Molecular regulations and lineage tracing in early Mouse embryogenesis

Wenyu Zhou

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Hannele Ruohola-Baker, Chair

Marshall S. Horwitz, Chair

Billie J. Swalla

Program Authorized to Offer Degree:

Biology

University of Washington

Abstract

Molecular regulations and lineage tracing in early Mouse embryogenesis

Wenyu Zhou

Chair of the Supervisory Committee:

Professor Marshall S. Horwitz

Department of Medicine

Professor Hannele Ruohola-Baker

Department of Biochemistry

This dissertation research aims to understand the regulations and developmental patterns of mouse early embryogenesis, in order to gain comprehensive knowledge in human early development to further the ultimate goal of designing future novel and effective regenerative therapies. The first part of this research focuses on transcriptional and metabolic regulations that take place in the embryo during the transition from pre- to post-implantation. By utilizing *in vitro* cell lines established from these two stages (pre-implantation: mESC, and post-implantation, EpiSC/hESC, respectively), we found a dramatic metabolic difference between them. We found that EpiSC/hESC are highly glycolytic, while ESC are bivalent in their energy production, dynamically switching from glycolysis to mitochondrial respiration on demand. Despite having a more developed and expanding mitochondrial content, EpiSC/hESC

have low mitochondrial respiratory capacity due to low cytochrome c oxidase (COX) expression. We further show that HIF1α is sufficient to drive ESC to a glycolytic Activin/Nodal-dependent EpiSC like stage. The second part of the research examines patterns of cell lineage development utilizing somatic mutations that inevitably arise with every cell division (a method dubbed "phylogenetic fate mapping"). We cataloged genomic mutations at an average of 110 mutation-prone polyguanine (polyG) tracts for about 100 cells clonally isolated from various corresponding tissues of two sibling littermates of a hypermutable mouse strain. We found that during mouse development, muscle and fat arise from a mixed pool of progenitor cells in the germ layer, but, in contrast, vascular endothelium in brain is derived from a smaller source of progenitor cells. We quantitatively demonstrated that development generally represents a combination of stochastic and deterministic events, as reflected in only partial conservation of cell lineage between individual mice.

Table of Contents

Chapter I: Review of early embryogenesis

Abstract

Mouse embryogenesis is a classic model for mammalian development. Due to the rapid embryonic growth process, it provides a unique platform to assist our understanding of human embryo development. This rapidity, coupled with recent advances in mouse genetics and new manipulation techniques, enables a deeper examination of the detailed molecular events and mechanistic steps involved in this process. Such knowledge will ultimately lead us to a better understanding of human embryo development, and pave the way for future effective regenerative therapies.

The fertilized egg divides and develops as it travels along the oviduct to the uterus. In a mouse this process takes 4-5 days. By the time it reaches the uterus, the egg has undergone many cell divisions to form a blastocyst, composed of an inner cell mass (ICM) inside the trophectoderm. After implantation, dramatic transformation of the architecture of the mouse embryo takes place. Within three days following implantation, the polar trophectoderm and the ICM develop into an elongated structure that is made up of the ectoplacental cone, the extraembryonic ectoderm, the epiblast and a layer of visceral endoderm. Gastrulation commences with the formation of the primitive streak, through which epiblast cells ingress to form the mesoderm and the endoderm. These two tissues and the ectoderm, the descendants of epiblast cells which do not pass through the primitive streak, constitute the three primary germ layers that contain

progenitors of all tissues.

The review in this chapter focuses on two complementary areas in this early stage of embryo development: 1) the changes and regulatory mechanisms that occur prior to and post of implantation stage, including transcriptional, epigenetic and metabolic characteristics which may be responsible for the distinct features between those two developmental stages; and 2) the developmental patterns following the gastrulation stage as revealed by cell lineage mapping techniques, including the unsolved challenges that ask for a new generation of fate mapping approaches.

Molecular regulation governing the transition of mESC-mEpiSC

Blastocyst formation begins at day 3 after fertilization in mouse. The blastocyst possesses an ICM, or embryoblast, which subsequently forms the embryo, and an outer layer of cells surrounding the ICM and the blastocyst cavity, called trophoblast, which later forms the placenta. To understand the regulation underlying this specific developmental stage, *in vitro* cell lines are frequently utilized as a model that allows flexible genetic manipulation and easy access. Mouse embryonic stem cells (mESC) are established from pre-implantation stages from pre-blastocyst (embryonic day 2.5 [E2.5]) to late blastocyst (E4.5). They were first derived from either dissociated morulae (Eistetter, 1989), intact blastocysts (Evans & Kaufman, 1981), or entire ICMs (Martin, 1981). mESC can also be obtained from eight cell embryo stage (Delhaise et al.., 1996) and preblastocyst embryos (Tesar, 2005). In contrast, mouse epiblast-

derived stem cells (EpiSC) are derived from the epiblast, a tissue of the post-implantation embryo that generates the embryo proper (Brons et al.., 2007; Tesar et al.., 2007) or recently isolated from pre-implantation embryos (Najm et al.., 2011). Those two mouse stem cell lines are distinct from each other in various aspects, including, but not limited to, the transcriptional signals controlling differentiation, the epigenetic state and metabolic profiles.

Transcriptional differences

Brons et al. and Tesar et al. first reported that independently derived EpiSC lines were similar to each other but distinct from mouse ES cells (Brons et al., 2007; Tesar et al., 2007). First of all, *Oct3/4* and *Nanog* expressed slightly (1-2.5 times) higher in mESC, and *Sox2* expressed at equivalent levels in mESC and EpiSCs. Secondly, genes associated with the ICM such as *Pecam1*, *Tbx3*, *Gbx2* and *Rex1 (Zfp42)* were expressed by mESC, but significantly decreased or not detected in EpiSCs. On the other hand, genes associated with the epiblast and early germ layers such as *Otx2*, *Eomes*, *Foxa2*, *brachyury (T)*, *Gata6*, *Sox17*, *Cer1, FGF5* and *Nodal* were expressed at higher levels in EpiSCs. Interestingly, genes associated with the germ line, including *Stella*, *Piwil2*, *Stra8* and *Dazl* were expressed by mESC, but were significantly decreased or not detected in EpiSCs. It is worth noting that this difference in the expression of germ cell specific genes is consistent with the difference between the ICM and the post-implantation epiblast. Primordial germ cell specification occurs in a precise temporal and spatial manner beginning with *Blimp1*

(Prdm1) expression in a small number of cells in the posterior, proximal epiblast at E6.25. The *Blimp1* expression is further up-regulated by *Stella* around E7.25 (Hayashi et al., 2007). When EpiSCs were treated with BMP4, a growth factor central to primordial germ cell differentiation (Lawson et al., 1999), *Stella* and *Blimp1* were both induced in a temporal manner consistent with their specification *in vivo*. These data show that germ cell differentiation can be induced in EpiSCs and illustrates the value of EpiSCs as the proximal source of cells with restricted fates.

Brons et al. compared the expression profiles of EpiSCs to mESC, as well as those of ICMs from mouse blastocysts to the late epiblast cells from post-implantation mouse embryos. The comparison showed that EpiSCs had fewer differences from dissected late epiblast than from ICM or mESC, suggesting that EpiSCs are transcriptionally similar to their embryonic tissue source. Interestingly, in such a comparison visualized in a multi-dimensional scaling plot, the expression profile of mESC is similar to that of ICM, and EpiSC is similar to epiblast. However, both mESC and EpiSC are not transcriptionally identical to ICM and epiblast, respectively. The differences between the cell lines in culture and the tissue they are derived from, suggest that both mESC and EpiSC each display unique phenotypic properties which are determined not only by the developmental stage of the embryo from which the cell line is derived, but also by the growth factor environment in which they were established *in vitro*.

Human ES cells (hESC) were first derived in 1998 from human embryos in blastocyst stage, which were cultured from embryos in cleavage stage produced by in vitro fertilization for clinical purposes (Thomson et al., 1998). hESC were long thought to be the homo sapiens equivalent to mouse ES cells, until mouse EpiSCs were successfully derived and found to be more similar to hESC in molecular level. Although EpiSC and hESC were established from embryos at different developmental stages in mouse and human, respectively, both of them depend on activin/Nodal signaling to maintain their pluripotent status (James et al..., 2005). Such a requirement sets EpiSC/hESC in sharp contrast to mESCs that depend on LIF for the maintenance of pluripotency (Daheron et al., 2004; Niwa et al., 1998). Comparisons between mESC, EpiSC and hESC illustrate that stem cell pluripotency is not a fixed ground state, but is strongly influenced by developmental and environmental context. In other words, pluripotency does not represent a single defined state; subtle stages of pluripotency display similarities and differences in measurable characteristics relating to gene expression and cellular phenotype.

Epigenetic profiles

Besides the transcriptional changes that have to take place when early blastocyst develops to post-implantation embryo, epigenetic responses are also essential for this transition. Epigenetic modifications not only contribute to the repression of inappropriate developmental programs in time and space, but also ensure heritability of existing or newly acquired phenotypic states. These extrinsic

and intrinsic regulators determine the developmental origin and subsequent propagation of pluripotent states *in vivo* and *in vitro* (Surani et al., 2007).

The epigenetic changes which take place between pre-implantation and post-implantation development are perhaps most relevant to the understanding of how ES cells are maintained in an undifferentiated state and how the early stages of their differentiation may be regulated (Surani et al., 2007). An important reasoning behind this argument comes from the fact that the development prior to the blastocyst stage is relatively less prone to perturbation, due to the partial regulation by maternally inherited transcriptional and epigenetic modifiers. In contrast, the transition to the post-implantation development is driven by epigenetic and genetic regulators that are transcribed by the embryo itself. Thus, it is speculated that pre-implantation development primarily involves both erasure and maintenance of epigenetic modifications, while post-implantation development primarily involves establishment of new epigenetic modifications.

Mouse ES cells are known to display “bivalent domains” across 56 large loci rich in highly conserved noncoding elements (Bernstein et al., 2006). Those bivalent domains consist of large regions of H3 lysine 27 methylation harboring smaller regions of H3 lysine 4 methylation, and are thought to silence developmental genes in ES cells while keeping them poised for activation. Despite of detailed analysis in mouse ES cells, a comprehensive comparison between mouse ES cells and EpiSC is

still missing, and further evidence are still required regarding whether the epigenetic changes occurring in the subtle pluripotent stages *in vitro* recapitulate the developmental events *in vivo* when pre-implantation embryo transits to post-implantation epiblast. It has only been known that *Stella* (Dppa3) was only associated with inductive H3K4me3 in mouse ES cells as compared to EpiSC and hESC, consistent with its restricted expression to mouse ES cells. In comparison, epiblast associated gene *Otx2* had the inductive mark in all three cell types but also the repressive H3K27me3 mark in mouse ES cells (Tesar et al., 2007).

Metabolic Features

Elaborate metabolic arrangement during the transition of pre- and post-implantation embryos

Despite significant advances using these approaches, the framework defining pluripotency in stem cells remains incompletely understood. This is in part due to the difficulty of correlating expression data to functional activity. Given that the function and integrity of a cell are affected by primary metabolism, a promising complementary approach is to directly explore the metabolic signatures that reflect the integrated function of multiple pathways operating within cells.

Unlike in cancer cells in which the metabolic mechanisms are well studied, the metabolic characteristics of pluripotent cells have not been intensively explored until recently, and it has become an emerging and important field toward the goal of better understanding the acquisition and maintenance of the pluripotency. While recent

studies mostly focus on the metabolic differences between pluripotent cells and their differentiated progenies, few of them examined the metabolic changes occurring in subtle developmental stages in the embryo. In this chapter, I will mainly review two fundamental metabolic processes in a pluripotent cell, glycolysis and mitochondrial oxidative phosphorylation (OXPHOS), and discuss the fine balance between those two processes that a cell utilizes to maintain its pluripotent state.

Mammalian cells have to rely on external substrates to generate metabolites to fulfill their function. Glucose is one of the major sources of carbohydrates utilized in most organisms, from bacteria to human. Glucose is transported into cells by glucose transporters where it is then phosphorylated by the hexose kinase. It is later split into three-carbon molecules that are either converted to glycerol for lipid synthesis or sequentially transformed to pyruvate through orderly reactions with specific kinases. Pyruvate could be further catalyzed into acetyl-CoA in the tricarboxylic acid (TCA) cycle. This is followed by the formation of citrate from acetyl-CoA and oxaloacetate, which begins a new round of TCA cycle that generates intermediate metabolites used for biosynthesis. The NADH and succinate, produced in the TCA cycle, are further oxidized, releasing electrons to pass through the electron transport chain to the ATP synthase for the generation of ATP. This process, called mitochondria oxidative phosphorylation (OXPHOS), takes place in the mitochondria and consumes oxygen.

The mammalian embryo develops in a hypoxic environment, especially during the

blastocyst formation and implantation. It was reported that oxygen tension in the monkey uterus is as low as 11 mm Hg (1.5% O_2); in hamsters and rabbits, intrauterine oxygen tension even further decreases during the peri-implantation period, to 37 mm Hg (5.3% O_2) and 24 mm Hg (3.5% O_2), respectively (Fischer & Bavister, 1993). In such a hypoxic environment, developing embryos rely on glycolysis using glucose as the predominant substrate for nutrients uptake at the blastocyst stage (Gott et al., 1990; Leese & Barton, 1984) and early post-implantation period (Ellington, 1987). Intriguingly, prior to the blastocyst stage, pyruvate uptake exceeds that of glucose in unfertilized and fertilized ova to the morula stage. It appears that, both in the mouse and human, there may be an initial blockade to glucose utilization in early embryos and that pyruvate is the major energy substrate (Leese & Barton, 1984; Leese et al., 1986; Wales et al., 1987). It is worth noting that human embryos consume pyruvate at a level 15-20 times higher than that of pre-implantation mouse embryos; in contrast, their glucose uptake at the blastocyst stage corresponds closely on the basis of their difference in volume (Hardy et al., 1989).

While the developing embryo sits in an environment with low oxygen tension, it still consumes oxygen as its mitochondria develop inside the cell. The number, morphology and function of mitochondria dramatically change at different developmental stages. In early embryo development, mitochondria in the 8-cell embryo reveal minimal matrix electron density. Elongating mitochondria with inner mitochondrial membranes arranged into transverse cristae appear, and the replication of mtDNA takes place in expanding blastocysts (Sathananthan & Trounson, 2000;

Shepard et al., 2000; Thundathil et al., 2005). In the development up to the blastocyst stage, the oxygen consumption by the mouse embryo is relatively constant; it increases most sharply at the blastocyst stage before declining at the early post-implantation stage (Houghton et al., 1996). The increase at the blastocyst stage is mainly contributed by OXPHOS-dependent oxygen consumption, while the OXPHOS-independent oxygen consumption stays constant (Trimarchi et al., 2000). OXPHOS has been shown to be required in mouse and rabbit embryos (Kane & Buckley, 1977; Thomson, 1967), from one-cell embryos to blastocyst stage, but interestingly not in rat (Brison & Leese, 1994).

Overall, in the early embryo development starting from fertilization to the formation of blastocyst and then post-implantation stage, the embryo holds an especially elaborate balance between glycolysis and OXPHOS to meet its energy and nutrient needs. OXPHOS takes a major role in the early development up to the morula stage, during which period the pyruvate is the dominant substrate. Followed is a particularly unique development stage, the formation of blastocyst, in which the first lineage commitment is made to either stay pluripotent as ICM, or differentiate and become trophoectoderm that ultimately develops to placenta. In this important stage, both activities of OXPHOS and glycolysis are high, and pyruvate starts losing its leading role to glucose as the main substrate consumed. When blastocyst further develops and implants into the uterus, it comes to the stage of post-implantation embryo, in which epiblast is specified and further develops into three germ layers as the fountain for all bodily tissues to form eventually. In this developmental window shortly before and

after the implantation, the OXPHOS is shut down in an unknown way in the developing embryo, retaining glycolysis as the major mechanism providing cells with energy and nutrients for further growth.

It is especially interesting to closely compare the metabolic states of the blastocyst before and after implantation, as they give rise to two different pluripotent cells, mouse ES cells and EpiSCs, respectively, and provide a platform to understand the difference and regulation behind the two subtle pluripotent stages. Studies are emerging to characterize the metabolic states of those two stages. However, the mechanisms behind the metabolic profiles, the link between metabolism and the transcriptional and epigenec regulations, and finally the impact of the metabolic state on the cellular phenotype are still largely unexplored. More specific questions include, but are not limited to, by what mechanism is it that pre-implantation blastocyst utilizes both OXPHOS and glycolysis, while post-implantation epiblast only replies on glycolysis? What are the transcriptional regulators involved/impacted in these two different metabolisms? What are the benefits of those metabolic differences to the developing embryo?

Hints from the comparison of pluripotency to differentiation

Although the studies that directly examine the metabolic mechanisms behind the two subtle pluripotent stages are sparse, in order to understand those unsolved questions, we might be able to get some hints on the possible pathways involved from the studies that examine the difference between pluripotent cells to their differentiated

progenies. The common thing between those two developmental processes is that both display a shifted balance between OXPHOS and glycolysis.

Embryonic stem cells are one kind of pluripotent cells with well studied metabolism as compared to differentiated cells. ES cells mostly rely on glycolysis for their metabolic needs, and they are able to increase the activities involved in glycolytic pathways and suppress OXPHOS at the same time to maintain their glycolytic metabolism. It is thought the increase of glycolysis could be through rearranged glycolytic phosphotransfer network (Chung et al., 2010; Kondoh et al., 2007). On the other hand, the OXPHOS through mitochondria is largely limited, due to the immature mitochondrial structure/mass (Cho et al., 2006), suppression of mitochondrial genes associated with mitochondrial replication (Facucho-Oliveira et al., 2007) and respiratory factors (Varum et al., 2009), as well as the repression of mitochondria-specific nuclear-encoded transcription factors (St John et al., 2005). It is well demonstrated that the specific metabolic profile of ES, high glycolytic but low OXPHOS activity, is beneficial for the maintenance of its pluripotency. Among others, the reduction of reactive oxygen species that are normally generated as byproduct of OXPHOS could protect ES cells from ROS induced stresses (Saretzki et al., 2008), damages (Lin et al., 2005) and differentiation (Crespo et al., 2010). In accordance with this observation, ES cells are also found to have elevated anti-oxidant activities (Cho et al., 2006; Saretzki et al., 2008; Varum et al., 2009). Overall, The metabolic features of ES cells are not merely consequences of upstream regulation, but impose functional impacts on the cellular states, as interruption of the glycolytic metabolism

of ES results in loss of pluripotency and/or impaired differentiation potentials (Mandal et al., 2011; Schieke et al., 2008; Yanes et al., 2010).

Induced pluripotent cells (iPS), which were discovered only recently, may present another pluripotent stage that is different from embryonic stem cells. Those pluripotent cells are induced from somatic differentiated cells using a few transcriptional factors, Oct4, Sox2, Nanog, Lin28 or c-Myc (Takahashi et al., 2007; Takahashi & Yamanaka, 2006; Wernig et al., 2007; Yu et al., 2007), or other stem cell specific microRNAs through either integrative or non-integrative means. Overall similarities have been reported comparing iPS cells to ES cells (Bock et al., 2011; Wernig et al., 2007), and besides the transcriptional and Epigenetic, the metabolic profile of iPS cells largely resembles that of ES cells (Armstrong et al., 2010; Prigione et al., 2010), including low mitochondrial mass, mtDNA copy number and the expression levels of genes associated with mitochondrial biogenesis. Interestingly, iPS cells, like ES cells, also have low mitochondrial ROS and a similar level of anti-oxidative defense. However, minor differences do exist, and it has been demonstrated that seemingly minor variations between iPS cells and ES cells may have significant phenotypic consequences (Lister et al., 2011; Stadtfeld et al., 2010). Metabolically as compared to ES cells, iPS cells display a mixed profile of glucose-related gene expression, mitochondrial morphology and oxygen consumption that is not quite that of ES cells (Varum et al., 2011). Furthermore, the levels of unsaturated fatty acid metabolites are lower in iPS cells compared to ES cells, while metabolites involved in

the S-adenosyl methionine cycle are all significantly elevated in iPS cells (Panopoulos et al., 2012).

Hidden links to cancer

Cancer cells are another kind of cells that share metabolic similarity with pluripotent cells, and bear great research interests. Those cells are well demonstrated to sustain enhanced aerobic glycolysis, linking to activation of oncogenes or loss of tumor suppressors (Cairns et al., 2011; Koppenol et al., 2011; Levine & Puzio-Kuter, 2010; Vander Heiden et al., 2009; Warburg, 1956). In addition to the glycolytic metabolism, some cancer cells are able to self-renew, like pluripotent cells and other adult stem cells (Reya et al., 2001). Furthermore, many pathways that regulate normal stem cell development are also classically associated with cancer (Taipale & Beachy, 2001).

Among them, there are a few individuals that are worth special attention. C-Myc, first discovered in Burkitt's lymphoma patients, is a transcriptional factor well demonstrated as a proto-oncogene. c-Myc is found essential for the activation of genes involved in glycolysis for cell growth and proliferation through direct targeting on them, which include enolase A, hexokinase II, lactate dehydrogenase A (Shim et al., 1997; Tavtigian et al., 1994), phosphofructokinase and glucose transporter (Osthus et al., 2000). Furthermore, c-Myc also directly stimulates genes involved in ribosome biogenesis (van Riggelen et al., 2010), nucleotide metabolism and DNA replication (Zeller et al., 2006). Importantly, c-Myc is one of the factors used to induce both

mouse and human pluripotent cells from somatic cells. Due to conflicting roles it plays in mouse and human and the successful generation of iPS cells without it, c-Myc is thought to act as a booster of reprogramming rather than a controller of maintenance of pluripotency (Takahashi et al., 2007). Indeed, later it was shown that c-Myc is important in establishing pluripotency and repressing fibroblast-specific gene expression in murine reprogramming cells (Sridharan et al., 2009). Given the role c-Myc has in cancer, glycolytic metabolism, and early pluripotency acquisition, it is possible and remains to be demonstrated that the c-Myc-induced metabolic switch also takes place in the early re-programming from somatic to pluripotent cells, and that this metabolic switch could be required for initial establishment of the re-programming potential.

Hypoxia inducible factor 1 alpha (HIF1α) is another interesting molecule that plays important roles connecting tumorigenesis, glycolytic metabolism and the development of embryonic/adult stem cells. HIF1α is stabilized under hypoxia, directly targets a transcriptional program that induces glycolysis and inhibits mitochondrial biogenesis of cancer cells, and has a crucial role in cancer progression and metastasis (Bertout et al., 2008; Semenza, 2012). Since much of mammalian embryogenesis occurs at a low oxygen tension as discussed above, HIF1a stabilized by low oxygen tension functions as a critical transcriptional factor in many developmental systems, including neural tube specification, cardiovascular differentiation and bone morphogenesis. Recently, HIF1a is shown to be crucial for the ESC-EpiSC transition and the maintenance of

embryonic and neural stem cells by interacting Activin/Nodal and Wnt signaling, respectively (Mazumdar et al., 2010; Zhou et al., 2012). With its interacting function in those diverse biological processes, it remains to be explored whether HIF induced metabolic changes play an important role for the maintenance or establishment/acquisition of pluripotency, as it does in the biology of cancer.

Developmental pattern of gastrulation

Summary of developmental steps involved in gastrulation

Following implantation, post-implantation embryos continue developing, getting ready for the next important event, gastrulation, the defining feature of metazoan development. It serves to specify seemingly equivalent pluripotent cells to three different fates, which, then, ultimately establish all of the tissues of the developing organism. After implantation in the uterus, the mouse embryo first develops uniquely as a cylindrical structure, in which the ICM epithelialize into a layer of epiblast cells and the formation of a central proamniotic cavity takes place. At this stage prior to the onset of gastrulation, the mouse conceptus consists of three tissue types: the extraembryonic ectoderm, extraembryonic endoderm and the embryonic ectoderm, epiblast. Like the chick, mouse embryo gastrulates by the recruitment of epiblast cells to a transient embryonic structure called the primitive streak. The formation of the primitive streak establishes the anterior-posterior (A-P) axis with the streak located on the posterior side of the embryo. With the definition of the A-P axis, the left-right axis becomes apparent. At the primitive streak, the epiblast cells undergo an epithelial to

mesenchymal transition, then ingress in-between the epiblast and endoderm to become incorporated into either the mesoderm or the definitive endoderm germ layers. The mesoderm is formed as a new tissue sheet expanding from both sides of the primitive streak (Bellairs, 1986; Hashimoto et al., 1987; Tam et al., 1993). At the conclusion of gastrulation, cell populations are allocated to different germ layers, and those that are required for the formation of specific organs or body parts are physically brought together. During gastrulation, the epiblast will also give rise to the primordial germ cells.

The gastrulation of the mouse embryo has another unique feature. The embryo initially develops ‘inside-out’, with the internally-located ectoderm surrounded by the mesoderm and endoderm. After gastrulation, this necessitates an inversion of the germ layers to bring the ectoderm to the outside of the embryo and the endoderm to the inside with the mesoderm in the middle (Kaufman, 1990). In contrast, most mammalian embryos, including humans, develop as planar structures with the epiblast on top and the endoderm below (O'Rahilly & Muller, 2010; Tamarin, 1983; Tarara et al., 1987). Despite this significant morphological difference between the mouse and most mammals, the finer details of the molecular and cellular aspects of germ layer formation are likely to be similar because the embryos of all of these species utilize the primitive streak as the conduit for germ layer formation during gastrulation.

Clonal fate mapping studies subsequently revealed that single, labeled cells within the

post-implantation epiblast could contribute to all three germ layers (Lawson et al., 1991). Despite this fact, the experiments also revealed a regional predisposition within the post-implantation 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 which are transplanted to a different region of a host epiblast adopt the predicted developmental fate of the area to which they are relocated (Tam & Zhou, 1996). These data show that post-implantation epiblast cells are pluripotent and subject to regional influences to allocate them to differentiated lineages.

Classic approaches to trace cell lineage

In order to understand how cells determine their developmental destiny, it is necessary to define the developmental route, by which particular cells give rise to all of their descendants. The simplest approach, first employed with tunicates more than a century ago, is to directly observe the cell divisions of an embryo as it grows (Clarke & Tickle, 1999; Stern & Fraser, 2001; Sulston et al., 1983). Determining embryonic lineages in higher organisms through direct observation is impossible, because the embryos, and the environment in which they develop, the uterus in mammals, are non-transparent. It is further complicated by the lengthy gestation, and the difficulty of accounting for the trillion or more cells that may be present by the time of birth. Consequently, attempts at mapping cell fate have relied on marking a single cell at some point during embryogenesis and then later identifying descendants of that cell

that also contain the marker.

The simplest way to tag a cell is to inject a dye. A variety of dyes, including fluorescence, HRP, carbocyanine (DiI), were utilized to study the developmental pattern of the epiblast (Lawson et al., 1991), primitive streak (Smith et al., 1994), notochordal plate (Sulik et al., 1994) and other structures of mouse embryo at the gastrulating stage (Tam et al., 1993). An alternative approach for marking a cell is to introduce a permanent genetic change heritable by daughter cells, establishing mosaic patterning that cannot be diluted through successive divisions (Nesbitt & Gartler, 1971; Stern & Fraser, 2001).

Development of gene transfer methods has facilitated cell fate analysis in mice. In some cases, cells are genetically marked through retrovirus-mediated transfer of a transgene encoding β-galactosidase or another colored marker (Lemischka, 1993; Soriano & Jaenisch, 1986). One approach has taken advantage of construction of a transgenic mouse containing a β-galactosidase-encoding lacZ marker gene functionally inactivated through partial tandem duplication (Eloy-Trinquet & Nicolas, 2002). Infrequent–spontaneous intra-allelic mitotic recombination regenerates the marker and, thereby, allows for tagging of clonal descendants. Another transgenic approach has utilized a mouse ("stop-EGFP") ubiquitously expressing a gene for enhanced green fluorescent protein (EGFP) inactivated with a stop codon whose reversion following mutagenesis similarly tags a clonally descended population of cells (Ro & Rannala, 2004). A similar inducible gene targeting technique allows the

tracking of daughter cells committed to certain specified lineages, which might not all be in the same place at a particular time (Kuhn et al., 1995; Sauer, 1998).

Collectively, the cell marking techniques described above have been informatively employed to analyze cell lineage during the embryo development in the mouse, as well as in many other developmental systems in other organisms. Although widely utilized, technical issues of these different techniques may require complicated experimental procedures and, in some cases, lead to inconclusive interpretation (Salipante & Horwitz, 2007). However, there is no doubt regarding their importance.

It should be emphasized that cell marking studies do not yield the same amount of information as direct observation. Instead, those studies reveal surprisingly scant detail about cell lineage in comparison to fate maps produced from direct observation. One important piece of knowledge lacked in the cell marking studies is the hierarchy of relationships between cells. For example, suppose that an embryonic cell is tagged with a marker and that, later, four cells are now found to be tagged with that marker. While it is clear that all four cells are descended from the original cell, there are actually a minimum of 15 different patterns through which the progenitor cell could have divided to yield those four descendants (Salipante & Horwitz, 2007).

Development as a Stochastic Process

Compared to the invariant lineage in worms (Sulston et al., 1983), lineage

specification in mammals and other organisms with more cells is not completely deterministic, and development is partly stochastic (Sternberg & Felix, 1997). The random component to development becomes apparent at two levels of organization (Salipante & Horwitz, 2007). First, the cells of the inner cell mass of early embryos are pluripotent, having the potential to differentiate all tissues of the body. Second, the differentiation process is also stochastic. Some have gone so far as to argue (Arias & Hayward, 2006; Moller & Pagel, 1998; Rao et al., 2002) that the ultimately stochastic nature of cell fate is an inescapable consequence of the transcription of genes, their epigenetic regulation, and the networks into which they assemble, because the molecular interactions that control them unavoidably reduce to quantum mechanical noise (Salipante & Horwitz, 2007). This is not to say that the randomness itself cannot be constrained through genetic controls. Development in higher animals likely represents a compromise between what can be efficiently scripted in the genome and the effort required to control the entropic noise intrinsic to its underlying biochemistry.

It may be possible to combine information across individuals in order to produce a fate map representative of a species as a whole. A “consensus” fate map could be constructed by combining information from separate organisms into a single phylogenetic reconstruction. A fate map derived in this way would reflect the probability that a particular embryonic lineage will pursue a particular cell fate, a view of development that is more compatible with the actual events of mammalian embryogenesis than is a single fate map derived from examination of just one individual (Lawson et al.., 1991). By comparing and contrasting the fate maps of

individual organisms, it should additionally be possible to identify precisely what cell lineage pathways are variable and which are conserved throughout development.

Recently, a new and systematic approach, phylogenetic fate mapping, has been proposed, tested and utilized in various cases to retrospectively generate a single cell resolution fate map of a mouse or human (Carlson et al., 2012; Salipante & Horwitz, 2006; Salipante et al., 2010; Salipante et al., 2008; Salk et al., 2009). The design of this approach is based on the fact that nearly every time a cell divides, mutations arise in its genome. The inheritance of mutations by daughter cells offers a record of cell divisions, in which the order that the mutations have occurred during development reflects a cell's ancestral embryonic relationships. Cells that exhibit the most closely related patterns of somatic mutation share the most recent common ancestry, analogous to the manner in which species with closely related genomes are separated by less evolutionary time than those containing more divergent sequences. In this approach, the cells of an adult mouse are treated as members of an asexual population, individualized by somatic mutation, yet descended from a common founder (the zygote). With such an approach, mouse gastrulation is thought to experience an early period of substantial cell mixing followed by more coherent growth of clones later (Salipante et al., 2010).

It is worth noting that the phylogenetic fate mapping technique not only reveals the hierarchical lineage ancestries that other classic cell marking studies cannot provide, but also generates a probabilistic landscape. Such a probabilistic view is, in fact,

compatible with the stochastic processes that underlie the biological basis of development itself.

References

Arias AM, Hayward P (2006) Filtering transcriptional noise during development: concepts and mechanisms. *Nat Rev Genet* **7:** 34-44

Armstrong L, Tilgner K, Saretzki G, Atkinson SP, Stojkovic M, Moreno R, Przyborski S, Lako M (2010) Human induced pluripotent stem cell lines show stress defense mechanisms and mitochondrial regulation similar to those of human embryonic stem cells. *Stem Cells* **28:** 661-673

Bellairs R (1986) The primitive streak. *Anat Embryol (Berl)* **174:** 1-14

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315-326

Bertout JA, Patel SA, Simon MC (2008) The impact of O2 availability on human cancer. *Nat Rev Cancer* **8:** 967-975

Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH, Gnirke A, Eggan K, Meissner A (2011) Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* **144:** 439-452

Brison DR, Leese HJ (1994) Blastocoel cavity formation by preimplantation rat embryos in the presence of cyanide and other inhibitors of oxidative phosphorylation. *J Reprod Fertil* **101:** 305-309

Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA, Vallier L (2007) Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448:** 191-195

Cairns RA, Harris IS, Mak TW (2011) Regulation of cancer cell metabolism. *Nat Rev Cancer* **11:** 85-95

Carlson CA, Kas A, Kirkwood R, Hays LE, Preston BD, Salipante SJ, Horwitz MS (2012) Decoding cell lineage from acquired mutations using arbitrary deep sequencing. *Nat Methods* **9:** 78-80

Cho YM, Kwon S, Pak YK, Seol HW, Choi YM, Park do J, Park KS, Lee HK (2006) Dynamic changes in mitochondrial biogenesis and antioxidant enzymes during the spontaneous differentiation of human embryonic stem cells. *Biochem Biophys Res Commun* **348:** 1472-1478

Chung S, Arrell DK, Faustino RS, Terzic A, Dzeja PP (2010) Glycolytic network restructuring integral to the energetics of embryonic stem cell cardiac differentiation. *J Mol Cell Cardiol* **48:** 725-734

Clarke JD, Tickle C (1999) Fate maps old and new. *Nat Cell Biol* **1:** E103-109

Crespo FL, Sobrado VR, Gomez L, Cervera AM, McCreath KJ (2010) Mitochondrial reactive oxygen species mediate cardiomyocyte formation from embryonic stem cells in high glucose. *Stem Cells* **28:** 1132-1142

Daheron L, Opitz SL, Zaehres H, Lensch MW, Andrews PW, Itskovitz-Eldor J, Daley GQ (2004) LIF/STAT3 signaling fails to maintain self-renewal of human embryonic stem cells. *Stem Cells* **22:** 770-778

Delhaise F, Bralion V, Schuurbiers N, Dessy F (1996) Establishment of an embryonic stem cell line from 8-cell stage mouse embryos. *Eur J Morphol* **34:** 237-243

Eistetter HR (1989) Pluripotent embryonal stem cells lines can be established from disaggregated mouse morulae. *Dev Growth Differ* **31:** 275-282

Ellington SK (1987) In vitro analysis of glucose metabolism and embryonic growth in postimplantation rat embryos. *Development* **100:** 431-439

Eloy-Trinquet S, Nicolas JF (2002) Clonal separation and regionalisation during formation of the medial and lateral myotomes in the mouse embryo. *Development* **129:** 111-122

Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292:** 154-156

Facucho-Oliveira JM, Alderson J, Spikings EC, Egginton S, St John JC (2007) Mitochondrial DNA replication during differentiation of murine embryonic stem cells. *J Cell Sci* **120:** 4025-4034

Fischer B, Bavister BD (1993) Oxygen tension in the oviduct and uterus of rhesus monkeys, hamsters and rabbits. *J Reprod Fertil* **99:** 673-679

Gott AL, Hardy K, Winston RM, Leese HJ (1990) Non-invasive measurement of pyruvate and glucose uptake and lactate production by single human preimplantation embryos. *Hum Reprod* **5:** 104-108

Hardy K, Hooper MA, Handyside AH, Rutherford AJ, Winston RM, Leese HJ (1989) Non-invasive measurement of glucose and pyruvate uptake by individual human oocytes and preimplantation embryos. *Hum Reprod* **4:** 188-191

Hashimoto K, Fujimoto H, Nakatsuji N (1987) An ECM substratum allows mouse mesodermal cells isolated from the primitive streak to exhibit motility similar to that inside the embryo and reveals a deficiency in the T/T mutant cells. *Development* **100:** 587-598

Hayashi K, de Sousa Lopes SM, Surani MA (2007) Germ cell specification in mice. *Science* **316:** 394-396

Houghton FD, Thompson JG, Kennedy CJ, Leese HJ (1996) Oxygen consumption and energy metabolism of the early mouse embryo. *Mol Reprod Dev* **44:** 476-485

James D, Levine AJ, Besser D, Hemmati-Brivanlou A (2005) TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132:** 1273-1282

Kane MT, Buckley NJ (1977) The effects of inhibitors of energy metabolism on the growth of one-cell rabbit ova to blastocysts in vitro. *J Reprod Fertil* **49:** 261-266

Kaufman MH (1990) Morphological stages of postimplantation embryonic development. In *Postimplantation Mammalian Embryos: A Practical Approach.*, Copp AJ, Cockroft, D.L. (ed), pp 81-91. Oxford: IRL Press

Kondoh H, Lleonart ME, Nakashima Y, Yokode M, Tanaka M, Bernard D, Gil J, Beach D (2007) A high glycolytic flux supports the proliferative potential of murine embryonic stem cells. *Antioxid Redox Signal* **9:** 293-299

Koppenol WH, Bounds PL, Dang CV (2011) Otto Warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer* **11:** 325-337

Kuhn R, Schwenk F, Aguet M, Rajewsky K (1995) Inducible gene targeting in mice. *Science* **269:** 1427-1429

Lawson KA, Dunn NR, Roelen BA, Zeinstra LM, Davis AM, Wright CV, Korving JP,

Hogan BL (1999) Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* **13:** 424-436

Lawson KA, Meneses JJ, Pedersen RA (1991) Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113:** 891-911

Leese HJ, Barton AM (1984) Pyruvate and glucose uptake by mouse ova and preimplantation embryos. *J Reprod Fertil* **72:** 9-13

Leese HJ, Hooper MA, Edwards RG, Ashwood-Smith MJ (1986) Uptake of pyruvate by early human embryos determined by a non-invasive technique. *Hum Reprod* **1:** 181-182

Lemischka IR (1993) Retroviral lineage studies: some principals and applications. *Curr Opin Genet Dev* **3:** 115-118

Levine AJ, Puzio-Kuter AM (2010) The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes. *Science* **330:** 1340-1344

Lin T, Chao C, Saito S, Mazur SJ, Murphy ME, Appella E, Xu Y (2005) p53 induces differentiation of mouse embryonic stem cells by suppressing Nanog expression. *Nat Cell Biol* **7:** 165-171

Lister R, Pelizzola M, Kida YS, Hawkins RD, Nery JR, Hon G, Antosiewicz-Bourget J, O'Malley R, Castanon R, Klugman S, Downes M, Yu R, Stewart R, Ren B, Thomson JA, Evans RM, Ecker JR (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471:** 68-73

Mandal S, Lindgren AG, Srivastava AS, Clark AT, Banerjee U (2011) Mitochondrial function controls proliferation and early differentiation potential of embryonic stem cells. *Stem Cells* **29:** 486-495

Martin GR (1981) 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

Mazumdar J, O'Brien WT, Johnson RS, LaManna JC, Chavez JC, Klein PS, Simon MC (2010) O2 regulates stem cells through Wnt/beta-catenin signalling. *Nat Cell Biol* **12:** 1007-1013

Moller AP, Pagel M (1998) Developmental stability and signalling among cells. *J Theor Biol* **193:** 497-506

Najm FJ, Chenoweth JG, Anderson PD, Nadeau JH, Redline RW, McKay RD, Tesar

PJ (2011) Isolation of epiblast stem cells from preimplantation mouse embryos. *Cell Stem Cell* **8:** 318-325

Nesbitt MN, Gartler SM (1971) The applications of genetic mosaicism to developmental problems. *Annu Rev Genet* **5:** 143-162

Niwa H, Burdon T, Chambers I, Smith A (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev* **12:** 2048-2060

O'Rahilly R, Muller F (2010) Developmental stages in human embryos: revised and new measurements. *Cells Tissues Organs* **192:** 73-84

Osthus RC, Shim H, Kim S, Li Q, Reddy R, Mukherjee M, Xu Y, Wonsey D, Lee LA, Dang CV (2000) Deregulation of glucose transporter 1 and glycolytic gene expression by c-Myc. *J Biol Chem* **275:** 21797-21800

Panopoulos AD, Yanes O, Ruiz S, Kida YS, Diep D, Tautenhahn R, Herrerias A, Batchelder EM, Plongthongkum N, Lutz M, Berggren WT, Zhang K, Evans RM, Siuzdak G, Izpisua Belmonte JC (2012) The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell Res* **22:** 168-177

Prigione A, Fauler B, Lurz R, Lehrach H, Adjaye J (2010) The senescence-related mitochondrial/oxidative stress pathway is repressed in human induced pluripotent stem cells. *Stem Cells* **28:** 721-733

Rao CV, Wolf DM, Arkin AP (2002) Control, exploitation and tolerance of intracellular noise. *Nature* **420:** 231-237

Reya T, Morrison SJ, Clarke MF, Weissman IL (2001) Stem cells, cancer, and cancer stem cells. *Nature* **414:** 105-111

Ro S, Rannala B (2004) A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. *EMBO Rep* **5:** 914-920

Salipante SJ, Horwitz MS (2006) Phylogenetic fate mapping. *Proc Natl Acad Sci U S A* **103:** 5448-5453

Salipante SJ, Horwitz MS (2007) A phylogenetic approach to mapping cell fate. *Curr Top Dev Biol* **79:** 157-184

Salipante SJ, Kas A, McMonagle E, Horwitz MS (2010) Phylogenetic analysis of developmental and postnatal mouse cell lineages. *Evol Dev* **12:** 84-94

Salipante SJ, Thompson JM, Horwitz MS (2008) Phylogenetic fate mapping: theoretical and experimental studies applied to the development of mouse fibroblasts. *Genetics* **178:** 967-977

Salk JJ, Salipante SJ, Risques RA, Crispin DA, Li L, Bronner MP, Brentnall TA, Rabinovitch PS, Horwitz MS, Loeb LA (2009) Clonal expansions in ulcerative colitis identify patients with neoplasia. *Proc Natl Acad Sci U S A* **106:** 20871-20876

Saretzki G, Walter T, Atkinson S, Passos JF, Bareth B, Keith WN, Stewart R, Hoare S, Stojkovic M, Armstrong L, von Zglinicki T, Lako M (2008) Downregulation of multiple stress defense mechanisms during differentiation of human embryonic stem cells. *Stem Cells* **26:** 455-464

Sathananthan AH, Trounson AO (2000) Mitochondrial morphology during preimplantational human embryogenesis. *Hum Reprod* **15 Suppl 2:** 148-159

Sauer B (1998) Inducible gene targeting in mice using the Cre/lox system. *Methods* **14:** 381-392

Schieke SM, Ma M, Cao L, McCoy JP, Jr., Liu C, Hensel NF, Barrett AJ, Boehm M, Finkel T (2008) Mitochondrial metabolism modulates differentiation and teratoma formation capacity in mouse embryonic stem cells. *J Biol Chem* **283:** 28506-28512

Semenza GL (2012) Hypoxia-inducible factors in physiology and medicine. *Cell* **148:** 399-408

Shepard TH, Muffley LA, Smith LT (2000) Mitochondrial ultrastructure in embryos after implantation. *Hum Reprod* **15 Suppl 2:** 218-228

Shim H, Dolde C, Lewis BC, Wu CS, Dang G, Jungmann RA, Dalla-Favera R, Dang CV (1997) c-Myc transactivation of LDH-A: implications for tumor metabolism and growth. *Proc Natl Acad Sci U S A* **94:** 6658-6663

Smith JL, Gesteland KM, Schoenwolf GC (1994) Prospective fate map of the mouse primitive streak at 7.5 days of gestation. *Dev Dyn* **201:** 279-289

Soriano P, Jaenisch R (1986) Retroviruses as probes for mammalian development: allocation of cells to the somatic and germ cell lineages. *Cell* **46:** 19-29

Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q, Plath K (2009) Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136:** 364-377

St John JC, Ramalho-Santos J, Gray HL, Petrosko P, Rawe VY, Navara CS, Simerly

CR, Schatten GP (2005) The expression of mitochondrial DNA transcription factors during early cardiomyocyte in vitro differentiation from human embryonic stem cells. *Cloning Stem Cells* **7:** 141-153

Stadtfeld M, Apostolou E, Akutsu H, Fukuda A, Follett P, Natesan S, Kono T, Shioda T, Hochedlinger K (2010) Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465:** 175-181

Stern CD, Fraser SE (2001) Tracing the lineage of tracing cell lineages. *Nat Cell Biol* **3:** E216-218

Sternberg PW, Felix MA (1997) Evolution of cell lineage. *Curr Opin Genet Dev* **7:** 543-550

Sulik K, Dehart DB, Iangaki T, Carson JL, Vrablic T, Gesteland K, Schoenwolf GC (1994) Morphogenesis of the murine node and notochordal plate. *Dev Dyn* **201:** 260-278

Sulston JE, Schierenberg E, White JG, Thomson JN (1983) The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol* **100:** 64-119

Surani MA, Hayashi K, Hajkova P (2007) Genetic and epigenetic regulators of pluripotency. *Cell* **128:** 747-762

Taipale J, Beachy PA (2001) The Hedgehog and Wnt signalling pathways in cancer. *Nature* **411:** 349-354

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131:** 861-872

Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:** 663-676

Tam PP, Williams EA, Chan WY (1993) Gastrulation in the mouse embryo: ultrastructural and molecular aspects of germ layer morphogenesis. *Microsc Res Tech* **26:** 301-328

Tam PP, Zhou SX (1996) 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

Tamarin A (1983) Stage 9 macaque embryos studied by scanning electron microscopy. *J Anat* **137 (Pt 4):** 765-779

Tarara R, Enders AC, Hendrickx AG, Gulamhusein N, Hodges JK, Hearn JP, Eley RB, Else JG (1987) Early implantation and embryonic development of the baboon: stages 5, 6 and 7. *Anat Embryol (Berl)* **176:** 267-275

Tavtigian SV, Zabludoff SD, Wold BJ (1994) Cloning of mid-G1 serum response genes and identification of a subset regulated by conditional myc expression. *Mol Biol Cell* **5:** 375-388

Tesar PJ (2005) Derivation of germ-line-competent embryonic stem cell lines from preblastocyst mouse embryos. *Proc Natl Acad Sci U S A* **102:** 8239-8244

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448:** 196-199

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM (1998) Embryonic stem cell lines derived from human blastocysts. *Science* **282:** 1145-1147

Thomson JL (1967) Effect of inhibitors of carbohydrate metabolism on the development of preimplantation mouse embryos. *Exp Cell Res* **46:** 252-262

Thundathil J, Filion F, Smith LC (2005) Molecular control of mitochondrial function in preimplantation mouse embryos. *Mol Reprod Dev* **71:** 405-413

Trimarchi JR, Liu L, Porterfield DM, Smith PJ, Keefe DL (2000) Oxidative phosphorylation-dependent and -independent oxygen consumption by individual preimplantation mouse embryos. *Biol Reprod* **62:** 1866-1874

van Riggelen J, Yetil A, Felsher DW (2010) MYC as a regulator of ribosome biogenesis and protein synthesis. *Nat Rev Cancer* **10:** 301-309

Vander Heiden MG, Cantley LC, Thompson CB (2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324:** 1029-1033

Varum S, Momcilovic O, Castro C, Ben-Yehudah A, Ramalho-Santos J, Navara CS (2009) Enhancement of human embryonic stem cell pluripotency through inhibition of the mitochondrial respiratory chain. *Stem Cell Res* **3:** 142-156

Varum S, Rodrigues AS, Moura MB, Momcilovic O, Easley CAt, Ramalho-Santos J, Van Houten B, Schatten G (2011) Energy metabolism in human pluripotent stem cells and their differentiated counterparts. *PLoS One* **6:** e20914

Wales RG, Whittingham DG, Hardy K, Craft IL (1987) Metabolism of glucose by human embryos. *J Reprod Fertil* **79:** 289-297

Warburg O (1956) On the origin of cancer cells. *Science* **123:** 309-314

Wernig M, Meissner A, Foreman R, Brambrink T, Ku M, Hochedlinger K, Bernstein BE, Jaenisch R (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448:** 318-324

Yanes O, Clark J, Wong DM, Patti GJ, Sanchez-Ruiz A, Benton HP, Trauger SA, Desponts C, Ding S, Siuzdak G (2010) Metabolic oxidation regulates embryonic stem cell differentiation. *Nat Chem Biol* **6:** 411-417

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin, II, Thomson JA (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318:** 1917-1920

Zeller KI, Zhao X, Lee CW, Chiu KP, Yao F, Yustein JT, Ooi HS, Orlov YL, Shahab A, Yong HC, Fu Y, Weng Z, Kuznetsov VA, Sung WK, Ruan Y, Dang CV, Wei CL (2006) Global mapping of c-Myc binding sites and target gene networks in human B cells. *Proc Natl Acad Sci U S A* **103:** 17834-17839

Zhou W, Choi M, Margineantu D, Margaretha L, Hesson J, Cavanaugh C, Blau CA, Horwitz MS, Hockenbery D, Ware C, Ruohola-Baker H (2012) HIF1alpha induced switch from bivalent to exclusively glycolytic metabolism during ESC-to-EpiSC/hESC transition. *EMBO J* **31:** 2103-2116

Chapter II: HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC-EpiSC transition

Abstract

The function of metabolic state in stemness is poorly understood. Mouse embryonic stem cells (ESC) and epiblast stem cells (EpiSC) are at distinct pluripotent states representing the inner cell mass (ICM) and epiblast embryos. Human embryonic stem cells (hESC) are similar to EpiSC stage. However, we now show a dramatic metabolic difference between these two ~~stages~~; EpiSC/hESC are highly glycolytic, while ESC are bivalent in their energy production, dynamically switching from glycolysis to mitochondrial respiration on demand. Despite having a more developed and expanding mitochondrial content, EpiSC/hESC have low mitochondrial respiratory capacity due to low cytochrome c oxidase (COX) expression. Similarly, *in vivo* epiblasts suppress COX levels. These data reveal that EpiSC/hESC have functional similarity to the glycolytic phenotype in cancer (Warburg effect). We further show that HIF1α is sufficient to drive ESC to a glycolytic Activin/Nodal-dependent EpiSC like stage. This metabolic switch during early stem cell development may be deterministic.

Introduction

Pluripotent embryonic stem cells are able to self-renew and differentiate into the three germ lineages. Unraveling the developmental mechanisms through which

pluripotency is maintained holds tremendous promise for understanding early animal development, as well as developing regenerative medicine and cell therapies. Mouse and human ES cells are isolated from the inner cell mass (ICM) of pre-implantation embryos (Brook & Gardner, 1997; Evans & Kaufman, 1981; Thomson et al., 1998), while EpiSC represent cells from the post-implantation epiblast, a later stage in development (Tesar et al., 2007). ESC and EpiSC are pluripotent, yet display distinct features in terms of gene expression, epigenetic modifications and developmental capacity following blastocyst injection. Though isolated from the inner cell mass, hESC are similar to EpiSC based on transcriptional and protein expression profiles and their epigenetic state. Thus, pluripotency does not represent a single defined state; subtle stages of pluripotency, with similarities and differences in measurable characteristics relating to gene expression and cellular phenotype, provide an experimental system for studying potential key regulators that constrain or expand the developmental capacity of embryonic stem cells.

ESC, often termed naïve pluripotent cells (Nichols & Smith, 2009), efficiently contribute to chimeric embryos, maintain both X chromosomes in an active state (XaXa) in female cells, and are relatively refractory in their potential to differentiate into primordial germ cells (PGCs) *in vitro*. EpiSC and hESC, or primed pluripotent cells, can give rise to differentiated teratomas, but EpiSC are highly inefficient in repopulating the ICM upon aggregation or injection into host blastocysts. These cells have variable and at times abnormal X-chromosome inactivation status (XiXa), and

are poised for differentiation into PGC precursors *in vitro* (Brons et al., 2007; Hayashi & Surani, 2009; Tesar et al., 2007). Naïve ESC can be cloned with high efficiency as packed dome colonies, and are stabilized by LIF/Stat3 (Smith et al., 1988). In contrast, EpiSC and hESC are characterized by flat colony morphology, relative intolerance to passaging as single cells, and a dependence on bFGF and TGFβ/Activin signaling rather than LIF/Stat3 (Bendall et al., 2007; Greber et al., 2010; James et al., 2005).

In order to understand how these pluripotent cells maintain their distinct abilities to self-renew and differentiate, global gene-expression, epigenetic modification, and protein expression profiling have been employed to identify key regulators. Despite significant advances using these approaches, the framework defining pluripotency in stem cells remains incompletely understood. This is, in part, due to the difficulty of correlating expression data to functional activity. Given that the function and integrity of a cell are affected by primary metabolism, a promising complementary approach is to directly explore the metabolic signatures that reflect the integrated function of multiple pathways operating within cells.

In the current study, we evaluated the bioenergetic profiles of ESC, hESC and EpiSC with respect to mitochondrial DNA copy number, cellular ATP levels, oxygen consumption rate and extracellular acidification rate. We show that while ESC are metabolically bivalent, EpiSC and hESC are almost exclusively glycolytic. We further show that hypoxia inducible factor 1α (HIF1α) is an important regulator in the

metabolic and functional transition from ESC to EpiSC. These results demonstrate a significant relationship between metabolic phenotype and pluripotent developmental stage that correlates with the underlying stem cell functional biology.

Results

EpiSC and hESC are metabolically distinct from ESC

To characterize the metabolic profiles of ESC, EpiSC and hESC, we initially measured two metabolic parameters: oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) under various conditions and treatments using three different experimental systems (Seahorse Extracellular Flux analyzer, Fig 1 and Perifusion Flow System, and Perifusion Microscopic System in Suppl. Fig 1). OCR mainly measures the level of mitochondrial respiration. ECAR correlates with glycolytic activity, since the major exported acid, lactic acid, is derived from pyruvate generated through glycolysis, recycling NADH to NAD^+ for utilization in glycolysis. We used two representative cell lines for each pluripotency stage (ESC: R1 and G4; EpiSC: EpiSC#5 and EpiSC#7; and hESC: H1 and H7), and measured the baseline OCR of these cells in minimal medium. Interestingly, we found that EpiSC and hESC both have low basal OCRs (normalized to cell number or protein level, Suppl. Table 4) compared to ESC (Fig 1a and c). In the presence of glucose, the ECARs for EpiSC and hESC are substantially higher than ESC (Fig 1b, e and f). This observation indicates a strong preference of EpiSC and hESC for glycolytic metabolism. The ECAR difference in ESC and EpiSC was confirmed by direct measurement of lactate

levels in conditioned media (Fig 2a). The ECAR difference could also partially result from other possible acid generation, including monocarboxylates and CO_2 produced from respiration. Furthermore, carbonyl cyanide 3-chlorophenylhydrazone (CCCP) was added in order to discharge the proton gradient, thereby allowing maximal turnover of the electron transport chain uncoupled from ATP synthesis. This analysis allows estimation of the maximal mitochondria reserve in the presence of glucose (Goldsby & Heytler, 1963; Heytler, 1963). A robust increase in OCR was detected in ESC in the presence of CCCP (Fig.1a and d; Suppl. Fig1a and b). However, very little or no increase in OCR was observed with EpiSC or hESC (Fig1a and d; Suppl. Fig.1a and b), indicating that these cell types have diminished mitochondrial functional reserves. The observed change in ECAR due to CCCP administration could be caused by increased glycolysis, or by increased CO_2 production from the TCA cycle. From calculations based on OCR and ECAR changes upon glucose addition, we further show that ATP production upon glucose addition is higher in EpiSC and hESC than ESC (Fig 1g), probably reflecting a higher glycolytic capacity in these cells. In contrast, cellular ATP content is lower in EpiSC than ESC (Fig 1h), suggesting a high ATP consumption rate in EpiSC. To compare different stages of ES cells in human, we used hESC H1 cells treated with sodium butyrate as a developmentally earlier stage (Ware et al., 2009). We observed that similar to EpiSC, hESC H1 cells also contain a lower steady-state level of ATP compared to an earlier pluripotent stage (Suppl. Fig 2a). Taken together, these results demonstrate a clear metabolic difference between ESC as compared to EpiSC and hESC: the latter two cells are alike in terms

of having lower mitochondrial respiration and higher glycolytic rate. These differences raise interesting questions as to how these metabolic changes occur and the impact of these differences on cellular pluripotency.

EpiSC and hESC are highly glycolytic

To further test the requirement for glycolysis in the two pluripotent stages, we cultured ESC, EpiSC and hESC with 2-deoxy-glucose (2-DG), a glucose analog that competes with glucose as a substrate for glycolytic enzymes and therefore acts as an inhibitor of glycolysis. In the presence of 2-DG, we observed that ESC grow more slowly, but maintain an ESC phenotype, forming domed cell colonies that stain with alkaline phosphatase (Fig 2b). However, EpiSC and hESC cannot survive in the presence of 2-DG (Fig 2b). In EpiSC and hESC, ECAR decreases to a greater extent than in ESC with addition of 2-DG (Fig 2c), however, unlike in ESC, the ability to increase respiration to compensate for decreased glycolysis is greatly diminished at the Epi-stage (both in EpiSC and hESC). A similar effect was observed using a lactate dehydrogenase inhibitor, oxamate (Fig 2d). In the presence of oxamate, pyruvate generated by glycolysis cannot be converted to lactate, but may be available for mitochondrial oxidation in the citric acid cycle, leading to an increase in mitochondrial respiration as observed in ESC. Our results showed that no increase in OCR was observed for EpiSC and hESC (Fig 2d). Taken together, these results indicate that glycolysis is essential for EpiSC and hESC bioenergetics due to their low mitochondrial respiratory capacity.

EpiSC and hESC have more mature mitochondria but lower mitochondrial respiration than ESC

Several additional lines of evidence further confirm that EpiSC and hESC have reduced mitochondrial respiration as compared to ESC. Treatment of these cells with oligomycin, an ATP synthase inhibitor (Chappell & Greville, 1961), resulted in similar residual OCR for ESC, EpiSC and hESC (Fig 3a). Since inhibition of mitochondrial ATP synthesis results in similar residual OCR, the higher OCR in ESC can be attributed to higher level of coupled mitochondrial respiration. FCCP treatment following Oligomycin resulted in higher OCR increase in ESC than EpiSC and hESC (Fig 3a), confirming higher level of maximal mitochondrial activity in ESC. Another mitochondrial uncoupler, 2,4-Dinitrophenol (DNP) (Krahl & Clowes, 1936) also gave similar results to CCCP (Fig 3b).

Lower mitochondrial respiration in EpiSC and hESC could be due to reduced numbers of mitochondria or reflect the developmental immaturity of mitochondria in these cells compared to ESC. To test this, we first examined morphology of mitochondria in EpiSC and hESC compared to ESC by electron microscopy. We observed that the majority of mitochondria in ESC are rounded to oval, displaying sparse and irregular cristae and an electron-lucent matrix, in contrast to the mitochondria of EpiSC and hESC, which are more elongated, and contain well-defined transverse cristae and a dense matrix (Fig 3c-e). Elongated mitochondria were observed about three and five times as frequently in EpiSC and hESC, respectively, as

compared to ESC (Fig 3f). This morphological assessment suggests that the mitochondria of EpiSC and hESC are more mature in appearance than ESC, consistent with their relatively later developmental stage. Similarly, significantly higher mitochondrial DNA (mtDNA) copy numbers were detected in EpiSC compared to ESC (Fig 3g), mtDNA copy number was also lower in hESC H1 cultured with sodium butyrate compared to H1 (Suppl. Fig 2b). These results stand in stark contrast to the lower respiratory activity of EpiSC and hESC relative to ESC. We also tested the possibility that diminished pyruvate oxidation by mitochondrial pyruvate dehydrogenase in EpiSC may cause the differences in mitochondrial respiration compared to ESC. Treatment with dichloroacetate, an inhibitor of pyruvate dehydrogenase kinases (Whitehouse et al.., 1974), increased respiration in ESC, but not EpiSC (Fig 3h).

Reduced mitochondrial respiration in EpiSC and hESC is attributable to a deficiency in ETC complex IV cytochrome c oxidase (COX)

In a search for other possible mechanisms accounting for the low mitochondrial respiration activity in EpiSC/hESC, we observed that EpiSC have lower mitochondrial membrane potential than ESC as measured by staining with tetramethylrhodamine methyl ester (TMRM) (Fig 4a), a dye that rapidly and reversibly equilibrates across membranes in a voltage-dependent manner (Ehrenberg et al., 1988). In agreement with a recent study (Folmes et al., 2011), we also observed that mouse embryonic fibroblasts (MEFs) have less TMRM staining than ESC. Lower mitochondrial membrane potential seen in EpiSC suggests that the mitochondrial

electron transport chain (ETC) may not operate sufficiently to generate an effective proton gradient. In order to identify mechanisms in mitochondrial ETC that could account for the lower membrane potential of EpiSC compared to ESC, we examined gene expression microarray data from these two types of cells (Tesar et al., 2007), and surprisingly, found that a majority of genes in mitochondrial complex I and IV are expressed at a lower level in EpiSC compared to ESC (Fig 4b; Suppl. Fig 3a and c). Notably, in the complex IV cytochrome c oxidase (COX) family, 20 out of a total of 22 nuclear-encoded genes are down-regulated in EpiSC ($p<0.005$) (Fig 4b). We further validated the significant reduction of key genes in these ETC components in EpiSC compared to ESC by quantitative PCR-assay (Fig 4c), and compared the expression abundance of these key genes as compared to β-Actin in mouse and human (Suppl. Table 1). Given the uniformly reduced expression of COX mRNAs in EpiSC, it is possible that translation and assembly of COX proteins are largely defective. To test whether COX activity is deficient in EpiSC and hESC, we prepared mitochondrial extracts from ESC and EpiSC, as well as two hESC lines, H1 and H7, to measure the COX activity *in vitro*. Indeed, there is about two-fold reduction in COX activity per microgram of mitochondrial protein in EpiSC as compared to ESC (Fig 4d). We also observed that hESC resemble EpiSC in having a low level of COX activity (Fig 4d). Since complex IV levels are limiting and have previously been shown to tightly regulate mitochondrial respiratory capacity (Villani et al., 1998), low complex IV activity in EpiSC and hESC could explain their low mitochondrial respiration activity relative to ESC.

We further found that expression of synthesis of cytochrome c oxidase 2 (SCO2), peroxisome proliferator-activated receptor γ coactivator-1β (PGC-1β) and Estrogen receptor-related receptor β (Esrrb) are significantly lower in EpiSC as compared to ESC (Fig 4e). SCO2 is required for the assembly of the COX complex IV and mutation of this gene in humans results in fatal cardioencephalomyopathy due to mitochondrial respiratory failure (Papadopoulou et al., 1999) (Other mitochondrial assembly factors were also examined in Suppl. Table 2). Similarly, PGC-1β controls mitochondrial oxidative metabolism by activating specific target genes that are key components of mitochondria, including those in the mitochondrial membrane and ETC (Lelliott et al., 2006; Sonoda et al., 2007). More specifically, PGC-1β could act as a ligand for Esrrb to control metabolism and energy balance (Kamei et al., 2003). Lower expression of SCO2, PGC-1β and Esrrb in EpiSC could contribute to the reduced mitochondrial respiration activity in these cells compared to ESC.

Lower mitochondrial COX genes in post-implantation epiblast in vivo

To test whether the metabolic differences between ESC and EpiSC reflect differences that exist *in vivo*, we compared our results with results obtained from high throughput deep sequencing of mRNA using the freshly dissected inner cell mass (ICM) of pre-implantation embryos and the epiblast of post-implantation embryos (Fig 5) (manuscript in preparation). In agreement with results of ESC and EpiSC cultured *in vitro*, our deep RNA-sequence results reveal a significantly lower level of COX

mRNA in the epiblast relative to the inner cell mass (Fig 5a-b, *in vivo*: $p<0.05$, *in vitro*: $p<0.01$ as compared to all other genes). Further, close examination reveals high correlation in the most significantly down-regulated COX genes (Fig 5c) and their regulators, PGC-1β and Esrrb *in vivo* versus *in vitro* (Fig 5d). These data confirm a dramatic down-regulation of mitochondrial COX genes during the transition from ICM to epiblast in vivo.

HIF1α is a key regulator of the pluripotent state

To understand the drivers of the acquisition of a highly glycolytic state in EpiSC, we searched gene expression signatures in *in vitro* microarray data and identified the characteristic HIF1α-driven gene expression profile in EpiSC but not in ESC (Suppl. Table 3). We validated three of the key HIF1α targets, PDK1, LDHA and PYGL, in EpiSC compared to ESC, and observed a 10-70 fold increase in expression levels of these HIF1α targets in EpiSC stage (Fig 6a). As a control, we observed the expected increase of Cer1 in EpiSC compared to the ESC stage. We further observed that HIF1α protein is present at a significantly higher level in EpiSC than in ESC (Fig 6b, Suppl. Fig 4). To test whether HIF1α is sufficient to induce the transition from ESC to EpiSC, we over-expressed or induced HIF1α in ESC transiently for three days in the presence of leukemia inhibitory factor (LIF). Importantly, expression of a non-degradable form of HIF1α by retroviral infection or induction of endogenous HIF1α by the chemical hypoxia inducer $CoCl_2$ both render ESC not only morphologically but also metabolically similar to EpiSC. We show that HIF1α stabilization through both

means (Fig 6c) significantly increase the percentage of EpiSC-like colonies in ESC culture in the presence of LIF (Fig 6d-h). Further, HIF1α over-expressing ESC have reduced mitochondrial respiration and higher glycolytic activity compared to control ESC (Fig 6i-k). Although transient over-expression of HIF1α for three days did not show changes at the molecular levels of key genes (data not shown), over-expression of HIF1α for a longer period (six days) does result in significant changes in the expression levels of key genes toward EpiSC-like stage, including lineage marker Cer1, glycolytic gene LDHA and two other metabolism related genes COX7a1 and Esrrb (Fig 6l and Suppl. Fig 5). These data suggest that HIF1α acts as a key regulator of the metabolic and phenotypic shifts from ESC to EpiSC.

Activin signaling is indispensable in the HIF-regulated transition from ESC to EpiSC

Activin is shown to be essential for maintenance of EpiSC in culture and withdrawal of Activin signaling results in EpiSC differentiation into neuroendoderm (Camus et al., 2006; James et al., 2005; Vallier et al., 2004). In contrast, ESC do not require Activin for pluripotency (Tesar et al., 2007); conversely, addition of Activin signaling results in a shift from the ESC towards EpiSC state (Guo et al., 2008; Hayashi et al., 2011; Fig 7a, Suppl. Fig 6). In this study we show that HIF1α activation switches ESC morphologically, metabolically and based on the expression signature towards an EpiSC-like state. To test whether HIF1α regulates this state switch through Activin signaling, we cultured ESC with LIF media containing $CoCl_2$ as well as an inhibitor of Activin signaling, SB431542, which specifically binds with Activin receptor-like

kinase (ALKi) (Inman et al., 2002). While chemical hypoxia alone induced the EpiSC-like state in 50% of the ESC colonies, no EpiSC-like induction was observed when the Activin pathway was repressed during chemical hypoxia (Fig 7a). These data show that HIF1α dependent induction of EpiSC-state requires Activin/Nodal signaling. During Activin induced ESC to EpiSC transition, we observed that HIF1α protein is stabilized (Fig 7b). Furthermore, we also observed significant changes in the expression levels of key metabolic genes, including up-regulation of glycolytic gene LDHA and down-regulation of genes regulating mitochondrial activity COX7a1 and Esrrb when ESC are cultured with Activin and FGF (Fig 7c and Suppl. Fig 5). No Cer1 up-regulation was observed in ESC cultured with Activin and FGF for three days, even though these cells displayed EpiSC-like metabolic signature. We therefore further analyzed the kinetics of key metabolic and lineage related genes in the course of ESC to EpiSC transition induced by Activin and FGF in culture (Hayashi et al., 2011). The analysis shows that changes in metabolic gene expression precede the changes in the expression of EpiSC lineage markers upon Activin treatment (Fig 7d). Together these data suggest that Activin signaling pathway is required during the ESC to EpiSC transition possibly by regulating key genes related the metabolism that are characteristic of EpiSC stage.

Discussion

In the present study, we demonstrate that a dramatic switch from a bivalent metabolism to an exclusively glycolytic metabolism takes place between two

pluripotent stages reflective of the pre-implantation ICM and post-implantation epiblast (Fig.7e). While ESC possesses functional mitochondrial respiration in minimal media and upon extrinsic induction, EpiSC and hESC are defective in mitochondrial function, mainly relying on glycolysis for cellular ATP demand. We found that EpiSC and hESC show low mitochondrial ETC complex IV activity, compromising the overall respiratory capacity of these cells. The down-regulation of complex IV in ICM to epiblast transition is also observed *in vivo*, suggesting that the ETC down-regulation in the epiblast stage has a tremendous beneficial value for the pluripotent cell population. Furthermore, EpiSC and hESC up-regulate key glycolytic genes, maximizing their anaerobic capacity to fulfill cellular energy demand.

Metabolic changes are associated with cellular differentiation. The choice between anaerobic metabolism and aerobic respiration may play an important role in determining specific lineage decisions (Bracha et al., 2010; Mandal et al., 2011; Roberts et al., 2009; Yanes et al., 2010). Among other changes, the number, morphology and function of mitochondria dramatically change at different developmental stages. In early embryo development, mitochondria in the 8-cell embryo reveal minimal matrix electron density. Elongating mitochondria with inner mitochondrial membranes arranged into transverse cristae appear, and the replication of mtDNA takes place in expanding blastocysts (Sathananthan & Trounson, 2000; Thundathil et al., 2005). It has been shown previously that undifferentiated ESC, compared to their differentiated progeny, have restricted oxidative capacity with low

mtDNA copy number and low mitochondrial mass (Cho et al., 2006). Consistent with these previous findings, our data show that compared to ESC, the advanced pluripotency state reflected in EpiSC leads to more mature mitochondria and higher mtDNA copy numbers. However, paradoxically we found that mitochondria in EpiSC are less active, and defective in aerobic respiration due to compromised COX activity. Low COX gene expression and low mitochondrial respiration are conserved in hESC, suggesting that low mitochondrial activity is beneficial for cells at this stage. One possibility is that since the Primordial Germ Cell (PGC) precursors--which are a necessity for the continuity of the species--are formed at the epiblast stage, the developing animal will minimize the potential harm generated by reactive oxygen species by blocking mitochondrial activities to protect the germ line. Recent findings support this hypothesis. Activin treatment for short period of time induces ESC to a stage potent for PGC differentiation (Hayashi et al., 2011). These cells show an EpiSC-like metabolic signatures, however, they do not show yet the canonical fate marker changes observed in EpiSC (Fig 7d), suggesting that the metabolic changes may be imperative for successful PGC formation.

COX activity has been shown to be a rate limiting factor in mitochondrial respiration (Villani et al.., 1998). The degradation of mitochondrial function through loss of COX activity is also evident in several pathological cases. COX is a specific intra-mitochondrial site of age-related deterioration (Dillin et al., 2002; Ren et al., 2010), and is currently considered an endogenous marker of neuronal oxidative metabolism

(Bertoni-Freddari et al., 2004), which when defective may be causal for Alzheimer's disease (Ojaimi et al., 1999). In the present study, we demonstrate that while EpiSC/hESC have a robust number of maturing mitochondria, the expression of COX genes is down-regulated, reducing the mitochondrial function in EpiSC/hESC. This work, through defining the metabolic differences between two pluripotent stages has revealed that the developing animal can modulate mitochondrial activity by regulating COX levels. While reduction of COX activity is previously shown to associate with pathological cases, the developing pluripotent stem cell can harness this reduction to its benefit, possibly to protect its pluripotent stage against oxidative stress. It will be important to reveal whether this same strategy is used in other developmental stages.

The PGC-1 family is involved in regulating mitochondrial activity. Compared to ESC, we found that EpiSC show lower PGC-1β expression, as well as a strong repression of its receptor ERR-β. Reduced expression of PGC-1β combined with ERR-β has been shown to result in reduced ERR-mediated transcription of nuclear-encoded mitochondrial genes, which would ultimately attenuate mitochondrial function (Shao et al., 2010). Direct comparison using microarray analysis reveals increased expression of PGC-1α in EpiSC (Suppl. Table 1), which is in accordance with its known role regulating mitochondrial biogenesis and replication (Wu et al., 1999) and could explain the increased mitochondria content observed in EpiSC. A critical aspect of PGC-1 co-activators is that they are highly versatile; different members of the PGC-1 family are shown to interact with distinct transcription factors by binding to

various receptors either in a ligand-dependent manner or without addition of a ligand (Lin et al., 2005). We speculate that in early embryonic development, PGC-1 members play different and crucial roles in mitochondrial biology: PGC-1α may regulate mitochondrial replication and biogenesis by activating mitochondrial and nuclear transcription, and the lack of PGC-1β may play a role in repressing the function of these newly generated mitochondria. Overall, the coordinated regulation by PGC-1 enables early embryonic cells to develop a sufficient number of mitochondria as a reservoir for the increased energy demands for future differentiation, while maintaining an anaerobic metabolism important for self-renewal and pluripotency (Gan et al., 2010; Varum et al., 2009).

Embryonic development takes place in a hypoxic environment (Fischer & Bavister, 1993; Lee et al., 2001), and HIF1α signaling has been shown to play an indispensable role in directing morphogenesis in the embryo and placenta (Dunwoodie, 2009). We have shown that HIF1α can play a role in pluripotency by regulating metabolic transition of the pluripotent cells before and after implantation. We demonstrate that HIF1α over-expression, not only induces morphological change reflective of the transition from ESC to EpiSC, but is also sufficient to enhance glycolysis at the expense of oxidative phosphorylation. These observations are consistent with the known function of HIF1α in glycolysis (Seagroves et al., 2001). Moreover, HIF1α can also induce active suppression of mitochondrial oxidative respiration. We observe lower mitochondrial respiratory activity in EpiSC, and HIF1α over-expression in ESC

phenocopies this metabolic shift.

We further reveal that metabolic changes during the ESC to EpiSC transition induced by HIF1α act through Activin/Nodal signaling. HIF1α has been shown to bind to HIF responsive element (HRE) on the Activin B promoter to directly regulate its expression (Wacker et al., 2009). HIF1α induces Activin receptor-like kinase (Garrido-Martin et al., 2010), which further mediates some hypoxia induced processes, such as angiogenesis (Lux et al., 2006). Moreover, activation of Activin/Nodal signaling is required to maintain pluripotent cells in culture as EpiSC, preventing the spontaneous differentiation process (Camus et al., 2006; James et al., 2005; Vallier et al., 2004), and recombinant Activin is sufficient to transit ESC towards EpiSC (Hayashi et al., 2011; Fig 7a and Suppl. Fig 6). It has also been shown that Activin/Nodal stabilizes HIF1α by decreasing prolyl hydroxylase 2 (Wiley et al., 2010). Accordingly, we observed HIF1α stabilization due to Activin induction in ESC. Given these observations, it is possible that a feedback loop exists between HIF1α and Activin/Nodal signaling during early embryonic development (Fig 7e).

We identify three transcriptional signaling pathways (PGC-1β, HIF1α and Activin/Nodal) that are involved in the dramatic metabolic change between pluripotency stages (Fig 7e). HIF1α is shown to negatively regulate PGC-1β by inhibiting c-Myc transcriptional activity (Zhang et al., 2007). Further, Activin/Nodal signaling is reported to affect metabolism and is suggested to directly down-regulate

PGC-1, down-regulating mitochondrial metabolism (Li et al., 2009). Interestingly, the changes in metabolic gene expression precede the changes in the expression of EpiSC lineage marker upon Activin treatment, suggesting that metabolic changes may be leading the process. We propose a regulatory network that controls the proper metabolic switch in early embryo development (Fig 7e). In this network, we envision HIF1α as a master regulator: it not only plays an important role in anaerobic metabolism by activating key glycolytic enzymes, but also actively represses mitochondrial activity through inhibition of PGC-1β. Moreover, HIF1α acts through Activin/Nodal signaling, to broaden its effect by inhibiting the differentiation process and to strengthen its suppressive role in PGC-1. It remains to be answered whether such a regulatory network is conserved in human embryo development, and what other intermediate regulators are involved in this network.

Cancer cells are frequently characterized by a glycolytic shift, known as the Warburg effect. HIF-1 and Myc, transcription factors linked to the Warburg effect, are integral to embryonic stem cell programs. The outcome of the Warburg effect is to increased metabolic flux of glucose carbons into biosynthetic precursors, fueling anabolic processes, and control of redox potential and ROS that are required for rapid tumor cell growth and division. The developmental suppression of oxidative phosphorylation in EpiSC/hESC may serve a similar function in preparation for embryonic growth and formation of germ cell layers.

Experimental Procedures

Cell Culture

Early passage (passage<40) of ESC and EpiSC were cultured on irradiated MEF feeder in 37◦C described previously (Tesar et al., 2007; Ying et al., 2008). Specifically, medium for ESC contained DMEM (Invitrogen), 5% ES cell-qualified fetal bovine serum (Atlas Biologicals), 1mM 2-mercaptoethanol (Sigma-Aldrich), 2mM pyruvate (Invitrogen), non-essential amino acids (Invitrogen) and 10^3 units/ml leukemia inhibitory factor (LIF; Millipore) with addition of GSK and MEK inhibitors (2i: GSKi: CHIR99021; MEKi: PD0325901. Stemgent). ESC were passaged every 2-3 day as a single-cell suspension using 0.25% trypsin/EDTA. Medium for EpiSC culture consisted of DMEM-F12 (Invitrogen), 20% knockout serum replacement (Invitrogen), 5ng/ml FGF2 (R&D Systems), 0.1mM 2-mercaptoethanol (Sigma-Aldrich), 2mM pyruvate (Invitrogen), non-essential amino acids (Invitrogen) and recombinant Activin A (Humanzyme). EpiSC were passaged every 2–3 days with Dispase (Invitrogen) and triturated into small clumps. hESC were cultured as EpiSC, but without addition of Activin A.

Oxygen consumption rate (OCR) and extracellular acidification rate (ECAR) measurements using SeaHorse Cellular Flux assays: SeaHorse plates were pre-treated by coating with 0.1% Gelatin and irradiated MEFs were seeded thereafter. About 24 hours before measurement, MEFs were lysed using a detergent solution 0.5% Triton and 0.034% (v/v) NH_4OH (Sigma-Aldrich) to retain their extracellular matrix and eliminate background OCR and ECAR. Cell density titrations were performed to

define the optimal seeding density for ESC (Suppl. Fig 7a) and EpiSC (Suppl. Fig 7b), and in following experiments, ESC and EpiSC were passaged and seeded in growth media described above onto pre-treated SeaHorse plate with 2-2.5×10^5 ESC or 0.8-1×10^5 EpiSC per XF24 well to ensure about 90% surface coverage at the time of experiment. Culture media were exchanged for base media (unbuffered DMEM (Sigma D5030) supplemented with 2 mM Glutamine) 1 hour before the assay and for the duration of the measurement. Substrates and selective inhibitors were injected during the measurements to achieve final concentration of: Glucose (0.5mM, 2.5mM and 7mM), CCCP (500nM), Oligomycin (2.5μM), 2,4-DNP (100 μM), DCA (20mM), 2-DG (50mM) and Oxamate (50mM) (All from Sigma-Aldrich), and CCCP, 2,4-DNP, 2-DG and Oxamate titrations were performed (CCCP, Suppl. Fig 8a; 2-DG, Suppl. Fig 8b; Oxamate, Suppl. Fig 8c and 2,4-DNP, Suppl. Fig 9, respectively). The OCR and ECAR values were further normalized to the number of cells present in each well, quantified by the Hoechst staining (HO33342, Sigma-Aldrich) as measured using fluorescence at 355nm excitation and 460nm emission. Normalization to the total protein amount in these cells was observed to result same normalization factor as the Hoechst staining (Suppl. Table 4). The baseline OCR and ECAR were defined as the average values measured from timepoint 1 to timepoint 5 (0-45 minutes) during the experiments. Changes in OCR and ECAR in response to substrates and inhibitors additions were defined as the maximal change after the chemical addition compared to the baseline. Due to variations in the absolute magnitude of OCR and ECAR measurements in different experiments, the relative OCR/ECAR levels were used to

compare and summarize independent biological replicates. Calculations were done as the ratio of OCR or ECAR values in EpiSC or hESC compared to ESC.

Mitochondrial membrane potential measurement: ESC and EpiSC were washed with DPBS, and 2ml of DMEM with 100nM TMRM (Invitrogen) were added to the culture plate for incubation at 37 ◦C for 30 minutes. Cells were further trypsinized and resuspended in DPBS for FACS analysis (BD FACS Canto II System). Channel PE was used to detect the fluorescent signal as stained by TMRM.

Mitochondrial DNA (mtDNA) copy number measurement: The ratio of mtDNA to genomic DNA was calculated by dividing copies of Co1 with copies of Gapdh in each experiment.The details of the assay were further described in Supplemental Procedures. Primers are listed in Supplemental Table 6 primer list.

ATP turnover and steady-state level measurement: ATP turnover was calculated directly from SeaHorse OCR and PPR measurements following formula: 1 ATP= 5× OCR areas under the curve + PPR areas under the curve. The steady-state level cellular ATP was measured following instruction specified in the ATP Determination Kit (Invitrogen). Briefly, cells were lysed with MPER extraction buffer (Thermo Scientific) in the presence of proteinase inhibitors. Total protein amounts in each reaction were quantified using BCA protein assay (Thermo Scientific).

Cytochrome C oxidase (COX, ETC complex IV) activity assay: ESC and EpiSC were collected and spun down as pellets. Cell lysis, protein extraction and activity measurement followed the instructions specified in Complex IV Rodent Enzyme Activity Microplate Assay Kit (MitoSciences). The details of the assay were further described in Supplemental Procedures.

Isolation of E3.5 ICM and E6.5 epiblast and RNA-seqencing: All embryos were recovered from C57BL/6 females. E3.5 blastocysts were flushed from the uterus of superovulated pregnant females. For the isolation of ICM, blastocysts were first placed in a rabbit anti-mouse polyclonal antibody (Rockland Immunochemicals) for 20 min at 37^{o}C and followed by guinea pig serum complement for 20-30 min at 37^{o}C. The lysed trophectoderm cells were removed and the isolated ICM was placed in lysis buffer. The derivation of epiblast from E6.5 post-implantation embryos has been described previously (Brons et al., 2007). The detailed RNA-sequencing procedures were described in Supplemental Procedures.

HIF over-expression by retro-viral infection and $CoCl_2$ induction: To obtain constitutively stable expression HIF1α protein, non-degradable HIF1α over-expressing plasmid (Addgene plasmid 19005) was used, in which two the proline sites of HIF1α cDNA were changed to alanine as described previously (Yan et al., 2007). Retrovirus made from the plasmid was infected into ESC in the presence of hexadimethrine bromide at 4ng/ml (Polybrene, Invitrogen) and was changed into

normal growth media containing LIF but without 2i after 24 hours. Alternatively, $CoCl_2$ (Sigma) was used as a chemical hypoxia inducer to stabilize HIF1α in ESC. For this, 100μM $CoCl_2$ was provided at the time of plating. ESC were cultured in normal growth medium containing LIF and $CoCl_2$ but without 2i for three days before SeaHorse assay or morphology examination. To induce HIF1α expression for a longer term, ESC were first cultured in normal growth media with LIF and $CoCl_2$ (or HIF1α viral expression) but without 2i for three days, and then switched to EpiSC media containing Activin and FGF for additional three days for further maturation.

HIF1α protein western blot: HIF1α protein stabilization in various pluripotent stages was examined using western blot following procedures specified previously (Zhou et al., 2011), and using HIF1α (ab2185, Abcam, Cambridge, MA) at 1:1000 dilution.

Activin/Nodal signaling inhibition: SB431542 (Stemgent) was maintained as a 20mM stock solution in DMSO (vehicle) and was provided at 20 μM to the cultures at the time of plating and every day thereafter with the media change. ESC were cultured in normal growth media as specified above with SB431542 for three days before morphology examination.

Details of lactate measurement of ESC, EpiSC and hESC, RNA isolation and gene expression by real-time, PCR Electron microscopy of mitochondria and quantification of elongated mitochondria were described in Supplemental Procedures as in Appendix

II.

References

Bendall SC, Stewart MH, Menendez P, George D, Vijayaragavan K, Werbowetski-Ogilvie T, Ramos-Mejia V, Rouleau A, Yang J, Bosse M, Lajoie G, Bhatia M (2007) IGF and FGF cooperatively establish the regulatory stem cell niche of pluripotent human cells in vitro. *Nature* **448:** 1015-1021

Bertoni-Freddari C, Fattoretti P, Giorgetti B, Solazzi M, Balietti M, Casoli T, Di Stefano G (2004) Cytochrome oxidase activity in hippocampal synaptic mitochondria during aging: a quantitative cytochemical investigation. *Ann N Y Acad Sci* **1019:** 33-36

Bracha AL, Ramanathan A, Huang S, Ingber DE, Schreiber SL (2010) Carbon metabolism-mediated myogenic differentiation. *Nat Chem Biol* **6:** 202-204

Brons IG, Smithers LE, Trotter MW, Rugg-Gunn P, Sun B, Chuva de Sousa Lopes SM, Howlett SK, Clarkson A, Ahrlund-Richter L, Pedersen RA, Vallier L (2007) Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448:** 191-195

Brook FA, Gardner RL (1997) The origin and efficient derivation of embryonic stem cells in the mouse. *Proc Natl Acad Sci U S A* **94:** 5709-5712

Camus A, Perea-Gomez A, Moreau A, Collignon J (2006) Absence of Nodal signaling promotes precocious neural differentiation in the mouse embryo. *Dev Biol* **295:** 743-755

Chappell JB, Greville GD (1961) Effects of oligomycin on respiration and swelling of isolated liver mitochondria. *Nature* **190:** 502-504

Cho YM, Kwon S, Pak YK, Seol HW, Choi YM, Park do J, Park KS, Lee HK (2006) Dynamic changes in mitochondrial biogenesis and antioxidant enzymes during the spontaneous differentiation of human embryonic stem cells. *Biochem Biophys Res Commun* **348:** 1472-1478

Dillin A, Hsu AL, Arantes-Oliveira N, Lehrer-Graiwer J, Hsin H, Fraser AG, Kamath RS, Ahringer J, Kenyon C (2002) Rates of behavior and aging specified by mitochondrial function during development. *Science* **298:** 2398-2401

Dunwoodie SL (2009) The role of hypoxia in development of the Mammalian embryo. *Dev Cell* **17:** 755-773

Ehrenberg B, Montana V, Wei MD, Wuskell JP, Loew LM (1988) Membrane potential can be determined in individual cells from the nernstian distribution of cationic dyes. *Biophys J* **53:** 785-794

Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292:** 154-156

Fischer B, Bavister BD (1993) Oxygen tension in the oviduct and uterus of rhesus monkeys, hamsters and rabbits. *J Reprod Fertil* **99:** 673-679

Folmes CD, Nelson TJ, Martinez-Fernandez A, Arrell DK, Lindor JZ, Dzeja PP, Ikeda Y, Perez-Terzic C, Terzic A (2011) Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell Metab* **14:** 264-271

Gan B, Hu J, Jiang S, Liu Y, Sahin E, Zhuang L, Fletcher-Sananikone E, Colla S, Wang YA, Chin L, Depinho RA (2010) Lkb1 regulates quiescence and metabolic homeostasis of haematopoietic stem cells. *Nature* **468:** 701-704

Garrido-Martin EM, Blanco FJ, Fernandez LA, Langa C, Vary CP, Lee UE, Friedman SL, Botella LM, Bernabeu C (2010) Characterization of the human Activin-A receptor type II-like kinase 1 (ACVRL1) promoter and its regulation by Sp1. *BMC Mol Biol* **11:** 51

Goldsby RA, Heytler PG (1963) Uncoupling of Oxidative Phosphorylation by Carbonyl Cyanide Phenylhydrazones. Ii. Effects of Carbonyl Cyanide M-Chlorophenylhydrazone on Mitochondrial Respiration. *Biochemistry* **2:** 1142-1147

Greber B, Wu G, Bernemann C, Joo JY, Han DW, Ko K, Tapia N, Sabour D, Sterneckert J, Tesar P, Scholer HR (2010) Conserved and divergent roles of FGF signaling in mouse epiblast stem cells and human embryonic stem cells. *Cell Stem Cell* **6:** 215-226

Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, Smith A (2009) Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**: 1063-1069

Hayashi K, Surani MA (2009) Self-renewing epiblast stem cells exhibit continual delineation of germ cells with epigenetic reprogramming in vitro. *Development* **136:** 3549-3556

Hayashi K, Ohta H, Kurimoto K, Aramaki S, Saitou M (2011) Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**:

519-532

Heytler PG (1963) uncoupling of oxidative phosphorylation by carbonyl cyanide phenylhydrazones. I. Some characteristics of m-Cl-CCP action on mitochondria and chloroplasts. *Biochemistry* **2:** 357-361

Inman GJ, Nicolas FJ, Callahan JF, Harling JD, Gaster LM, Reith AD, Laping NJ, Hill CS (2002) SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol Pharmacol* **62:** 65-74

James D, Levine AJ, Besser D, Hemmati-Brivanlou A (2005) TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in human embryonic stem cells. *Development* **132:** 1273-1282

Kamei Y, Ohizumi H, Fujitani Y, Nemoto T, Tanaka T, Takahashi N, Kawada T, Miyoshi M, Ezaki O, Kakizuka A (2003) PPARgamma coactivator 1beta/ERR ligand 1 is an ERR protein ligand, whose expression induces a high-energy expenditure and antagonizes obesity. *Proc Natl Acad Sci U S A* **100:** 12378-12383

Krahl ME, Clowes GH (1936) Studies on Cell Metabolism and Cell Division : Ii. Stimulation of Cellular Oxidation and Reversible Inhibition of Cell Division by Dihalo and Trihalophenols. *J Gen Physiol* **20:** 173-184

Lee YM, Jeong CH, Koo SY, Son MJ, Song HS, Bae SK, Raleigh JA, Chung HY, Yoo MA, Kim KW (2001) Determination of hypoxic region by hypoxia marker in developing mouse embryos in vivo: a possible signal for vessel development. *Dev Dyn* **220:** 175-186

Lelliott CJ, Medina-Gomez G, Petrovic N, Kis A, Feldmann HM, Bjursell M, Parker N, Curtis K, Campbell M, Hu P, Zhang D, Litwin SE, Zaha VG, Fountain KT, Boudina S, Jimenez-Linan M, Blount M, Lopez M, Meirhaeghe A, Bohlooly YM, Storlien L, Stromstedt M, Snaith M, Oresic M, Abel ED, Cannon B, Vidal-Puig A (2006) Ablation of PGC-1beta results in defective mitochondrial activity, thermogenesis, hepatic function, and cardiac performance. *PLoS Biol* **4:** e369

Li L, Shen JJ, Bournat JC, Huang L, Chattopadhyay A, Li Z, Shaw C, Graham BH, Brown CW (2009) Activin signaling: effects on body composition and mitochondrial energy metabolism. *Endocrinology* **150:** 3521-3529

Lin J, Handschin C, Spiegelman BM (2005) Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab* **1:** 361-370

Lux A, Salway F, Dressman HK, Kroner-Lux G, Hafner M, Day PJ, Marchuk DA,

Garland J (2006) ALK1 signalling analysis identifies angiogenesis related genes and reveals disparity between TGF-beta and constitutively active receptor induced gene expression. *BMC Cardiovasc Disord* **6:** 13

Mandal S, Lindgren AG, Srivastava AS, Clark AT, Banerjee U (2011) Mitochondrial function controls proliferation and early differentiation potential of embryonic stem cells. *Stem Cells* **29:** 486-495

Nichols J, Smith A (2009) Naive and primed pluripotent states. *Cell Stem Cell* **4:** 487-492

Ojaimi J, Masters CL, McLean C, Opeskin K, McKelvie P, Byrne E (1999) Irregular distribution of cytochrome c oxidase protein subunits in aging and Alzheimer's disease. *Ann Neurol* **46:** 656-660

Papadopoulou LC, Sue CM, Davidson MM, Tanji K, Nishino I, Sadlock JE, Krishna S, Walker W, Selby J, Glerum DM, Coster RV, Lyon G, Scalais E, Lebel R, Kaplan P, Shanske S, De Vivo DC, Bonilla E, Hirano M, DiMauro S, Schon EA (1999) Fatal infantile cardioencephalomyopathy with COX deficiency and mutations in SCO2, a COX assembly gene. *Nat Genet* **23:** 333-337

Ren JC, Rebrin I, Klichko V, Orr WC, Sohal RS (2010) Cytochrome c oxidase loses catalytic activity and structural integrity during the aging process in Drosophila melanogaster. *Biochem Biophys Res Commun* **401:** 64-68

Roberts LD, Virtue S, Vidal-Puig A, Nicholls AW, Griffin JL (2009) Metabolic phenotyping of a model of adipocyte differentiation. *Physiol Genomics* **39:** 109-119

Sathananthan AH, Trounson AO (2000) Mitochondrial morphology during preimplantational human embryogenesis. *Hum Reprod* **15 Suppl 2:** 148-159

Seagroves TN, Ryan HE, Lu H, Wouters BG, Knapp M, Thibault P, Laderoute K, Johnson RS (2001) Transcription factor HIF-1 is a necessary mediator of the pasteur effect in mammalian cells. *Mol Cell Biol* **21:** 3436-3444

Shao D, Liu Y, Liu X, Zhu L, Cui Y, Cui A, Qiao A, Kong X, Chen Q, Gupta N, Fang F, Chang Y (2010) PGC-1 beta-regulated mitochondrial biogenesis and function in myotubes is mediated by NRF-1 and ERR alpha. *Mitochondrion* **10:** 516-527

Smith AG, Heath JK, Donaldson DD, Wong GG, Moreau J, Stahl M, Rogers D (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336:** 688-690

Sonoda J, Mehl IR, Chong LW, Nofsinger RR, Evans RM (2007) PGC-1beta controls

mitochondrial metabolism to modulate circadian activity, adaptive thermogenesis, and hepatic steatosis. *Proc Natl Acad Sci U S A* **104:** 5223-5228

Tesar PJ, Chenoweth JG, Brook FA, Davies TJ, Evans EP, Mack DL, Gardner RL, McKay RD (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448:** 196-199

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM (1998) Embryonic stem cell lines derived from human blastocysts. *Science* **282:** 1145-1147

Thundathil J, Filion F, Smith LC (2005) Molecular control of mitochondrial function in preimplantation mouse embryos. *Mol Reprod Dev* **71:** 405-413

Vallier L, Reynolds D, Pedersen RA (2004) Nodal inhibits differentiation of human embryonic stem cells along the neuroectodermal default pathway. *Dev Biol* **275:** 403-421

Varum S, Momcilovic O, Castro C, Ben-Yehudah A, Ramalho-Santos J, Navara CS (2009) Enhancement of human embryonic stem cell pluripotency through inhibition of the mitochondrial respiratory chain. *Stem Cell Res* **3:** 142-156

Villani G, Greco M, Papa S, Attardi G (1998) Low reserve of cytochrome c oxidase capacity in vivo in the respiratory chain of a variety of human cell types. *J Biol Chem* **273:** 31829-31836

Wacker I, Sachs M, Knaup K, Wiesener M, Weiske J, Huber O, Akcetin Z, Behrens J (2009) Key role for activin B in cellular transformation after loss of the von Hippel-Lindau tumor suppressor. *Mol Cell Biol* **29:** 1707-1718

Ware C, Wang L, Mecham BH, Shen L, Nelson AM, Bar M, Lamba DA, Dauphin DS, Buckingham B, Askari B, Lim R, Tewari M, Gartler SM, Issa JP, Pavlidis P, Duan Z, Blau CA (2009) Histone deacetylase inhibition elicits an evolutionarily conserved self-renewal program in embryonic stem cells. *Cell Stem Cell* **4**: 359-369

Wiley M, Sweeney KR, Chan DA, Brown KM, McMurtrey C, Howard EW, Giaccia AJ, Blader IJ (2010) Toxoplasma gondii activates hypoxia-inducible factor (HIF) by stabilizing the HIF-1alpha subunit via type I activin-like receptor kinase receptor signaling. *J Biol Chem* **285:** 26852-26860

Wu Z, Puigserver P, Andersson U, Zhang C, Adelmant G, Mootha V, Troy A, Cinti S, Lowell B, Scarpulla RC, Spiegelman BM (1999) Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1α. *Cell* **98:** 115-124

Yan Q, Bartz S, Mao M, Li L, Kaelin WG (2007) The hypoxia-inducible factor 2alpha N-terminal and C-terminal transactivation domains cooperate to promote renal tumorigenesis in vivo. *Mol Cell Biol* **27:** 2092-2102

Yanes O, Clark J, Wong DM, Patti GJ, Sanchez-Ruiz A, Benton HP, Trauger SA, Desponts C, Ding S, Siuzdak G (2010) Metabolic oxidation regulates embryonic stem cell differentiation. *Nat Chem Biol* **6:** 411-417

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

Zhang H, Gao P, Fukuda R, Kumar G, Krishnamachary B, Zeller KI, Dang CV, Semenza GL (2007) HIF-1 inhibits mitochondrial biogenesis and cellular respiration in VHL-deficient renal cell carcinoma by repression of C-MYC activity. *Cancer Cell* **11:** 407-420

Zhou W, Dosey TL, Biechele T, Moon RT, Horwitz MS, Ruohola-Baker H (2011) Assessment of Hypoxia inducible factor levels in cancer cell lines upon hypoxic induction using a novel reporter construct. *PLoS ONE* 6(11): e27460

Figure Legends

Figure 1: EpiSC and hESC show similar metabolic profiles. EpiSC and hESC have lower mitochondrial respiration activity and higher glycolytic rate than ESC. Oxygen consumption rate (OCR) (A, C and D) and extracellular acidification rate (ECAR) (B, E and F) measured in Seahorse extracellular flux assay are shown. To test energetic preference when Glucose is present, 0.5mM glucose, 2.5mM, 7mM glucose and 500nM CCCP, a mitochondria uncoupler, were injected during the experiment with the same time intervals; A and B are traces from a representative SeaHorse run. (G) ATP production was calculated directly from OCR and ECAR measured in Seahorse assay following ATP produced=OCR×time-course×5 + ECAR×time-course. (H) ESC have higher ATP content as the steady-state level than EpiSC. Results were

summarized from at least three independent biological experiments, each consisting of independent cell plating on five Seahorse microplate wells. The error bars represented the standard error of the mean, as calculated by sample standard deviation divided by the square root of the sample size. * indicates $p<0.05$, ** indicates $p<0.01$.

Figure 2: EpiSC are highly glycolytic compared to ESC. (A) EpiSC produce higher level of lactate, indicative of higher pyruvate generated through glycolysis. (B) ESC retain pluripotency in the presence of 2-deoxyglucose (2-DG), while EpiSC and hESC die upon 2-DG addition; Images for both bright fields and Alkaline Phosphatase (AP) staining are shown. (C) 2-DG results in higher ECAR reduction but lower OCR increase in EpiSC and hESC than ESC, confirming the high glycolysis and low mitochondria reserve present in EpiSC. (D) The addition of Oxamate, an inhibitor of lactate dehydrogenase, results in higher ECAR reduction but lower OCR increase in EpiSC and hESC than ESC, confirming the high glycolysis and low mitochondria reserve present in EpiSC. Results were summarized from two independent experiments, each consisting of independent cell plating on five Seahorse microplate wells. The error bars represented the standard error of the mean. * indicates $p<0.05$, ** indicates $p<0.01$.

Figure 3: Lower mitochondrial respiration in EpiSC, compared to ESC, is not due to mitochondrial immaturity, low mitochondria number or lack of pyruvate accessibility to mitochondria. (A) The differences in OCR between ESC, EpiSC and hESC were

abolished when Oligomycin was present, confirming that the difference in the aerobic respiration was caused by difference in the oxidative phosphorylation. FCCP treatment following Oligomycin resulted in higher OCR increase in ESC than EpiSC and hESC, confirming higher level of mitochondrial electron transport chain (ETC) activity in ESC. (B) Another mitochondrial uncoupler 2,4-DNP confirms that EpiSC and hESC have lower mitochondrial activity; Results were summarized from two independent experiments, and the error bars represented the standard error of the mean. (C-F) Electron microscopy shows that EpiSC and hESC contain more elongated mitochondria than ESC; Twenty EM images containing 194 and 296 mitochondria were used, respectively, for ESC and EpiSC mitochondrial quantification, and thirty images containing 212 mitochondria were used for hESC mitochondrial quantification. (G) EpiSC and hESC have higher mitochondrial DNA copy number than ESC. (H) ESC but not EpiSC increase OCR in response to DCA. Seahorse results were summarized from two independent experiments, each consisting of independent cell plating on five Seahorse microplate wells. The error bars represented the standard error of the mean. * indicates $p<0.05$, ** indicates $p<0.01$.

Figure 4: Expression of nuclear-encoded mitochondrial IV COX genes is lower in EpiSC compared to ESC, resulting in lower complex IV activity in EpiSC. (A) EpiSC have lower mitochondrial membrane potential than ESC and MEFs measured by

TMRM staining. (B) Heat-map of microarray expression data of mitochondrial complex IV COX gene cluster is shown, where EpiSC clearly demonstrate lower expression in these genes than ESC. (C) Validation of Cox8c, Cox6b2, Cox7a1 and NqoI by quantitative PCR assays. (D) Complex IV isolated from EpiSC or hESC (H1 and H7 lines) show lower activity *in vitro* than that of ESC; the activity of Complex IV isolated from MEFs is shown as comparison. (E) Potential regulators involved in mitochondria respiration, SCO2, PGC-1β and Esrrb, express at lower levels in EpiSC. Results were summarized from three independent biological samples, and the error bars represented the standard error of the mean. * indicates $p<0.05$, ** indicates $p<0.01$.

Figure 5: Deep RNA-sequence analysis of freshly dissected inner cell mass and epiblast reveals high similarity in mitochondrial COX gene expressions *in vivo* and *in vitro*. Expression levels of COX genes only are compared to that of all genes in the *in vivo* dataset (A) and in the *in vitro* dataset (B) to illustrate the down-regulation of COX genes (*in vivo*: $p<0.05$; *in vitro*: $p<0.01$). (C) The comparison of the most significantly down-regulated COX genes *in vivo* and *in vitro* data is shown. (D) High correlation of *in vivo* and *in vitro* data in the expression levels of PGC-1β and Esrrb is shown. Biological triplicates were used for each embryonic stage, the lysate comprised approximately 50 embryos in each experiment. * indicates $p<0.05$, ** indicates $p<0.01$.

Figure 6: HIF signature is present in EpiSC as glycolysis regulator. (A) PDK1, LDHA and PYGL that are involved in glycolysis and also HIF targets are expressed at higher levels in EpiSC; Cer1 is expected to increase in EpiSC and used in comparison. Quantitative PCR validation are shown; (B) EpiSC expresses HIF1α protein under 21 O_2, while little expression is shown in ESC; (C) HIF1α expression is confirmed in ESC treated with $CoCl_2$ and HIF1α viral expression; (D-H) HIF1α over-expressing ESC transiently for three days (ESC+HIF1α and ESC+$CoCl_2$ D3) form flat monolayer EpiSC-like colonies (shown by arrow), which are not present in control ESC (ESC+EV empty vector and ESC+DMSO); The normal-looking ESC labeled by stars and quantification of the percentage of EpiSC-like colonies are shown; ESC+HIF1α and ESC+$CoCl_2$ are metabolically different from ESC by having (I) lower baseline OCR and higher ECAR increase upon both 0.5mM Glucose (J) and 2.5mM Glucose addition (K), resembling EpiSC. (L) Significant changes resembling EpiSC in Cer1, LDHA, Cox7a1 and Esrrb expression were observed in ESC+HIF1α and ESC+$CoCl_2$. Results were summarized from two independent experiments, each consisting of independent cell plating on five Seahorse microplate wells. The error bars represented the standard error of the mean. * indicates $p<0.05$, ** indicates $p<0.01$.

Figure 7: (A) Quantification of the percentage of EpiSC-like colonies in ESC, ESC+$CoCl_2$, ESC+Activin and ESC+$CoCl_2$+ALKi show that Activin/Nodal signaling is required for $CoCl_2$ induced ESC to EpiSC transition. (B) ESC+Activin express

HIF1α as confirmed by western blot. (C) Significant changes resembling EpiSC in LDHA, Cox7a1 and Esrrb expression were observed in ESC+Activin. (D) Expression kinetics are shown for key metabolic genes (LDHA, PDK1, PYGL, Cox7a1 and Esrrb) and lineage related genes (Cer1, Zfp42, Dppa2, Dppa3 and Klf4) in ESC to EpiSC transition induced by Activin and FGF in culture, microarray from Hayashi et al. was examined; (E) Illustration of the regulatory network in the metabolic transition from ESC to EpiSC: higher HIF activity in EpiSC increases glycolysis. HIF down-regulates PGC-1β expression, reducing COX expression and ultimately resulting in a low mitochondrial respiration in EpiSC.

For Supplemental Figures and additional information, please refer to the publication: Zhou W, Choi M, Margineantu D, Cavanaugh C, Hesson J, Sweet I, Horwitz MS, Hockenbery D, Ware C and Ruohola-Baker H. HIF1α induced switch from bivalent to exclusively glycolytic metabolism during ESC to EpiSC/hESC transition. EMBO Journal. Mar 23; 31(9): 2103-16.

Chapter III: Somatic Mutations Quantify Random Contributions to Mouse Development

Abstract

The *C. elegans* cell fate map, wherein the lineage of each of its approximately 1000 cells is visibly charted beginning from the zygote, has proven remarkably insightful for understanding developmental biology. However, development in nematodes is invariant from one specimen to the next, whereas in mammals, cell lineages are assigned probabilistically, and development exhibits considerable variation between even constitutionally genetically identical individuals. Consequently, a single defined cell fate map applicable to all individuals cannot exist. To determine the extent to which patterns of cell lineage are conserved between different mice, we have employed the recently developed method of "phylogenetic fate mapping" to compare cell fate maps of sibling littermates. In this approach, somatic mutations arising in individual cells of an embryo are used to retrospectively deduce lineage relationships through phylogenetic and—as newly investigated here—related analytical approaches based on genetic distance. We have cataloged genomic mutations at an average of 110 mutation-prone polyguanine (polyG) tracts for about 100 cells clonally isolated from various corresponding tissues of two sibling littermates of a hypermutable mouse strain. We find that during mouse development, muscle and fat arise from a mixed pool of progenitor cells in the germ layer, but, in contrast, vascular endothelium in brain is derived from a smaller source of progenitor cells. Additionally, formation of

tissue primordia is then followed by establishment of left and right lateral compartments, with restricted cell migration between those divisions. We quantitatively demonstrate that development generally represents a combination of stochastic and deterministic events, as reflected in only partial conservation of cell lineage between individual mice. Such information may ultimately prove useful for determining how chance influences normal development and gives rise to birth defects.

Introduction

Mouse gestation takes about 20 days(SILVER 1995), and, although cell cycle length is variable, embryonic cells divide about twice per day (MAC AULEY *et al.*. 1993). It can therefore be surmised that about 40 or so mitotic generations transpire between fertilization and birth—a value similar to other estimates derived from different assumptions (FRUMKIN *et al.*. 2005). If all embryonic cell divisions produced 2 daughter cells that both subsequently divided, then a newborn mouse should be composed of 2^{40} ($\approx 10^{11}$) cells. Given that the mass of a cell is about 10^{-12} kg(WOLFF and PERTOFT 1972), a newborn mouse would weigh about 10 g. Actual measurements are nearer to just 1 g(SILVER 1995). However, each of the 2 daughter cells may experience different fates; both daughter cells do not always divide, nor do they do so at the same time. Along with the effects of apoptosis, this accounts for the fact that a newborn mouse has fewer cells than anticipated if embryonic cell proliferation were to proceed exponentially.

In fact, asymmetric cell divisions are evident in the *C. elegans*‘ cell fate map, in which the lineage of every cell in the worm, beginning from the zygote, is charted(SULSTON *et al.*. 1983). Based on the cell fate map, it becomes apparent that sometimes one daughter cell continues to proliferate while the other ceases to divide and undergoes terminal differentiation or death. There are then only 2 types of proliferative cell divisions, distinguishable by how they are graphed on the lineage tree: one type in which both daughter cells divide and the other where only one daughter cell continues to divide. If only the first of these 2 possibilities were to hold constant—that daughter cells constitutively divide—then there would only be one possible cell lineage tree, a symmetric one with each node bifurcating at every branch. However, the addition of the second possible type of cell division—in which one of the 2 daughter cells ceases to further divide—adds significant complexity to the repertoire of potential cell lineage trees and consequently to the different types of tissue and body plans that can be created during embryogenesis.

For any given number of n cells in an embryo there are a surprisingly large possible number ($(2n-3)!/2^{n-2}(n-2)!$) of potential cell lineage histories(SALIPANTE and HORWITZ 2007). For an embryo with 4 cells there are 15 different possible fate maps, for 8 cells there are 135,135, and for 16 cells the number exceeds 10^{15}. For the thousand or so cells of the adult worm(SULSTON *et al.*. 1983), the number of potential different lineage histories is, by all practical measures, infinite. Yet, all individual

worms invariantly develop identically; the cell fate map remains constant from one *C. elegans* specimen to the next(SULSTON *et al.*. 1983).

For many animals, however, including mice and other mammals, there does not exist a single, defined fate map in which the same developmental plan is followed by all individuals of that species. Instead, development is partly stochastic(STERNBERG and FELIX 1997). In contrast to *C. elegans*, any given cell from an early embryo is totipotent and can adopt any of a number of different possible cell fates. Commitment to any particular lineage is probabilistic (as reviewed (SALIPANTE and HORWITZ 2007)). A striking illustration of the variable development occurring between even genetically identical individuals of the same species is evident in cloned animals, where size, blood cell indices and serum markers, skin type, hair growth patterns, blood vessel branching and even the number of teats all show considerable heterogeneity~~, even among constitutionally genetically identical individuals~~(ARCHER *et al.*. 2003). Similar examples include variable heart valve morphology(SANS-COMA V 2012), craniofacial structure(BILLINGTON *et al.*. 2011), and numbers of neurons(AIREY *et al.*. 2006; WILLIAMS *et al.*. 1996) and cortical brain patterning(AIREY *et al.*. 2006) among isogenic strains of rodents. These studies indicate that while genetic background and environment contribute to variation, at least some differences are not genetically determined but are rather inescapable consequences of developmental noise.

Here we attempt to measure the extent to which random versus deterministic factors shape development. We employ an approach that we have dubbed "phylogenetic fate mapping", previously developed by our group(CARLSON *et al.*. 2011; SALIPANTE and HORWITZ 2006; SALIPANTE *et al.*. 2010; SALIPANTE *et al.*. 2008) and others(FRUMKIN *et al.*. 2008; FRUMKIN *et al.*. 2005; NAVIN *et al.*. 2011; SHIBATA and TAVARE 2007; WASSERSTROM *et al.*. 2008a; WASSERSTROM *et al.*. 2008b), in which cell lineage histories are inferred from somatic mutations. We have dissected single cells from corresponding tissues of 2 mouse littermates, expanded the cells clonally *ex vivo*, cataloged length-altering mutations at dozens of polyguanine (polyG) repeat mutational hotspots dispersed throughout the genome, and determined the order in which mutations have arisen, toward the goal of reconstructing cellular lineages. For the purpose of maximally extracting somatic genetic information, we have additionally introduced a technical refinement in which studies are conducted in DNA repair-deficient hypermutable mouse strains and have also evaluated new methods of inferring cellular ancestry based on genetic distance, in addition to those based on phylogenetics.

Results

Mutation profiles of single cells

We have previously carried out phylogenetic fate mapping studies utilizing the developmentally normal "Immortomouse" strain, which expresses a conditional SV40 T-antigen oncogene and conveniently allows for derivation of conditionally-immortalized cell lines (JAT *et al.*. 1991; SALIPANTE *et al.*. 2010) from clonally

expanded single cells. To obtain larger numbers of informative mutations, here we have taken the additional step of breeding the Immortomouse into hypermutable strains, deficient both in the lagging-strand DNA polymerase delta proof-reading (ALBERTSON *et al.*. 2009; GOLDSBY *et al.*. 2002) and MLH1 DNA mismatch repair (EDELMANN *et al.*. 1996) activities.

We successfully isolated and cultured as conditionally immortalized clonal cell lines about 100 single cells dissected from various tissues at similar locations from two adolescent (5 week) female mouse littermates (here identified as "mouse 1" and "mouse 2"). We harvested cells representing vascular endothelial tissue from the brain, preadipocytes from abdominal fat, and fibroblasts from hindlimb muscles (Supplemental Table 1). In addition to mutations developing somatically during the lifetime of the mouse, mutations can also arise during *ex vivo* clonal expansion; however, they are expected to randomly populate only a few cells in each clone and because they are unique to each isolate are unlikely to confound inferences of lineage. We therefore assume that the most frequent alleles in a clone represent genotypes of the original single cell from which the clone is derived (SALIPANTE *et al.*. 2010; SALIPANTE *et al.*. 2008)_ENREF_16. As an additional measure to control for mutations arising during *ex vivo* clonal expansion, for several isolates, we split each clone after just a few passages into two separate cultures and independently genotyped and analyzed each member of the pair to insure that separately they produced equivalent results (*vide infra*).

To ascertain somatically-acquired mutations in each of the single cell clones, we genotyped an average of 110 polyG loci per clone and identified somatic mutations that either shortened or lengthened the polyG tract (genotyping data for mouse 1 and 2 in Supplemental Tables 2 and 3 (respectively), Microsoft Excel files). Combining data from all cells harvested, each mouse exhibits an average of 0.5 mutant alleles/polyG locus/cell (Fig. 1a), which is more than one hundred-fold greater than we previously observed (0.003 mutations/locus/cell) using mice with intact DNA repair machinery(SALIPANTE *et al.*. 2010). On average, for each cell (Fig. 1b), more than one third of the 110 polyG markers (mouse 1: 36.7%, mouse 2: 34.4%) exhibited a somatic mutation. It is worth noting that the SV40 T-antigen originates from a strain (mixture of CBA/Ca and C57BL/10) different from the one (C57BL/6J) than it is crossed into and that contains the MLH1 and DNA polymerase delta deficient alleles. Littermates therefore carry differing amounts of strain-specific DNA from each parent, most likely including at loci encoding other DNA fidelity factors as well as polyG markers. The similarity in mutation profiles between the two individuals suggests that the genetic effects induced by the deficiency in polymerase proof-reading domain and mismatch repair genes are unlikely to be influenced by differences between mouse strains.

We next experimentally assayed the mutation frequency at polyG loci. From each mouse we selected one muscle fibroblast and one preadipocyte cell line and isolated

12 single cells that were each passaged for a defined number (20) of doublings. For each of the 48 subclones, we genotyped 110 polyG loci and identified mutations that were not found in the parental cell line from which the subclones were derived. We calculate that mouse 1 muscle fibroblasts and preadipocytes exhibit equal mutation rates, with a mean of 0.010 mutations/division/polyG locus, while mouse 2 displays similar values (P=0.248), with an average of 0.012 and 0.013 mutations/division/locus for muscle fibroblasts and preadipocytes, respectively (Supplemental Table 4, with the genotyping data from which it is found in Supplemental Table 5, a Microsoft Excel file). These results do not indicate that mutation rates vary with cell type or between individuals and support the notion that mutations can be used to infer cell lineage histories in different tissues from different mice.

Quantifying mitotic history of tissues

Cells within the body all originate from the zygote. We approximated the genotype of the zygote as being the alleles of each locus occurring most frequently among the clones examined. Because mutations arise with regular frequency during mitosis, a measure of the genetic distance separating individual cells from the zygote is expected to be proportional to the number of mitoses those cells have undergone since conception. We calculated genetic distance for tissues based on the mean number of pairwise allelic differences for the polyG markers, adjusting for missing data (data for mouse 1 and 2 in Supplemental Tables 6 and 7 (respectively), Microsoft Excel files). Measuring this distance from the zygote for cells in each mouse suggests that

fibroblasts from hindlimb muscle and preadipocytes from abdominal fat have undergone a similar number of mitoses, yet it is significantly fewer than those of vascular endothelial cells derived from the brain (Fig. 1c). One potential explanation for this observation is that it simply takes fewer cell divisions from the point at which muscle and fat differentiation begins until their development is complete, compared to what is required for the formation of blood vessels in the brain. Alternatively, it is possible that these tissues all arise at a similar point during development, but that muscle and fat originate from a larger group of progenitor cells than vascular endothelium. In the latter scenario, endothelial cells of blood vessels would require relatively more cell divisions before committing to specified lineages in order to produce the large numbers of cells required during the tissue maturation process.

To distinguish between these two possibilities, we compared the pairwise genetic distance among single cell clones within each tissue as well as to the zygote. For muscle and fat the distance between cells within each tissue is greater than their distance to the zygote (Table 1). In contrast, vascular endothelial cells demonstrate that they are about as distant from each other as they are to the zygote. Since we isolated a similar number of cells from those tissues (Supplemental Table 1), we minimized possible bias introduced by unequal sampling. The pattern observed in muscle fibroblasts and preadipocytes may be interpreted as showing that during organogenesis, these cells form a population of mixed lineages bearing various genotypes, instead of from a few closely related progenitors. Following organogenesis,

mutations continue to accumulate in descendant cells derived from the mixed founder population, with the result that cells within an organ are more dissimilar to each other than they are to the zygote. Contrastingly, for brain vascular endothelial cells, organogenesis appears to initiate from a limited number of progenitors, and cells within the tissue appear to undergo a large number of cell divisions in order to fully commit to the specific lineage. In this case, the genetic distance of cells from the zygote is much greater and is comparable to the average distance of single cells within the same type of tissue.

Notably, in both mice, we observed that relationships are in general much closer for cells in the same type of tissue than they are for cells in different types of tissue (Table 1). An interpretation of this observation is that the fate of progenitor cells are specified early in embryogenesis and remain committed during the remainder of development. It appears that cell migration between different primordial tissues is rare; otherwise, genetic distances within tissues would be similar to those between different types of tissues.

This notion also applies when examining the relatedness of left-sided tissues to their right-sided counterparts (Table 1). Interestingly, we found that the distance between contralateral tissues of the same type is generally larger than it is for the distance between the same types of ipsilateral tissues; however, the genetic distance for contralateral tissues of the same type is still smaller than the average distance between

unrelated types of tissues. This finding suggests that establishment of left and right polarity takes place after specification of lineages to individual tissues, and, subsequently, cells largely develop constrained to either side.

Reconstruction of lineage relationships by distance-based methods

We next evaluated whether genetic distance information could be used to infer lineage relationships between tissues. We used two approaches (one based on the eBURST algorithm and another utilizing network analysis) for deriving clonal relationships between tissues and cells from genetic distance calculations.

We first adapted the eBURST algorithm(FEIL and ENRIGHT 2004), which was originally developed to display clonal relationships among bacterial populations for the purpose of investigating infectious disease epidemiology. An advantage of eBURST analysis is that it may more sensitively detect clonal relationships in cases where there is insufficient genetic information to establish phylogeny. However, the algorithm is designed to interpret genotypes arising in haploid genomes. An additional limitation is its inability to analyze datasets as large as those generated in our study. To avoid these problems, our modified eBURST algorithm calculates relative genetic distances from pairwise comparisons of genotypes, connects isolates with related genotypes into groups and clonal complexes and identifies the founding genotype of each clonal complex. Analysis using the modified eBURST algorithm suggests that muscle fibroblasts and fat preadipocytes are clonally related (mouse 1 shown in Fig.

2a, mouse 2 in Supplemental Fig. 1), in agreement with the above findings indicating that muscle fibroblasts and preadipocytes share a common population of progenitor cells. Only under such circumstances, is it possible for descendants of closely related lineages to localize and develop in physically separated tissues. For most clones, modified eBURST analysis does not detect meaningful relationships between other cell types. Nevertheless, given the fact that we examined only a small proportion of the cells present in any tissue, we are largely limited to detecting relationships between cells that are only separated by a few cell divisions. (Based on assumptions described in the Methods section, we estimate that the modified eBURST algorithm is limited to detecting clonal relationships of cells separated by $\leq$12 mitoses.) Intriguingly, modified eBURST analysis revealed in both mice several connections between single cell clones derived from distant tissues (such as from contralateral tissues), suggesting that cell migration may occur during development, such that spatially separated cells share similar mutation profiles. Overall, however, most cells from spatially isolated tissues did not exhibit such a relationship, suggesting that cell migration appears to be rare during development, at least within the tissues we studied.

We then also examined for similarities among cells through use of network analysis (Fig. 2b), which offers a complementary approach for identifying ancestral relationships based on genetic distance(HUSON 1998). In mouse 1, muscle fibroblasts and preadipocytes are most genetically similar, consistent with the findings reported above. The same close relationship between fibroblasts and preadipocytes appears in

mouse 2, at least on the right side of the body; however, not all relationships in mouse 1 are preserved in mouse 2. To compare the overall similarity of tissue relationships between the two mice, we measured distances between the same pairs of tissues in both mice and calculated Pearson correlation coefficients (Fig. 2c, based on data in Supplemental Table 8, a Microsoft Excel file). This analysis demonstrates that the relatedness of different tissues to the zygote is largely the same in both mice (Pearson correlation coefficient=0.789, R^2=0.622, and P=0.0067), but the relatedness between any two different tissues in the pair of mice follows no discernible pattern (Supplemental Table 9). We reconcile these observations by proposing that in different individuals, tissues develop at similar times with similar sizes of progenitor cell populations, but that the lineage composition of those progenitor cells is randomly assigned. Although the overall coefficient index for all pairs of tissues demonstrates that tissue relationships between these individuals are far from perfectly correlated, it is nonetheless non-random; in other words, the overall pattern represented in two mouse littermates reflects a combination of deterministic and stochastic developmental events.

Phylogenetic reconstruction of lineage relationships

In order to more specifically infer lineage relationships among cells from each mouse, we used the genotypes of individual cells, and also collectively the mean genotype across tissues, to infer phylogenetic trees. We first computed genotypes for particular tissues based on the most frequent alleles found in all cells from the same type of

tissue. Phylogenetic reconstruction of the different tissues (Fig. 3a) demonstrates that, among all the types of tissue investigated in this study, vascular endothelial cells from the left and right sides of the brain share the most recent common progenitor and are therefore most closely related (as was found above in analyses based on mitotic distances). Fibroblasts from the left and right kidney are also closely related. Notably, these relationships are conserved in both individual littermates. Other tissues demonstrate more variable relationships, however. In comparing the two mice, for instance, kidney podocytes are closest to vascular endothelial cells in mouse 1 while they are most similar to preadipocytes in mouse 2. Despite the findings from the distance-based analysis, we failed to discern any relationship between preadipocytes and muscle fibroblasts using phylogenetic inference. This may be a consequence of using the zygote genotype as an outgroup in the phylogenetic reconstruction. It is possible that the similarity of genotypes of muscle, preadipocytes and zygote, demonstrated by distance-based clustering and network analyses, pose difficulties in resolving those groups from one another when employing phylogenetic analysis and, consequently, does not produce an informative tree structure.

When applying phylogenetic analysis to individual cells (as opposed to the composite genotype produced from cells of the same tissue type, as shown in Fig. 3a), the number of somatic mutations identified was insufficient to produce well-supported bifurcating trees through phylogenetic reconstruction (mouse 1 shown in Fig. 3b and mouse 2 in Supplemental Fig. 2); half of terminal branches cannot be fully resolved

and appear as polytomies. Employing even a low threshold of 50% Bayesian posterior probability yielded a tree in which all branches correspond to terminal bifurcations of pairs of cells, without revealing complex internal branching structures. Although this topology is limiting, there are nevertheless several noteworthy findings contained in the phylogeny. First, internal control clones that were split from the same parental clone in culture are largely paired together with high confidence (mouse 1: 16/18 paired with an average of 0.99 posterior probability; mouse 2: 26/28 paired with an average of 0.97 posterior probability), indicating neither that mutations occurring during *ex vivo* expansion nor that errors in determining marker genotypes are of sufficient magnitude to influence phylogenetic reconstructions. Second, pairs of single cell clones from different tissue origins occur frequently (mouse 1: 9/14; mouse 2: 8/11). Compared to pairs of phylogenetically related cells derived from the same tissue, pairs of phylogenetically related cells from dissimilar types of tissues exhibit longer branches connecting them to their most recent common progenitor, confirming observations from our earlier studies(SALIPANTE *et al.*. 2010). Reassuringly, phylogenetically related pairs of cells from different tissues also had higher degrees of genetic similarity in our distance-based analyses and similarly formed statistically significant connections in the modified eBURST and network analyses. Altogether, the paired patterns of single cell clones in the phylogenetic reconstruction are consistent with cell mixing and migration occurring during embryogenesis. Yet, cell mixing and migration appear restricted to certain developmental stages and/or certain types of tissue, because, by and large, cells develop in a constrained space that is

likely defined by interactions with neighboring cells and surrounding tissue architecture.

Patterns of cell growth inferred from the shape of the tree

The topology of a phylogenetic tree is shaped by the process through which it has grown(KIRKPATRICK and SLATKIN 1993; SHAO 1990). For example, if a lineage bifurcates, but only one of the subsequent two cell lines persists, then the shape of the tree will be asymmetric at that branch. For a tree produced from composite genotypes representing cells of the same tissue type (as in Fig. 3a), these properties translate to the probability that progenitor cells will give rise to distinct tissue types. We therefore examined the topology of phylogenetic reconstructions for nonrandom shapes. We first generated 5×10^4 random trees containing the same number (14) of branches as the experimentally-derived phylogenies depicted in Fig. 3a and measured their degree of asymmetry. The results are shown in the histogram in Fig. 3c, in which asymmetry is measured by the *N-bar* statistic (AGAPOW and PURVIS 2002). (We also measured asymmetry using a different statistic, Colless' imbalance statistic I_c (COLLESS 1982), which produced similar results, Supplemental Fig. 3.) Although the trees shown in Fig. 3a are symmetric, they correspond to a Bayesian estimate of the single best tree. To get a sense of the range of the shapes of trees that are compatible with the experimental data for mouse 1, we therefore collected the 5×10^4 best-fitting trees (of 2.5×10^5 total) produced by the phylogenetic analysis, measured their asymmetry, and superimposed the result on the values for the randomly generated trees (Fig. 3c, which

shows symmetry measured by *N-bar*, and Supplemental Fig. 3, which shows symmetry measured by I_c). Compared to randomly generated trees, possible phylogenies best fitting the experimental data are much more symmetric. We reject a trivial explanation that the symmetry arises from polytomies, where the tree cannot be resolved, because the posterior probabilities support the inferred structures. The most obvious biological explanation for a symmetric tree is that there is no variation in speciation and/or extinction rates for different branches of the tree. With respect to embryogenesis, this implies that distinct types of tissue, represented by individual clades in the phylogenetic tree, each have a similar probability of descending directly from the zygote, at the root of the tree. Overall, this observation suggests that a population of pluripotent cells in the early embryo contributes to different lineages without bias and that the determination of lineage commitment during development is itself a stochastic event.

Discussion

Comparison of tissue relationships in two sibling mice reveals details about how well overall patterns of development are conserved between different individuals. Results from our distance-based analysis point to a stochastic model of development, in which progenitors of different tissues and their exact genetic composition are randomly determined. Additionally, the highly symmetric shape of reconstructed cell lineage trees in these mice, generated by phylogenetic inference using mutations accumulating in single cells, similarly supports the apparently stochastic nature of

lineage differentiation occurring during embryogenesis.

Ever since Waddington first proposed a probabilistic model for how gene regulation modulates development in 1957(WADDINGTON 1957), stochastic contributions to cell fate determination have been repeatedly demonstrated in studies employing various linage tracing techniques, including dye injection (LAWSON *et al.*. 1991), retroviral marking (SORIANO and JAENISCH 1986), and chimeras formed from embryonic stem (ES) cells obtained from mixtures of differently pigmented mouse strains (SABURI *et al.*. 1997). For example, with respect to the latter, sibling littermates exhibit variable patterns of pigmentation, indicating that, at least in skin, mature tissues are randomly derived from primordial progenitors. Yet, the simple fact that most mice (and other individuals within a species) are patterned more-or-less the same suggests that there are limits to stochastic effects occurring during differentiation. A goal of our study was to determine where and when such restrictions might occur.

Developmental stochasticity has been mathematically modeled and experimentally demonstrated to be an inescapable consequence of gene transcription (KALMAR *et al.*. 2009; SWAIN *et al.*. 2002), epigenetic gene regulation(FEINBERG and IRIZARRY 2010) and protein interaction(BATADA *et al.*. 2004). Ultimately, these processes reflect the inherent noise in the networks into which genes and their products assemble, as governed by statistical and quantum mechanics(HUANG *et al.*. 2007; RAO *et al.*. 2002; WANG *et al.*. 2011). However, this is not to say that development is solely a random

process, as our data also indicate that during lineage specification, the timing and numbers of progenitor cell populations appear to be conserved between individuals.

An immediate question is how and why certain developmental events occur predictably while others appear to be random. Although our study does not provide direct clues, it is reasonable to speculate that such a balance between stochasticity and determinism is an evolutionary consequence that defines one species and distinguishes it from another but that at the same time allows for beneficial diversity within a species, promoting survival of at least some individuals in the face of a continually changing environment. This interpretation is somewhat analogous to the concept of genetic "buffering," in which populations may tolerate otherwise deleterious mutations in genes in order to maintain higher genetic diversity and thereby expedite the rate of adaption(DEPRISTO *et al.*. 2005). Overall, our study offers genetic evidence to separate variable developmental events from conserved ones, and therefore delineates a model in which development represents the sum of what can be efficiently specified in the genome balanced against the effort required to control entropic noise intrinsic to the underlying biochemistry.

One of the most significant events during development is gastrulation, when the single-layered blastula reorganizes into the three classic germ layers, which subsequently give rise to specialized cell types. Given that muscle, fibroblasts and fat share a common mesodermal origin, significant effort has focused on deciphering

genetic mechanisms determining lineage commitment of progenitor cells to one cell type or the other (HU *et al.*. 1995; SEALE *et al.*. 2008). However, the relative timing of lineage determination and the ultimate source of progenitors of muscle and fat are still unknown. In this study, by inferring from similarity in somatic mutations in individual single cell clones isolated from various tissues, we show that muscle fibroblasts and preadipocytes are similar in genetic composition and therefore separate into discrete lineages at a similar time during development. Our data suggest that both of these tissues may descend from a pool of progenitor cells with mixed lineages, instead of from a single or a few progenitor cells with similar mitotic histories. This notion resonates with recent discoveries of postnatal mesenchymal stem cells (MSCs), a type of cell that holds the potential to differentiate into multiple lineages in muscle, fat and bone tissues, and which have been located as nonhematopoietic cells in bone marrow (CAPLAN 1991; KUZNETSOV *et al.*. 2001; PITTENGER *et al.*. 1999), pericytes encircling capillaries and microvessels (CRISAN *et al.*. 2008), adipose tissue(ZUK *et al.*. 2002) and indeed from almost every postnatal connective tissue (BIANCO *et al.*. 2008). Given such a diversity of postnatal MSCs in various anatomical locations, it is reasonable to speculate that they could be derived from precursors with different genetic composition. We therefore propose a developmental model in which at the early three germ layer stage, there might be a large pool of progenitor cells within mesoderm that possess multiple lineage differentiation potentials, yet that themselves arise from proliferative growth and can be distinguished from each other by the mitotic mutations they bear. Such a mixed pool of progenitor cells gives rise to

precursors that initiate formation of muscle, fat and other cell types. While most of these progenitor cells differentiate and contribute to tissue formation, a few of them might persist as multipotent cells in these tissues postnatally through continuous self-renewal, providing a stem cell source for regeneration.

Another finding pertains to the establishment of lateral compartmentalization during mouse development. We conclude that the formation of tissue primordia is followed by the very early establishment of the left and right sagittal compartments within various tissues. Subsequently, cells mainly develop in their left or right territory with restricted cell migration in-between. Among individuals, such a development scheme could vary in terms of where exactly progenitor cells come from; however, the overall timing of lineage determination and the size of the founder population are largely conserved. At later stages of development, some tissues (for example, muscle and fat, as studied in our case) arise from a mixed pool of progenitor cells in the germ layer, while other tissues (for instance, vascular endothelium in brain, also as we have shown here), are derived from a single or at least limited population of progenitor cells. The phenomenon that we describe may become manifest in human disorders caused by somatic mutations with restricted laterality. For example, Proteus syndrome has been recently found to result from somatic mutations arising during embryonic development in *AKT1*(LINDHURST *et al.*. 2011); a feature of Proteus syndrome can be hemihypertrophy(JOSHI *et al.*. 2005), in which there is overgrowth of multiple tissues in a mosaic pattern but affecting only one side of the body, either right or left, with

respect to the sagittal plane.

Our studies initiate an investigation into differentiating between conserved and variable features of mammalian development. A considerable amount of experimentally-derived molecular genetic information (based on several hundred thousand PCR reactions) was required to generate the mutational data required for analysis here. Yet, the conclusions that can be drawn from studies based on just 2 simultaneously studied individuals are necessarily limited. We look forward to technological advancements that will facilitate identification of mutations for the purposes of inferring cell lineage. Along those lines, we have recently demonstrated how deep sequencing holds promise in this regard(CARLSON *et al.*. 2011). As cell fate maps become available for greater numbers of cells at increasingly higher resolution, and from multiple specimens of the same species, it should become easier to distinguish genetically determined variation from effects attributable to uncontrollable and random events occurring during embryogenesis. Such information could prove particularly valuable in sorting out birth defects where, for some, *de novo* single gene and chromosomal mutations are increasingly recognized as causative, yet for others, older concepts relating to disruptions of developmental events (without necessarily invoking genetic factors) still hold sway: a case in point being the “Robin Sequence”, in which multiple genetic and idiopathic factors contribute to human mandibular birth defects(EVANS *et al.*. 2011).

Experimental Procedures

Mouse strains: *Pold1*$^{+/e}$*Mlh1*$^{+/\Delta}$ mice were obtained from B. Preston (University of Washington)(CARLSON *et al.*. 2011). The DNA polymerase delta gene *Pold1* retained an inactive exonuclease domain due to a single point mutation (D400A)(ALBERTSON *et al.*. 2009; GOLDSBY *et al.*. 2002), while the mismatch repair gene *Mlh1* was dysfunctional due to the deletion of exon 2(EDELMANN *et al.*. 1996). In order to obtain desired cell replication capability *in vitro*, we employed *H-2K*^{b}-tsA58 transgenic mice ("Immortomouse" strain), cells from which can be conditionally immortalized as driven by an interferon-inducible and temperature sensitive form of the simian virus 40 large tumor antigen gene(JAT *et al.*. 1991). Homozygous *H-2K*^{b}-tsA58 transgenic mice were separately bred to heterozygously deficient *Pold1*$^{+/e}$ *and Mlh1*$^{+/\Delta}$ mouse lines. The resulting lines were crossed to each other that were then mated amongst themselves to produce the mutant mice *Pold1*$^{+/e}$*Mlh1*$^{\Delta/\Delta}$*H-2K*^{b}-tsA58$^{+/-}$ for our study.

Cell isolation and culture: Kidney, abdominal fat tissue, muscles from the hindlimb and brain were dissected separately from two 5 week-old female *Pold1*$^{+/e}$*Mlh1*$^{\Delta/\Delta}$*H-2K*^{b}-tsA58$^{+/-}$ mice. Tissues were minced and cells were separated by digestion with 0.5mM EDTA, 15 U/ml papain (Roche) and 200ug/ml Dnase I (Roche). To release cells from brain tissue slurries, samples were passed through Potter-Elvehjem tissue grinders. Fat and muscle were subjected to vigorous pipetting. Kidney tissue was broken down by filtering tissues through a 40-mesh screen. Cells were seeded into 15 cm culture dishes at dilutions yielding well-separated single cells, and clones arising

from those single cells were further isolated using cloning cylinders followed by deposition into single wells. Cells were cultured in DMEM/F12 media (Gibco/Invitrogen) containing 20% fetal bovine serum (Gibco/Invitrogen), 200 ng/ml mouse interferon gamma (R&D Systems) and penicillin G (100 U/ml) plus streptomycin (100 µg/ml) at 33℃ with 5% CO_2 and 5% O_2 in a humidified incubator.

Mutational analysis: Clones were expanded to approximately 10^6 cells, and DNA was extracted using ArchivePure DNA Cell/Tissue Kit (5prime). 2 ng of DNA was used in each 5 µl PCR reaction consisting of 1 µM of oligonucleotide primers, 200 nM dNTPs, 0.05 U Taq DNA polymerase in 1× manufacturer-supplied buffer (Qiagen). For each pair of primers, the forward primer was fluorescently tagged while the reverse primer was tailed with 5'-GTTTCTT-3', as detailed in (SALIPANTE *et al.*. 2010). Primers used in the study are listed as in Supplemental Table 10. PCR products were diluted in 10 µl of Hi-Di Formamide (ABI/Life Technologies) with GeneScan 500 ROX Size Standard (ABI/Life Technologies) and subject to capillary electrophoresis on a 3730xl DNA Analyzer (ABI/Life Technologies). All reactions were carried out in 384-well plates and liquid handling was performed on a Matrix Platemate 2×3 Pipetting Workstation (Thermo Scientific).

Genotype interpretation: Results generated by the 3730xl DNA Analyzer were imported into GeneMaker 1.4 (Softgenetics) for automated fragment alignment and size calling. To minimize "stutter" artifacts from PCR amplification of repetitive

sequences, manual size calling was further performed on each locus to ensure accuracy. Specifically, homozygous or heterozygous alleles were defined based on three parameters: I_{1H}, I_{2H}, I_{3H}, corresponding to the fluorescent intensity (U) of the highest, second- and the third-highest signals, respectively. Homozygote genotypes were assigned when $\left|(I_{1H}-I_{2H})-(I_{1H}-I_{3H})\right| \leq 10^4$ U (e.g. 106/106); heterozygote genotypes were assigned when $\left|(I_{1H}-I_{2H})-(I_{1H}-I_{3H})\right| \geq 10^4$ U and I_{2H} (or I_{3H}) > $0.8I_{1H}$ (e.g. 106/105), while signals with patterns falling in-between were assigned ambiguously (marked as "X", e.g. 106/X). Alleles were further assigned as being derived from one parent or the other so as to minimize the number of mutations required to generate the observed genotypes. Genotypes of zygote and individual tissues were defined as the most frequent alleles of all single cell clones as a whole or that of single cell clones from corresponding tissue types, respectively.

Genetic distance calculation: Briefly, alleles of each pair of samples on each locus were compared and a distance was obtained by dividing the sum of minimal difference in length across all the loci by the number of loci examined. Loci that have more than one "X" (missing data, as defined above) in a pair of single cell clones were not considered in the calculation. For pairwise comparison of tissues, all pairwise distances of single cell clones within compared tissues were averaged, and the significance was calculated by Students' T-test against averaged distance of single cell clones of all tissues. The pairwise distances among single cell clones are further presented in a network. The most significant similarities among single cell clones are

emphasized, with grey lines connecting a pair of significantly related single cell clones. Each single cell clone is presented as a dot with different color indicative of its tissue origin while tissue layout is assigned so as to reflect its relative anatomical location on the anteroposterior axis. The diameter of circle for tissues correlates the averaged distance within tissues. Details of the algorithm are presented in Supplementary Methods. The analysis was performed using a computer program (Supplemental Methods) written in the Python programming language.

Modified eBURST clustering analysis: The eBURST algorithm has been employed to address clonal relationships of bacterial populations (BERES *et al.*. 2010; FEIL 2004; FEIL *et al.*. 2004; FEIL *et al.*. 2000). In our adaptation, an empirical threshold value was assigned, and only isolates having smaller distance were grouped clonally. The founding genotype was defined as the one that exhibited the smallest distances to the largest number of other members in the same group. In our modified eBURST algorithm, because markers were randomly selected from throughout the genome without respect to location within genes or other functional elements, mutations from different loci are weighed equally, and the relative distances of genotypes therefore represent the relatedness of the genotypes. A distance of 0.2 was used as the threshold, since this is equivalent to the distance of cells separated by 15 cell divisions, based on the observed mutation rate of 0.013 mutations/division/locus in the hypermutable mouse strain used in this study. (Distance value=mutation rate X number of cell

divisions X number of loci genotyped, in this case, 0.2=0.013X15X1) Our modified eBURST analysis was performed using a computer program (Supplemental Methods) written in the Python programming language.

Phylogenetic reconstruction: Phylogenetic trees of cells isolated from the two mice were constructed using Bayesian inference as implemented in MrBayes 3.1(HUELSENBECK *et al.*. 2001; RONQUIST and HUELSENBECK 2003). The standard data type was used and alleles on each locus were converted to a single digit from 0-9 according to their mutation patterns. A uniform distribution on the interval (0.05, 50) was used in the model of Gamma-shaped rate variation across sites, and the parameter of the symmetric Dirichlet distribution was fixed to infinity. The Metropolis-coupled Markov Chain Monte Carlo method (MCMC)(HASTINGS 1969; METROPOLIS and ULAM 1949) was used to approximate the posterior probabilities of trees. MCMC samples from the first 5-6$\times 10^7$ generations were discarded, and samples from subsequent 2-3$\times 10^6$ generations were included for tree reconstruction.

Measurement and statistical test of the shape of phylogenetic trees: 10^4 random trees with the specified number of branches were generated using COMPONENT (http://taxonomy.zoology.gla.ac.uk/rod/cpw.html). Two measures that summarize the shape of a phylogenetic tree, *N-bar*(KIRKPATRICK and SLATKIN 1993) and Colless' imbalance statistic I_c (COLLESS 1982), were calculated using the software package TreeStat (http://tree.bio.ed.ac.uk/software/treestat/). Distributions of *N-bar* or I_c values

of random trees were overlaid to the distribution of these values of reconstructed phylogenetic trees with high posterior probability using graphing functions in Microsoft Excel.

Figure Legends

Figure 1. Somatic polyG mutation profiles of two mouse littermates. a. Histogram showing how many different mutant alleles are identified for each marker across all of the approximately 100 cells genotyped for each mouse. b. Histogram showing the number of all polyG marker mutations detected per cell for each mouse. c. Average genetic distance of different types of tissues to the zygote for each mouse.

Figure 2. Lineage relationships inferred from methods based on genetic distance. a. Modified eBURST analysis, showing "population snapshot" of single cell clones in Mouse 1. Clusters of related single cell clones and individual unlinked clones are displayed as a single modified eBURST diagram by using the distance value (D) of 0.2 as cut-off. Clusters of linked single cell clones correspond to complexes that share highly similar mutational profiles. Each single cell clone is represented as a dot with color indicative of its tissue origin. (Mouse 2 shown in Supplemental Fig. 1.) b. Network representation depicting mutational similarities among single cell clones between both mice. Significant similarities between single cell clones for mouse 1 are shown with grey connecting lines. Each single cell clone is depicted as a dot with different colors indicative of tissue origin while the layout on the graph reflects relative anatomical location on the anteroposterior axis. The diameter of the circles correlates with the average distance within tissues. Orange lines show relationships that are conserved in mouse 2. c. Scatter plot of distance between equivalent pairs of tissue, comparing mouse 1 to mouse 2. Distances of specific tissues to the zygote are colored orange; a trend line indicates their correlation. Among these comparisons, the

distances between individual tissues to the zygote are largely conserved between the two mice.

Figure 3. Phylogenetic reconstruction of tissues and single cell clones in both mice. a. Phylogenetic tree of tissues, with mouse 1 in black and mouse 2 in orange, overlaid. Numbers indicate Bayesian posterior probabilities. b. Phylogenetic tree of single cell clones in mouse 1. Only branch structures with larger than 50% posterior probability are shown. Pairs of single cell clones from the same parental cell are marked with asterisk. (Mouse 2 shown in Supplemental Fig. 2.) c. Distribution of *N-bar* symmetry statistic for mouse 1 tissue trees with highest posterior probabilities compared to random trees with the same number of branches, showing a symmetric nature of the actual tissue trees. Examples of symmetric (left) and asymmetric trees (right) are shown. (Mouse 2 shown in Supplemental Fig. 3.)

Supplemental Material

Supplemental Figure 1. Modified eBURST analysis, showing "population snapshot" of single cell clones in Mouse 2.

Supplemental Figure 2. Phylogenetic tree of single cell clones in mouse 2.

Supplemental Figure 3. Distribution of I_c symmetry statistic for mouse 1 tissue trees with highest posterior probabilities compared to random trees.

Supplemental Table 1. Sources and numbers of cells isolated from each mouse.

Supplemental Table 2. Genotype data for mouse 1, Microsoft Excel file.

Supplemental Table 3. Genotype data for mouse 2, Microsoft Excel file.

Supplemental Table 4. Mutation frequency among clonal isolates *in vitro*.

Supplemental Table 5. Genotype data supporting Supplemental Table 4.

Supplemental Table 6. Genetic distance data for mouse 1, Microsoft Excel file.

Supplemental Table 7. Genetic distance data for mouse 2, Microsoft Excel file.

Supplemental Table 8. Pairwise genetic distance comparisons between mouse 1 and 2.

Supplemental Table 9. Statistical analysis supporting tissue correlations in Supplemental Table 8.

Supplemental Table 10. PCR primers for markers.

Supplemental Methods. Algorithms for genetic distance calculations.

Supplemental Software – Program for calculation of genetic distance.

Supplemental Software – Modified eBURST program.

Supplemental Table 3: Pearson correlation coefficient values for different kinds of

comparison among tissues.

For detailed Supplemental figures and additional information, please refer to the publication: Zhou W, Tan Y, Crist EM, Anderson DJ, Ruohola-Baker H, Salipante SJ and Horwitz MS. Use of Somatic Mutations to Quantify Random Contributions to Mouse Development. Submitted. BMC Genomics.

References

AGAPOW, P. M., and A. PURVIS, 2002 Power of eight tree shape statistics to detect nonrandom diversification: a comparison by simulation of two models of cladogenesis. Syst Biol **51:** 866-872.

AIREY, D. C., F. WU, M. GUAN and C. E. COLLINS, 2006 Geometric morphometrics defines shape differences in the cortical area map of C57BL/6J and DBA/2J inbred mice. BMC Neurosci **7:** 63.

ALBERTSON, T. M., M. OGAWA, J. M. BUGNI, L. E. HAYS, Y. CHEN *et al.*., 2009 DNA polymerase {varepsilon} and {delta} proofreading suppress discrete mutator and cancer phenotypes in mice. Proc Natl Acad Sci U S A.

ARCHER, G. S., S. DINDOT, T. H. FRIEND, S. WALKER, G. ZAUNBRECHER *et al.*., 2003 Hierarchical phenotypic and epigenetic variation in cloned swine. Biol Reprod **69:** 430-436.

BATADA, N. N., L. A. SHEPP and D. O. SIEGMUND, 2004 Stochastic model of protein-protein interaction: why signaling proteins need to be colocalized. Proc Natl Acad Sci U S A **101:** 6445-6449.

BERES, S. B., R. K. CARROLL, P. R. SHEA, I. SITKIEWICZ, J. C. MARTINEZ-GUTIERREZ *et al.*., 2010 Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. Proc Natl Acad Sci U S A **107:** 4371-4376.

BIANCO, P., P. G. ROBEY and P. J. SIMMONS, 2008 Mesenchymal stem cells: revisiting history, concepts, and assays. Cell Stem Cell **2:** 313-319.

BILLINGTON, C. J., JR., B. NG, C. FORSMAN, B. SCHMIDT, A. BAGCHI *et al.*., 2011 The molecular and cellular basis of variable craniofacial phenotypes and their genetic rescue in Twisted gastrulation mutant mice. Dev Biol **355:** 21-31.

CAPLAN, A. I., 1991 Mesenchymal stem cells. J Orthop Res **9:** 641-650.

CARLSON, C. A., A. KAS, R. KIRKWOOD, L. E. HAYS, B. D. PRESTON *et al.*., 2011 Decoding cell lineage from acquired mutations using arbitrary deep sequencing. Nat Methods.

COLLESS, D. H., 1982 Review of phylogenetics: The theory and practice of phylogenetic systematics. Systematic Zoology **31:** 100-104.

CRISAN, M., S. YAP, L. CASTEILLA, C. W. CHEN, M. CORSELLI *et al.*., 2008 A perivascular origin for mesenchymal stem cells in multiple human organs. Cell Stem Cell **3:** 301-313.

DEPRISTO, M. A., D. M. WEINREICH and D. L. HARTL, 2005 Missense meanderings in sequence space: a biophysical view of protein evolution. Nat Rev Genet **6:** 678-687.

EDELMANN, W., P. E. COHEN, M. KANE, K. LAU, B. MORROW *et al.*., 1996 Meiotic pachytene arrest in MLH1-deficient mice. Cell **85:** 1125-1134.

EVANS, K. N., K. C. SIE, R. A. HOPPER, R. P. GLASS, A. V. HING *et al.*., 2011 Robin sequence: from diagnosis to development of an effective management plan. Pediatrics **127:** 936-948.

FEIL, E. J., 2004 Small change: keeping pace with microevolution. Nat Rev Microbiol **2:** 483-495.

FEIL, E. J., and M. C. ENRIGHT, 2004 Analyses of clonality and the evolution of bacterial pathogens. Curr Opin Microbiol **7:** 308-313.

FEIL, E. J., B. C. LI, D. M. AANENSEN, W. P. HANAGE and B. G. SPRATT, 2004 eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol **186:** 1518-1530.
FEIL, E. J., J. M. SMITH, M. C. ENRIGHT and B. G. SPRATT, 2000 Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data. Genetics **154:** 1439-1450.

FEINBERG, A. P., and R. A. IRIZARRY, 2010 Evolution in health and medicine Sackler colloquium: Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proc Natl Acad Sci U S A **107 Suppl 1:** 1757-1764.

FRUMKIN, D., A. WASSERSTROM, S. ITZKOVITZ, T. STERN, A. HARMELIN *et al.*., 2008 Cell lineage analysis of a mouse tumor. Cancer Res **68:** 5924-5931.

FRUMKIN, D., A. WASSERSTROM, S. KAPLAN, U. FEIGE and E. SHAPIRO, 2005 Genomic variability within an organism exposes its cell lineage tree. PLoS Comput Biol **1:** e50.

GOLDSBY, R. E., L. E. HAYS, X. CHEN, E. A. OLMSTED, W. B. SLAYTON *et al.*., 2002 High incidence of epithelial cancers in mice deficient for DNA polymerase delta proofreading. Proc Natl Acad Sci U S A **99:** 15560-15565.

HASTINGS, W. K., 1969 Monte Carlo sampling methods using Markov chains and

their applications. Biometrika **57:** 97-109.

HU, E., P. TONTONOZ and B. M. SPIEGELMAN, 1995 Transdifferentiation of myoblasts by the adipogenic transcription factors PPAR gamma and C/EBP alpha. Proc Natl Acad Sci U S A **92:** 9856-9860.

HUANG, S., Y. P. GUO, G. MAY and T. ENVER, 2007 Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. Dev Biol **305:** 695-713.

HUELSENBECK, J. P., F. RONQUIST, R. NIELSEN and J. P. BOLLBACK, 2001 Bayesian inference of phylogeny and its impact on evolutionary biology. Science **294:** 2310-2314.

HUSON, D. H., 1998 SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics **14:** 68-73.

JAT, P. S., M. D. NOBLE, P. ATALIOTIS, Y. TANAKA, N. YANNOUTSOS *et al.*., 1991 Direct derivation of conditionally immortal cell lines from an H-2Kb-tsA58 transgenic mouse. Proc Natl Acad Sci U S A **88:** 5096-5100.

JOSHI, U., J. A. VAN DER SLUIJS, G. J. TEULE and R. PIJPERS, 2005 Proteus syndrome: a rare cause of hemihypertrophy and macrodactyly on bone scanning. Clin Nucl Med **30:** 604-605.

KALMAR, T., C. LIM, P. HAYWARD, S. MUNOZ-DESCALZO, J. NICHOLS *et al.*., 2009 Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biol **7:** e1000149.

KIRKPATRICK, M., and M. SLATKIN, 1993 Searching for evolutionary patterns in the shape of a phylogenetic tree. Evolution **47:** 1171-1181.

KUZNETSOV, S. A., M. H. MANKANI, S. GRONTHOS, K. SATOMURA, P. BIANCO *et al.*., 2001 Circulating skeletal stem cells. J Cell Biol **153:** 1133-1140.

LAWSON, K. A., J. J. MENESES and R. A. PEDERSEN, 1991 Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. Development **113:** 891-911.

LINDHURST, M. J., J. C. SAPP, J. K. TEER, J. J. JOHNSTON, E. M. FINN *et al.*., 2011 A mosaic activating mutation in AKT1 associated with the Proteus syndrome. N Engl J Med **365:** 611-619.

MAC AULEY, A., Z. WERB and P. E. MIRKES, 1993 Characterization of the unusually rapid cell cycles during rat gastrulation. Development **117:** 873-883.

METROPOLIS, N., and S. ULAM, 1949 The Monte Carlo method. J Am Stat Assoc **44:** 335-341.

NAVIN, N., J. KENDALL, J. TROGE, P. ANDREWS, L. RODGERS *et al.*., 2011 Tumour evolution inferred by single-cell sequencing. Nature **472:** 90-94.

PITTENGER, M. F., A. M. MACKAY, S. C. BECK, R. K. JAISWAL, R. DOUGLAS *et al.*., 1999 Multilineage potential of adult human mesenchymal stem cells. Science **284:** 143-147.

RAO, C. V., D. M. WOLF and A. P. ARKIN, 2002 Control, exploitation and tolerance of intracellular noise. Nature **420:** 231-237.

RONQUIST, F., and J. P. HUELSENBECK, 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:** 1572-1574.

SABURI, S., S. AZUMA, E. SATO, Y. TOYODA and C. TACHI, 1997 Developmental fate of single embryonic stem cells microinjected into 8-cell-stage mouse embryos. Differentiation **62:** 1-11.

SALIPANTE, S. J., and M. S. HORWITZ, 2006 Phylogenetic fate mapping. Proc Natl Acad Sci U S A **103:** 5448-5453.

SALIPANTE, S. J., and M. S. HORWITZ, 2007 A phylogenetic approach to mapping cell fate. Curr Top Dev Biol **79:** 157-184.

SALIPANTE, S. J., A. KAS, E. MCMONAGLE and M. S. HORWITZ, 2010 Phylogenetic analysis of developmental and postnatal mouse cell lineages. Evol Dev **12:** 84-94.

SALIPANTE, S. J., J. M. THOMPSON and M. S. HORWITZ, 2008 Phylogenetic fate mapping: theoretical and experimental studies applied to the development of mouse fibroblasts. Genetics **178:** 967-977.

SANS-COMA V, F. M., FERNANDEZ B, DURAN AC, ANDERSON RH, ARQUE JM, 2012 Genetically alike Syrian hamsters display both bifoliate and trifoliate aortic valves. J Anat **Epub ahead of print**.

SEALE, P., B. BJORK, W. YANG, S. KAJIMURA, S. CHIN *et al.*., 2008 PRDM16 controls a brown fat/skeletal muscle switch. Nature **454:** 961-967.

SHAO, K. T., 1990 Tree Balance. Systematic Biology **39:** 266-276.

SHIBATA, D., and S. TAVARE, 2007 Stem cell chronicles: autobiographies within genomes. Stem Cell Rev **3:** 94-103.

SILVER, L. M., 1995 *Mouse Genetics. Concepts and Applications*. Oxford University Press, Oxford.

SORIANO, P., and R. JAENISCH, 1986 Retroviruses as probes for mammalian development: allocation of cells to the somatic and germ cell lineages. Cell **46:** 19-29.
STERNBERG, P. W., and M. A. FELIX, 1997 Evolution of cell lineage. Curr Opin Genet Dev **7:** 543-550.

SULSTON, J. E., E. SCHIERENBERG, J. G. WHITE and J. N. THOMSON, 1983 The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev Biol **100:** 64-119.

SWAIN, P. S., M. B. ELOWITZ and E. D. SIGGIA, 2002 Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci U S A **99:** 12795-12800.

WADDINGTON, C. H., 1957 *The Strategy of the Genes*. Allen & Unwin, London.

WANG, J., K. ZHANG, L. XU and E. WANG, 2011 Quantifying the Waddington landscape and biological paths for development and differentiation. Proc Natl Acad Sci U S A **108:** 8257-8262.

WASSERSTROM, A., R. ADAR, G. SHEFER, D. FRUMKIN, S. ITZKOVITZ *et al.*., 2008a Reconstruction of cell lineage trees in mice. PLoS One **3:** e1939.

WASSERSTROM, A., D. FRUMKIN, R. ADAR, S. ITZKOVITZ, T. STERN *et al.*., 2008b Estimating cell depth from somatic mutations. PLoS Comput Biol **4:** e1000058.

WILLIAMS, R. W., R. C. STROM, D. S. RICE and D. GOLDOWITZ, 1996 Genetic and environmental control of variation in retinal ganglion cell number in mice. J Neurosci **16:** 7193-7205.

WOLFF, D. A., and H. PERTOFT, 1972 Separation of HeLa cells by colloidal silica density gradient centrifugation. I. Separation and partial synchrony of mitotic cells. J Cell Biol **55:** 579-585.

ZUK, P. A., M. ZHU, P. ASHJIAN, D. A. DE UGARTE, J. I. HUANG *et al.*., 2002 Human adipose tissue is a source of multipotent stem cells. Mol Biol Cell **13:** 4279-4295.

Appendix I Assessment of Hypoxia Inducible Factor levels in cancer cell lines upon hypoxic induction using a novel reporter construct

Abstract

Hypoxia Inducible Factor (HIF) signaling pathway is important for tumor cells with limited oxygen supplies, as it is shown to be involved in the process of proliferation and angiogenesis. Given its pivotal role in cancer biology, robust assays for tracking changes in HIF expression are necessary for understanding its regulation in cancer as well as developing therapies that target HIF signaling. Here we report a novel HIF reporter construct containing tandem repeats of minimum HIF binding sites upstream of eYFP coding sequence. We show that the reporter construct has an excellent signal to background ratio and the reporter activity is HIF dependent and directly correlates with HIF protein levels. By utilizing this new construct, we assayed HIF activity levels in different cancer cell lines cultured in various degrees of hypoxia. This analysis reveals a surprising cancer cell line specific variation of HIF activity in the same level of hypoxia. We further show that in two cervical cancer cell lines, ME180 and HeLa, the different HIF activity levels observed correlate with the levels of hsp90, a cofactor that protects HIF against VHL-independent degradation. This novel HIF reporter construct serves as a tool to rapidly define HIF activity levels and therefore the therapeutic capacity of potential HIF repressors in individual cancers.

Introduction

Hypoxia is well known to fundamentally regulate many aspects of cell biology. Most of the effects of hypoxia involve the hypoxia inducible factor (HIF), a highly conserved and crucial oxygen regulated heterodimeric transcription factor composed of an alpha (α) and a beta (β) subunit. Both of these subunits belong to the PER-ARNT-SIM (PAS) group in basic-helix-loop-helix (bHLH) family of transcription factors [1]. Two genes encoding mammalian HIFα subunits (HIF1α, HIF2α) are well studied: HIF1α is ubiquitously expressed whereas HIF2α exhibit more restricted tissue distribution [2]. In normoxia, HIFα undergoes prolyl hydroxylation and binds to an ubiquitin E3-ligase, the von Hippel-Lindau (VHL) protein, which leads to polyubiquitination and rapid proteosomal degradation of HIFα [3,4]. Under hypoxia, HIFα hydroxylation is inhibited, resulting in accumulation of HIFα and formation of HIFα-HIFβ heterodimers. The heterodimers further complex with the coactivator p300, and bind to the promoters of HIF target genes to induce gene expression [5]. In addition to hypoxia, many other pathways can affect HIF stabilization [6,7]. Cofactors, such as PACF P300/CBP associated factor [8] and hsp90 [9,10], also help facilitate HIF stabilization and enhance HIF activities. Hsp90 has shown to protect HIFα against VHL-independent degradation that can occur in hypoxia [11].

The well studied HIF target genes include those that are involved in oxygen delivery and cell proliferation, such as the vascular endothelial growth factor (*VEGF*) [12,13] and *p21* [14]. In addition, HIF facilitates adaptation to oxygen deprivation by regulating genes involved in glucose uptake and metabolism, such as carbonic

anhydrase (*CA9*) [15], which maintains cellular pH_i homeostasis under hypoxia. It was recently reported that HIF and its target gene, *Oct4*, are responsible for hypoxia induced cancer stem cell phenotype that is thought to drive the progression and aggressiveness in certain tumors [16]. Given its pivotal role in angiogenesis and tumor progression, HIF is a therapeutically attractive target and blocking HIF, especially when combined with conventional therapies, has shown beneficial effects [17,18].

To examine HIFs' temporal and spatial expression in tissues, several direct and indirect reporter systems are developed in order to track HIF protein expression *in vivo*. Either full length HIF cDNA, or a fragment under oxygen-dependent regulation has been linked to fluorescent protein [19], or firefly luciferase [20] for constructing HIF-fusion proteins. Alternatively, since HIF fusion protein studies do not reveal whether HIF complex is transcriptionally active, promoter based reporters have also been developed. Typically, 5-8 repeats of the hypoxia response elements (HREs) (5'-GCCCTACGTGCTGTCTCACACAGC-3') from the 3' enhancer region of human Epo gene, or the HRE from VEGF (5′-CACAGTGCATACGTGGGCTCCAACAGGTCCTCT-3′) are linked in tandem with a minimal promoter to drive the expression of a downstream reporter gene [21,22]. It is worth noting that in these constructs, HRE contains not only the HIF-1α or HIF2α consensus binding sites (5'-CGTG-3' and 5'-TRCGTG-3', respectively), but also *Epo* or *VEGF* promoter specific sequences. Recently *Oct4* has shown to be induced by

HIF under hypoxia [16], however, *Oct4* promoter only contains three repeats of CGTG, the actual HIF binding site but not the HRE sequences observed either in *Epo* or *VEGF* promoter.

In order to maximize the specificity and sensitivity of the reporter construct, a strategy of using the most primitive transcription factor binding site in tandem in a reporter has been successfully utilized previously in the case of Wnt-pathway analysis [23,24]. In the present study, we utilized a similar strategy to build up a promoter based reporter, only incorporating the minimal HIF1α and HIF2α binding sites together (CGTGTACGTG) in tandem in the promoter. We show that this new HIF reporter with HIF binding repeats (HBR) has a good signal to background ratio and signal dynamics in deoxygenation and reoxygenation. We also demonstrate that the signal is HIF dependent as revealed by HIF RNAi studies, and it correlates with the cellular HIF protein level in different cell lines. By utilizing our new construct, we show that HIF activity levels vary significantly in different cancer cell lines cultured in the same degree of hypoxia. We further reveal that in two cervical cancer cell lines, the differences in HIF activity levels correlate with the level of the HIF cofactor, Hsp90, which protects HIF against VHL-independent degradation.

Results

We constructed a lentiviral plasmid, in which the expression of enhanced yellow florescent protein (eYFP) was under the regulation of twelve tandem repeats of minimal HIF-binding sites (CGTGTACGTG), followed by a minimal human

thymidine kinase (TK) promoter (12U-HBR). Also, a 770bps of β-globin intron sequences was incorporated between the TK promoter and eYFP for better transcription of eYFP (Figure 1a, Figure S1). To assay the construct's function *in vitro*, we transduced the virus into HeLa cells and cultured them in either normoxic (20% O_2) or hypoxic (2% O_2) condition. After 24 hours, we observed that cells under hypoxia uniformly expressed eYFP (~70% of eYFP positive), while those under normoxia only gained fluorescence with greatly reduced intensity (~6% of eYFP positive, Figure 1b and Figure S2).

To further optimize the signal to background ratio, we reduced the number of HIF-binding sites, thereby creating HIF-reporters with 3 tandem binding sites (3U-HBR) or with 6 sites (6U-HBR) (Figure 1a). When analyzing in HeLa cells, we observed that 6U-HBR construct achieved a better signal-to-background ratio than 3U-HBR and 12U-HBR constructs (33.4, 5.8 and 15.1, respectively. Figure 1c). To compare the specificity and sensitivity to other hypoxia reporters, we specifically obtained one of the widely used HIF reporter 5HRE_hCMV_Luc [25]. After transient transfected 5HRE_hCMV_luc reporter in HeLa cells, we observed 50-100 fold greater luciferase activity under 2% O_2 as compared to cells growing in 20% O_2, which was consistent with the original finding. In comparison, our HBR_eYFP reporter shows about 30-40 fold increase in 2% O_2, an increase that falls in a similar scale as seen with 5HRE_hCMV_luc reporter. With the similarity in signal intensity, however, the two HIF reporter constructs are not directly comparable, given that they have distinct

backbone sequences, different minimal promoters and reporter genes, all of which could render them distinct behavior and dynamics under hypoxia. In this study, 6U-HBR was used for further optimization and characterization.

We next characterized the dynamics of the construct in the process of deoxygenation and reoxygenation. When transferred from normoxia to hypoxia (2%), 6U-HBR-HeLa cells had rapid response (Figure 2a); however, when transferred from hypoxia back to normoxia, the fluorescent intensity decreased slowly (72 hours to reach the basal level; Figure 2b). For a construct with better ability to timely reflect the turnover of HIF, we modified eYFP and generated a destabilized version by attaching mouse ornithine decarboxylase 422-461 domain, a region responsible for rapid degradation of the protein [26], to the C-terminus of eYFP, therefore creating the construct of 6U-destabilized-HBR (6UD-HBR). We observed that 6UD-HBR had a reduced fluorescence but similar dynamics as seen with 6U-HBR (Figure 2a). When transferred back to normoxia, the fluorescent signal of 6UD-HBR cells showed dramatic reduction reaching the lowest level in 10 hours. Based on these observations, 6U-HBR is more sensitive in detecting changes in the process of deoxygenation while 6UD-HBR better reflects changes in the process of reoxygenation.

To confirm HIF-dependency of the reporter, we performed RNAi experiments targeting HIFs under hypoxia. When knocking down three HIFs (HIF1α, HIF1β and HIF2α), 6U-HBR-HeLa cells under hypoxia had greatly reduced fluorescence, close

to the level seen under normoxia (Figure 3a-f, approximately equal numbers of cells observed in a-d). Furthermore, we showed that over-expression of non-degradable HIF1α or HIF2α alone was sufficient to activate the expression of the construct (Figure 4a and 4b). To test whether the signal induced by HIFα over-expression (OE) was directly caused by the canonical activity of HIFα, we introduced RNAi against the essential cofactor of HIFα, HIF1β, in HIFα OE cells. We showed that when transfected with HIF1β RNAi, the eYFP intensity induced by HIFα OE was greatly reduced (Figure 4b-d). It is worth noting that in the RNAi experiments performed in a high-throughput platform, cells showed equivalent intensity across wells in the same treatment group (Figure 4b-d). These experiments demonstrate the applicability and sensitivity of the 6U-HBR construct in a high-throughput platform. The sensitivity of the construct to both HIF1α and HIF2α was further confirmed by the results when knocking down individual HIFα or HIFs with various combinations in HeLa cells (Figure S3). Altogether, we show that the reporter is directly sensing HIF signaling in the cells.

In mammalian embryogenesis and development, tissues develop in a microenviroment with various levels of oxygen. It has been shown that hypoxia, on one hand, is essential for heart formation [27,28,29] and endochondrial bone formation [30,31,32,33]; however, on the other hand, it impairs adipose tissue development [34,35] and skin myofibroblast differentiation [36]. Given the critical role of HIF signaling in hypoxia, one fundamental question is whether cells gain

tissue-specific responses to their low oxygen environment by differentially regulating HIF pathway. To determine how cells from different tissue origins respond to hypoxia in terms of HIF levels and activities, we examined 5 6U-HBR carcinomas cell lines, namely, 786+VHL from renal cell adenocarcinoma, A549 from lung carcinoma, ME180 from epidermoid carcinoma in cervix, U251 from glioma, and HeLa from cervical adenocarcinoma. Cells were incubated in different oxygen levels (20%, 5%, 3%, 2%, 1% or 0.3%) for 4 days, and the average 6U-HBR fluorescence was determined by FACS analysis. These cells might have been expected to behave similarly under hypoxic conditions due to the fact that they have adapted to culture conditions; however, we observed that they responded distinctly, under the same degree of oxygen. For instance, ME180 and HeLa both had little response at 3% of oxygen, however, when oxygen level was below 3%, ME180 exhibited dramatic increase of the signal, while HeLa had relatively low increase (Figure 5). On the other hand, the signal of U251 increased slowly at 5% and 3% of oxygen, more dramatically in 2% and 1% of oxygen, and distinctly continued to increase even in 0.3% of oxygen (Figure 5).

We further verified that 6U-HBR signals in these cell lines correlated with intracellular HIF protein levels. Since the 1% oxygen environment exhibited the greatest difference in 6U-HBR signals among ME180, U251 and HeLa cells, we used this level of oxygen to assay HIF mRNA and protein. While HIF1α mRNA levels were similar among these cell lines, HIF2α mRNA levels varied (Figure S4), which

was in accordance with a previous study [2]. For protein level, we first showed that four hours in hypoxia induced higher HIF1α protein levels in ME180, U251 and A549 than in HeLa and 786-O+VHL (Figure 6a-b). Since it has been previously reported that in prolonged hypoxia, HIF1α protein degrades while HIF2α exhibits minimal change [37], we hypothesized that not only the total HIF protein levels but also the stabilization dynamics of the proteins are responsible for different 6U-HBR reporter activities in these cell lines. We therefore analyzed the differences in degradation kinetics of HIFα proteins in all the five cancer cell lines in 1% of oxygen. This analysis revealed that HIF1α in HeLa cells is less stable during prolonged hypoxia than in any other cell types analyzed (Figure 6c and d). When comparing HIF2α levels in 1% hypoxia versus normaxia, the difference among these five cell lines is diminished in terms of total protein levels and degradation patterns (Figure S5). However, while no HIF1α protein was observed in 786-O+VHL cells, HIF2α protein was induced in hypoxia in this cell lines, explaning the responsiveness of this cell line to the 6U-HBR reporter (Figure S5; Figure 5). Altogether, the data show that higher levels of more stable HIFα proteins are observed in ME180, A549 and U251 compared to HeLa and 786-O+VHL cells, supporting the findings revealed by 6U-HBR reporter that these five cell lines display distinct HIF activities in the same degree of hypoxia.

We chose HeLa and ME180 for further study since they are both cervical cancer cell lines that exhibit very distinct responses to hypoxia. In order to understand the

regulation of HIFα in these two cervical cancer cell lines, we analyzed mRNA microarray data previously performed in these cell lines in 2% hypoxia [38]. Interestingly, we identified hsp90 mRNA differentially expressed between HeLa and ME180. We first validated this finding by real-time PCR analysis and showed that ME180 has higher levels of hsp90 both in normoxia and hypoxia (Figure 7). To determine whether higher levels of hsp90 in ME180 could be causal for higher HIF activity, we reduced hsp90 binding to HIF1α in ME180 by incubating cells with 17-allylamino-demethoxygeldamycin (17-AAG) in 2% hypoxic conditions. 17-AAG inhibits ATP binding to hsp90, thereby preventing the interaction of hsp90 and its target protein [39]. Since hsp90 interaction with HIF1α is shown to increase HIF1 stability in hypoxia [11], 17-AAG should reduce hsp90 dependent HIF1 activity. Accordingly, 6U-HBR ME180 incubated with 17-AAG showed lower levels of 6U-HBR reporter activity than observed in the controls (Figure 8a). Furthermore, the mRNA expression of HIF1α target gene CA9 was significantly reduced (Figure 8b), suggesting that HIF activity was reduced in ME180 when hsp90 was rendered inactive. These data support the hypothesis that the difference in hsp90 levels is causal for the difference in HIF activity observed in the two cervical cancer cell lines. We excluded the possibility of reduction of HIF activities caused by general transcriptional regulation in the cells by showing stable expression of another endogenous housekeeping gene in both cell types (WYHAZ, Figure 8c).

Discussion

Here we create and characterize a novel HIF reporter construct, in which eYFP

expression is regulated by tandem repeats of minimal HIF1α and HIF2α binding sites. By employing cells that are infected with lentivirus of the HIF-HBR construct, we show that cells with 6 tandem repeats as the promoter in the construct (6U-HBR) give better signal-to-background ratio than those with 3 or 12 repeats. Also, in terms of reporter kinetics, the 6U-HBR construct achieves a stronger signal in the process of deoxygenation, while cells with 6U-destabilized-HBR construct more accurately reflect the reduction of HIF levels in the process of reoxygenation. Moreover, through siRNA studies against HIF1α, HIF2α and HIF1β as well as over-expression studies of ndHIF1α and ndHIF2α, we demonstrate that the construct is HIF specific and is able to respond to both HIF1α and HIF2α. Utilizing 5 different cancer cell lines infected with 6U-HBR construct, we observe that different cancer cell lines have distinct eYFP signals under the same oxygen levels, which correlate with the total amount of HIF1 α and HIF2 α proteins in these cell lines.

We utilized the reporter to explore difference in the hypoxic response of HeLa and ME180 cervical cancer cell lines. We observed that these two cancer cell lines have distinct 6U-HBR signals which correlate with the amount of HIFα proteins in response to the same level of hypoxia. Similarly, the different behaviors between HeLa and ME180 under hypoxia are also observed in angiogenic growth factor expression [40,41] as well as in proteasome, histone deacetylase [42] and Hsp90 action [43]. Among them, Hsp90 is of particular interest. In accordance to previous analysis [43], we show that ME180 has higher levels of Hsp90 mRNA than HeLa

both under normoxia and hypoxia. Higher basal level of HIF cofactor Hsp90 in ME180 could contribute to the observed higher HIF activity, since hsp90 binding to HIF has been reported to protect HIFα against VHL-independent degradation [9,11]. Low oxygen levels inactivate PHD/VHL dependent degradation of HIF, resulting in stabilized and active HIF transcription factor. However, HIFα is eventually degraded in hypoxic conditions. This VHL-independent degradation of HIF1α has shown to be RACK1 dependent [11]. Hsp90 competes with RACK1 binding to HIF1α, thereby protecting HIF1α against hypoxic degradation. In present study, we show that ME180 cells have higher levels of hsp90 and more stable HIFα proteins compared to HeLa (Figure 6). We further show that reduction of hsp90 in ME180 diminishes the difference of HIF level and activity as compared to HeLa. These data suggest that different Hsp90 levels could be causal for the different HIF activity levels observed in the two cervical cancer cell lines under the same level of hypoxia. It will be interesting to test in the future whether oxygen/PDH/VHL-independent HIFα degradation is in general the key regulator of the HIF activity differences observed in cancer cell lines under the same level of hypoxia.

It is known that cervical cancer cells are often infected with an oncogenic-type human papillomavirus (HPV), and HeLa is transformed by HPV18 while ME180 by HPV68. HPV of different types vary in the expression and regulation of the two primary HPV oncogenes, E6 and E7 [44], that are sufficient to alter HIF-1α level. HPV11 E6 and HPV31 E6 can stimulate HIF-1α expression as a consequence of ubiquitin-dependent

degradation of p53 [45] while HPV16 E7 can interact with a cullin-2 ubiquitin ligase complex that mediates VHL-dependent HIFα degradation[46]. It will be interesting to reveal in the future whether different responses to hypoxia observed in HeLa and ME180 are associated with the type of oncogenic HPV they are infected with.

In this study we report a novel HIF reporter construct containing tandem repeats of minimum HIF binding sites upstream of eYFP coding sequence. We show that the signal of the reporter is HIF dependent and correlates with the cellular HIF protein levels in different cell lines. By utilizing this new construct, we find a surprising variation of HIF activity in different cancer cell lines under the same level of hypoxia. This novel HIF reporter construct may serve as a tool to rapidly define HIF activity levels and therefore therapeutic capacity of potential HIF repressors in individual cancers.

Materials and Methods

Cells, tissue culture and hypoxia induction: HeLa and ME180 (cervical carcinoma), A549 (lung carcinoma), U251 (glioma) cells were from the American Type Culture Collection (Rockville, MD). 786-O cells transfected with a wild-type VHL were obtained from Dr. William G. Kaelin Jr. (Dana-Farber Institute, Boston, MA). Cancer cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with penicillin, streptomycin and 10% fetal bovine serum (FBS, Invitrogen, Carlsbad, CA). Cells were passed with Trypsin/EDTA (Invitrogen, Carlsbad, CA) when they reached 80% confluence.

Hypoxia induction: cells were cultured in multi-gas incubators (Sanyo, San Diego, CA). Nitrogen gas was supplied to the chambers in order to induce a controlled reduced percentage of oxygen. For normoxia, cells were cultured in incubators containing 5% CO_2 and atmospheric concentration of O_2, approximately 20% to 21% O_2. Throughout this paper "normoxia" was referred as 20% O2.

Lentiviral plasmid construction: we generated HIF reporter construct with HIF1α consensus binding site CGTG followed HIF2α site TACGTG together as one unit of binding site for HIFα factors and repeated them in tandem 12 times (12U-HBR), with 5 base pairs of nucleotides of various combinations spacing between. Induction element CACAG, a DNA element necessary for hypoxic induction, was evenly inserted into the whole sequences three times to assist in the induction process. The final HIF-binding sequences were directly synthesized by Integrated DNA Technologies, Inc, Coralville, Iowa (Figure S1). TK minimal promoter followed by 770bps β-globin intron sequences and eYFP was amplified from pBARL plasmid (courtesy of Dr. Randall Moon laboratory, University of Washington, Seattle, WA) and linked to HIF-binding sequence by fusion PCR. The HIF-TK-eYFP sequences were further ligated into a lentiviral plasmid pRRL-cPPT-X-PRE-SIN [47] (courtesy of Dr. William Osborne laboratory, University of Washington, Seattle, WA) between ClaI and XhoI sites. 6U-HBR and 3U-HBR were constructed in the same way, except the HIF-binding sequences only contain six binding repeats (6U-HBR) or three

repeats (3U-HBR). To build up 6U-HBR-destabilized version, the mouse ornithine decarboxylase 422-461 domain sequences were first amplified from Plasmid M38 TOP-dGFP (courtesy of Dr. Randall Moon laboratory) and then inserted into 6U-HBR construct through MluI site following the 3' of eYFP sequence.

Lentivirus production and stable transduction of cell lines: lentiviral plasmids were transfected into FT293 cell lines (Invitrogen, Inc, Carlsbad, CA) by calcium phosphate transfection to produce lentiviral particles. Specifically, transferring plasmid (12U-HBR, 6U-HBR, 3U-HBR, or 6U-HBR-destabilized construct), packaging plasmid, VSVG plasmid and REV plasmid were transfected together at 23:15:8:11.5 ratio, and 25μg of mixed plasmids were transfected per 150mm plate with 1mL of 2X HBS and 0.1mL 2.5M $CaCl_2$. After transfection, media were changed after 24 hours and viral supernatants were harvested and filtered after 72 hours. To obtain concentrated virus, the viral supernatants were further centrifuged at 6100 rpm for 17 hours at 4℃. Precipitate was resuspended in 1X TBS (50 mM Tris.HCl, pH 7.4 and 150 mM NaCl) and then stored in -80℃ before use. To transduce cell lines with virus, appropriate amount of virus were directly added into the media with the presence of hexadimethrine bromide at 4ng/ml (Polybrene, Invitrogen Inc, Carlsbad, CA) and media was changed after 24 hours.

Fluorescence-activated cell sorting (FACS) analysis: cells to be analyzed were trypsinized from plates, centrifuged down and then fixed in 4% of paraformaldehyde

for at least 30 minutes before the analysis. Standard settings were applied in cell sorting with appropriate channel voltages and at least 30,000 cells were analyzed for each experiment. FITC channel was used to detect eYFP fluorescence. FlowJo (Tree Star, Inc. Ashland, OR) was employed to visualize data and FITC values were defined as the geometric mean of florescence intensity of eYFP positive population for hypoxic samples. For samples incubated in 20% O_2, the intensities were defined as the geometric mean of florescence intensity of eYFP of the total population (eYFP negative).

Non-degradable HIF (ndHIF) over expression: to obtain constitutively stable expression HIFα protein, ndHIF over-expressing plasmids (Addgene plasmid 19005 and 19006) were used, in which two the Proline sites of HIF cDNA were changed to Alanine as described previously [48]. Retrovirus made from those plasmids were infected into HeLa and A549 cell lines in the presence of hexadimethrine bromide at 4ng/ml (Polybrene, Invitrogen Inc. Carlsbad, CA) and media was changed after 24 hours. Over-expression of HIF1α and HIF2α was confirmed by western blots as shown in Figure S6.

siRNA against HIF assay: siRNAs against HIF factor were gift of Dr. Zhan Zhang [38]. siRNAs were transient transfected into cells on 6-well plate with Lipofectamine 2000 (Invitrogen Inc. Carlsbad, CA) following the Lipofectamine 2000 instruction. Normally, ~60% reduction of gene expression is obtained after two days (Figure S3b).

Specifically, siRNAs and Lipofectamine were transfected at 100pmol: 2.5µl per well. Media was changed after 24 hours and cells were cultured under hypoxia thereafter. Cells were fixed for FACS after 24 hours in hypoxia. In the high through-put experiment, 6U-HBR HeLa cells stably expressing ndHIF1α/ndHIF2α (ndHIFα OE) or empty vector (EV) were plated in 384-well plate. HIF1β siRNA or scramble siRNA was transfected into ndHIFα OE cells and EV cells in 384-well plate using RNAiMAX reagent (Invitrogen Inc, Carlsbad, CA). Operetta high content imaging system (PerkinElmer, San Jose, CA) was used for microscopy images and fluorescence quantification.

HIF protein western blots: for Western blotting analysis, cells were washed with DPBS and directly lysed on culture dish using homogenizing buffer consisting of 20mM Tris-HCl (pH 7.5), 150mM NaCl, 15% Glycerol, 1% Triton, 3% SDS, 25mM β-glycerolphosphate, 50mM NaF, 10mM NaPyrophosphate, 0.5% Orthovanadate, 1% PMSF (all chemicals are from Sigma-Aldrich, St. Louis, MO), 25U Benzonase® Nuclease (EMD Chemicals, Gibbstown, NJ) and protease inhibitor cocktail (Complete Mini, Roche Applied Science, Germany). The protein concentration of each sample was determined by BCA protein assay system (Thermo Scientific, Rockford, IL). 20µg of protein extracts were loaded, separated by 7.5% SDS-PAGE, and transferred to polyvinylidene difluoride membranes (Hybond-N+, Amersham Pharmacia Biotech, Buckinghamshire, England). Membranes were blocked with 5% nonfat dry milk for at least 60 minutes at room temperature, and incubated overnight

at 4°C with HIF1α antibody (BD biosciences, San Jose, CA) diluted at 1:2000, or HIF2α antibody (Abcam, Cambridge, MA) diluted at 1:1000 (or HIF2α antibody from Novus Biologicals diluted at 1:5000). Finally, after blots had been incubated for 1 hour with horseradish peroxidase-conjugated secondary antibodies, they were visualized by enhanced chemiluminescence (Millipore Corp, Billerica, MA). Protein expression levels in the gel were quantified by densitometry implemented in Image-J (National Institutes of Health, Bethesda http://rsb.info.nih.gov/ij/, 1997-2009).

mRNA quantification among different cell lines: RNA was extracted from cells adherent on plates using Trizol reagent (Invitrogen Inc. Carlsbad, CA) and cDNA was synthesized with Omniscript RT kit (Qiagen Inc. Valencia, CA). To determine the normalization factor, geNorm algorithm was used to seek appropriate endogenous housekeeping genes for comparison among different cell lines [49]. HPRT1 and YWHAZ were chose as the endogenous control genes for normalization. mRNA level of HIF1α, HIF2α, Hsp90 and endogenous control genes were quantified in triplicates of 20μl of PCR reactions using SyberGreen system (Applied Biosystems, Foster City, CA), each with 25ng cDNA. All of the reactions were performed in 7300 real time PCR system (Applied Biosystems, Foster City, CA) using default settings. The following primers were used in the quantification: *HIF1α*_F: 5'-TCCATGTGACCATGAGGAAA-3' and *HIF1α*_R: 5'-CCAAGCAGGTCATAGGTGGT -3'; *HIF2α*_F: 5'-CCACCAGCTTCACTCTCTCC-3' and *HIF2α*_R: 5'-TCAGAAAAGGCCACTGCTT-3'; *hsp90*_F: 5'-

TCTGGAAGATCCCCAGACAC-3', *hsp90*_R: 5'-AGTCATCCCTCAGCCAGAGA-3'.

Knockdown of hsp90 through 17-AAG: Cells were incubated in 2% hypoxia with 100nM 17-AAG in media for 24 hours and then replaced by fresh media. Cells were fixed for FACS or harvested for mRNA quantification after 48 hours in hypoxia.

Acknowledgement

We thank members of the Ruohola-Baker lab for helpful discussions. We thank Dr. William Osborne for providing the lenti-viral plasmids and Kelly Kernan for help on lenti-virus preparations. We thank Dr. Zhan Zhang and Jeremy Wechsler for guidance and advice on HIF western blots. We also thank Jeffrey Boyd for assistance on FACS analysis.

References

1. Wang GL, Jiang BH, Rue EA, Semenza GL (1995) Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. Proc Natl Acad Sci U S A 92: 5510-5514.
2. Jain S, Maltepe E, Lu MM, Simon C, Bradfield CA (1998) Expression of ARNT, ARNT2, HIF1 alpha, HIF2 alpha and Ah receptor mRNAs in the developing mouse. Mech Dev 73: 117-123.
3. Maxwell PH, Wiesener MS, Chang GW, Clifford SC, Vaux EC, et al.. (1999) The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. Nature 399: 271-275.
4. Ohh M, Park CW, Ivan M, Hoffman MA, Kim TY, et al.. (2000) Ubiquitination of hypoxia-inducible factor requires direct binding to the beta-domain of the von Hippel-Lindau protein. Nat Cell Biol 2: 423-427.
5. Pouyssegur J, Dayan F, Mazure NM (2006) Hypoxia signalling in cancer and approaches to enforce tumour regression. Nature 441: 437-443.
6. Pollard PJ, Briere JJ, Alam NA, Barwell J, Barclay E, et al.. (2005) Accumulation

of Krebs cycle intermediates and over-expression of HIF1alpha in tumours which result from germline FH and SDH mutations. Hum Mol Genet 14: 2231-2239.
7. Bardos JI, Ashcroft M (2005) Negative and positive regulation of HIF-1: a complex network. Biochim Biophys Acta 1755: 107-120.
8. Xenaki G, Ontikatze T, Rajendran R, Stratford IJ, Dive C, et al.. (2008) PCAF is an HIF-1alpha cofactor that regulates p53 transcriptional activity in hypoxia. Oncogene 27: 5785-5796.
9. Isaacs JS, Jung YJ, Mimnaugh EG, Martinez A, Cuttitta F, et al.. (2002) Hsp90 regulates a von Hippel Lindau-independent hypoxia-inducible factor-1 alpha-degradative pathway. J Biol Chem 277: 29936-29944.
10. Katschinski DM, Le L, Schindler SG, Thomas T, Voss AK, et al.. (2004) Interaction of the PAS B domain with HSP90 accelerates hypoxia-inducible factor-1alpha stabilization. Cell Physiol Biochem 14: 351-360.
11. Liu YV, Baek JH, Zhang H, Diez R, Cole RN, et al.. (2007) RACK1 Competes with HSP90 for Binding to HIF-1α and Is Required for O2-Independent and HSP90 Inhibitor-Induced Degradation of HIF-1α. Molecular Cell 25: 207-217.
12. Shweiki D, Itin A, Soffer D, Keshet E (1992) Vascular endothelial growth factor induced by hypoxia may mediate hypoxia-initiated angiogenesis. Nature 359: 843-845.
13. Plate KH, Breier G, Weich HA, Risau W (1992) Vascular endothelial growth factor is a potential tumour angiogenesis factor in human gliomas in vivo. Nature 359: 845-848.
14. Carmeliet P, Dor Y, Herbert JM, Fukumura D, Brusselmans K, et al.. (1998) Role of HIF-1alpha in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis. Nature 394: 485-490.
15. Wykoff CC, Beasley NJ, Watson PH, Turner KJ, Pastorek J, et al.. (2000) Hypoxia-inducible expression of tumor-associated carbonic anhydrases. Cancer Res 60: 7075-7083.
16. Covello KL, Kehler J, Yu H, Gordan JD, Arsham AM, et al.. (2006) HIF-2alpha regulates Oct-4: effects of hypoxia on stem cell function, embryonic development, and tumor growth. Genes Dev 20: 557-570.
17. Staab A, Loeffler J, Said HM, Diehlmann D, Katzer A, et al.. (2007) Effects of HIF-1 inhibition by chetomin on hypoxia-related transcription and radiosensitivity in HT 1080 human fibrosarcoma cells. BMC Cancer 7: 213.
18. Wang Y, Liu Y, Malek SN, Zheng P (2011) Targeting HIF1alpha Eliminates Cancer Stem Cells in Hematological Malignancies. Cell Stem Cell 8: 399-411.
19. Berchner-Pfannschmidt U, Frede S, Wotzlaw C, Fandrey J (2008) Imaging of the hypoxia-inducible factor pathway: insights into oxygen sensing. Eur Respir J 32: 210-217.
20. Safran M, Kim WY, O'Connell F, Flippin L, Gunzler V, et al.. (2006) Mouse model for noninvasive imaging of HIF prolyl hydroxylase activity: assessment of an oral agent that stimulates erythropoietin production. Proc Natl Acad Sci U S A 103: 105-110.

21. Shibata T, Akiyama N, Noda M, Sasai K, Hiraoka M (1998) Enhancement of gene expression under hypoxic conditions using fragments of the human vascular endothelial growth factor and the erythropoietin genes. Int J Radiat Oncol Biol Phys 42: 913-916.
22. Post DE, Van Meir EG (2001) Generation of bidirectional hypoxia/HIF-responsive expression vectors to target gene expression to hypoxic cells. Gene Ther 8: 1801-1807.
23. Biechele TL, Moon RT (2008) Assaying beta-catenin/TCF transcription with beta-catenin/TCF transcription-based reporter constructs. Methods Mol Biol 468: 99-110.
24. Biechele TL, Adams AM, Moon RT (2009) Transcription-based reporters of Wnt/beta-catenin signaling. Cold Spring Harb Protoc 2009: pdb prot5223.
25. Shibata T, Giaccia AJ, Brown JM (2000) Development of a hypoxia-responsive vector for tumor-specific gene therapy. Gene Ther 7: 493-498.
26. Ghoda L, van Daalen Wetters T, Macrae M, Ascherman D, Coffino P (1989) Prevention of rapid intracellular degradation of ODC by a carboxyl-terminal truncation. Science 243: 1493-1495.
27. Lee YM, Jeong CH, Koo SY, Son MJ, Song HS, et al.. (2001) Determination of hypoxic region by hypoxia marker in developing mouse embryos in vivo: a possible signal for vessel development. Dev Dyn 220: 175-186.
28. Krishnan J, Ahuja P, Bodenmann S, Knapik D, Perriard E, et al.. (2008) Essential role of developmentally activated hypoxia-inducible factor 1alpha for cardiac morphogenesis and function. Circ Res 103: 1139-1146.
29. Harvey RP (2002) Patterning the vertebrate heart. Nat Rev Genet 3: 544-556.
30. Amarilio R, Viukov SV, Sharir A, Eshkar-Oren I, Johnson RS, et al.. (2007) HIF1alpha regulation of Sox9 is necessary to maintain differentiation of hypoxic prechondrogenic cells during early skeletogenesis. Development 134: 3917-3928.
31. Provot S, Zinyk D, Gunes Y, Kathri R, Le Q, et al.. (2007) Hif-1alpha regulates differentiation of limb bud mesenchyme and joint development. J Cell Biol 177: 451-464.
32. Schipani E, Ryan HE, Didrickson S, Kobayashi T, Knight M, et al.. (2001) Hypoxia in cartilage: HIF-1alpha is essential for chondrocyte growth arrest and survival. Genes Dev 15: 2865-2876.
33. Schipani E (2006) Hypoxia and HIF-1alpha in chondrogenesis. Ann N Y Acad Sci 1068: 66-73.
34. Macfarlane CM (1997) In vitro influence of sublethal hypoxia on differentiation of the 3T3-L1 preadipose cell line and its physiological implications. Life Sci 60: 1923-1931.
35. Sahai A, Patel MS, Zavosh AS, Tannen RL (1994) Chronic hypoxia impairs the differentiation of 3T3-L1 fibroblast in culture: role of sustained protein kinase C activation. J Cell Physiol 160: 107-112.
36. Modarressi A, Pietramaggiori G, Godbout C, Vigato E, Pittet B, et al.. (2010) Hypoxia impairs skin myofibroblast differentiation and function. J Invest

Dermatol 130: 2818-2827.
37. Kong X, Alvarez-Castelao B, Lin Z, Castano JG, Caro J (2007) Constitutive/hypoxic degradation of HIF-alpha proteins by the proteasome is independent of von Hippel Lindau protein ubiquitylation and the transactivation activity of the protein. J Biol Chem 282: 15498-15505.
38. Mathieu J, Zhou W, Zhang Z, Wang AJ, Heddleston JM, et al.. (2011) HIF induces human embryonic stem cell markers in cancer cells. Cancer Research 71: 4640-4652.
39. Schulte TW, Neckers LM (1998) The benzoquinone ansamycin 17-allylamino-17-demethoxygeldanamycin binds to HSP90 and shares important biologic activities with geldanamycin. Cancer Chemother Pharmacol 42: 273-279.
40. Chiarotto JA, Hill RP (1999) A quantitative analysis of the reduction in oxygen levels required to induce up-regulation of vascular endothelial growth factor (VEGF) mRNA in cervical cancer cell lines. Br J Cancer 80: 1518-1524.
41. Pilch H, Schlenger K, Steiner E, Brockerhoff P, Knapstein P, et al.. (2001) Hypoxia-stimulated expression of angiogenic growth factors in cervical cancer cells and cervical cancer-derived fibroblasts. Int J Gynecol Cancer 11: 137-142.
42. Lin Z, Bazzaro M, Wang MC, Chan KC, Peng S, et al.. (2009) Combination of proteasome and HDAC inhibitors for uterine cervical cancer treatment. Clin Cancer Res 15: 570-577.
43. Schwock J, Pham NA, Cao MP, Hedley DW (2008) Efficacy of Hsp90 inhibition for induction of apoptosis and inhibition of growth in cervical carcinoma cells in vitro and in vivo. Cancer Chemother Pharmacol 61: 669-681.
44. Barbosa MS, Vass WC, Lowy DR, Schiller JT (1991) In vitro biological activities of the E6 and E7 genes vary among human papillomaviruses of different oncogenic potential. J Virol 65: 292-298.
45. Nakamura M, Bodily JM, Beglin M, Kyo S, Inoue M, et al.. (2009) Hypoxia-specific stabilization of HIF-1alpha by human papillomaviruses. Virology 387: 442-448.
46. Huh K, Zhou X, Hayakawa H, Cho JY, Libermann TA, et al.. (2007) Human papillomavirus type 16 E7 oncoprotein associates with the cullin 2 ubiquitin ligase complex, which contributes to degradation of the retinoblastoma tumor suppressor. J Virol 81: 9737-9747.
47. Barry SC, Harder B, Brzezinski M, Flint LY, Seppen J, et al.. (2001) Lentivirus vectors encoding both central polypurine tract and posttranscriptional regulatory element provide enhanced transduction and transgene expression. Hum Gene Ther 12: 1103-1108.
48. Yan Q, Bartz S, Mao M, Li L, Kaelin WG (2007) The hypoxia-inducible factor 2alpha N-terminal and C-terminal transactivation domains cooperate to promote renal tumorigenesis in vivo. Mol Cell Biol 27: 2092-2102.
49. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al.. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3:

RESEARCH0034.

Figure legend

Figure 1. Schematic representation of HIF-HBR reporter construct. 3, 6 or 12 tandem units of HIF binding repeats (3U, 6U or 12U-HBR) and a TK minimal promoter together regulate the eYFP expression. b-c) Optimization of the number of HIF binding repeats. b) 6U-HBR HeLa cells turn eYFP on in 2% but not in 20% oxygen. Equal amount of cells ($5X10^4$) were plated in both culture dishes and microscopy and FACS analysis were performed after 2 days; c) the signal-to-background ratio is represented by FACS analysis. HeLa cells were transduced with HIF-HBR reporter lentiviral particles for 24hrs and then grown under normoxia (20% O2) or hypoxia (2% O2) for 48 hours before FACS analysis. Ratios are of geometric means of eYFP intensity in 2% to 20% oxygen.

Figure 2. Dynamics of 6U-HBR and 6UD-HBR HeLa cell lines. $5X10^4$ cells were plated onto 35mm plates and then cultured under either normoxia (20% O_2) or hypoxia (2% O_2). At different time points, cells were harvested and fixed for FACS analysis. Mean and error bars are from three biological repeats.

Figure 3. 6U-HBR construct is HIF-dependent. 6U-HBR HeLa cells transfected with siRNAs against three HIFs (HIF1α, HIF1β and HIF2α) are eYFP negative under hypoxia, demonstrated by both microscopy (upper panel a-d) and FACS results (lower panel e and f). Equal amount of cells ($5X10^4$) were plated in culture dishes and

microscopy and FACS analysis were performed after 2 days after siRNA transfection.

Figure 4. Sensitivity of 6U-HBR construct. a) 6U-HBR construct in A549 is activated either with ndHIF1α or ndHIF2α, but not with empty vector construct. b-d) HIF1β siRNA alone largely reduces fluorescent signals induced by ndHIF1α/ndHF2α over-expression (ndHIFα OE) in HeLa cells, indicating that the eYFP signals are reversible by manipulating levels of HIF factors. Both microscopy images (b) and fluorescent quantification (c and d) are shown.

Figure 5. Cancer cells with 6U-HBR show different levels of HIF activity under hypoxia. Five cancer cell lines infected with 6U-HBR lentivirus were cultured under various degree of hypoxia, including 20%, 5%, 3%, 2%, 1% and 0.3% of O_2. Their eYFP intensities were determined after 4 days in culture by FACS analysis. The degrees of hypoxia are presented as log of relative O_2 level to normoxia (20% O_2).

Figure 6. HIF protein levels vary in different cancer cells under 1% O_2. a) HIF1α in ME180, U251, A549, HeLa and 786-O+VHL cells under 1% O_2 for 4 hours. b) Quantification for HIF1α protein shown in a). Quantification was performed on scanned film images obtained from separate western blots. c) HIF1α levels and d) quantification in ME180, U251, A549, HeLa and 786-O+VHL cells under 1% O_2 at three time points: 4 hours, 12 hours and 48 hours, compared to the levels under 20% O_2 as controls.

Figure 7. Hsp90 mRNA levels are different in ME180 and HeLa both under hypoxia and normoxia. Cells were either incubated under 1% hypoxia or 20% normoxia for 48 hours before quantification the hsp90 mRNA level by real-time PCR. The error bars are presented as sample error of the mean (SEM) from three independent biological experiments.

Figure 8. 17-AAG reduces the difference in HIF levels and its transcriptional activity between ME180 and HeLa. Results of three independent experiments are summarized and the error bars are presented as sample error of the mean (SEM). a) Fluorescence intensities of HeLa in DMSO, ME180 in DMSO and ME180 in 100 nM 17-AAG in 2% hypoxia are normalized to that in 20% normoxia. b) The fold changes of CA9 mRNA levels normalized to 20% normoxia in those cells suggest that the difference, in terms of HIF activity, is reduced in the presence of 17-AAG. c) The stable expression of another housekeeping gene, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide (YWHAZ), demonstrates the generally normal transcriptional activities in those cells treated either with DMSO or 17-AAG.

Supporting Information Legends

Supplemental Figure 1: Promoter sequence of 12U-HBR construct. The minimal TK promoter is not shown.

Supplemental Figure 2: a) 6U-HBR HeLa cells turn eYFP on in 2% but not in 20% oxygen. Equal amount of cells ($5X10^4$) were plated in both culture dishes and FACS analysis were performed after 2 days. The percentage of eYFP positive cells under theese conditions are shown. b) The patterns of 6U-HBR HeLa cells in response to different levels of hypoxia. The percentage of eYFP positive cells and the mean intensity of each sample were shown.

Supplemental Figure 3: 6U-HBR construct is HIF-dependent. a) siRNAs assays against various combinations of HIFs demonstrate that the construct is HIF-dependent. HeLa cells were incubated in 2% O_2 for one day after siRNA transfection; b) Validation is shown for knockdown experiments of HIF1β expression by HIF1β siRNAs. Mean and error bars are from three biological repeats

Supplemental Figure 4: HIFα mRNA levels in ME180, U251 and HeLa cancer cell lines. Delta Ct is calculated directly from Ct values of endogenous genes (HPRT1 and YWHAZ) minus that of HIF1α or HIF2α.

Supplemental Figure 5: a) HIF2α in ME180, U251, A549, HeLa and 786-O+VHL cells under 1% O_2 for 4 hours. b) HIF2α levels in ME180, U251, A549, HeLa and 786-O+VHL cells under 1% O_2 at three time points: 4 hours, 12 hours and 48 hours, compared to the levels under 20% O_2 as controls. HIF2α antibody (Novus

Biologicals,CO, NB100-122) was used.

Supplemental Figure 6: Over-expression of HIF1α and HIF2α by retroviral infection was confirmed in western blots. HIF2α antibody (Abcam, Cambridge, MA, ab20654) was used.

For detailed figures and supplemental data, please refer to publication: Zhou W, Dosey T, Biechele T, Moon RT, Horwitz MS and Ruohola-Baker H. Assessment of Hypoxia Inducible Factor levels in cancer cell lines upon hypoxic induction using a novel reporter construct. PLoS One, 2011 Nov 23; 6(11).

Appendix II Supplemental Experimental Procedures for Chapter II

Oxygen consumption measurements in perifusion system: Oxygen consumption rate was also measured in a flow culture system that concomitantly measures OCR while collecting outflow fractions for subsequent measurement for other substrate (Sweet et al.., 2002a; Sweet et al.., 2002b). Briefly, the flow culture system includes a pump, gas equilibrator, a glass cell perifusion chamber and a fraction collector. Oxygen consumption was calculated as the flow rate multiplied by the difference between inflow and outflow oxygen tension, measured based on phosphorescence lifetimes of an oxygen-sensitive polymerized platinum porphyrin with a fluorometer TauTheta (Boulder CO). The inflow oxygen tension was measured in the absence of cells in the chamber before and after each experiment, and for the calculations, oxygen content of media equilibrated with 5% carbon dioxide/balance air was assumed to be 217 nmol/ml. The dye consists of dimethylsiloxane-bisphenol A-polycarbonate blockcopolymer (5 g; GE) and platinum tetrapentafluorophenyl porphyrin (25 mg; Porphyrin Products) dissolved in 100 ml dichloromethane, and was painted inside the perifusion chamber 3 mm above the cells and excited by pulses from an ultraviolet light-emitting diode (395 nm; HUUV-5102; Roithner). In some experiments, cells were plated on a cover slip within a rectangle area (15 by 20 mm) on the upstream side of an area where the polymerized platinum porphyrin dye was affixed and allowed to attach for 18 hours. Subsequently, the cover slip was placed into a

temperature-controlled, 100-mL perifusion dish (Bioptechs, Butler, PA) that was mounted onto the stage of a Nikon Eclipse TE-200 inverted microscope, and KRB was pumped through the dish at a flow rate of 50 ml/min. Fluorescent emission of the dye was imaged at 610 nm by a Photometrics Cool Snap EZ camera (Photometrics, Tucson, AZ) during excitation at 546 nm with a Xenon lamp. Results are expressed as OCR after first converting the signal to oxygen concentration by use of calibration curve (determined at the end of each experiment using tanks of gas with varying oxygen levels.

RNA-Seq Library Preparation, Sequencing, and Alignment: Total RNAs were extracted using QIAGEN Micro RNeasy kit and amplified following protocols of WT-Ovation Pico kit (Nugen). The quality and quantity of RNAs were determined on an Agilent 2100 Bioanalyzer (Agilent Technologies). cDNA amplification products were made into double-stranded cDNA using WT-Ovation Exon Module kit (Nugen). 3 ug of cDNA products was sheared by sonication (Bioruptor UCD-200) for 2 x 15 min at 4oC, H speed (320 W). cDNA fragments were then blunt-ended with End-It DNA End-Repair kit (Epicentre Biotechnologies) and A-tail was added to the end of the fragments. Solexa_1 and Solexa_2 adaptors were ligated to each end, and the fragments were subjected to size selection by gel electrophoresis. 200-300 bp fragments were recovered and enriched by 18 cycles of PCR amplification. Cluster generation of cDNA library was performed on a cBot Cluster Station (Illumina, Inc.) and the samples were sequenced on Genome Analyzer IIx (Illumina, Inc.). The 36-bp

reads obtained from each sample were matched to the Mouse genome (mm 9, NCBI Build 37) using Bowtie (Langmead et al.., 2009) and reads that aligned uniquely were used in the downstream analysis. Unambiguously mapped reads were used to generate transcript or gene counts. Differential expression was assessed using transcript counts as inputs to DESeq (Anders and Huber, 2010). The transcripts with an adjusted $P > 0.05$ were considered to be differentially expressed.

Cytochrome C oxidase (COX, ETC complex IV) activity assay: ESC and EpiSC were collected and spun down as pellets. Cell lysis, protein extraction and activity measurement followed the instructions specified in Complex IV Rodent Enzyme Activity Microplate Assay Kit (MitoSciences). Briefly, cells were lysed in 1x lauryl maltoside buffer to maintain intact protein activity. Mitochondrial COXs were enriched in the reaction by binding to Complex IV specific antibodies coated on the microplate. The oxidative capacity of COX in the presence of cytochrome C was monitored with 550nm absorbance for 2 hours. Rates of oxidation were calculated by taking absorbance values between 20-60 minutes following Cytochrome C addition, when the decrease of OD values fell into a linear range. Total protein amount in the reactions was determined using the Nanodrop 2000 UV-VIS Spectrophotometer (Thermo Scientific).

Mitochondrial DNA (mtDNA) copy number measurement: The number of copies of mtDNA Co1 DNA and genomic Gapdh DNA were quantified using a standard curve

with ten-fold dilutions from 10^2 to 10^7 copies, and the ratio of mtDNA to genomic DNA was calculated by dividing copies of Co1 with copies of Gapdh in each experiment. Each 10µl reaction contained 0.5-2.0 ng of DNA extract, 1x SYBR green mix, and 300nM of each primer. Reactions were performed using an Applied Biosystems 7300 real time PCR system using 95°C for 10 minutes followed by 50 cycles at 95°C for 10 seconds, 53°C for 15 seconds for mt-Co1 or 60°C for 15 seconds for Gapdh, and 72°C for 28 seconds. Fluorescence was measured during the last step of each cycle using the FAM/SYBR channel. Primers are listed in Supplemental Table 2 primer list.

Electron microscopy of mitochondria and quantification of elongated mitochondria: ESC and EpiSC were washed with DPBS (Invitrogen) and fixed directly on the culture dish with ½ Karnovsky's fixative. Fixed cells were scraped off the plate and resuspended in 1ml ½ Karnovsky's fixative. Pellet wash, dehydration and embedding were handled at the Fred Hutchinson cancer research center electron microscopy core. Briefly, Pellets were further washed six times in 0.1M Cacodylate buffer, in between the pellets were soaked in 1% Osmium Tetroxide in 0.1M Cacodylate buffer for two hours. Pellets were dehydrated during transferred in graded series of ethanol at 50%, 70%, 95%, two times in 100% and finally in 2X propylene oxide for 15 minutes. Pellets were embedded in 1:1 PO/Epon resin overnight with Eppendorf tubes capped. At least 20 images were taken for both cell types at 2500X magnification for further quantification. Elongated mitochondria were defined as the ones with an outline

length 2 times longer than the width. The number of elongated mitochondria was counted manually in a blind manner by at least three individuals.

Lactate measurement of ESC, EpiSC and hESC: Approximately $2X10^6$ ESC, $0.8X10^6$ EpiSC and $2X10^6$ hESC were seeded onto 96-well plate and incubated overnight in their normal growth media. Before the assay, cells were washed in DPBS and were incubated in 150µl of assay media (DMEM/F12, Invitrogen) for 30 minutes. The assay media was further diluted 1/30-1/100 to reach a lactate concentration in the linear detecting range of the 0-25 µM. Lactate concentration of the diluted assay media was measured in a 50µl colorimetric reaction, consisting of 100µM Amplex Red (Invitrogen), 1X Reaction Buffer (Invitrogen), 200U/µl horseradish peroxidase (Sigma-Aldrich) and 2U/ml lactate oxidase (Sigma-Aldrich). The reactions were incubated at room temperature for 30 minutes and read using absorbance at 560nm. The lactate concentration was further normalized to the number of cells present in each well, quantified by the Hoechst staining(HO33342, Sigma-Aldrich) as measured using fluorescence at 355nm excitation and 460nm emission.

RNA isolation and gene expression by real-time PCR: RNA was extracted from cells adherent on plates using Trizol reagent (Invitrogen) and cDNA was synthesized with the Omniscript RT kit (Qiagen). To determine the normalization factor, geNorm algorithm was used to seek appropriate endogenous housekeeping genes for comparison among different cell lines (Vandesompele et al.., 2002). Beta-actin and

H2afz were chose as the endogenous control genes for normalization. 20µl PCR reactions were performed in triplicate using the SyberGreen system (Applied Biosystems), each with 25ng cDNA. All of the reactions were performed in the 7300 real time PCR system (Applied Biosystems) using default settings. Primers used in our study are listed in Supplemental Table 6 primer list.

Microarray analysis: Microarray data were retrieved from the Gene Expression Omnibus website (www.ncbi.nlm.nih.gov/geo) under accession number GSE7902.

References:

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

Facucho-Oliveira, J.M., Alderson, J., Spikings, E.C., Egginton, S., and St John, J.C. (2007). Mitochondrial DNA replication during differentiation of murine embryonic stem cells. J Cell Sci *120*, 4025-4034.

Facucho-Oliveira, J.M., and St John, J.C. (2009). The relationship between pluripotency and mitochondrial DNA proliferation during early embryo development and embryonic stem cell differentiation. Stem Cell Rev *5*, 140-158.

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

Li, N., Alam, J., Venkatesan, M.I., Eiguren-Fernandez, A., Schmitz, D., Di Stefano, E., Slaughter, N., Killeen, E., Wang, X., Huang, A*., et al.*. (2004). Nrf2 is a key transcription factor that regulates antioxidant defense in macrophages and epithelial cells: protecting against the proinflammatory and oxidizing effects of diesel exhaust chemicals. J Immunol *173*, 3467-3481.

Mamo, S., Gal, A.B., Bodo, S., and Dinnyes, A. (2007). Quantitative evaluation and selection of reference genes in mouse oocytes and embryos cultured in vivo and in vitro. BMC Dev Biol *7*, 14.

Matoba, S., Kang, J.G., Patino, W.D., Wragg, A., Boehm, M., Gavrilova, O., Hurley, P.J., Bunz, F., and Hwang, P.M. (2006). p53 regulates mitochondrial respiration. Science *312*, 1650-1653.

Moon, E.J., Sonveaux, P., Porporato, P.E., Danhier, P., Gallez, B., Batinic-Haberle, I., Nien, Y.C., Schroeder, T., and Dewhirst, M.W. (2010). NADPH oxidase-mediated reactive oxygen species production activates hypoxia-inducible factor-1 (HIF-1) via the ERK pathway after hyperthermia treatment. Proc Natl Acad Sci U S A *107*, 20477-20482.

Odet, F., Duan, C., Willis, W.D., Goulding, E.H., Kung, A., Eddy, E.M., and Goldberg, E. (2008). Expression of the gene for mouse lactate dehydrogenase C (Ldhc) is required for male fertility. Biol Reprod *79*, 26-34.

Sweet, I.R., Cook, D.L., Wiseman, R.W., Greenbaum, C.J., Lernmark, A., Matsumoto, S., Teague, J.C., and Krohn, K.A. (2002a). Dynamic perifusion to maintain and assess isolated pancreatic islets. Diabetes Technol Ther *4*, 67-76.

Sweet, I.R., Khalil, G., Wallen, A.R., Steedman, M., Schenkman, K.A., Reems, J.A., Kahn, S.E., and Callis, J.B. (2002b). Continuous measurement of oxygen consumption by pancreatic islets. Diabetes Technol Ther *4*, 661-672.

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 Biol *3*, RESEARCH0034.

Wang, S.C., Myers, S., Dooms, C., Capon, R., and Muscat, G.E. (2010). An ERRbeta/gamma agonist modulates GRalpha expression, and glucocorticoid responsive gene expression in skeletal muscle cells. Mol Cell Endocrinol *315*, 146-152.

Yano, M., Watanabe, K., Yamamoto, T., Ikeda, K., Senokuchi, T., Lu, M., Kadomatsu, T., Tsukano, H., Ikawa, M., Okabe, M.*, et al.*. (2011). Mitochondrial dysfunction and increased reactive oxygen species impair insulin secretion in sphingomyelin synthase 1-null mice. J Biol Chem *286*, 3992-4002.

Figure 1

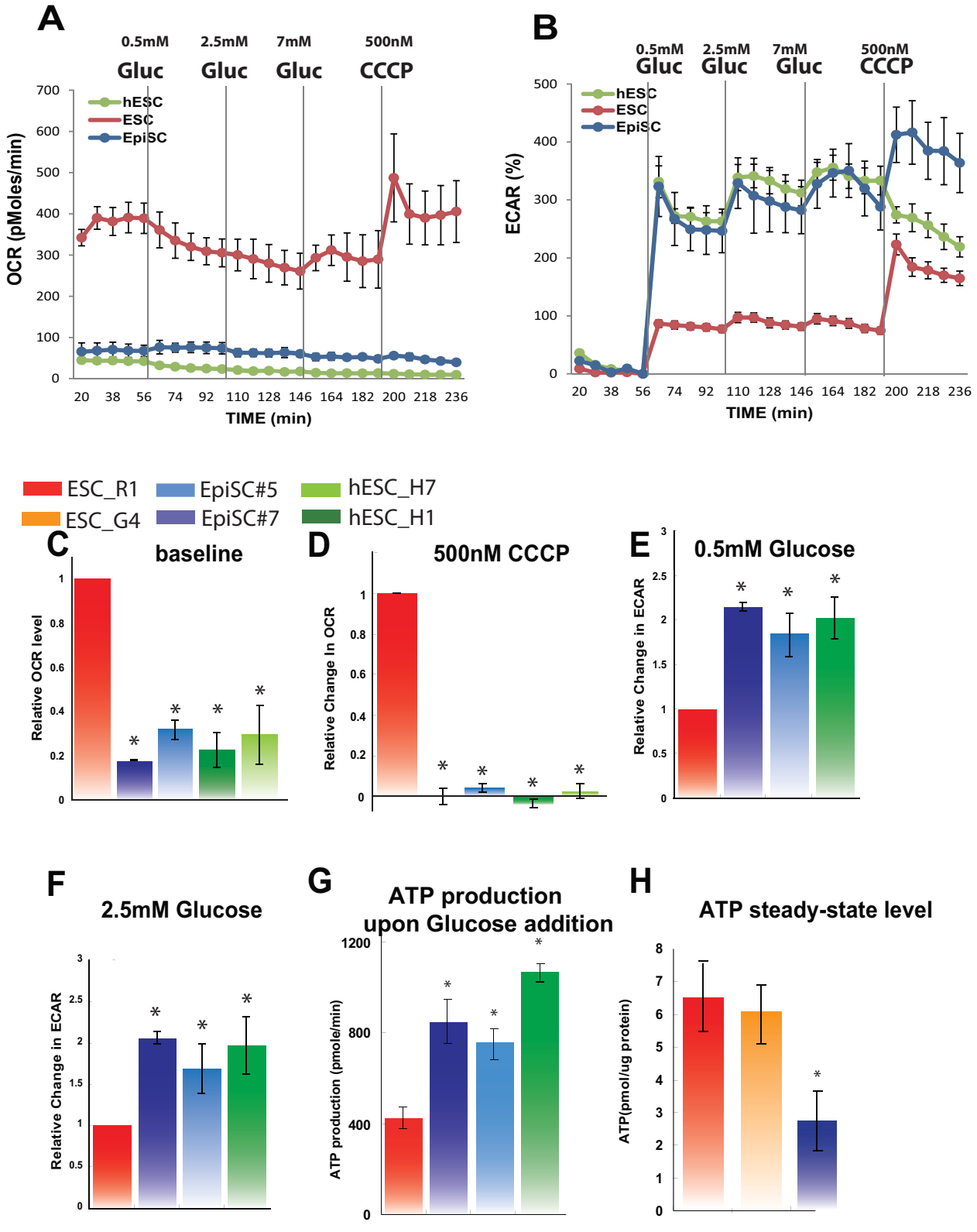

A
0.5mM Gluc
2.5mM Gluc
7mM Gluc
500nM CCCP
hESC
ESC
EpiSC
OCR (pMoles/min)
TIME (min)
B
ECAR (%)
TIME (min)
ESC_R1
ESC_G4
EpiSC#5
EpiSC#7
hESC_H7
hESC_H1
C
baseline
Relative OCR level
D
500nM CCCP
Relative Change In OCR
E
0.5mM Glucose
Relative Change in ECAR
F
2.5mM Glucose
Relative Change in ECAR
G
ATP production upon Glucose addition
ATP production (pmole/min)
H
ATP steady-state level
ATP(pmol/ug protein)

Figure 2

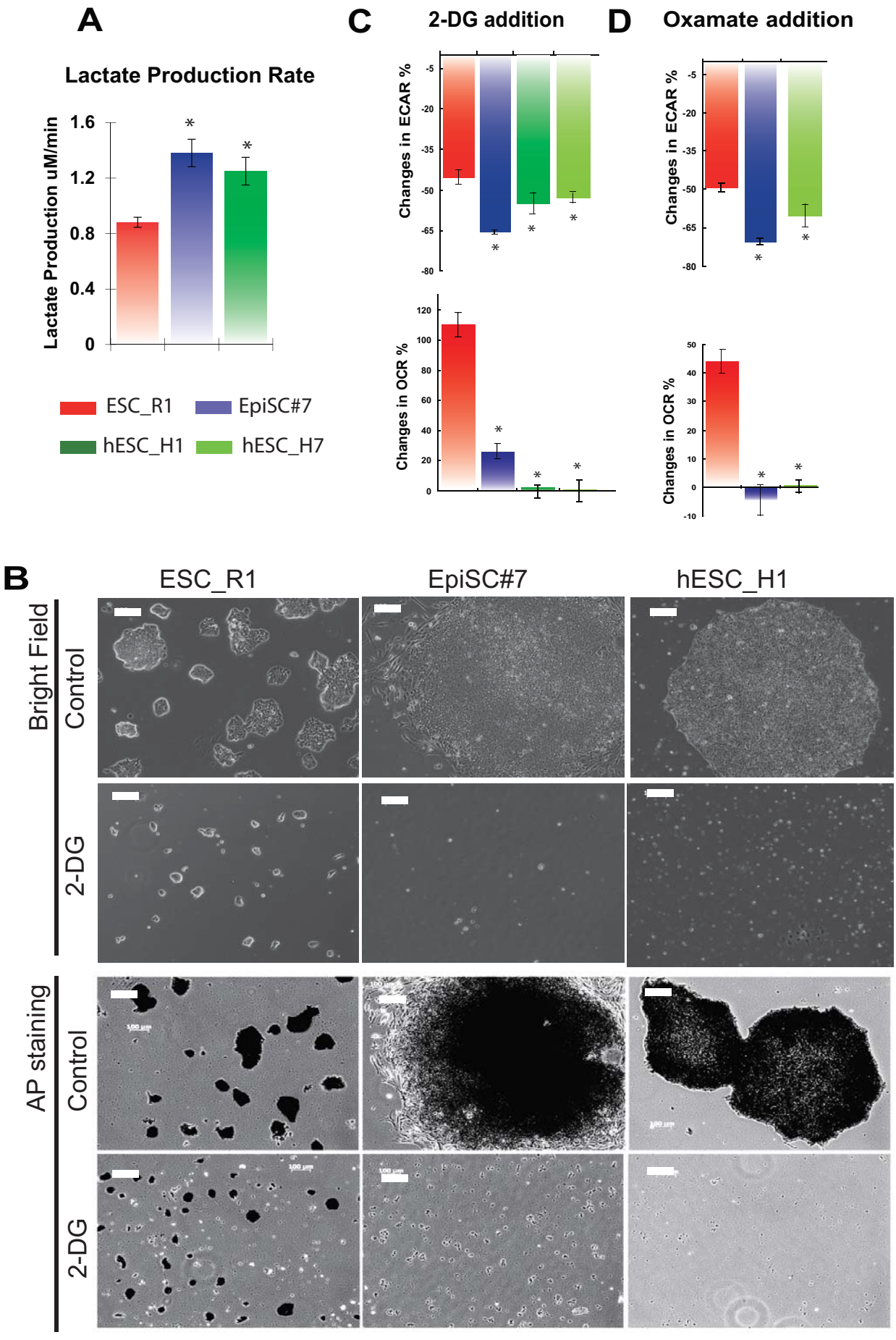
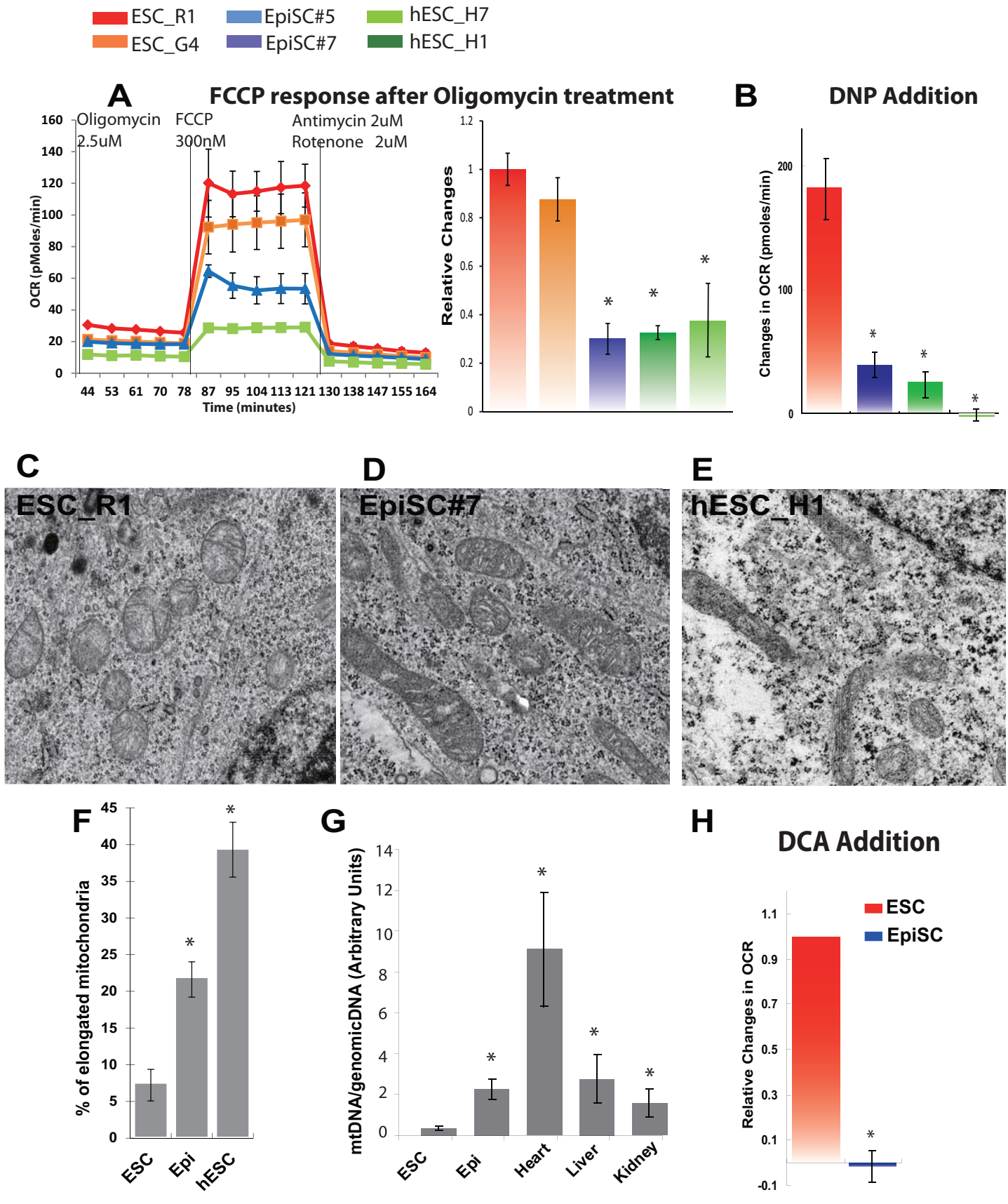
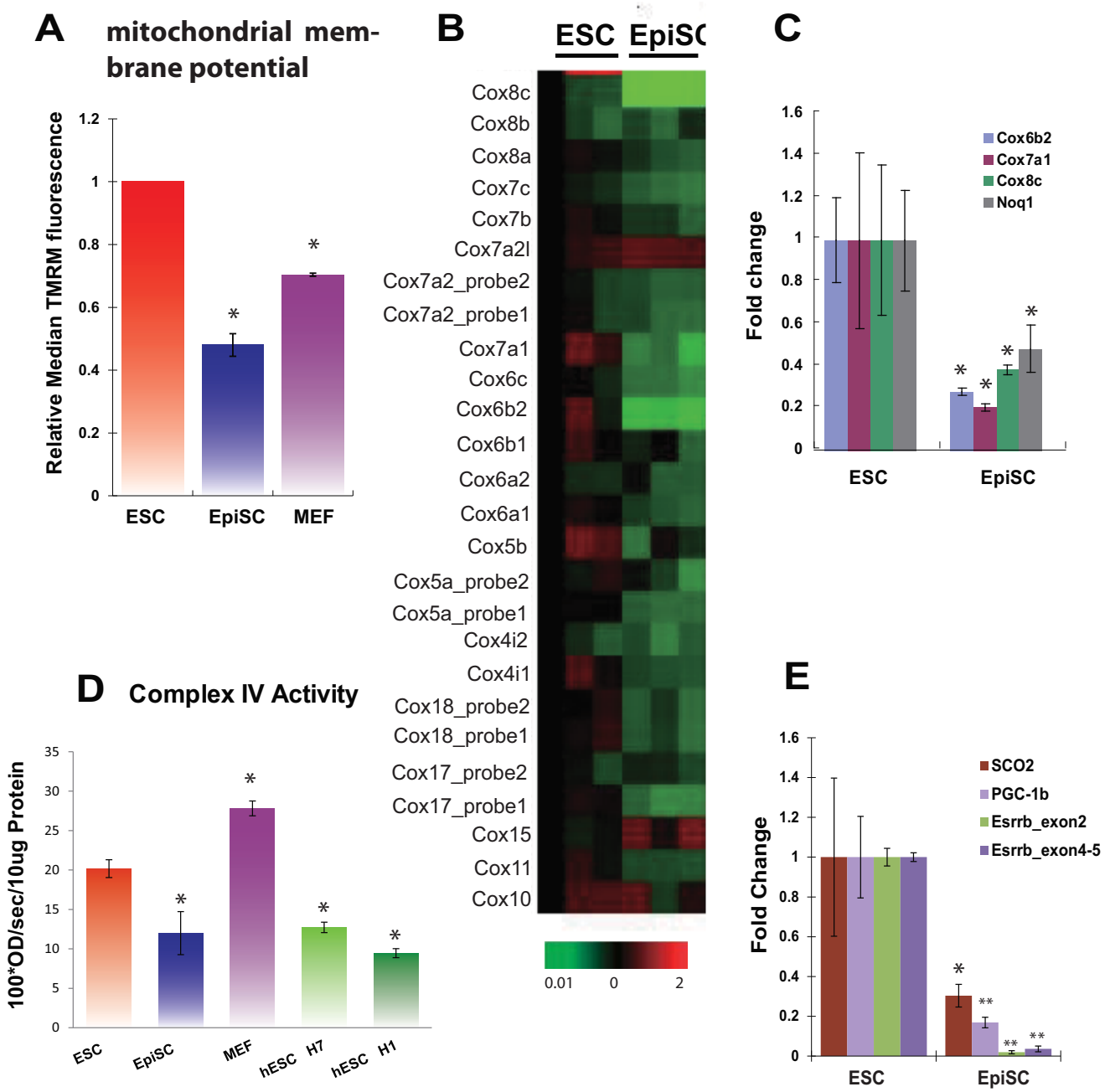

A
Lactate Production Rate
Lactate Production uM/min
1.6
1.2
0.8
0.4
0
ESC_R1
EpiSC#7
hESC_H1
hESC_H7
C
2-DG addition
Changes in ECAR %
-5
-20
-35
-50
-65
-80
Changes in OCR %
120
100
80
60
40
20
0
D
Oxamate addition
Changes in ECAR %
Changes in OCR %
50
40
30
20
10
0
-10
B
ESC_R1
EpiSC#7
hESC_H1
Bright Field
Control
2-DG
AP staining
Control
2-DG

Figure 3


ESC_R1
ESC_G4
EpiSC#5
EpiSC#7
hESC_H7
hESC_H1
A
FCCP response after Oligomycin treatment
Oligomycin 2.5uM
FCCP 300nM
Antimycin 2uM Rotenone 2uM
OCR (pMoles/min)
Time (minutes)
Relative Changes
B
DNP Addition
Changes in OCR (pmoles/min)
C
ESC_R1
D
EpiSC#7
E
hESC_H1
F
% of elongated mitochondria
ESC
Epi
hESC
G
mtDNA/genomicDNA (Arbitrary Units)
ESC
Epi
Heart
Liver
Kidney
H
DCA Addition
Relative Changes in OCR
ESC
EpiSC


A
mitochondrial membrane potential
Relative Median TMRM fluorescence
1.2
1
0.8
0.6
0.4
0.2
0
ESC
EpiSC
MEF
B
ESC
EpiSC
Cox8c
Cox8b
Cox8a
Cox7c
Cox7b
Cox7a2l
Cox7a2_probe2
Cox7a2_probe1
Cox7a1
Cox6c
Cox6b2
Cox6b1
Cox6a2
Cox6a1
Cox5b
Cox5a_probe2
Cox5a_probe1
Cox4i2
Cox4i1
Cox18_probe2
Cox18_probe1
Cox17_probe2
Cox17_probe1
Cox15
Cox11
Cox10
0.01
0
2
C
Fold change
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Cox6b2
Cox7a1
Cox8c
Noq1
ESC
EpiSC
D
Complex IV Activity
100*OD/sec/10ug Protein
35
30
25
20
15
10
5
0
ESC
EpiSC
MEF
hESC H7
hESC H1
E
Fold Change
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
SCO2
PGC-1b
Esrrb_exon2
Esrrb_exon4-5
ESC
EpiSC

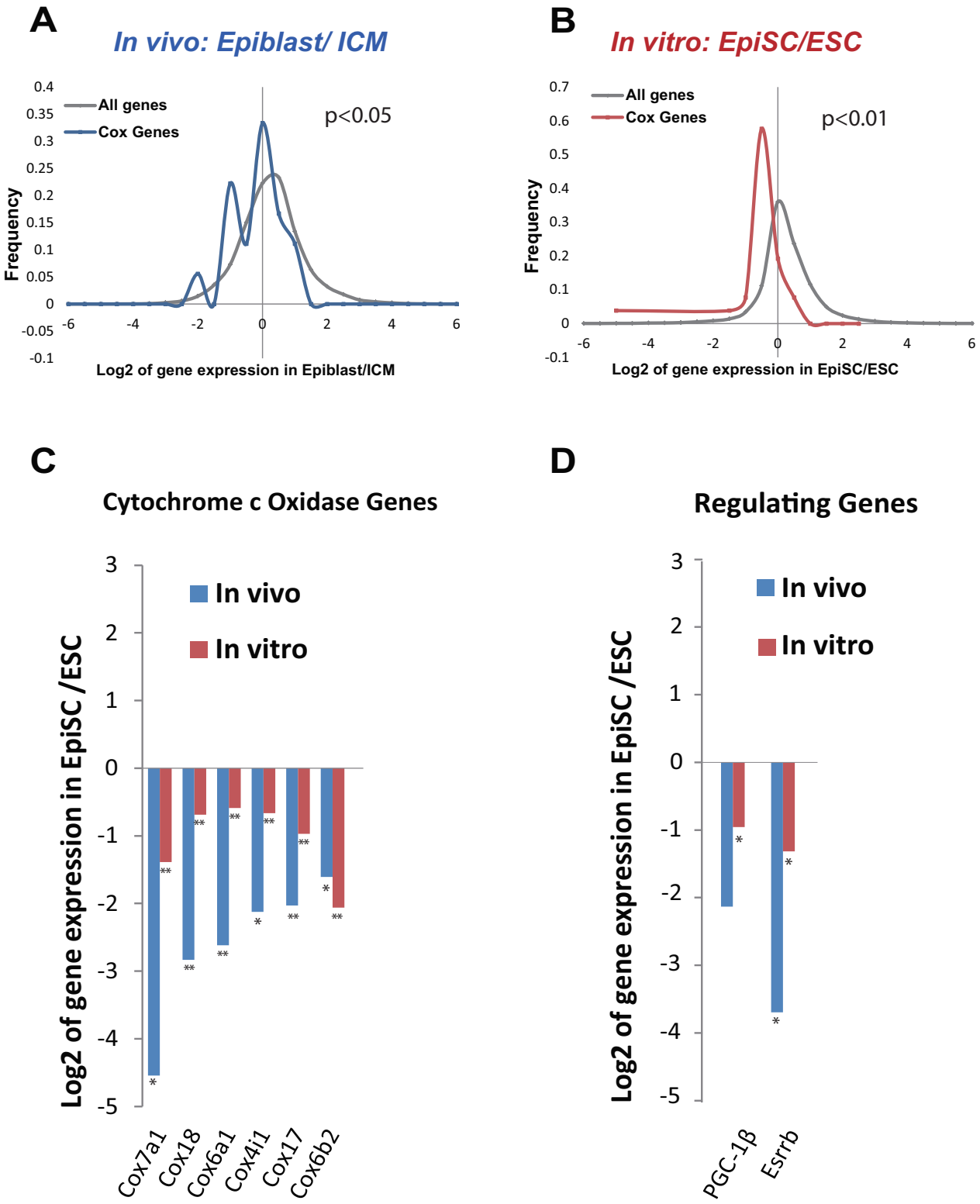

A
In vivo: Epiblast/ ICM
All genes
Cox Genes
p<0.05
Frequency
Log2 of gene expression in Epiblast/ICM
B
In vitro: EpiSC/ESC
All genes
Cox Genes
p<0.01
Frequency
Log2 of gene expression in EpiSC/ESC
C
Cytochrome c Oxidase Genes
In vivo
In vitro
Log2 of gene expression in EpiSC /ESC
Cox7a1
Cox18
Cox6a1
Cox4i1
Cox17
Cox6b2
D
Regulating Genes
In vivo
In vitro
Log2 of gene expression in EpiSC /ESC
PGC-1β
Esrrb

Figure 6

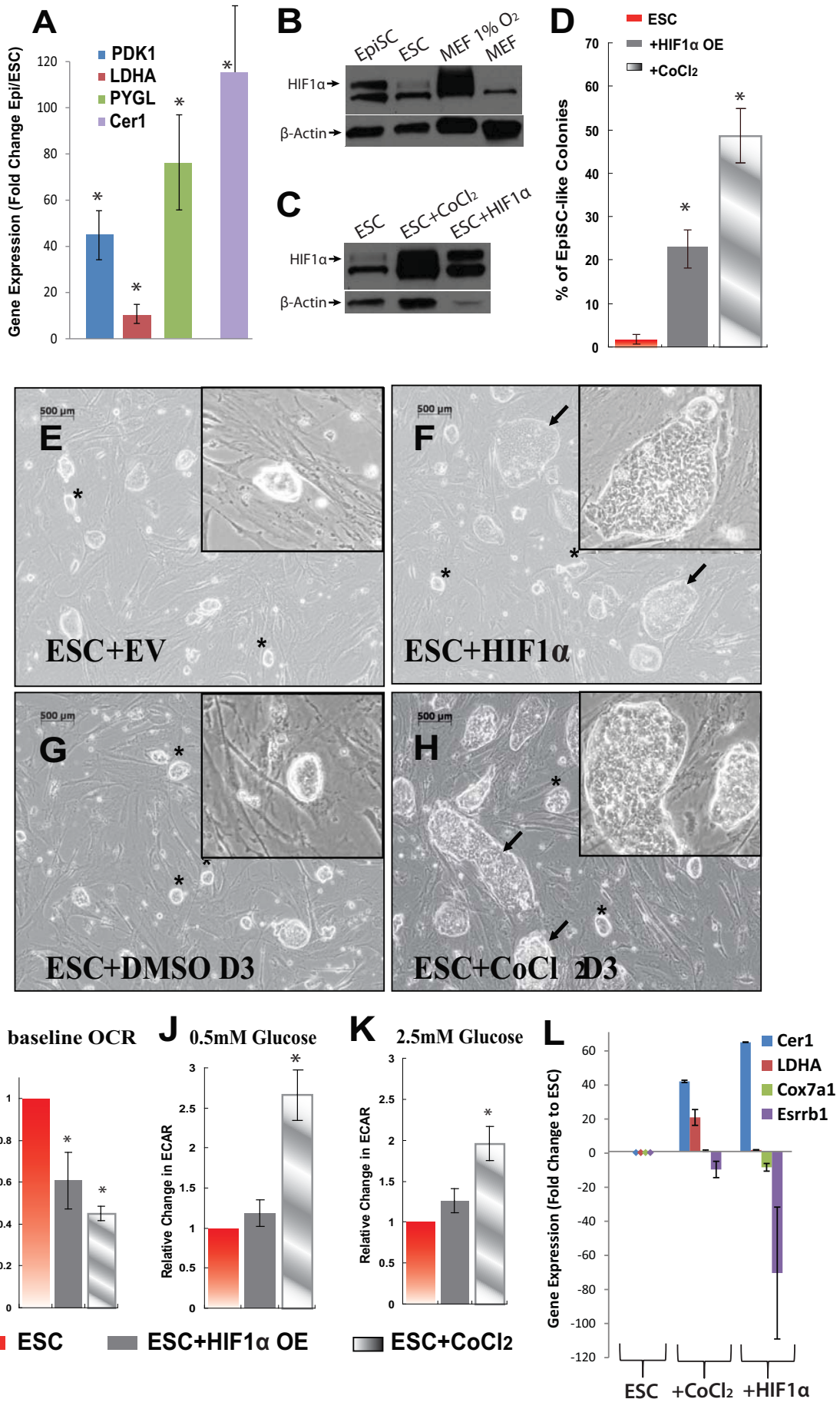

A
PDK1
LDHA
PYGL
Cer1
Gene Expression (Fold Change Epi/ESC)
B
EpiSC
ESC
MEF 1% O2
MEF
HIF1α
β-Actin
C
ESC
ESC+CoCl2
ESC+HIF1α
HIF1α
β-Actin
D
ESC
+HIF1α OE
+CoCl2
% of EpiSC-like Colonies
E
ESC+EV
F
ESC+HIF1α
G
ESC+DMSO D3
H
ESC+CoCl 2D3
500 μm
I
baseline OCR
Relative Change In OCR
J
0.5mM Glucose
Relative Change in ECAR
K
2.5mM Glucose
Relative Change in ECAR
ESC
ESC+HIF1α OE
ESC+CoCl2
L
Cer1
LDHA
Cox7a1
Esrrb1
Gene Expression (Fold Change to ESC)
ESC
+CoCl2
+HIF1α

Figure 7

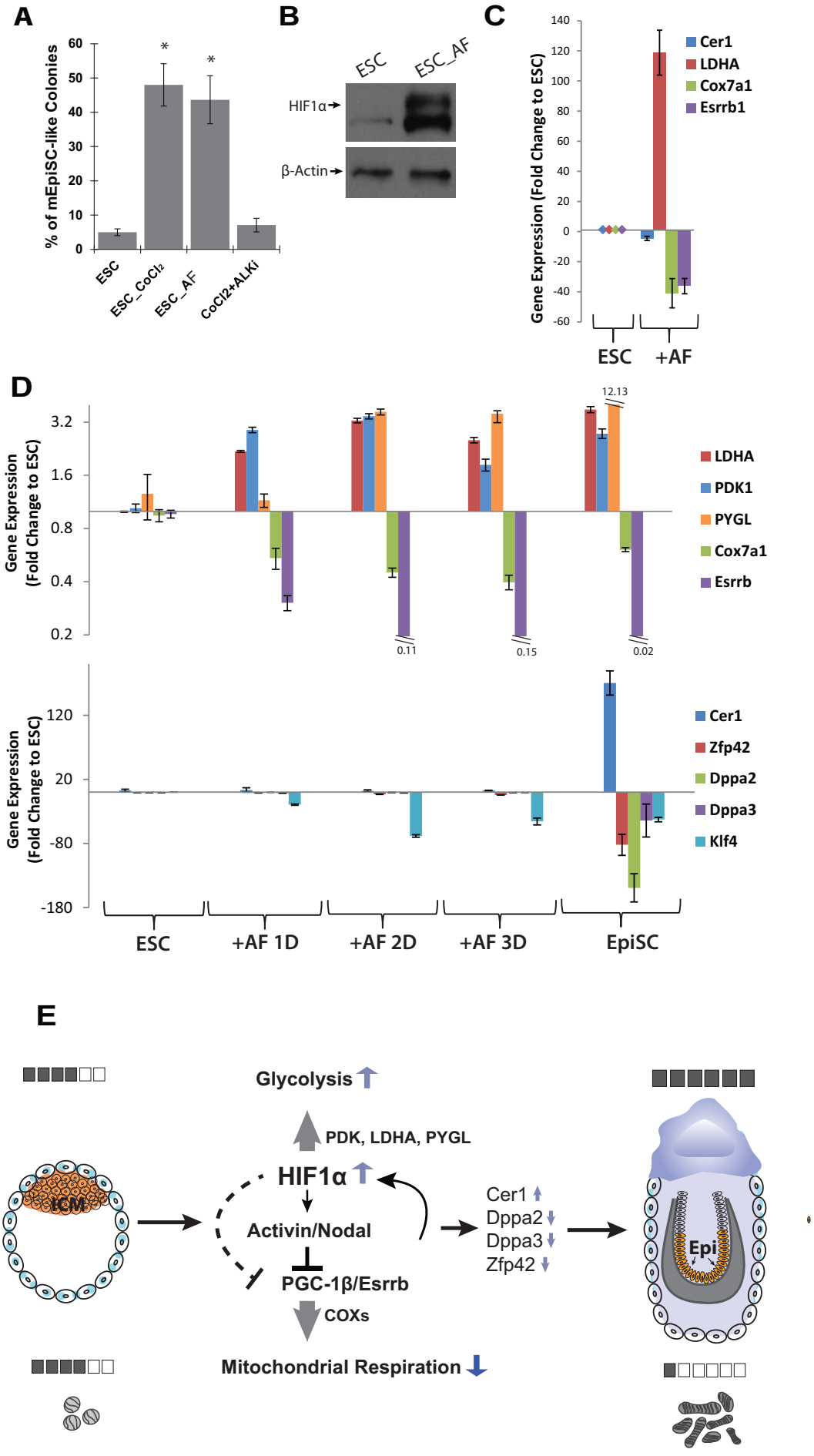
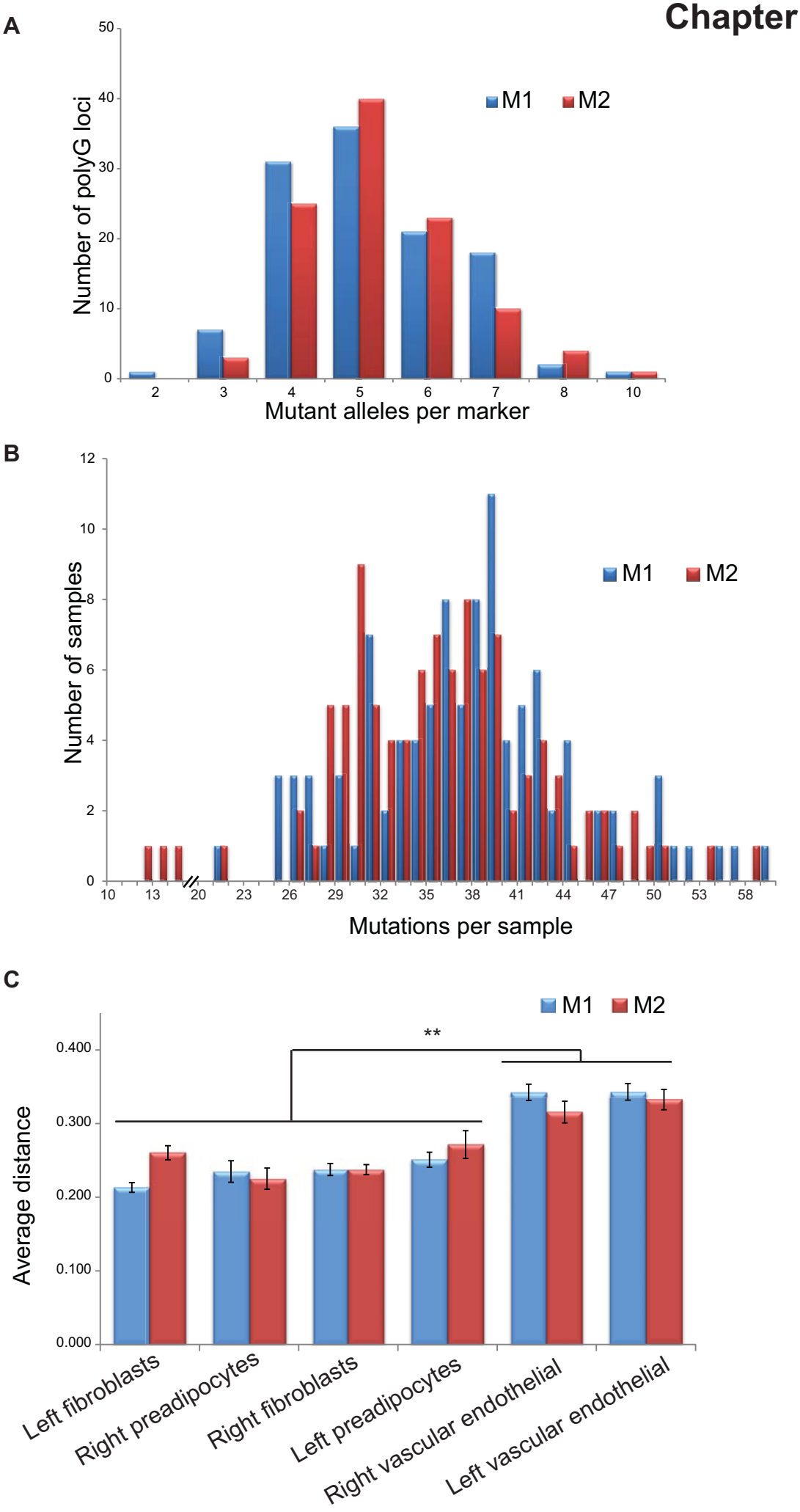
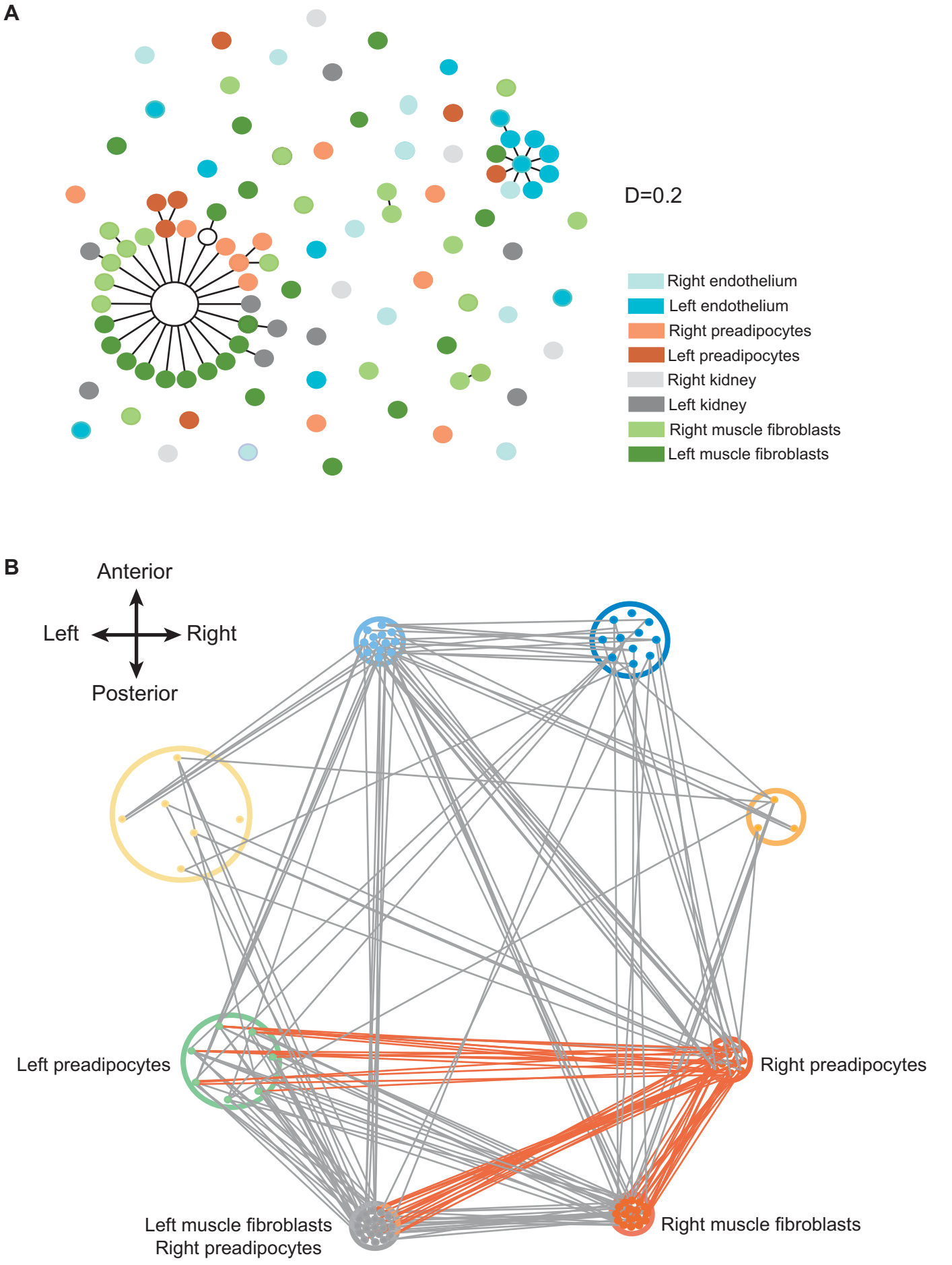
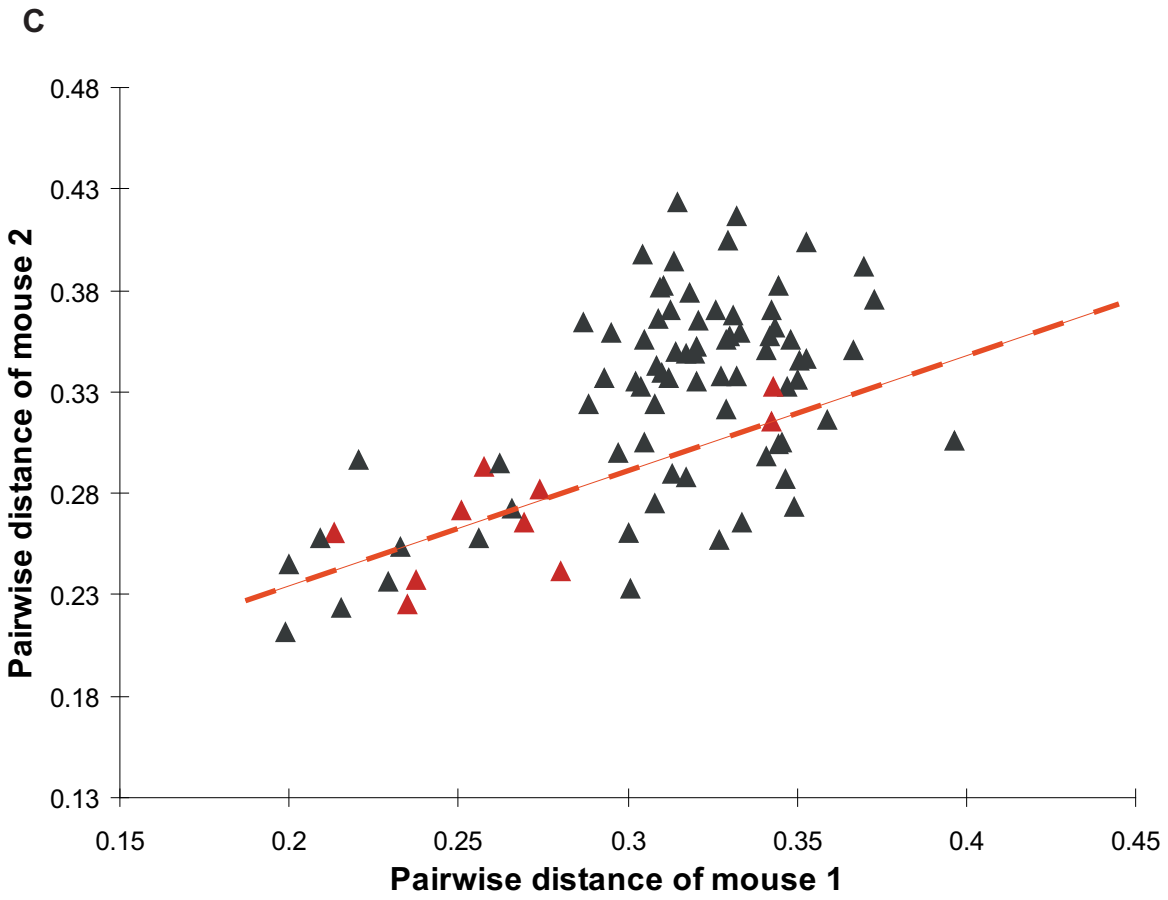
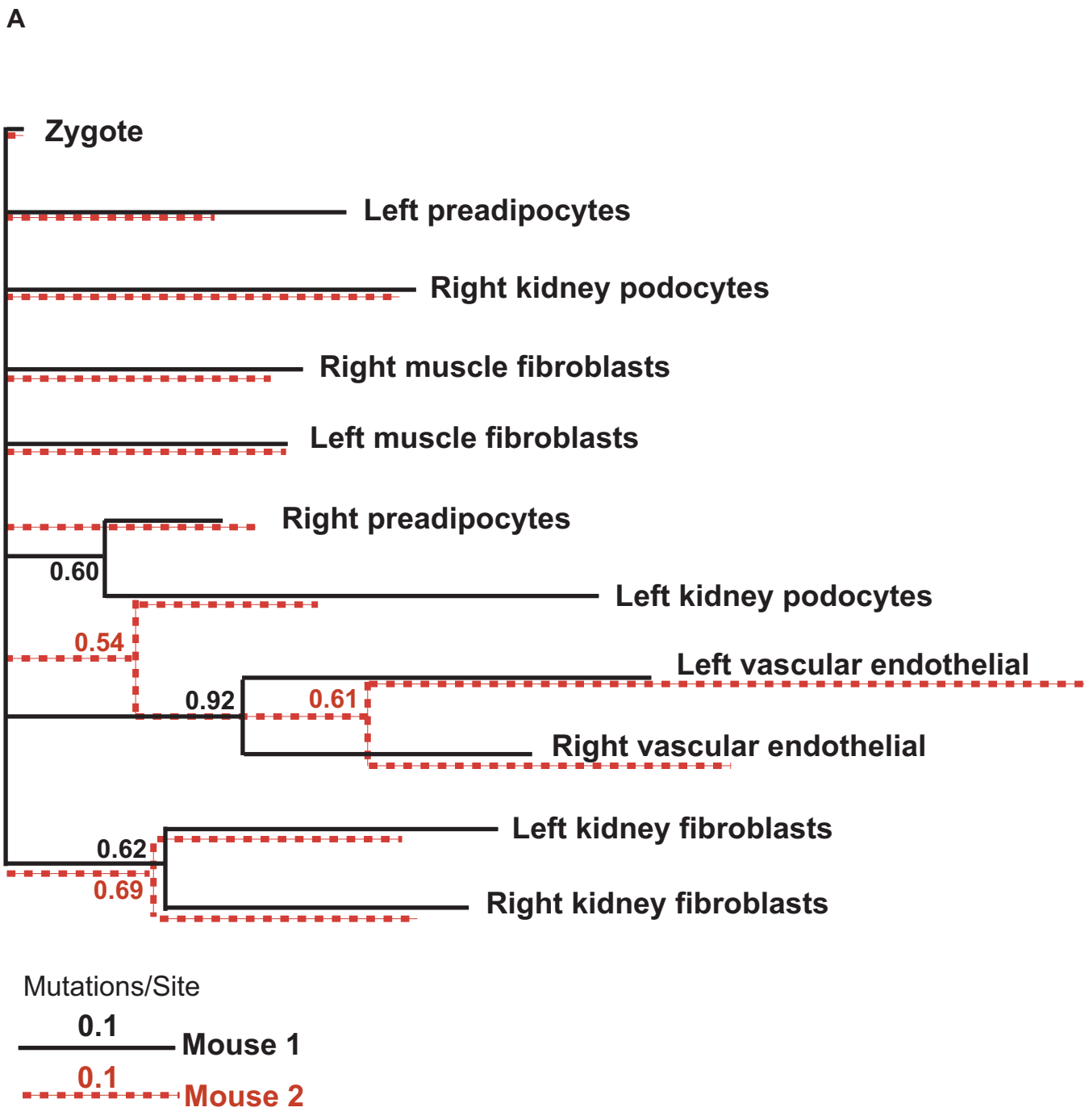

A
% of mEpiSC-like Colonies
ESC
ESC_CoCl2
ESC_AF
CoCl2+ALKi
B
ESC
ESC_AF
HIF1α
β-Actin
C
Gene Expression (Fold Change to ESC)
Cer1
LDHA
Cox7a1
Esrrb1
ESC
+AF
D
Gene Expression (Fold Change to ESC)
LDHA
PDK1
PYGL
Cox7a1
Esrrb
12.13
0.11
0.15
0.02
Cer1
Zfp42
Dppa2
Dppa3
Klf4
ESC
+AF 1D
+AF 2D
+AF 3D
EpiSC
E
Glycolysis
PDK, LDHA, PYGL
HIF1α
Activin/Nodal
PGC-1β/Esrrb
COXs
Mitochondrial Respiration
ICM
Cer1
Dppa2
Dppa3
Zfp42
Epi

Figure 1

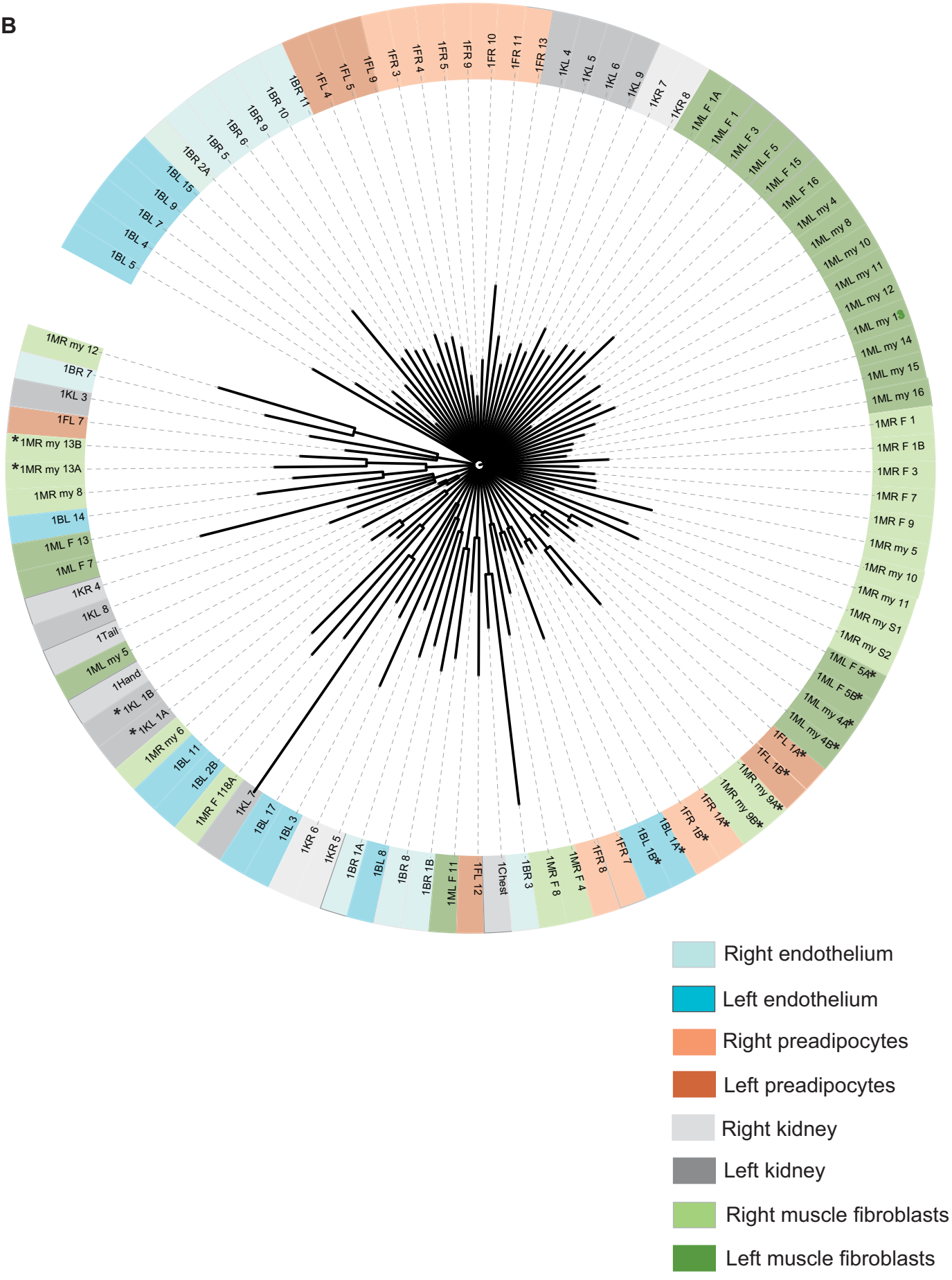

A
Number of polyG loci
M1
M2
Mutant alleles per marker
B
Number of samples
M1
M2
Mutations per sample
C
M1
M2
**
Average distance
Left fibroblasts
Right preadipocytes
Right fibroblasts
Left preadipocytes
Right vascular endothelial
Left vascular endothelial

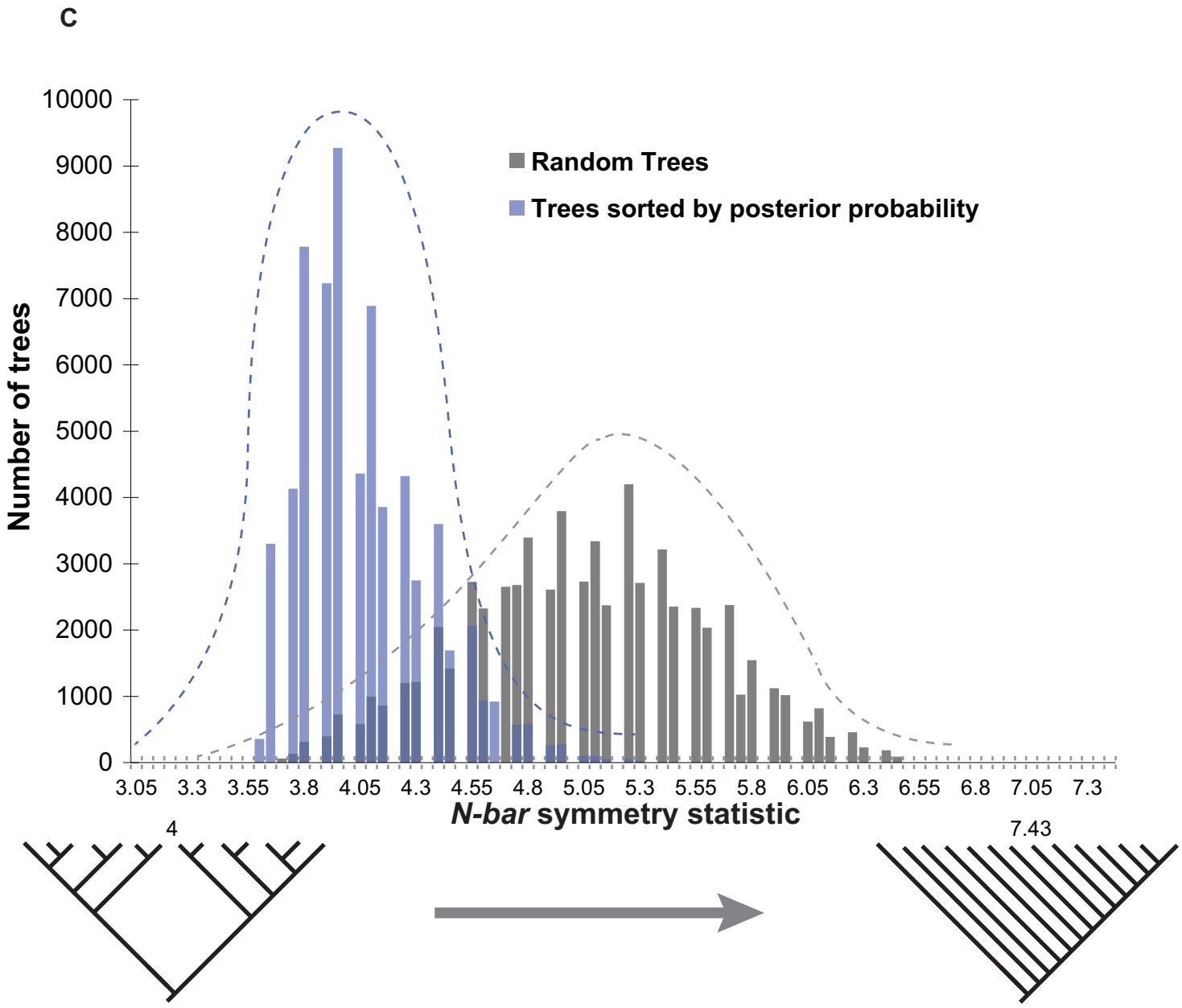

A
D=0.2
Right endothelium
Left endothelium
Right preadipocytes
Left preadipocytes
Right kidney
Left kidney
Right muscle fibroblasts
Left muscle fibroblasts
B
Anterior
Left
Right
Posterior
Left preadipocytes
Right preadipocytes
Left muscle fibroblasts
Right preadipocytes
Right muscle fibroblasts

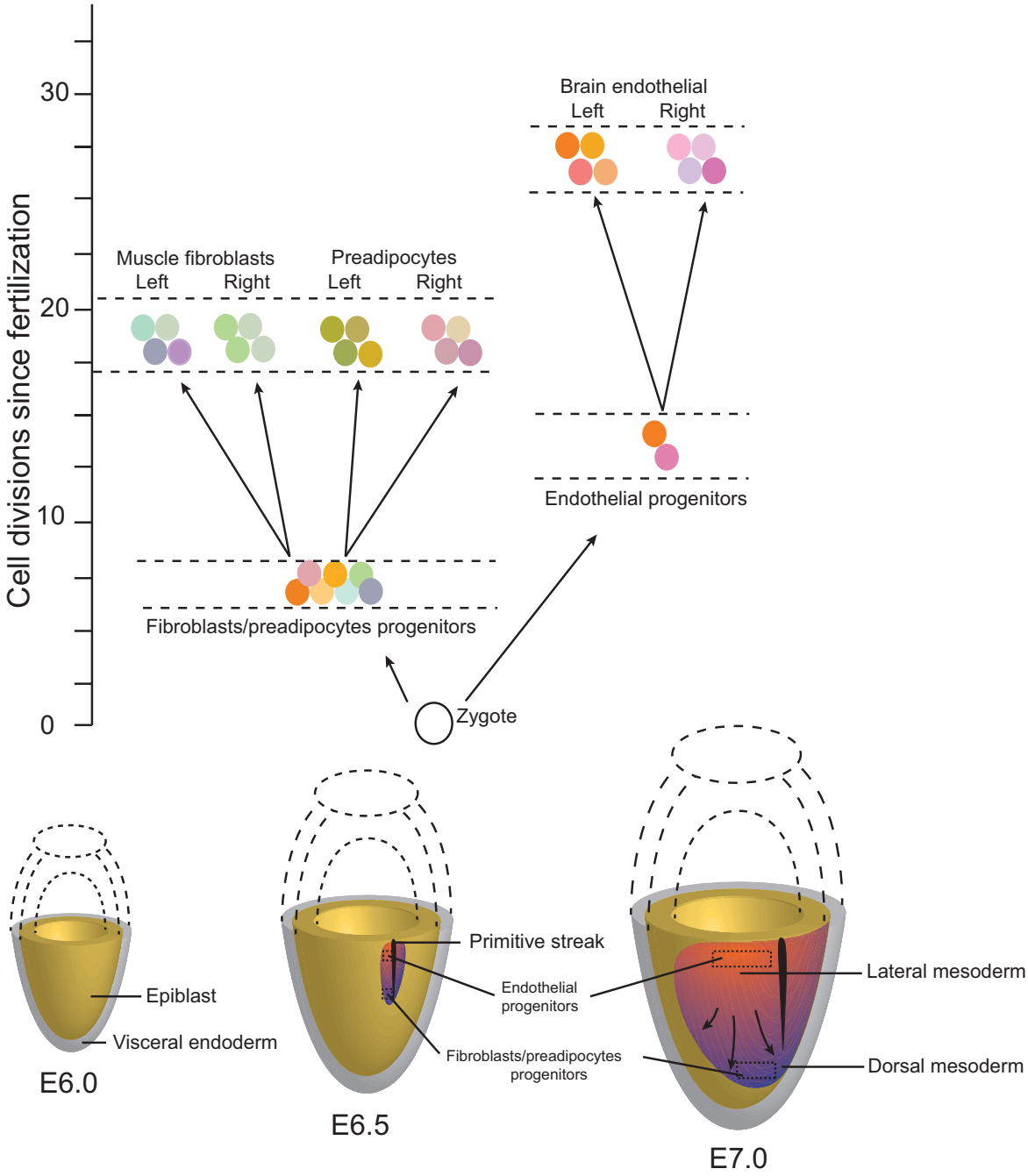

C
0.48
0.43
0.38
0.33
0.28
0.23
0.18
0.13
Pairwise distance of mouse 2
0.15
0.2
0.25
0.3
0.35
0.4
0.45
Pairwise distance of mouse 1


A
Zygote
Left preadipocytes
Right kidney podocytes
Right muscle fibroblasts
Left muscle fibroblasts
Right preadipocytes
0.60
Left kidney podocytes
0.54
Left vascular endothelial
0.92
0.61
Right vascular endothelial
Left kidney fibroblasts
0.62
0.69
Right kidney fibroblasts
Mutations/Site
0.1
Mouse 1
0.1
Mouse 2

B


1MR my 12
1BR 7
1KL 3
1FL 7
*1MR my 13B
*1MR my 13A
1MR my 8
1BL 14
1ML F 13
1ML F 7
1KR 4
1KL 8
1Tail
1ML my 5
1Hand
* 1KL 1B
* 1KL 1A
1MR my 6
1BL 11
1BL 2B
1MR F 118A
1KL 7
1BL 17
1BL 3
1KR 6
1KR 5
1BR 1A
1BL 8
1BR 8
1BR 1B
1ML F 11
1FL 12
1Chest
1BR 3
1MR F 8
1MR F 4
1FR 8
1FR 7
1BL 1B*
1BL 1A*
1FR 1B*
1FR 1A*
1MR my 9B*
1MR my 9A*
1FL 1B*
1FL 1A*
1ML my 4B*
1ML my 4A*
1ML F 5B*
1ML F 5A*
1MR my S2
1MR my S1
1MR my 11
1MR my 10
1MR my 5
1MR F 9
1MR F 7
1MR F 3
1MR F 1B
1MR F 1
1ML my 16
1ML my 15
1ML my 14
1ML my 13
1ML my 12
1ML my 11
1ML my 10
1ML my 8
1ML my 4
1ML F 16
1ML F 15
1ML F 5
1ML F 3
1ML F 1
1ML F 1A
1KR 8
1KR 7
1KL 9
1KL 6
1KL 5
1KL 4
1FR 13
1FR 11
1FR 10
1FR 9
1FR 5
1FR 4
1FR 3
1FL 9
1FL 5
1FL 4
1BR 11
1BR 10
1BR 9
1BR 6
1BR 5
1BR 2A
1BL 15
1BL 9
1BL 7
1BL 4
1BL 5
Right endothelium
Left endothelium
Right preadipocytes
Left preadipocytes
Right kidney
Left kidney
Right muscle fibroblasts
Left muscle fibroblasts


C
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Number of trees
Random Trees
Trees sorted by posterior probability
3.05
3.3
3.55
3.8
4.05
4.3
4.55
4.8
5.05
5.3
5.55
5.8
6.05
6.3
6.55
6.8
7.05
7.3
N-bar symmetry statistic
4
7.43


Cell divisions since fertilization
30
20
10
0
Brain endothelial
Left
Right
Muscle fibroblasts
Left
Right
Preadipocytes
Left
Right
Endothelial progenitors
Fibroblasts/preadipocytes progenitors
Zygote
Epiblast
Visceral endoderm
E6.0
Primitive streak
Endothelial progenitors
Fibroblasts/preadipocytes progenitors
E6.5
Lateral mesoderm
Dorsal mesoderm
E7.0