases in a more realistic way than currently used animal models (Semb, 2005). They could also allow more reliable tests for developing new drugs, especially in regard to toxicity and side effects that in many cases have been shown to be species specific (Gregory *et al.*, 2005). Furthermore, they stand as a state of the art model to investigate master genes and signalling pathways that either guide differentiation or help maintaining the "stemness" state in the human blastocyst. Eventually, they may be extensively employed to investigate and study the more differentiated stages of early human development.

1.2 hESCs as a Model for Early Human Development

1.2.1 A model for the human embryo. As the manipulation of a human embryo was not permitted by law in many countries (and still is not) and unethical, so far the extent to which analysis of early developmental stages has been carried out was limited to spontaneous mutations and genetically inherited diseases in the population that was analysed using either the classic or inverted genetic approach (Dvash and Benvenisty, 2004). Animal models have been utilised as an alternative, allowing the use of well defined and reproducible experimental conditions (Darling, 1996). Being a rather close model to humans, having a quick reproductive rate and easy-to-manipulate embryos, the mouse has been chosen as a model organism to mimic pathologies and understand mammalian development. Although the use of inbred mouse strains can remove genetic variability, this does not reflect the genetic diversity in humans. Often rodents also undergo crucially different developmental mechanisms (Dvash and Benvenisty, 2004). Although supernumerary, Artificial Reproductive Technology- (ART) derived embryos could provide a model for human development, their limited supply, likely abnormalities and 14 day limit of culture (in the UK) suggest hESCs could be an alternative choice in order to unravel the early development of the human embryo (Allegrucci *et al.*, 2005).

1.2.2 A model for human gastrulation. To date, many data are available about gastrulation and its regulation from the mouse, in particular about germ layer formation and primitive streak creation in the developing embryo (Tam and Loebel, 2007). This not the case for the

human embryo for which little is known. Many features of preimplantation development are shared amongst mammals. In all mammals the formation of the blastocyst generates two different populations of cells, the trophectoderm and the ICM (Gilbert, 2003). As a consequence of delamination of cells from ICM, the hypoblast is formed to give rise to the extraembryonic endoderm and the yolk sac of the foetus. The remaining cells, forming the epiblast, will give rise to all the tissues of the proper embryo and to the amniotic cavity (Gilbert, 2003). Gastrulation then begins to form the anteroposterior axis of the embryo through cell migration. By the end of this process ectoderm, mesoderm and endoderm will be formed as a consequence of transcription factor mediated activation of differentiation genes.

1.2.3 hEB formation: mimicking the human gastrula. In order to model gastrulation in the human embryo, differentiating hESCs into embryoid bodies (EBs) could represent an interesting choice to investigate germ layer formation and cell lineage differentiation (Itskovitz-Eldor *et al.*, 2000). hEBs can be formed by the *in vitro* aggregation of hES cells in suspension cultures. During hEB formation, hESCs are removed from adherent culture and placed in an appropriate differentiation medium where they can start to form clusters and follow lineage specific cell fates (Itskovitz-Eldor *et al.*, 2000). During differentiation, stem cells can give rise to populations belonging to any of the three germ layers and tissue types (Itskovitz-Eldor *et al.*, 2000). This process mimics some of the *in vivo* developmental stages of the embryo (Mossman et al., 2005). As hEBs develop, they pass through different stages presenting, at first, endodermal cell formation and then undergoing a process of cavitation and expansion that results in blastocyst like fluid-filled cystic EBs (Itskovitz-Eldor *et al.*, 2000). Although hEB differentiation is spontaneous and disorganised

compared to the developing blastocyst, they represent a potentially useful model for early human development in which analysis of temporally and spatially regulated expression of germ layer specific differentiation genes can be assessed and investigated.

1.2.4 Ectodermal differentiation and associated markers. Ectodermal differentiation is very common in spontaneously differentiating hESCs and has been suggested to be a default pathway for these cells (Reubinoff *et al.*, 2000; Reubinoff *et al.*, 2001). This can occur when Noggin-like BMP antagonists are present in culture and therefore inhibit endomesodermal differentiation (Pera *et al.*, 2004). Many kinds of neuroepithelial lineage-derived cells have been obtained using varying culture methods, including co-culture (Kawasaki *et al.*, 2002; Mizuseki *et al.*, 2003) or specific growth factor (all *trans*-Retinoic Acid (RA), bFGF, Epidermal Growth Factor (EGF)) media supplementation (Nistor *et al.*, 2005). In these ways, Tyrosine Hydroxylase positive (TH+) neurons and dopaminergic subsets (Schulz *et al.*, 2003; Schulz *et al.*, 2004), *NURR1* and *LMX1b* positive (Kawasaki *et al.*, 2002), and Pax family genes expressing (Perrier *et al.*, 2004) neuronal cells have all been derived. Although these genes and related proteins can be used to detect specific population of neuron-like progenitors in culture, these are not early ectodermal markers. In respect to this, SRY Box containing (*SOX)* family genes have been found to play key roles in cell fate determination and regulation (Zhao *et al.*, 2004) and are known to be expressed throughout the entire neural development. Mouse *Sox3* is expressed in primitive ectoderm before the primitive streak stage of the embryo (Wood and Episkopou, 1999) and *SOX1* has been found active in the ectodermal germ layer after the neural plate formation (Zhao *et al.*, 2004). Interestingly, *SOX1* expression has been found in hESC derived ectoderm at the

Primitive Anterior Neuroepithelia (PAN) stage 10-11 days post differentiation and activated later in both the anterior and posterior neural plate and tube, reaching a peak of expression at 14-16 days post differentiation (Pankratz *et al.*, 2007) (**Fig. 1-3**). This clearly makes this protein an important marker to investigate ectodermal formation in differentiating hEBs.

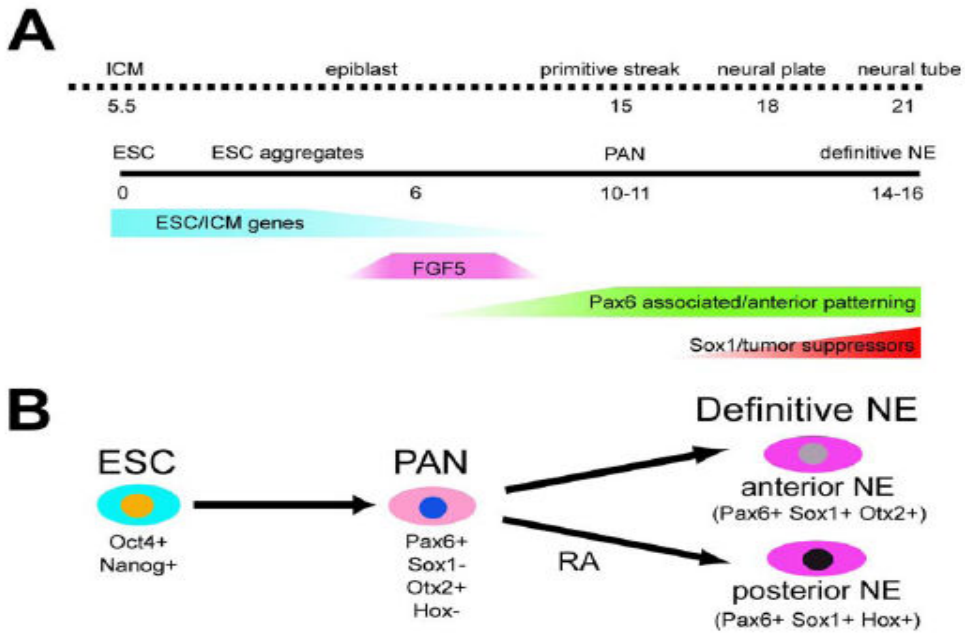

Figure 1-3. Ectodermal differentiation and associated gene expression.

A) Temporal comparison between events occurring *in vivo* in the human embryo and the ones taking place *in vitro* concerning hESCs differentiation towards the neuroectodermal lineage. The epiblast originates from the ICM in the developing blastocyst at 5.5 dpc; during gastrulation the primitive streak is formed (15 dpc) and, afterwards, neural plate and neural tube develop (18-21 dpc) amongst the neuroectodermal cell population. *In vitro*, hESCs are isolated at 5.5. dpc from the blastocyst and reach the PAN stage at 10-11days from the beginning of the differentiation process and finally give rise to definitive NE 3-6 days later. This, at first, goes along with the expression of pluripotent genes that are then downregulated to allow the expression of lineage specific ones: firstly *FGF5* (5-7 days after differentiation has begun), then *PAX6* and eventually *SOX1*.
B) Model of neuroectodermal development in humans: hESCs expressing pluripotent genes like *NANOG* and *OCT4* give rise to PAN *PAX6*+, *SOX1*-, *OTX2*+ and *HOX*- cells. These cells then form the definitive NE: anterior NE is created by default, whereas posterior NE develops after RA stimulation. While anterior NE is *PAX6*+, *SOX1*+ and *OTX2*+, posterior NE expresses *PAX6*, *SOX1* and *HOX*.
Key: ICM = Inner Cell Mass. Dpc = Days post coitum. ESC = Embryonic Stem Cells. PAN = Primitive Anterior Neuroepithelia. NE = NeuroEpithelia. FGF5 = Fibroblast Growth Factor 5. PAX6 = Paired bOX gene 6. SOX1 = SRY bOX-containing gene 1. HOX = Homeobox-containing gene. OTX2 =Orthodenticle homologue 2. RA = all-*trans* Retinoic Acid.
Source: Pankratz *et al.*, 2007.

1.2.5 Mesodermal differentiation and associated markers. Mesoderm is another germ layer that is formed by hESCs prior to differentiation into muscle, kidney, lung, bone and hematopoiteic progenitor cells. Scientists have obtained cardiomyocytes (Mummery *et al.*, 2002; Mummery *et al.*, 2003), alveolar pneumocytes and airway epithelial tissues (Denham *et al.*, 2006; Coraux *et al.*, 2005), keratinocytes (Green *et al.*, 2003) and myeloid and erythroid progenitors (Chadwick *et al.*, 2003). Mesoderm derived cells have been found to express specific fate related transcription factors like Nkx2.5, GATA4, MEF3 (Kehat *et al.*, 2001; Xu *et al.*, 2002), Protein C (Denham *et al.*, 2006), p63, cytokeratin14 and basonuclin (Green *et al.*, 2003). Beside these markers of mesodermal derivatives, early ones like Brachyury (T, belonging to the T-box gene family), Goosecoid and Mixl1 have been highlighted in respect to mesodermal lineages in the mouse (Gaunt *et al.*, 1993; Hart *et al.*, 2002). Human T expression has been assessed in hESCs as well as in differentiating hEBs (Cai *et al.*, 2006) and mice homozygous mutant for these genes showed gastrulation inhibition and primitive streak abnormalities (Yamada, 1994; Robb *et al.*, 2000). The fact that T is one of the earliest marker of mesoderm formation, and that its absence affect posterior mesoderm formation and notochord differentiation (Berry *et al.*, 1999) makes this transcription factor a useful marker to identify mesoderm development in hEBs.

1.2.6 Endodermal differentiation and associated markers. Despite the lack of many specific markers for early endoderm progenitors (Trounson, 2006), definitive endoderm has been induced in mouse (Kubo *et al.*, 2004) as well as in human (Assady *et al.*, 2001) ESCs by enrichment of spontaneous differentiating cells (Brolen *et al.*, 2005) or by culturing in Insulin Transferrin Selenium Fibronectin (ITSF) medium

(Segev *et al.*, 2004). Both pancreatic like β-cells (Segev *et al.*, 2004) and hepatocyte progenitors (Rambhatla *et al.*, 2003) have been produced, demonstrating the expression of genes like albumin, α-1-antitrypsin, cytokeratin8, insulin-2 (Segev *et al.*, 2004; Rambhatla *et al.*, 2003). As for early endodermal markers, SOX17 represents a transcription factor of the SOX family that has been demonstrated to function as an endoderm determinant in the mouse, where negative homozygotes of this gene showed gut endoderm deficiency (Kanai-Azuma *et al.*, 2002). SOX17 can therefore be employed to distinguish endoderm-oriented differentiation in hEBs.

1.3 hESCs: a Model to Study Foetal Origin of Adult Diseases

1.3.1 Developmental origin of adult diseases. Since the mid '70s, many scientists have begun to investigate possible effects of maternal nutrition on preconceptual, periconceptual, prenatal and post natal stages of mammalian development (Ravelli *et al.*, 1976; Barker and Osmond, 1986). Based on epidemiological evidence, a link was hypothesised to occur between suboptimal maternal nutrition and irregular foetal development leading to a predisposition to a range of adult pathologies such as cardiovascular diseases and obesity (Barker, 1994; Harding, 2004). Since then, many experiments conducted in mice and rats have

clearly shown how dietary restrictions in protein or calories during early, mid or late gestation can cause hypertension (Langley and Jackson, 1994), impaired nephrogenesis (Langley-Evans *et al.*, 1999) and glucose intolerance leading to diabetes (Ravelli *et al.*, 1998; Painter *et al.*, 2005) in the offspring. All together, these findings seem to correlate *in utero* nutrient availability to the onset of chronic, degenerative diseases in the progeny later in life. Accumulating evidence about this hypothesis has lead the scientific community to formulate a Developmental Origin of Health and Disease (DOHaD) theory that strongly associates nutritional programming during embryonic and foetal development to the offspring health status (Gluckman and Hanson, 2004; Gillman, 2005). Many additional pathologies such as stroke, asthma (Tantisira and Weiss, 2001), cancer and osteoporosis (Newnham *et al.*, 2002; Ozanne *et al.*, 2004) have been also correlated with DOHaD. Dietary components can affect multiple processes in the mammal developing embryo and foetus such as gene expression, cell cycle and hormone metabolism modifying the incidence and extent of a disease (Trujillo *et al.*, 2006).

1.4 DNA Methylation and Nutritional Programming

1.4.1 DNA methylation. Epigenetic mechanisms can control gene expression via either histone/chromatin or DNA modifications without any nucleotide sequence alteration (Henikoff and Matzke, 1997). One of these processes is represented by DNA methylation, an epigenetic mechanisms that has been associated with gene silencing (Waterland and Jirtle, 2003; Allegrucci *et al.*, 2005). These changes are stable and inheritable and can thus provide an explanation for how external factors can alter gene expression without causing a mutation in the genetic code (Paoloni-Giacobino and Chaillet, 2004) (**Fig. 1-4**). The most studied epigenetic mechanism is DNA methylation that concerns the addition of a methyl group to designated "C" residues (at carbon 5) belonging to CpG dinucleotides (Paoloni-Giacobino and Chaillet, 2004). CpG rich regions known as CpG islands have been shown to regulate the expression of many housekeeping genes and to play a critical role in mammalian development acting on expression level dosage control of a small gene subset called imprinted genes (Thurston *et al.*, 2007).

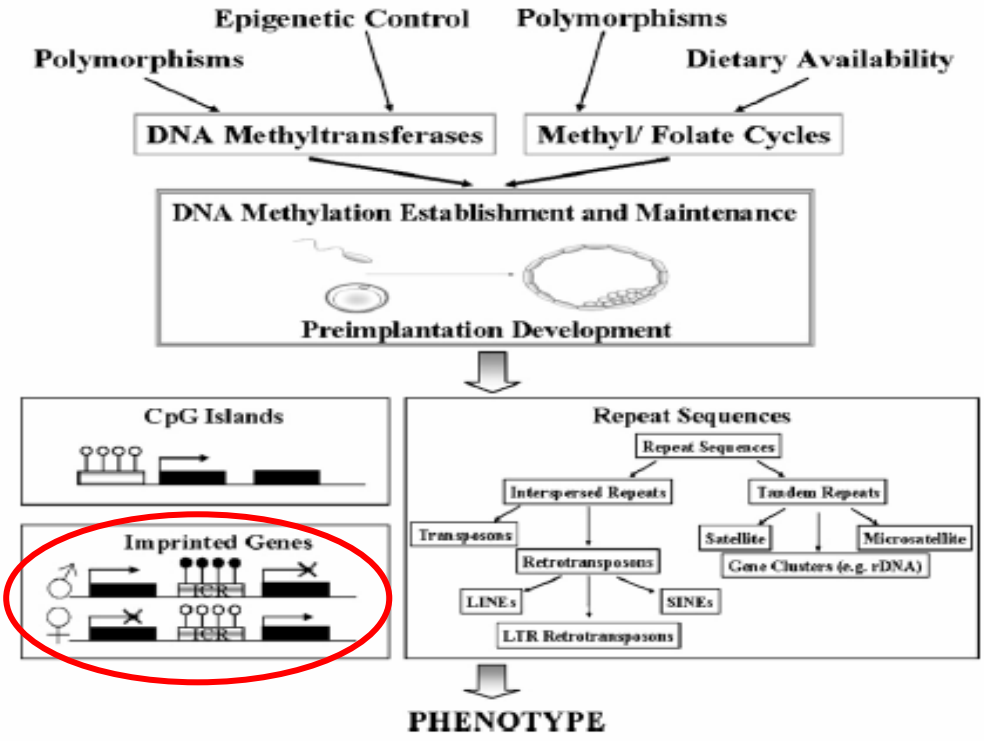

Figure 1-4. DNA methylation establishment during mammalian development.

DNA methylation is established by DNMTs in mammals and influenced by many environmental and genotypic factors such as dietary availability of methyl substrates, methyl cycle or DNMT enzymes' polymorphisms, other epigenetic controls such as histone modifications. DNA methylation is an active process in the early embryo, where the imprinting is erased and re-established to be maintained in the adult life and inherited by all dividing cells. This process can affect and regulate gene expression, covalently modifying specific elements in the genome: promoter specific CpG islands, imprinted genes' ICRs, repeat sequences such as transposons, LTR retrotransposons, LINEs, SINEs, satellite and microsatellite DNA, gene clusters. These modifications will eventually affect the phenotype of each individual, modifying his transcriptome.
Key: DNMTs = DNA MethylTransferases. CpG = Cytosine-phospho-Guanosine. ICR = Imprinting Control Region. LTR = Long Terminal Repeats. LINE = Long Interspersed Element. SINE = Short Interspersed Element.
Source: Thurston *et al.*, 2007.

1.4.2 DNA methylation pattern establishment. DNA methylation is established and maintained in the cell by the action of at least three families of DNA methyl transferase enzymes (Dnmts) (1, 2 and 3) cloned and classified in the mouse according to their C-terminal catalytic domain. Dnmt1 exists in three isoforms, Dnmt1o is oocyte specific, Dnmt1s is somatic cell specific, and Dnmt1p is pachytene spermatogonia specific. The two germ cell specific isoforms (1o and 1p) are thought to be important for imprinted gene allele specific methylation pattern in the developing gametes (Mertineit *et al.*, 1998). All these isoforms present methyltransferases activity and high specificity for hemimethylated replicating DNA and they are so called maintenance methyltransferases (Bestor, 1992). On the other hand, Dnmt2/DNMT2 has been identified in mammals (Yoder and Bestor, 1998) and is ubiquitously expressed in somatic cells as well as in mESCs. This DNMT, anyway, seems to target RNA rather than DNA, as it has been shown able to catalyse the methylation of tRNA (Okano *et al.*, 1998) even if some proposed it could be involved in DNA repair (Pradhan and Esteve, 2003). Dnmt3/DNMT3 is a family that consists of three already identified isoforms highly expressed in ESCs (Chen *et al.*, 2003) and called Dnmt3a, Dnmt3b and Dnmt3l with the last lacking a catalytic activity. Dnmts 3a and 3b are involved in *de novo* methylation during the post implantation stage of mammalian devolpment (Buryanov and Shevchuk, 2005), targeting different types of sequences in the genome: Dnmt3a was demonstrated to be responsible for *de novo* methylation of major satellite repeats and imprinted genes in specific CpG rich regions (Chen *et al.*, 2003; Oka *et al.*, 2006), whereas Dnmt3b seems to carry out methylation of minor satellite repeats (Okano *et al.*, 1999). Dnmt3a isoform 2 and a splice variant, Dnmt3a-β, have been shown to be preferentially expressed by mESCs (Weisenberger *et al.*, 2002). Dnmt3b isoform 2 and isoform 3 have been identified as the most

abundant isoforms expressed in hES cell lines, adding evidence of their role in early development. On the other hand, Dnmt3l has been demonstrated necessary as a cofactor for the activity of the other two related enzymes and crucial for the correct establishment of imprinted gene methylation during gametogenesis (Hata *et al.*, 2002). Human *DNMT3L* instead, seems to be expressed only post fertilisation suggesting different functions or a timing difference in activation (Young and Beaujean, 2004). The expression of this enzyme has also been demonstrated to be regulated by promoter methylation on target CpG sites in hESCs and somatic cells (Aapola *et al.*, 2004). Dnmt enzymes are known to be expressed in a strictly regulated fashion in order to maintain a correct methylation imprinting throughout the mammalian cell cycle. Thus, Dnmt1/DNMT1 is highly expressed in adult somatic cells and its inhibition was shown to cause aberrant methylation patterns often leading to tumorigenesis and developmental abnormalities.

1.4.3 Epigenetic reprogramming in the developing embryo. Although adult somatic mammalian cells possess a stable DNA methylation pattern (Surani, 2001), the embryonic genome clearly undergoes major and important changes in DNA methylation during development (Reik *et al.*, 2001; Lee *et al.*, 2002) (**Fig. 1-5**). In the mouse, the post fertilisation zygote undertakes an extensive both passive and active demethylation that continues up to the blastocyst stage when the majority of the CpG sites are unmethylated except for some of those belonging to imprinted genes and specific repeat sequences like transposons (Razin and Shemer, 1995). Notably, this general demethylation process is regarded as species-specific presenting some distinguishable features in different organisms studied (Young and Beaujean, 2004). After implantation, genome-wide *de novo* methylation is

restored establishing inheritable adult patterns and epigenetic signature, and then passed unaltered to the daughter cells upon each cell cycle (Buryanov and Shevchuk, 2005). This process has been demonstrated essential for a normal development with the depletion of the enzymes involved resulting in embryonic death as shown by Li *et al.* (1992) in the mouse and by Stancheva and Meehan (2000) in *Xenopus,* or in disease onset in the adult as shown by Wijmenga *et al.*, (1998) in humans. In fact, they demonstrated an existing correlation between centromeric Instability of Chromosomes 1, 9, and 16 and Facial anomalies (ICF) syndrome and extensive hypomethylation, especially in satellite repeat rich regions. During genome re-methylation, an asymmetry is established in the blastocyst resulting in a hypermethylated ICM and a hypomethylated trophoblast, at least in the mouse (Santos *et al.*, 2002). All these active processes of methylation establishment and maintenance are carried out by DNTMs (Meehan, 2003).

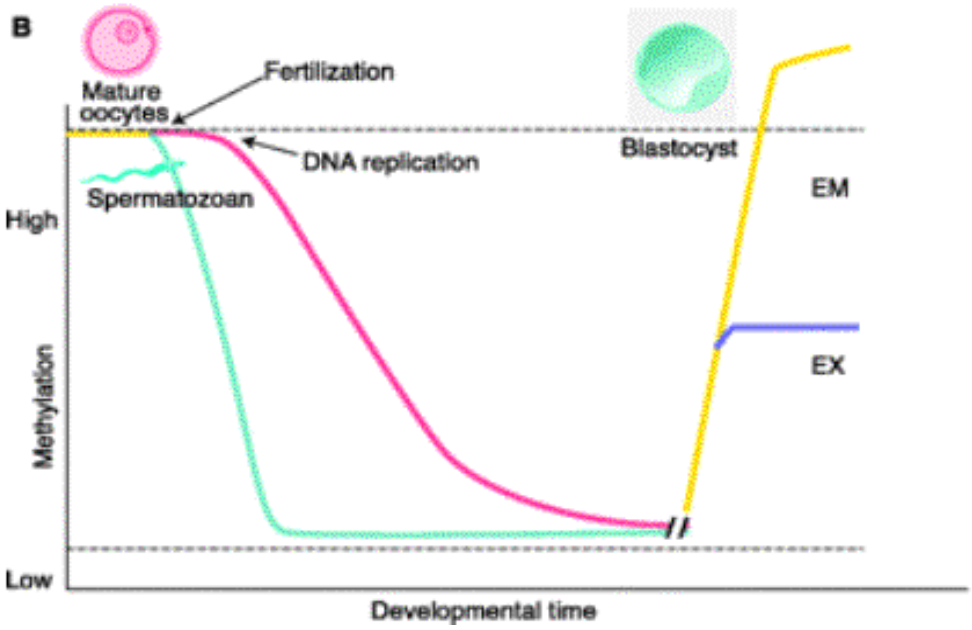

Figure 1-5. Epigenetic reprogramming in the mammalian embryo.

After fertilisation, a mammalian zygote contains sperm-derived as well as oocyte-derived DNA that present epigenetic feature and DNA methylation typical of mature germ cells. This signature has to be erased to re-establish a new, embryo-specific methylation pattern. While oocyte-derived DNA undergoes passive demethylation by extrusion of Dnmt1 from the nucleus and consequent, progressive loss of methylation on newly synthesised strands, sperm-derived DNA undertakes a faster, active demethylation process that is completed in a few cell cycles. Methylation is then restored in the developing blastocyst by *de novo* DNMTs Dnmt3a and Dnmt3b along with the formation of differentially met ylated areas: hypomethylation marks EX derivatives whereas hypermethylation identifies EM tissues.
Key: Dnmt = DNA Methyltransferase. EX derivatives = Extra Embryonic derivatives. EM tissues = Embryonic tissues.
Source: © Terese Wislow, Caitlin Duckwall, http://stemcells.nih.gov/info/scireport/

1.4.4 Methyl/folate cycle. Cellular reactions involving methylation are carried out by different enzymes that utilise a derivate of adenosine, S-Adenosyl Methionine (SAM), as the universal methyl donor (Van den Veyver, 2002); DNMT mediated DNA methylation does not represent an exception (McCabe and Caudill, 2005). The production and effective availability of this compound is regulated by the combined methionine and

folate cycles that entail aminoacids like Glycine (Gly), Serine (Ser), Methionine (Met) and cofactors like selenium, vitamin B6, B12 and folate (**Fig. 1-6**). These cycles aim to produce methyl donors reconverting S-Adenosyl Homocysteine (SAH), produced by the demethylation of SAM, to methionine and then back to SAM to make it available again. A crucial step of this cycle is represented by the conversion of homocysteine to Met by the enzyme Met Synthase. This enzyme requires Methyl-Tetrahydrofolate (Me-THF) as a substrate and vitamin B12 (cobalamin) as a cofactor for the methylation reaction to produce Met (Wolthers and Scrutton, 2007). Furthermore, the enzyme THF Reductase needs vitamin B6 as a cofactor for Me-THF to be produced again from THF (Aguilar *et al.*, 2004), creating a methyl cycle dependence on folate. DNMTs then, provided with SAM, methylate DNA on CpG sites causing the accumulation of SAH that has to be converted by the enzyme SAH hydrolase (SAHh) to adenosine and homocysteine for the cycle to continue. As a consequence, alteration in the concentration of any of these cofactors/substrates could lead to an impairment of the cycle and affect DNA methylation; thus, because of the active epigenetic remodelling happening in the developing embryo, this effect could be even greater on fetal development and contribute to the onset of related diseases in the adult (Allegrucci *et al.*, 2005).

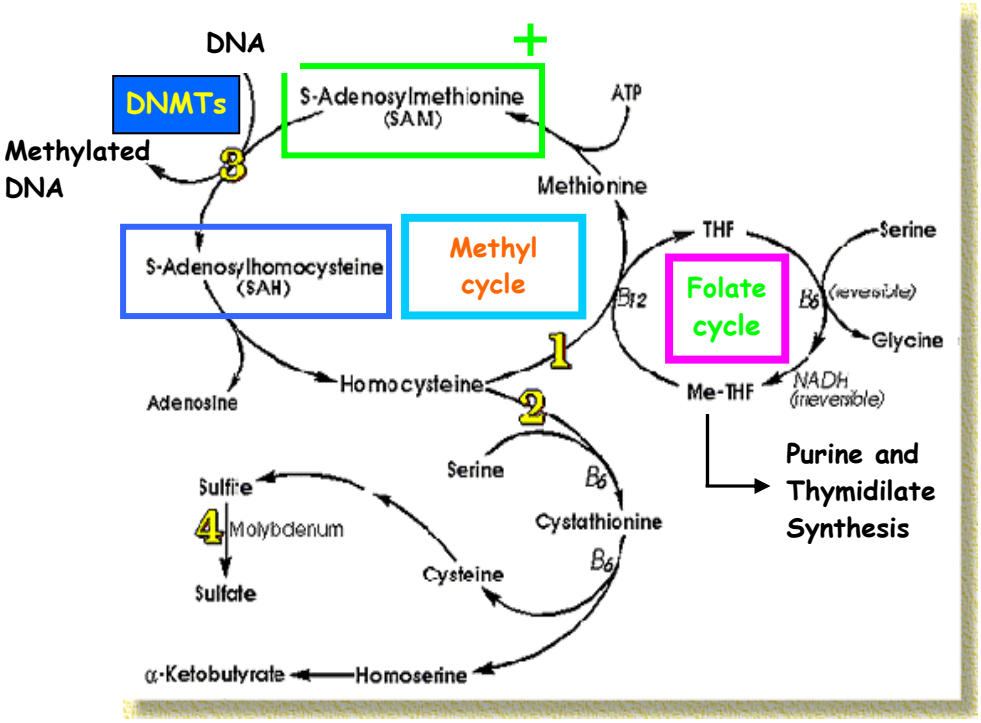

Figure 1-6. Methyl/folate cycle in mammalian cells.

Methyl and folate cycle are two intimately connected biochemical cycles that have the purpose of producing the universal methyl donor, SAM, and Me-THF for methylation and biological synthesis that require a carbon unit, respectively. Maintaining a balance in the cycle is essential to ensure correct establishment of DNMT-mediated (3) DNA methylation pattern as well as purine synthesis for the DNA replication process. In order for cycle enzymes to do so, folate and important cofactors such as vitamin B6 and B12 are needed to catalyse cycle reactions and reconvert SAH into SAM. Since the only source for these compounds is dietary intake, adequate levels of these components are required in a balanced diet. In methyl cycle, Met is converted to SAM by MAT; DNMTs utilise SAM to methylated "C" residue on target CpG sites, producing SAH. In order for SAH to be reconverted to SAM, the enzyme SAHh and MS (1) (that needs vitamin B12 as a cofactor) produce Ado and Homocysteine that is finally reconverted to Met, for the cycle to restart. As Me-THF is an essential methyl donor for MS to produce Met from Homocysteine, a constant supply of THF, synthesised by diet-derived folate, is crucial. In addition, Homocysteine can exit the cycle and be either converted to Cysteine as a source for intracellular sulphur or produce Homoserine and α-ketobutyrate.
Key: SAM = S-Adenosyl-Methionine. Me-THF = Methyl-TetraHydroFolate. DNMT (3) = DNA MethylTransferase. MAT = Methionine-Adenosyl-Transferase. CpG = Cytosine-phospho-Guanosine. SAH = S-Adenosyl-Homocysteine. SAHh = S-Adenosyl-Homocysteine hydrolase. MS (1) = Methionine Synthase. Ado = Adenosine. Cys = Cysteine. (2) = Cystathionine-β-Synthase. (4) = Sulphur metabolism pathway. ATP = AdenosineTriPhospate.
Source: http://www.ceri.com/sam.gif.

1.4.5 Epigenetic consequences of early nutritional programming. A perturbation in the methyl/folate cycle and in the involved metabolites' concentration has been shown to have profound effects on the availability of methyl groups to be employed for DNA and histone methylation (Scott, 1999). Interestingly, a correlation between early embryo epigenetic defects and aberrant methylation caused by methyl/folate cycle disruption has been suggested (Steele *et al.*, 2005). These defects are thought to rise from an altered maternal diet during early stages of pregnancy and believed to be reproduced by in vitro ART manipulation of embryos when non physiological concentration of these metabolites are present in the media employed (Steele *et al.*, 2005). Notably, maternal dietary restriction in rats has revealed profound effects on glucocorticoid receptor, 11β-hydroxysteroid dehydrogenase and on peroxisomal proliferators-associated receptors in all the three germ layer derivatives of the foetuses (Bertram *et al.*, 2001); alteration of these genes has been correlated to cardiovascular and metabolic diseases (Lillycrop *et al.*, 2005). Folate, particularly, has been demonstrated essential to avoid neural tube defect in the embryo (Berry *et al.*, 1999). As folate is exclusively derived from dietary intake and only present in poor quantities in edible vegetables (Suitor and Bailey., 2000), a further supplementation is often recommended especially for pregnant women (De Bree *et al.*, 1997). On the other hand, supplementation of maternal diet with methyl donors has been shown to permanently restore DNA methylation on epigenetically susceptible loci during early embryonic development (Waterland and Jirtle, 2003). Waterland and Jirtle (2003), in fact, using a specific diet containing different levels of methyl donors, demonstrated major changes in the methylation status of an Intracisternal A Particle (IAP) retrotransposon located nearby the *agouti* gene in mice. Furthermore, this has been shown to be able to influence and affect the

expression of the gene provoking phenotypic variation in the tissues examined. Hence, they proved how epigenetic regulation of gene expression by DNA methylation can be deeply influenced in the offspring by early maternal nutrition and by different levels of methyl donors (**Fig. 1-7**).

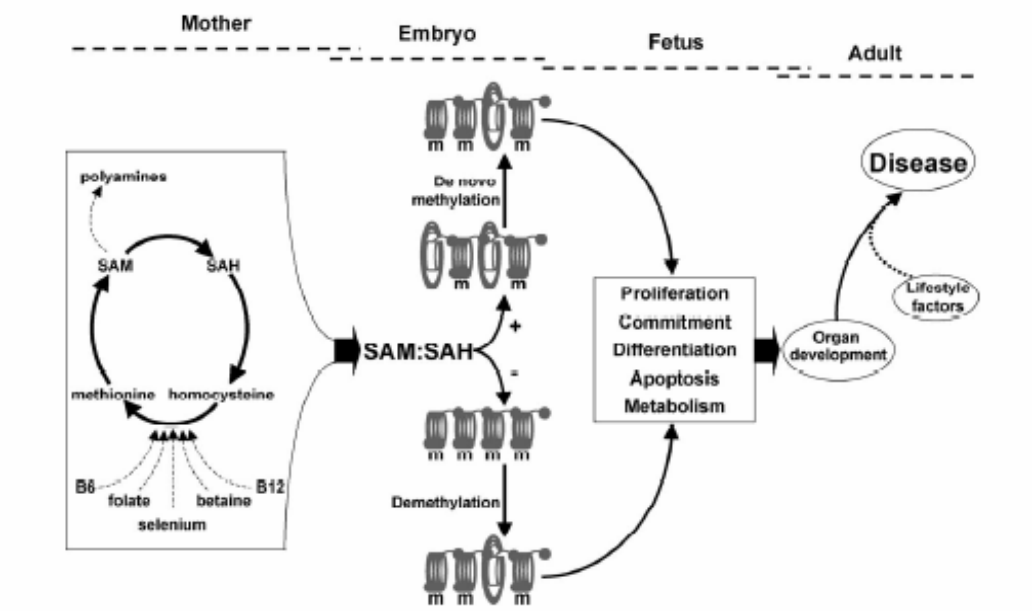

Figure 1-7. Methyl/folate cycle influence on DOHaD.

Altered levels of methyl cycle substrates in the mother can alter SAM:SAH ratio in the embryo and can consequently cause either hyper- or hypomethylation in its genome. This can therefore affect proliferation, commitment, differentiation, apoptosis and metabolism in general during foetal life. Eventually, disruption in this essential cellular mechanisms and lifestyle factors can have profound effects on diseases' onset in adulthood.
Key: DOHaD = Developmental Origin of Health and Disease. SAM = S-Adenosyl-Methionine. SAH = S-Adenosyl-Homocisteine. m = methyl group. + = SAM:SAH ratio >1. - = SAM:SAH ratio <1.
Source: Allegrucci *et al.*, 2005.

1.4.6 Nutritional programming and genomic imprinting. After the first discovery of mammalian genomic imprinting (McGrath and Solter, 1984; Surani *et al.*, 1984), this has been defined as an epigenetically regulated differential expression of a gene associated with parent-of-origin inheritance (Hitchins and Moore, 2002), resulting in a monoallelic (either paternal or maternal) expression of that gene. Imprinted genes are thought to regulate key processes such as cell cycle, growth and differentiation (Li *et al.*, 1999). Strikingly, despite being thought to affect only a few genes (Thurston *et al.*, 2007), the establishment of proper imprinting has been found vital for a correct development of the mammalian embryo and its disruption produced relevant growth disorders such as Large Offspring Syndrome (LOS), Beckwith-Wiedemann Syndrome (BWS) and Angelman syndrome (Hitchins and Moore, 2002). Nutritional programming and altered availability of methyl/folate cycle metabolites not only has effects on housekeeping genes but also on imprinted genes (Harding *et al.*, 2001). This could affect embryonic and foetal growth as well as placental transport system (Reik *et al.*, 2003). Disruption or deregulation of methylation could lead to imprinting aberrations such as Loss Of Imprinting (LOI), resulting in biallelic expression of the gene (Feinberg *et al.*, 2002), or gene silencing of both alleles (Waterland and Jirtle, 2004). Moreover, it has already been shown how ART induced imprinting abnormalities could be considered as a cause for anomalous foetal growth and LOS in the sheep (Young *et al.*, 2001). Besides, murine embryo culture has provided additional clues concerning the association between culture media and epigenetic defects (Khosla *et al.*, 2001).

.

2. Aim of the Project

A link has been hypothesised to occur between suboptimal maternal nutrition and irregular foetal development leading to a predisposition to a range of adult pathologies. Perturbations in the methyl/folate cycle and in the involved metabolites' concentration have been shown to have profound effects on the availability of methyl groups to be employed for DNA and histone methylation. As a clear connection between dietary intake of methyl group donors and epigenetic defects has been demonstrated both *in vivo* and *in vitro*, this project aims at proving that *in vitro* induced disruption in the methyl/folate cycle could lead to an abnormal SAM:SAH ratio and thus to an aberrant DNA methylation pattern in the pre implantation embryo. In order to achieve this, hESCs have been employed as a potential model for the human blastocyst as well as hEBs for the differentiating gastrula. Particularly, suitable methyl substrate-deficient media have been designed to analyse their effects on DNA methylation of human imprinted genes and their allelic expression. In addition, hEBs have been examined to investigate differentiation abnormalities and impaired germ layer formation. This could add new evidence to the hypothesis of a correlation between the availability of dietary methyl donor nutrients and epigenetic programming in the developing embryo.