Analysis of the Up-stream Region of SOX2, Embryonic Stem Cells Specific Gene, in the General Population

by

RAHMAN ALI

Department of Biochemistry
Faculty of Biological Sciences
Quaid-i-Azam University
Islamabad, Pakistan
2012

Analysis of the Up-stream Region of SOX2, Embryonic Stem Cells Specific Gene, in the General Population

A dissertation submitted in the partial fulfillment of the requirements for the degree of

MASTER OF PHILOSOPHY

by

RAHMAN ALI

in

BIOCHEMISTRY/MOLECULAR BIOLOGY

Department of Biochemistry
Faculty of Biological Sciences
Quaid-i-Azam University
Islamabad, Pakistan
2012

CERTIFICATE

This thesis, submitted by **Mr. Rahman Ali** to the Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan, is accepted in its present form as satisfying the thesis requirement for the Degree of Master of Philosophy in Biochemistry/Molecular Biology.

Supervisor: Dr. Amir Ali Khan

External Examinar: Dr. M. Jawad Hassan

Chairman: Dr. Salman Akbar Malik

Dated: 05-03-2012

In the name of Allah, The Most Gracious, The Most Merciful

Allah is the Light of the heavens and the earth. The parable of His Light is as (if there were) a niche and within it a lamp, the lamp is in glass, the glass as it were a brilliant star, lit from a blessed tree, an olive, neither of the east (i.e. neither it gets sun-rays only in the morning) nor of the west (i.e. nor it gets sun-rays only in the afternoon, but it is exposed to the sun all day long), whose oil would almost glow forth (of itself), though no fire touched it. Light upon Light! Allah guides to His Light whom He wills. And Allah sets forth parables for mankind, and Allah is All-Knower of everything.

(An-Nur 35)

Dedicated

To

My Beloved Parents

Whose prayers make me able to excel in my life

And

My Respected Supervisor

Dr. Amir Ali Khan

Who is a figure of inspiration in my life

CONTENTS

List of Figures

List of Tables

List of abbreviations

%	Percentage
°C	Degree centigrade
A	Adenine
AFP	Alpha-fetoprotein
bFGF	Basic fibroblasts growth factor
BIO	Bromoindirubin oxime
BMP	Bone morphogenetic proteins
Bp	Base pair
C	Cytosine
cDNA	Complementary deoxyribonucleic acid
CGBP	CpG island binding proteins
Chip	Chromatin immunoprecipitation
CM	Conditioned medium
Cm	Centi meter
C-Myc	cellular myelocytomatosis oncogene
DNA	Deoxy ribonucleic acid
Dnmt	DNA Methyle transferase
dNTP's	Deoxy nucleotide triphosphates
ERK	Extracellular-signal-regulated kinases
ESC	Embryonic stem cells
FBXO15	F-box protein 15
FGF-2	Fibroblast growth factor-2
FGF4	Fibroblast growth factor-4

FLK-1	Fetal liver kinase-1
G	Guanine
Gm	Gram
GSK-3	Glycogen synthetase kinase-3
HCl	Hydrochloric acid
HDACs	Histone deacetylase
hESC	Human embryonic stem cells
HMG	High mobility group
ICM	Inner cell mass
Id	Inhibitor of differentiation
IPSCs	Induced pluripotent stem cells
IVF	In vitro fertilization
Kb	Kilo base
KLF4	Krupppel-like factor4
KPK	Khyber Pakhtoon khwa
LEFTY-1	Left-right determiantion factor-1
LEFTY-2	Left-right determination factor-2
LIF	Leukemia inhibitory factor
M	Molar
MAPK	Mitogen-activated protein kinase
MEFs	Murine embryonic fibroblasts
mESC	Mouse embryonic stem cells
Mg	Milli gram
$MgCl_2$	Magnesium chloride
Ml	Milliliter

Mm	Milli meter
mM	Milli molar
NaCl	Sodium chloride
NCAM	Cell adhesion molecule, neural, 1
Nm	Nano meter
NPCs	Neural progenitor cells
NSCs	Neural stem cells
OCT-4	Octamer-binding transcription factor 4
OD	Optical density
OSKM	OCT4, SOX2, KLF4, MYC
PcG	Polycomb group proteins
PCR	Polymerase chain reaction
pH	Negative log of hydrogen ions concentration
PI3k	Phosphoinositide 3-kinase
POU	Pituitary specific-octmaer-unc-86 transcription factor
REX1	RNA exonuclease-1
Rpm	Revolution per minute
RR	Ready reaction mix
SC	Stem cell
SDS	Sodium dodesyl sulfate
SMAD1	Mothers against decapentaplegic homolog1
SNP	Single nucleotide polymorphism
SOX2	SRY (sex determining region Y)-box 2
SRY	Sex-determination region-Y
STAT3	Signal tranducer and activation of transcription

T	Thymine
Ta	Annealing temperature
Taq	Thermus aquaticus
TDGF1	Teratocarcinoma-derived growth factor 1
TE	Trophectodermal
TERT	Telomerase reverse transcriptase
TGF	Transforming growth factor
TGF-b1	Transforming growth factor-b1
TGF-β	Transforming growth factor-B
THY1	Thymocyte antigen-1
Tris	Hydroxymethyl aminomethane
UTF1	Undifferentiated embryonic cell transcription factor1
UV	Ultraviolent
v/v	Volume by volume

Acknowledgments

First and foremost, all attributes to Almighty ALLAH, The Gracious and The most Merciful, who blessed me with the perfect circumstances throughout my life, and bestowed me with the potential, ability and an optimal environment to accomplish my M.phil project.

I found no words in the dictionary to express my appreciation to my ***parents,*** *whose sacrifices, love and their moral and social support helped me most not only during my M.phil but in my entire life, without them I could not even think to reach the level of education, which I have today.*

I find it a perfect opportunity to express my feelings of deep thanks and admirations to a person of extreme dedication in the filed of Stem Cells, my supervisor, Professor ***Dr. Amir Ali Khan,*** *who provided me the opportunity to work in his excelled and utmost supervision. His patience, valuable suggestions and kind interest enabled me to complete my research work. He is an inspiration for me and I am honored to be in his supervision. May Almighty ALLAH bless him with perfect health and potential to carry on his valuable work. Ameen*

Specil thanks to Professor ***Dr. Waseem Ahmad*** *for providing the lab facility through out my research work, and I never felt hesitation to work in his lab.*

I present my heartfelt thanks to the chairperson ***Dr. Salman Akbar Malik,*** *Department of Biochemistry Quaid-i-Azam University Islamabad, who is a very kind and dedicated teacher. He is maxim for all students.*

I feel it a privilege to have Sir Salman Basit as one of my seniors, who has been very kind to me and guided me at his best during my research work. He has helped me a lot during my experimental work and provided technical guidance in a friendly way.

I wish to express my heartiest thanks to my batch fellows Zulfiqar Ahmad Barki, Irfan-u-llah Farhad, my juniors Faiz-ur-Rahman and Imran-ullah Yosuf Zai for their pleasant company and smiling attitudes.

I would like to express my sincere thanks to my friend Sultan Shahid for his valuable and memorable company. During my four years of university life I shared every thing with him and he never let me feel alone. The time we spent together was a beautiful moment of life. May Allah almighty bless him with a lot of happiness in his life (Ameen). Special thanks to my worthy friends Sajid ali, Khalid Khan, Ijaz Ali, Fawad Ali and Said Akbar for their support and valuable company. I wish them a brilliant Future.

No words in any dictionary can acknowledge the sacrifices, love and moral support given to me by my brothers. My studies and success is wholly and solely attributed to my parents and all my brothers. Their desires inspire me to excel further in my studies. I owe my deep admirations to my brothers Janat Gul, Redi Gul, Nawad Ali, Khalid khan, Ijaz Ali and my cousin Akbar Ali and Inzer Gul, who supported me not only during my research work but throughout my life. They treated me not only as a brother but behaved like good friends. I love and wish them every happiness of life.

In the end I want to present my indebtedness to all those hands that prayed for my betterment and serenity.

Rahman Ali

Abstract

Human embryonic stem cells (hESCs) are derived from the inner cell mass of the blastocyst, an early-stage embryo, which have the capability of differentiating into multiple cell lineages. Human embryonic stem cells posses the ability to self renew under appropriate conditions to give rise to equivalent daughter cells allowing indefinite propagation in culture. These characteristics make embryonic stem cells valuable tools in the scientific research, especially in regenerative therapy for treating a variety of human diseases. The undifferentiated state (ploripotency) of embryonic stem cells is maintained by the action of ES cell-specific transcription factors including *OCT4*, *NANOG*, and *SOX2*. These three factors bind to a large number of genes many of which are co- occupied by at least two of these three factors including the *OCT4*, *NANOG*, and *SOX2* genes themselves, thus suggesting auto and reciprocal regulation amongst themselves. *SOX2* can act synergistically with *OCT3/4* to activate *Oct–Sox* enhancers, which regulate the expression of pluripotent stem cell specific genes, including *Nanog*, *OCT3/4* and *SOX2* itself. Nanog, along with *OCT4* and *SOX2*, are core transcription factors which regulate the pluripotency and self-renewal of embryonic stem cells. In vitro studies have shown that *SOX2* is required by ES cells for its *Oct–Sox* enhancer activity and it is likely that, *SOX2* is responsible for the activation of these *Oct–Sox* enhancers. Thus the essential function of *SOX2* is to stabilize ES cells in a pluripotent state by maintaining the requisite level of *OCT3/4* expression. This proves that these genes are crucial for the early human development.

In present study upstream promoter region of *SOX2* gene was analyzed for the existence of any sequence variant (polymorphism), in the two distinct population of Pakistan; Pakhtoon and Punjabi. Genomic DNA was extracted from blood samples collected from these individuals was PCR amplified by using specific primer set for the promoter region of *SOX2* gene, followed by subsequent sequence analysis to find any sequence variant. Four out of 70 members showed a G to C transversion in the upstream promoter region of *SOX2* while the remaining six members possessed normal sequence. Amplified DNA products in which sequence variant were detected were treated with specific restriction enzyme (PmeI) to confirm polymorphism. Agarose gel electrophoresis after 16 hours of enzyme treatment revealed that the restriction enzyme failed to digest the region in which variation was detected, suggesting that the

sequence variation was not a polymorphism, but a mistake in the DNA sequencing process. The lack of polymorphism shows that the sequence in the upstream promoter region of *SOX2* gene is highly conserved and the evolution has not permitted any variation in this specific gene.

Introduction

Introduction

Stem Cells

Stem cells are those types of cells having the remarkable properties of renewing themselves and may be differentiate into other type of cells. For example a haematopoietic stem cell residing in bone marrow able to specialize into blood cells. These differentiated cell types will have special functions, for example they are able to produce antibodies, act like scavengers against infections and transport gases. Stem cells have the remarkable properties to develop into a variety of cell types in the human body, serving as a repair system by being able to divide without limit to replenish other cells (Fontes and Thomson, 2009).

Stem cells have been known to be isolated from pre-implantation embryos, fetuses, adults and the umbilical cord and under certain conditions these undifferentiated stem cells can be pluripotent having the ability to give rise to cells from all three germ layers (ectoderm, mesoderm and endoderm) or multipotent (having the ability to give rise to a limited number of other specialized cell types).

There are two main types of stem cells on the basis of their differentiation potential. Adult stem cells (multipotent) and embryonic stem cells (pluripotent).

Adult stem cells

Adult stem cells are undifferentiated cells, found throughout the body. These cells are also known as somatic stem cells and can be found in juvenile as well as adult animals and humans. Adult Stem cells are multipotent and only differentiate into their specific lineages (Fontes and Thomson, 2009; Wianny *et al.*, 2011). The specific roles of adult stem cells in a living organism are to maintain and repair the tissue in which they are found. Adult stem cells have been identified in many organs and tissues, including brain, bone marrow, peripheral blood, blood vessels, skeletal muscle, skin, teeth, heart, gut, liver, ovarian epithelium, and testis. The specific area of each tissue where they reside is called a stem cell niche (Wright, 2000). Two well known examples of adult stem cells are haematopoietic and mesenchymal stem cells.

The production and maintenance of the blood stem cells and their proliferation and differentiation into the cells of peripheral blood is known as hematopoisis.The

hematopoietic stem cells are derived early in the embryogenesis. These stem cells can be obtained from bone marrow, peripheral blood, umbilical cord blood, and fetal liver (Fontes and Thomson, 2009).

Mesenchymal stem cells are found in the nonhematopoietic bone marrow stroma. These stem cells are involved in the regeneration of mesenchymal tissues, such as bone, cartilage, muscle, ligament, tendon, adipose tissue, and stroma (Fontes and Thomson, 2009; Caplan, 1994).

Embryonic stem cells

Embryonic stem cells (ES) are pluripotent cells derived from pre-implantation embryos (wianny *et al.*, 2011). These stem cells have indefinite potential and ability to differentiate into all adult cell types (Evans and Kaufman, 1981). There are about 220 types of cells and tissues found in human bodies which arise from a single zygote.

In the early 1970, a promising start of the pluripotent stem cell derivation was carried out from early embryos. At that time it was found that tetracarcinoma developed after grafting early mouse embryos below the kidney capsule or in the testis of adult mice (Solter *et al.*, 1970; Stevens, 1970). The tetracarcinoma contained undifferentiated pluripotent cells known as embryonal carcinoma (EC) cells. These embryonal carcinoma cells have the potential that they could be expanded continuously in culture. They can also be differentiated in vitro in suspension cultures (Martin and Evans, 1975) or in vivo by grafting in histocompatible mice and were also capable to participate in normal embryogenesis by injection into early an embryo (Papaioannou *et al.*, 1975). Unfortunately, the efficiency of chimerism found with most EC cell line was significantly low, therefore the full development capacity could not be attained since these cells did not colonise the germ line. When the same conditions was implemented that were used for successful isolation and expansion of EC cells, the first mouse embryonic stem cells lines were isolated from the early mouse embryos 10 years later by two groups (Evans and Kaufman, 1981; Martin, 1981).

Derivation of Human Embryonic Stem Cells (hESC)

After the successful isolation of mouse embryonic stem cells (mESC), the scientists were able to grow human blastocysts in the lab (Fong *et al*,. 1994). Afterward human embryonic stem cells (hESCs) derivation was carried out by using IVF (in-vitro

fertilization) method. IVF is a type of reproductive technology that is specially used to obtain a zygote by fusing sperm and egg in a test tube or petri dish (in vitro) and the zygote then starts dividing and becoming an embryo.

By applying this procedure, first the zona pellucida was removed with pronase and then 21 zona-free blastocyst from 9 IVF patients were cultured intact as a whole embryo culture on irradiated human adult oviductal epithelial fibroblasts. After 7 to 11 days nineteen of these produced healthy ICM lumps. These lumps were mechanically separated from the peripheral trophectodermal (TE) cells and underlying feeder cells with hypodermic needles, trypsinized and passaged further on fresh irradiated human adult fallopian tubal feeders. In the first and second passages hESC cell colonies were produced. These hESCs of the first two passages showed typical hESC morphology, stained positive for alkaline phosphatase and had normal karyotypes. These cells differentiated from the third passage onwards (Fong *et al*,. 1994).

Thomson *et al* (1988) used immunosurgery for the complete separation of the ICM and cultured the intact ICM on irradiated murine embryonic fibroblasts (MEFs) instead of human feeders. Parts of each hESC colony were then cut mechanically with hypodermic needles and the cell clusters grown on fresh irradiated MEFs instead of cell dissociation into single cells with trypsin (Thomson *et* al,. 1988). By using such a procedure some workers successfully produced the first hESC line. Later Reubinoff *et al* used immunosurgery, mitomycin-C treated MEFs and a similar mechanical transfer method to derive and propagate hESC lines that could be spontaneously differentiated into neuronal cells in vitro (Fong *et al*,. 2000), as shown in the figure 1.1. Amit and Iskovitz and Suss-toby e*t al* later confirmed that the "whole embryo culture" was successfully worked as well as the immunosurgery protocol to produce hESc line (Amit and Itskovitz, 2002). Bongso *et al* reported in his work (using human feeders instead of MEFs) that hESC lines can be successfully derived and propagated on fetal, neonatal and adult human feeders (Fong *et al*,. 2000).

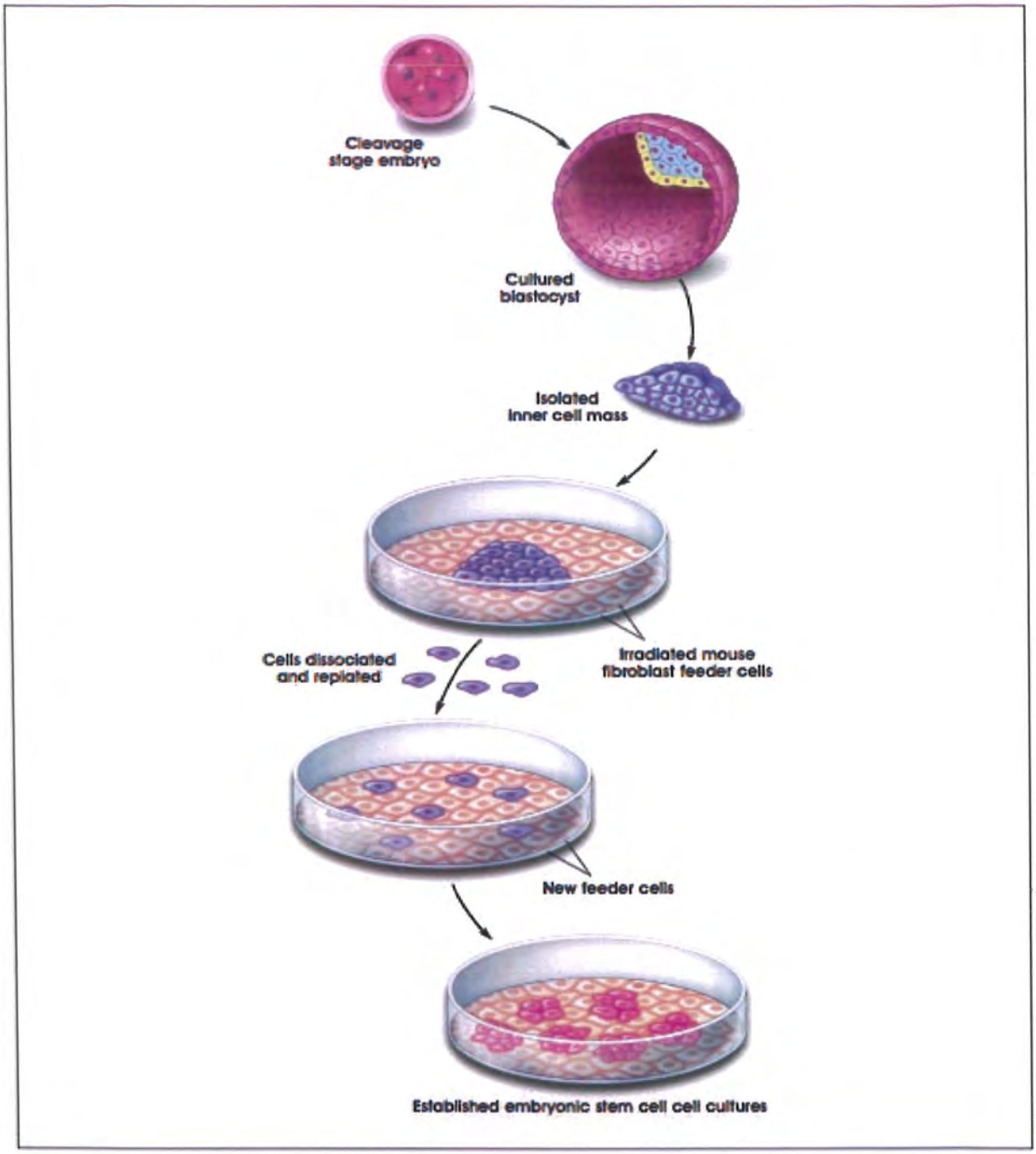

Figure 1.1: Derivation of human embryonic stem cells

Undifferentiated Growth of hESC

Human embryonic stem cells (hESC) like other types of stem cells have the properties to self-renew without differentiation. But they are also pluripotent and capable of differentiating into cells of three germ layers of the embryo. There may be some problem occurs in propagating and maintaining undifferentiated hES cells in vitro. Thus the hESC require a specialized growth environment to retain its undifferentiated state. This specialized types of growth environment are constantly undergoing revision and improvements (Bondar *et al*,. 2004).

As mentioned human ESCs are usually grown on a feeder cells layer. Mitotically inactivated mouse embryonic fibroblast (MEFs) have been shown to be used successfully to support hES cell growth and maintenance. Feeder cells are mitotically inactivated using irradiation or through incubation with mitomycine C. Humn ES grown on these feeders inactivated by irradiation or mitomycin C shows similar morphology and remain pluripotent in both cases. (Bondar *et al*,. 2004).

Feeders have been found to secrete soluble factors in culture medium that are important for maintenance of undifferentiated hES cells. According to some reports, undifferentiated hES cells that are grown in the absence of feeder cells layer are shows differentiation unless conditioned medium from mouse fibroblast culture is used to supplement the hES culture medium (Inokuma *et al*,. 2001). The exact identities of the secreted factors necessary for hES cells maintenance have not yet been identified, various proteins have been found in the mediunm conditioned by mouse fibroblasts feeders as candidates proteins (JWE and Bodnar, 2002).

Properties of undifferentiated growth

Before addressing factors that are involved in the growth and undifferentiated hES cells, it is necessary to point out the qualities that make it possible to distinguish between differentiated and undifferentiated hESC. The following are some properties of undifferentiated hES cells, including cell and colony morphology, cell surface antigens, gene expression, and pluripotency are described.

Undifferentiated state maintains the hES cell morphology and density which in turn promotes healthy pluripotent cell growth, as undifferentiated hES cells have a distinct morphology when viewed under the microscope (Itskovitz *et al*,. 1988; Pera *et al*,.

2000). Every individual cell within the colony will have a large nucleus and prominent nucleoli. Undifferentiated cell are tightly packed within the colony and maintain a defined border at the periphery of the colony. Also the morphology of undifferentiated hES colony is sharply contrasted by that of spontaneously differentiating hES colonies which may acquires many different forms. Differentiating colonies of hES cells may appear to lose their tight border, and cells within the colony begin to enlarge, flatten, and separate.

Human gene *OCT-4* encoding a POU domain transcription factor, is highly expressed in undifferentiated hES cells and is necessary for the cells to maintain the pluripotent state (Draper JS *et al*,. 2002; Inokuma *et al*,. 2001). As hES cells differentiates, *OCT-4* gene expression decreases to low levels (Maor *et al*,. 2001). When the expression of *OCT-4* decreases, the expression of the three embryonic germ layers gene representative such as *AFP*, *FLK-1*, and *NCAM* increases. Some other genes may also expressed in undifferentiated hES, some of which may function in the maintenance of pluripotency include *FGF4, SOX2*, and *NANOG* (Meijer *et al*,. 2004). Unlike mouse embryonic stem (mES) cells, undifferentiated hES cells do not express SSEA-1. Undifferentiated hES cells also positively stain for alkaline phosphatase and demonstrate telomerase activity (Rosler *et al*,. 2003).

Stem Cells Markers

Since the undifferentiated growth of human embryonic stem cells should be validated. For this purpose, several surface markers have been suggested to be used for the undifferentiated, pluripotent state of ESCs. For example *OCT4, NONOG* and *SOX2* are the most commonly involved in this process.

OCT4 is strongly expressed by human and mouse ESCs and is shown to be down regulated upon differentiation. It provides an excellent marker for undifferentiated ESCs and also *OCT4* is required for maintenance of the undifferentiated state of hESCs which differentiate towards trophectoderm if *OCT4* expression is knocked down by RNA interference (Ying *et al*., 2003).

NANOG is another gene expressed strongly in both mice and humans ESCs. This gene is discovered recently and is certainly required for maintenance of the undifferentiated state of mESCs (Robertson *et al*,. 2003). It is also regulated upon the differentiation of hESCs.

Some other genes, including *SOX2*, *REX1*, *TERT*, *UTF1* and *FGF4*, have also been known to be expressed by hESCs and to be used as markers for the undifferentiated, pluripotent state. For this purpose several DNA microarray expression studies of hESCs have also been reported and analyses of these have been used to confirm several genes as specific stem cell related marker genes (Jaenisch *et al.*, 2003; Miura *et al.*, 2004).

Molecular pluripotency of hESCs

The mechanism that regulates the stem cell self-renewal and pluripotency is mostly unknown. Some investigation into the molecular and cellular mechanism of stem cell self-renewal and pluripotency help to meet these challenges and provide the necessary tools for the regenerative potential of ES cells for the therapeutical purpose (Liu *et al*,. 2007).

Pluripotency of human ES cells is regulated by some signaling pathways. Some recent studies (in vitro and in vivo) have revealed that there are several genetic regulators that may play important roles in the human and mouse ES cells self-renewal and pluripotency process. For example extra signaling factors, transcription factors, microRNA, genes implicated in chromosomal stability, and DNA methylation etc.

Signalling Pathways

The following are the mainly involved pathways in human ES cells pluripotency;

MAPK-ERK Pathway

This signalling pathway plays an important role in the self-renewal of hESCs. It has been reported that this pathway is activated by the bFGF (basic fibroblast growth factor) in the medium. FGF ligands and their receptors are highly expressed in human embryonic stem cells. The fibroblast growth factors family has over related members including FGF1, FGF2, FGF and FGF8. This family has been shown to activate tyrosine kinase pathways (Dvorak *et al.*, 2005).Inhibition of this signalling pathway or removal of FGF resultes in differentiation of human embryonic stem cells (Mitalipov *et al.*, 2006), because bFGF has a key role in the activation of MAPK-ERK pathway. Human ES cells are most commonly cultured in the presence of bFGF either on fibroblast feeder layers or in the fibroblast-conditioned medium (Amit *et al*,. 2000: Xu *et al*,. 2001).

PI3K (Phosphoinositide 3-kinase) Pathway

This pathway has the importance in the proliferation, survival, and maintenance of pluripotency in ES cells. Activation of PI3K promotes the ES cell proliferation (Takahashi *et al.*, 2005). It has been shown that the inhibition of PI3K and AKt induces differentiation of human ES cells in the presence feeder cells, suggesting that PI3K/Akt signalling is necessary for the maintenance of ES cell pluripotency (Paling *et al.*, 2004).

Wnt/β-catenin signalling Pathway

Wnt pathway may be involved in the maintenance of pluripotency of human ES cells. This pathway is activated by 6-bromoindirubin-3-oxime (BIO), a specific inhibitor of glycogen synthase kinase-3 (GSK-3), which maintains the undifferentiated phenotype in ES and also sustains the expression of the ES cells specific surface markers. Wnt signalling pathway has been shown to improve the c-Myc level which also the target gene of STAT3 (Signal transducer and activation of transcription). Components of the Wnt signalling pathway are present in human ESCs, although levels of different receptors varied between undifferentiated and differentiated populations. Wnt is believed to stimulate Human ESC proliferation (Pankratz *et al.*, 2007).

TGFβ (transforming growth factor-B) Pathways

TGFβ is known to be a prototypic member of large super family related growth and differentiation factors. This family may have more than 40 members including TGFβ, Activin, Nodal, and bone morphogenetic proteins (BMPs), which are all associated with ES cells. TGFβ transduces signals from the membrane to the nucleus by binding a heteromeric complex of serine/threonine kinase receptors which are known as TGFβ type 1 and type II receptors. There is high affinity of activin for the type II receptors but do not bind to the type I receptors in the absence of type II receptors, whereas BMPs have higher affinity for their type I receptors than for type II (Valdimarsdottir and Mummery, 2005).

BMP4:

It is a member of the TGFβ super family, promote the differentiation of hESCs. Its activation is not required during the undifferentiated state of human ES cells, and it may be inhibited.

TGFβ /activin/nodal:

TGFβ and its correlate factors are shown to be highly expressed in undifferentiated human Es cells. Activin A (which is a member of the TGFβ family) is secreted by mouse embryonic feeders and the respective culture medium is enriched with activin A and maintains human ES cells undifferentiated without condition medium (CM) or STAT3 activation (Beattie *et al.*, 2005).

Notch signalling Pathway

Notch signaling has the important role in maintenance of the stem cells features. Notch related molecules are highly expressed in the NESCs and hESCs. When this pathway is inhibited, the self-renewing activity and the proliferation potential may be significantly reduced in the resulting stem cells (Johnston *et al*,. 2001).

In figure 1.2 the summarized roles of the different pathways are mentioned.

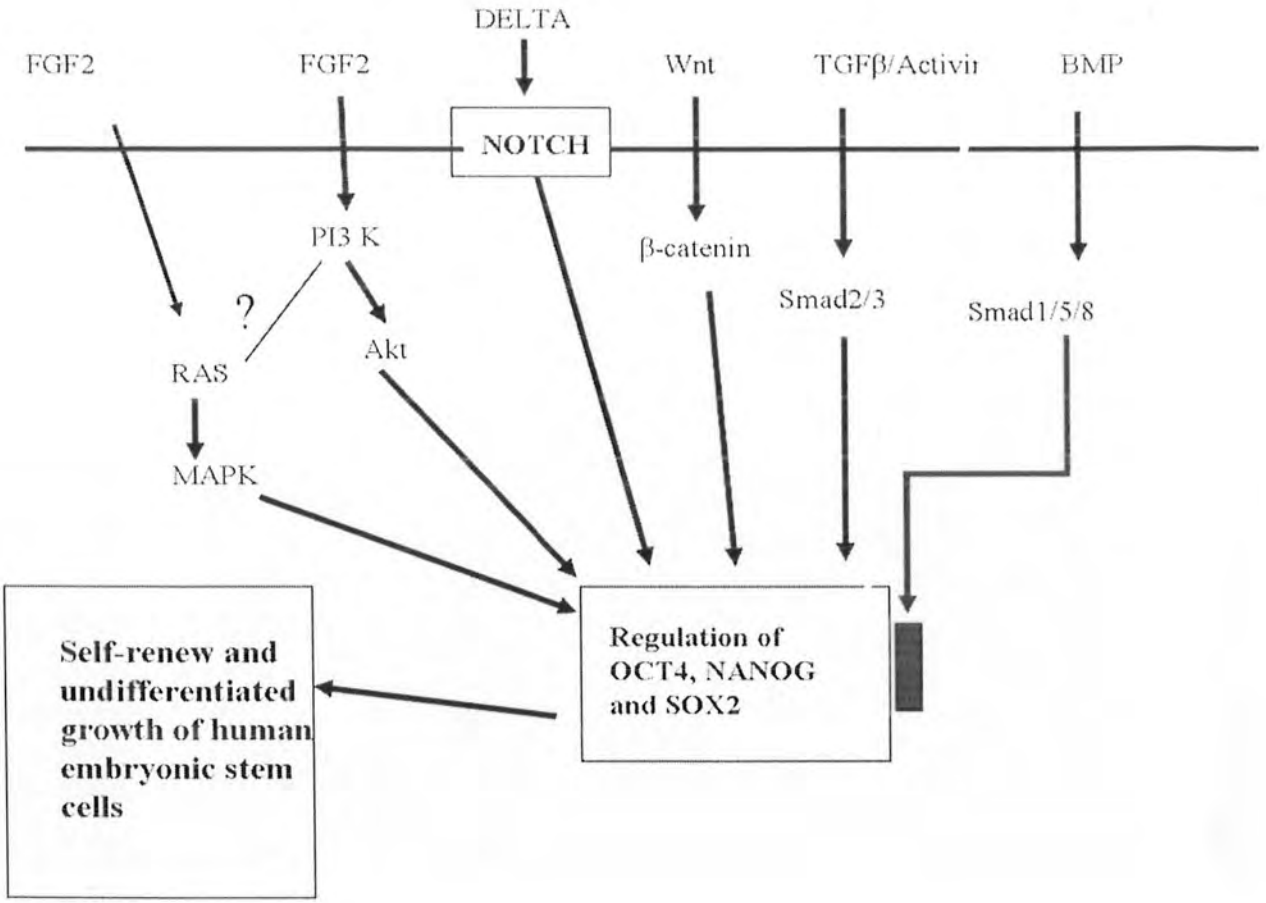

Figure 1.2: Self-renewal and undifferentiated growth of ES cells regulated by different signalling pathways.

Transcription factors

After complete description of signalling pathways, now the question arises that how these pathways maintain the pluripotency of hESCs. There are certain other factors that play a central role in this process. According to recent studies there is a transcriptional regulatory circuitry, responsible for the ES cells self-renewal and differentiation. This circuitry involves the transcription factor *Oct4*, *Sox2, Nanog* etc. Some of these factors may express specifically in pluripotent cells. These transcription factors are switched on/off through environment signals. When these gene are expressed the self renewal gene are activated and the differentiated genes are repressed, that's why the ES maintain their pluripotency (Liu *et al*,. 2007).

OCT4 is the most known member of these factors which functions to maintain pluripotency both *in vivo* and *in vitro*. This is a POU domain transcription factor that is specifically expressed in all pluripotent cells. The expression of *OCT4* is known to be regulated by a proximal enhancer and promoter in the epiblast and a distal enhancer and promoter at all other stages in the pluripotent cell lineage (Perrett *et al*., 2008).

NANOG is another member of the group of transcription factors whose functions are deemed essential for the process of self-renewal in human ESCs. *NANOG* is a NK2-family homeobox transcription factor and it acts by transcriptional activation, achieved by binding to homeobox domains in downstream target genes. Analogous to *OCT3/4* and *SOX2* and *NANOG* expression is high in human ESCs and is down regulated as cells differentiate. Transcription of *NANOG* is regulated by the binding of *OCT3/4* and *SOX2* to the *NANOG* promoter.

SOX2 gene encodes a member of the SRY-related HMG box (SOX) family of transcription factors involved in the regulation of embryonic development and the determination of cell fate, also this gene plays an important role in the maintenance of embryonic stem (ES) cell self-renewal and pluripotency. *SOX2* is expressed in the brain, retina, tongue, lungs, esophagus and stomach, and plays a key role in the differentiation and morphogenesis of these organs (Otsubo *et al.*, 2011).

SOX2 Gene

SOX2 is a gene located on chromosome 3q26.33 (Alvaro *et al*., 2010), which encodes a member of the SRY related HMG-box (SOX) family of transcription factors which

is involved in the regulation of embryonic development and determination of cell fate (Otsubo *et al.*, 2011). The *SOX2* gene belongs to a family of genes called *SOX* (SRY (sex determining region Y)-Box) genes. It is a single-exon gene encoding a protein with 317 residues which consists an N-terminal domain, a DNA binding high-mobility group domain and a transcriptional activation domain in the C-terminal *SOX2* protein (Wange *at al.*, 2007). The *SOX2* gene provides information for making a protein that plays a key role in the formation of different tissues and organs during embryonic development (Lengerke *et al.*, 2011).

SOX2 protein is especially involved in the development of eyes; also this protein regulates the activity of some other genes by attaching the specific regions of DNA that's why *SOX2* is called a transcription factor.

Pluripotency Governed by *SOX2* via Regulation of *OCT3/4*

SOX2 acts in combination with *OCT3/4* in pluripotency of human and mouse embryonic stem cells. The expression of most pluripotency-associated genes including *FGF4, UTF1, FBXO15, LEFTY1* and *NANOG* is regulated by an enhancer containing *OCT3/4* and *SOX2* binding motifs which is highly active in undifferentiated ES cells but not in differentiated cells. *OCT3/4* and *SOX2*, which bind independently to their respective binding motifs, act synergistically to activate these enhancers. Furthermore, the Oct-Sox enhancers are important in promoting the expression of *OCT3/4* and *SOX2* themselves, suggesting that these two transcription factors are regulated by a positive-feedback loop. This concept was supported by chromatin immunoprecipitation (ChIP) studies in both human and mouse ES cells, suggesting that *OCT3/4* and *SOX2* cooperatively activate or repress a set of genes (including *OCT3/4, SOX2,* and *NANOG*) through their SOX-enhancers which in turn results in their cooperative regulation of additional downstream genes. To determine the role of *SOX2,* Masui *et al,* established an inducible SOX2-null ES cell line. *SOX2* was indispensable for maintaining ES-cell pluripotency because SOX2-null ES cells differentiated primarily into trophoectoderm-like cells. *SOX2*, however, was not required for the activation of Oct–Sox enhancers. These enhancers remained active even after the depletion of *SOX2* protein. These findings suggest that *SOX2* regulates the expression of *OCT3/4* through the regulation of multiple transcription factors and plays an important role in ES cells (Masui *et al.*, 2007).

Induced Pluripotency

A cell (Mammalian cell) can be directly reprogrammed into induced pluripotent stem cell by enforcing expression of a few embryonic transcription factors (Stadtfeld and Hochedlinger, 2010; Maherali *et al*,. 2007). They are similar to pluripotent embryonic stem cells (ESCs) and can be derived from adult somatic cells (Yamanaka, 2007; Saha and Jaenisch, 2009). Therefore these IPSCs have a great potential for regenerative medicine. ESCs and IPSCs have similarities in much phenotypic behavior. For example cell morphology, expression of pluripotent markers, terotoma formation, ability to differentiate into germ layers etc. (Okita *et al*,. 2007; Wernig *et al*,. 2007).

Yamanaka and Takahashi converted mouse somatic cell to induced pluripotent cell by using only four transcription factors (*Oct4*, *Sox2*, *Klf4*, and *c-Myc*). These iPSCs closely resemble ESCs as they can restore the pluripotency associated transcriptional circuitry and epigenetic behavior (Maher ali and Hochedlinger, 2008). These four types of transcription factors are not stringently necessary in iPSC reprogramming, but some of them can be replaced with other factors (Takahashi and Yamanaka, 2006). Plath and colleagues find out more information about the roles of the four reprogramming factors. They suggested that *c-Myc* promoted the most prominent ESC like expression out of these four factors when expressed individually in fibroblasts, and *c-Myc* function more predominantly during reprogramming (Sridharan *et al*,. 2009).

Besides these four transcription factors (OSKM) as discussed above, Thomson and colleagues were able to reprogram human fibroblast with some other set of transcription factors including *OCT4*, *SOX2*, *NANOG*, and *LIN28*. This indicates that *KLF4* and *c-MYC* could be substitude with *NANOG* and *LIN28* (Yu *et al*,. 2007).

Numerous studies have demonstrated that these four types of transcription factors (OSKM) can be used for reprogramming of different types of somatic cells, including neural progenitor cells (NPCs), keratinocytes from the ectoderm, progenitor B cells from the mesoderm (Hanna *et al*,. 2008) and also the stomach cells and hepatocytes from the endoderm (Aoi *et al*,. 2008).

Scholer and colleagues further reduced the number of factors required for the reprogramming (Kim *et al*,. 2009). *Oct4* alone is also sufficient to reprogram neural stem cells (NSCs) to iPSCs. This finding also pointed out the role of *Oct4* in the

reprogramming of somatic cells, which generally do not express or express very low levels of *Oct4*. No transcription factor has yet been found to be able to replace the *Oct4* in reprogramming (Lengner *et al*,. 2008).

IPS mechanism proves that epigenetics is crucial for this procedure. Further more epigenetics play a part in the human development as well.

Chromatin and Epigenetic Modification

Embryonic stem cells differentiation from the pluripotent to developmentally more restricted state is accompanied by global changes in the genome expression patterns. There are some genes which are active in earlier stages and then gradually silenced at developmentally later stages and subsets of cell type-specific gene turned on. This is due to the result of active expression of transcription factors in concern with the chromatin remodeling and modification, includes covalent histone modification, DNA methylation of CpG dinucleotides and micrRNA etc (Liu *et al*,. 2007).

Chromatin and chromatin modification

To know more about the ES cells pluripotency many researcher have studied the chromatin. Some features of chromatin including nuclear architecture, chromatin structure, and histone modifications in ES cells are different from the somatic cells. For example ES cells chromatin shows characteristics of loosely euchromatin such as an abundance of acetylated histone modification and increase accessibility to nucleases (Boyer *et al*., 2006). ES cells undergo gene-specific remodeling of chromatin structure during in vitro differentiation. Histone deacetylase (HDACs) and methyl-CpG-binding protein (MECPs) are also expressed in ES cells and their levels regulated as cells undergo differentiation (Rao, 2004). As chromatin of pluripotent cell nuclei is an open conformation, some recent studies have shown that lineage-specific genes replicated earlier in pluripotent cells than differentiated cells and also having high levels of acetylated H3K9 and methylated H3K4. This modification is also combined with H3K27 trimethylation which is important for prevention of gene expression in ES cells (Azuara *et al*., 2006; Bernstein *et al*., 2006).

DNA methylation of CpG islands

DNA methylation of CpG islands is another common mechanism of gene silencing. This is required for induction of differentiation of ES cells, which was shown in

experiments with ES cells deficient either in the DNA mathyltransferases (Dnmt1, both Dnmt3a and Dnmt3b), or the CpG island-binding protein (CGBP) that binds to non-methylated DNA. These cells show severe DNA hypermathylation and a complete differentiation block (Jackson *et al.*, 2004; Carlone *et al.*, 2005). Hypermethylation at promoter region in differentiated cells typically results in decreased transcription of downstream genes. Hypomethylation in ES cells allows cells to maintain high level of gene expression and thus keeping them in pluripotent state. Further more, aberrant DNA methylation has been often reported to cause various human diseases.

micro-RNA

microRNA are thought to be a large family of small non-coding RNAs, which consist of of more than 200 known members in the mammalian genome. They are involved in many biological processes. For example, cell cycle regulation, apoptosis, cell differentiation and maintenance of stemness (Ambros, 2004; Griffiths-Jones, 2004). Various sets of these micrRNA are specifically expressed in pluripotent ES cells but not in differentiated embryonic bodies or in adult tissues, which suggest a role in the self-renewal of ES cells. When ES cells differentiate, they down regulate the stem cells maintenance genes and activate lineage-specific genes (Houbaviy *et al.*, 2003; Suh *et al.*, 2004). When loss of mature micrRNA occur in Dicer1 null mouse ES cells, failure of the respective mouse ES cells to differentiate into the three germ layers has been shown to occure. This suggests the importance of microRNA for ES cells pluripotency (Cheng *et al.*, 2005; Kanellopoulou *et al.*, 2005).

Single Nucleotide Polymorphism (SNP)

Single nucleotide polymorphism or 'SNP' is the variation in the sequence of the DNA among the individual of the same species or between paired chromosomes in an individuals. It is the most common variation occurs when a single nucleotide (A, T, C or G) in the genome is altered. For example SNP might change the DNA sequence AAGGCTAA to ATGGCTAA (Y. Liu *et al*,. 2010).

SNPs may occur both in coding and non coding regions of the genome. About 99% of the human

DNA sequences are the same and it may occur only 1% of the human population, some may have no effect on the cell function but some of them have been suggested

by the scientists that they predispose human to diseases. They are also evolutionarily stable (not changing much from generation to generation), making them easier to follow in population studies. Further more SNPs may also act as biological markers and due to their high density, they are ideal for studying the inheritance of genomic regions (Suzuki *et al*,. 2001).

Objective of the Research

Given the importance of the *Sox2* in maintaining the undifferentiated state and pluripotency, we intended to examine the variation in the promoter of *Sox2* gene in different population of Pakistan. Different ethnic groups of the general population (Pakhtoon and non-Pakhtoon origins) were visited to and their blood samples were collected for the current study. For this purpose we had to analyze upto 400 base pairs of its up-stream promoter region.

Materials and Methods

Material and Methods

Subjects

The families presented here for research purpose were mostly visited at their residential areas. During blood samples collection they were also interviewed to obtain information about their family background. For this purpose pakhtoon and non-Pakhtoon normal individuals were selected Table 2.1. Peripheral blood samples of 4-5 ml were collected from the available participants for the present study in ethylenediaminetetraacetic acid tubes (BD Vacutainer® K3 EDTA, Franklin Lakes NJ, USA) and stored at 4°C.

Extraction of Genomic DNA

Genomic DNA from peripheral blood, stored in EDTA tubes, was extracted using Kit method and standard Phenol-Chloroform method (Sambrook *et al.*, 1989).

Kit method

GF-1 Blood DNA Extraction Kit,(Vivantis, Malaysia) was used for genomic DNA extraction. The method applied for the genomic DNA extraction as provided in the protocol. In a 1.5 ml microcentrifuge tube (Axygen, Union, USA) 200 µl of buffer BB (blood lysis buffer provided in the kit) was added into a 200 µl of blood sample. After thoroughly mixing by pulsed-vortexing, 20µl of proteinase k was added and mixed immediately. Microcentrifuge tube containing these two solutions was then incubated at 65°C for 10 min. 200µl of absolute ethanol (BDH, Poole, England) was added; a homogeneous solution was obtained after thoroughly mixing. The sample was then transferred into a column assembled in a clean collection tube (provided in the kit).After centrifugation in a microcentrfuge (Eppendorf, Hamburg, Germany) at 5,000 rpm for 1 min, flow through was discarded. 500 µl of wash buffer1 provided in GF-1 Blood DNA Extraction Kit (vivantis, Malaysia) was added to the column and centrifuged at 5,000 rpm for 1 min, again through flow was discarded. Now the column was resuspended with 500 µl wash buffer2 (provided) and centrifuged at 5,000 rpm for 1 min, through flow was discarded and again 500µl of wash buffer2 was added to the column and centrifuged at maximum speed (14,000 rpm) for 3 minutes. The column was placed in a clean

microcentrifuge tube (Axygen, Union, USA).Now 100 µl of preheated Elution buffer, a DNA dissolving buffer provided in the kit was added and kept for 2 min.By centrifugation at 5,000 rpm for 1 min DNA was eluted and stored at 4°C.

Phenol-Chloroform method

In a 1.5 ml microcentrifuge tube (Axygen, Union, USA) equal volume (750 µl) of the blood and solution A [0.32 M Sucrose (BDH,Poole, England), 10 mM Tris of pH 7.5 (BDH, Poole England), 5 mM $MgCl_2$ (Sigma-Aldrich, St Louis, MO, USA)], 1% v/v Triton X-100 (Sigma-Aldrich, St Louis, MO, USA) were added and kept at room temperature for 10-15 minutes. The tube was then centrifuged for 1 minute at 13,000 rpm in a microcentrifuge (Eppendorf, Hamburg, Germany) and after discarding the supernatant, the pellet was resuspended in 400 µl of solution A. Centrifugation was repeated and after discarding the supernatant, the nuclear pellet was resuspended in 400 µl of solution B [10 mM Tris pH 7.5, 400 mM NaCl (BDH, Poole, England), 2 mM EDTA of pH 8.0 (BDH, Poole, England), 12 µl of 20% SDS (BDH, Poole, England)] and 6 µl of (20 mg/ml) proteinase K (Sigma-Aldrich, St Louis, MO, USA) and incubated at 37°C overnight. On the following day, 0.5 ml of a fresh mixture of equal volumes of solution C comprising solely Phenol (BDH, Poole, England) and D (24 volumes of Chloroform and 1 volume of Isoamyl alcohol (BDH, Poole, England) were added to the tube, mixed thoroughly and centrifuged for 10 minutes at 13,000 rpm. The aqueous phase was transferred to a new microcentrifuge tube and equal volume of solution D was added and recentrifuged at 13,000 rpm for 10 minutes. The aqueous phase was placed in a new tube and 55 µl sodium acetate (3 M, pH6) and equal volume of chilled isopropanol (BDH, Poole, England) was added. Tube was then inverted several times to precipitate the DNA and centrifugation was carried out again at 13,000 rpm for 10 minutes. The supernatant was discarded and the DNA pellet was washed with chilled 70% ethanol (BDH, Poole, England) and dried in vacuum concentrator 5301 (Eppendorf, Hamburg, Germany) at 37°C. After evaporation of residual ethanol, DNA was dissolved in 150 µl of DNA dissolving buffer composed of Tris-EDTA (Sigma-Aldrich, St Louis, MO, USA) and stored at 4°C.

DNA Quantification

Genomic DNA was quantified by taking optical density (OD) at 260nm in GeneRay UV-photometer

(Biometra,Goettingen, Germany) and subsequently diluted to 40-50 ng/μl for Polymerase chain reaction (PCR reaction).

Polymerase Chain Reaction

Polymerase Chain Reaction was performed using 40 ng of human genomic DNA (1μl) in 25 μl of reaction mixture containing 0.3 μl (10-20 pmol) of each primer, 0.2 μl (one unit) of Taq DNA polymerase, 0.5 μl (0.2 mM) of dNTP mix, 1.5 μl (1mM) $MgCl_2$, 2.5 μl of $10XNH_4SO_4$ buffer (MBI Fermentas, Life Sciences, York, UK) and 18.7 μl of PCR water. The reaction mixture was centrifuged for few seconds for thorough mixing. The standard thermal cycle conditions used included one cycle of denaturation at 96°C for 5 minutes, followed by 40 cycles of denaturation at 96°C for one minute, primers annealing at 50-60°C for one minute and primer extension or polymerization at 72°C for one minute, and final extension for 10 minutes at 72°C. This included initial denaturation cycle for one minute at 95°C, 48 cycles with 30 seconds denaturation at 95°C, 30 seconds annealing with progressively lowering temperature from 70 to 53°C at a rate of 1°C every third cycle and a primer extension of 40 seconds at 72°C, followed by 15 additional cycles with an annealing temperature of 58°C and final extension at 72°C for ten minutes (Frey *et al.*, 2008). The PCR was performed in T3000 thermocycler (Biometra,Gottingen, Germany).

Gel Electrophoresis

Agarose gel electrophoresis

Products of PCR were analyzed on 2% a garose gel, which was prepared by dissolving 2 g of high melting point agarose (Sigma-Aldrich, St Louis, MO, USA) in 100 ml 1X Tris-Borate-EDTA (Tris 89.1 mM, Borate 88.9 mM, EDTA 2.5 mM) buffer. To facilitate visualization of DNA after electrophoresis, 10 μl ethidium bromide (10 mg/ml) (Sigma-

Aldrich, St Louis, MO, USA) was added to the gel. Before loading to wells of the gel, 4.0 µl of PCR amplified products were mixed with equal volume of bromophenol blue dye (0.25% bromophenol blue in 40% Sucrose solution). Electrophoresis was performed at 100 volts for half an hour in a horizontal gel electrophoresis apparatus (Bio-Rad, Hercules, USA) having 1XTBE buffer. PCR amplified products were visualized by placing the gel on UV Transilluminator (Biometra, Gottingen, Germany).

Genomic DNA Amplification for Sequencing

To search for the variation in the promoter of *SOX2* gene among the different individuals of the general population of different areas, its promoter region about 400 bp were PCR-amplified, using primers sequences selected in that region. Primer set used for PCR with their optimal annealing temperature sequences given in Table 2.2.

First sequencing PCR

First sequencing PCR reaction is carried out at same conditions as described in Polymerase Chain Reaction. For first sequencing PCR reaction 1 µl (100 ng) genomic DNA of different individual, 2.5 µg (ng/µl) of forward and reverse primers of sox2 gene (promoter region), 3 µl $MgCl_2$ (25 mM), 5 µl 10X buffer (100 mM tris-HCl, 500 mM KCl with P^H 8.3), 1 µl dNTPs (10 mM), 0.2 µl (1 unit) of DNA polymerase (Fermentas, York, UK) with 34.8 µl of PCR water makes up a total of 50 µl reaction mixture for first sequencing PCR reaction.Polymearse chaion reaction was performed following conditions as described under polymerase chain reaction.

First purification

The amplified PCR product was tested on 2% agarose gel and purified by PureLink™ PCR purification kit (Invitrogen, CA, USA). Two hundred microliters of binding solution (H1, Concentrated Guanidine, HCl, EDTA, Tris-HCl and isopropanol) was added to amplification reaction mixture in eppendorf, the eppendorf containing amplification reaction mixture and H1 solution was applied to a spin at 13,000 rpm for two minutes in a microcentrifuge (Eppendorf, Hamburg, Germany), after spin the whole reaction mixture of eppendorf was transferred to a cartridge containing silica based membranes (where the double stranded DNA is selectively adsorbed) a column was also attached to bottom of

cartridge. 500 µl of alcohol containing H2 solution or washing solution (NaCl, EDTA, Tris-HCl) was added to cartridge that specifically removes unreacted primers and dNTPs. The cartridge was subjected to a spin of two minutes at 13,000 rpm in microcentrifuge (Eppendorf, Hamburg, Germany). As a result of centrifugation wash buffer settles down in column along with undesired elements and purified double stranded DNA was trapped in silica based membrane of cartridge. This trapped DNA was eluted by adding elution buffer (10 mM Tris-HCl, 0.1 mM EDTA with P^H 8.0) earlier placed at 65°C in a heater in cartridge. The cartridge was again centrifuged in a microcentrifuge (Eppendorf, Hamburg, Germany) at 13,000 rpm for 1 minute, with an empty eppendorf placed beneath the cartridge for collection of purified DNA eluted from cartridge.

Second sequencing PCR

The purified PCR products were subjected to cycle sequencing using ABI Prism Big Dye Terminator Cycle Sequencing Ready Reaction Kit v3.1 (PE Applied Biosystems). 10 µl sequencing PCR reaction contained 25 ng DNA template, 10 pmol forward or reverse primer, 1 µl 5X sequencing buffer, 1 µl ready reaction mix (RR) and 6 µl distilled water. Thermo-cycling conditions for sequencing included initial denaturation at 96°C for 3 minutes, followed by 30 cycles of denaturation at 96°C for 30 seconds, primer annealing at 50-60°C for 30 seconds, primer extension or polymerization at 72°C for 4 minutes and final extension for 10 minutes at 72°C.

Second purification

The PCR products were purified by ethanol precipitation in 1.5 ml microcentrifuge tube, containing 16 µl of distilled water and 64 µl 100% ethanol. Tubes were kept at room temperature for 10 minutes, and centrifuged at 13,000 rpm for 15 minutes. Supernatant was discarded and 200 µl of 70% ethanol was added into the tubes. After thorough mixing, tubes were again centrifuged at 13,000 rpm for 10 minutes. Supernatant was discarded and the pellet was resuspended in 15 µl of Hi-Di Formamide and transferred into 0.5 ml septa tubes to be directly sequenced in an ABI Prism 310 Automated DNA Sequencer (PE, Applied Biosystems, Foster City, CA, USA). Sequencing data was compared with sequences from National Center of Biotechnology Information (NCBI)

database to identify the causative variants through bioedit, sequence alignment software with bioedit editor version 6.0.7.

Variation Analysis

DNA sequence and chromatograms were obtained from ABI Prism 310 Automated DNA Sequencer, which was compared with wild type genomic sequence of *SOX2* gene obtained from Ensemble Genome browser to identify any nucleotide change (www.ensemble.org/index.htm).

Digestion with Restriction Enzyme

Restriction enzyme was used to digest the variation observed in the sequence of promoter area of *SOX2* of different individuals. MssI (PmeI) was found as a specific restriction enzyme through NeB Cutter software which cut the sequence at the respective nucleotides where variation was observed. Enzyme (PmeI) source is Methylobacterium species of Dd 5-732, (Fermentas, Life Sciences, York, UK Lot# 00084391) and the protocol for digestion of PCR products directly after amplification was used as provided by the supplier.

In a 1.5 ml microcentrifuge tube (Axygen, Union, USA), 10 µl of PCR reaction mixture (0.1-0.5 µg of DNA), 18 µl of Nuclease-free water, 2 µl of 10x buffer B (Fermentas, USA), 2 µl of MssI were added. After mixing the tube was spin down for few seconds and then incubated at 37 °C for 16 hours. The digested product was run on agarose gel to see the bands and compare it with the bands of normal PCR product.

Table 2.1: List of initial individuals selected for present study with their ethnic group, gender and residential areas. F. stands for female and M. for male individuals.

S.No.	Ethnic group	Gender	Age (years)	Residential Area
1.	Pakhtoon	M	26	Malakand Agency, K.P.K.
2.	Pakhtoon	M	26	Upper Dir, K.P.K.
3.	Non-Pakhtoon	M	26	DI Khan, K.P.K.
4.	Pakhtoon	M	25	Distt. Mardan, K.P.K.
5.	Pakhtoon	M	25	Malakand Agency, K.P.K.
6.	Non-Pakhtoon	F	25	Sialkot, Punjab.
7.	Non-Pakhtoon	F	25	Lahore, Punjab.
8.	Pakhtoon	F	26	DI Khan, Punjab
9.	Non-Pakhtoon	F	26	Sargodha, Punjab.
10.	Pakhtoon	F	26	Peshawar, KPK

Table2.2: List of primer sequences used to amplify *SOX2* gene promoter region.

No	Primer name		Sequence		Product	Ta
1	Sox2_Pro_F	5'	CGTCCCATCCTCATTTAAGC	3'	400bp	53 °C
2	Sox2_Pro_R	5'	GGTTTCTAGCGACCAATCAG	3'		

F = forward or left primer, R = reverse or right primer, bp = base pairs, Ta= optimal annealing temperature, ^{0}C = degree centigrade

Results

Results

The individuals selected for the current study were from different areas of Pakistan. A total of seventy (70) samples were used to carry out the current study, but primarily we only performed on the 10 samples. Details are given in the table 2.1. The individuals selected for the blood samples collection were mostly the members of Pakhtoon and Punjabi families of different areas including Dir, Malakand Agency, Swat, Dera Ismail Khan, District Mardan, Sargodha, and Chakwal.

The DNA was extracted and the sequence of the *Sox2* promoter region was sequenced. Only four of all these samples showed sequence variations (Polymorphism) when their sequencing was carried out and compared with wild type genomic sequence of *Sox2* gene promoter region obtained from Ensemble Genome browser. But the remaining samples possessed no variation and their sequences were the same as downloaded from the Ensemble.

The following are the detail description and result of those samples where we found a single nucleotide change in the promoter region of *sox2* gene.

Sample 1

This male individual lives in Bat khela, Malakand Agency, a well known area in North of Pakistan. Basically this is the province, Khyber Pakhtoon Khwa (KPK) where almost Pathan are resides. Blood sample of this person was collected for the current study to find out the sequence variation of *Sox2* promoter region if exist. To know about the family of this individual, this person has a normal family back ground having no abnormality. Basically they have average body weight and height.

Genomic DNA of this individual's blood sample was extracted (Figure: 3.1, 3.2), and the upstream promoter region (sequence shown in the figure 3.3) was Pcr amplified (as shown in Figure: 3.4). After genomic sequencing, variation was observed in the sequence of *Sox2* promoter region at 741 bp upstream region from the transcription start site, as in the wild type sequence the 'G' nucleotide was replaced with 'C' in the sequence obtained after sequencing of the respective gene region (Figure: 3.5A). This analysis was performed through BioEdit software.

Sample 4

As listed in the table 2.1, the area to which the blood donors basically belong is Dist. Mardan, KPK of Pakistan, where almost Pathans resides. As mentioned for the previous individuals, this person is also having normal family back ground, possessing normal characteristics of life. Blood sample was there fore collected for the current study.

As usual genomic DNA extraction was performed, bands were obtained on agarose gel (Figure: 3.1, 3.2). *Sox2* up-stream region was PCR amplified (Figure: 3.4) to proceed the process further towards our main objectives. After amplification genomic sequencing was performed to see any nucleotide change in the whole sequence of *Sox2* up-stream region. While aligning through BioEdit software a single nucleotide change was observed at the same region i.e. 741 bp up-stream region from the transcription start site, as it was observed for the sample 1. Here also the 'G' nucleotide of the wild type genomic sequence was replaced by 'C' after sequencing (Figure 3.5B).

Sample 5

This member also resides in the same area of Malakand agency as explained for the sample 1 (shown in the table: 2.1). Family has normal back ground history as; they are all basically normal individuals, having normal body height and weight. Blood sample was collected to study the *Sox2* up-stream region status.

Genomic DNA extraction was performed for further analysis, the bands were successfully observed on agarose gel (Figure: 3.1, 3.2), following by PCR to amplify the respective region (Figure: 3.3). PCR product (Figure: 3.4) was subjected to sequence the *Sox2* up-stream region to see if there is any variation in any of the nucleotide of the whole genomic sequence. In this case also the variation was observed in the same area of the *Sox2* up-stream region. The nucleotide 'G' of wild type genome was replaced by 'C' according to the data obtained after analyzing through BioEdit (Figure: 3.5C).

Sample 10

This sample was taken from a female individual basically resident of Peshawar (also shown in the table 2.1), a well known city in KPK, Pakistan. This individual belong a Pathan family, having normal features as discussed for the above members. They have a normal body structure, weight and height. Blood sample was taken from this person for the current study.

Genomic DNA extraction was successfully performed phenol chloroform method. The DNA bands were analyzed as shown in the figure (Figure: 3.1, 3.2), then the DNA was subjected to amplify the *Sox2* promoter area through PCR (Figure: 3.4), to see the nucleotide change in the sequence of the respective gene part, the genomic sequencing was carried out. The sequenced data was further analyzed through BioEdit. Interestingly the variation was found again exactly at the same nucleotide at 741 bp up-stream region of the transcription start site, and the 'G' was replaced by 'C' nucleotide as shown in the figure 3.5D.

DNA Extraction (Comparison between Kit and Manual Method)

Genomic DNA was extracted by both i.e. Kit and Phenol Chloroform (Manual) method. During the analysis of DNA bands on the agarose gel, some comparative points were noted as follow:

- DNA extracted by following kit method, having bands less visible on agarose gel as compared to the bands of DNA we extracted by performing with manual method (shown in the figure 3.1, 3.2). The reason is that kit method provides us less concentrated DNA while that of manual method having more concentration.
- Kit method of DNA extraction involves more centrifugation steps due to which degradation of DNA may occurs while manual procedure involves less centrifugation i.e. less chances of DNA degradations.
- The kit method have a beneficial feature as it take less time (1-2 hours) to extract the genomic DNA while, by performing through manual method it takes 2-3 days to complete the whole extraction process.

Hence the manual DNA extraction method gives us more yields of DNA and more bright bands on agarose gel as compared to the Kit method, so later on this method was followed for further DNA extraction.

Samples 1-10 (kit method)

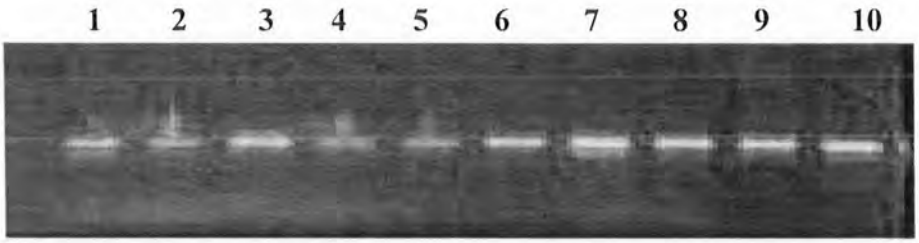

Figure 3.1: Electropherogram of ethidium bromide stained 1 % agarose gel of Genomic DNA, extracted by Kit method for sample 1-10.The serial number 1 to 10 indicates the individual number as listed in the table 2.1

Samples 1-10 (Phenol-chloroform Method)

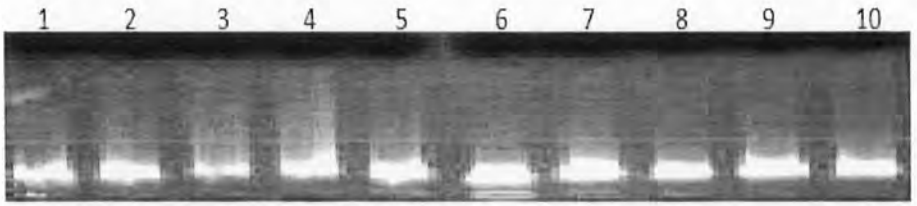

Figure 3.2: Electropherogram of ethidium bromide stained 1 % agarose gel of Genomic DNA, extracted by manual method for sample 1-10.The serial number 1 to 10 indicates the individual number as listed in the table 2.1

About 400 bp of up-stream promoter region of *SOX2* was PCR amplified by using two sets of Primers selected in the respective region, as shown in the figure 3.3. The sequence highlighted shows the primers sequences.

FP → TACCCTGCACCAAAAAGTAAATCAATATTAAGTTTAAAGAA

AAAAAAACCCACGTAGTCTTAGTGCTGTTTACCCACTTCCTTCGAAAAGGCGTGTGGTGT

GACCTGTTGCTGCGAGAGGGGATACAAAGGTTTCTCAGTGGCTGGCAGGCTGGCTCTGGG

AGCCTCCTCCCCCTCCTCGCCTGCCCCCTCCTCCCCCGGCCTCCCCCGCGCGGCCGGCGGC

GCGGGAGGCCCCGCCCCCTTTCATGCAAAACCCGGCAGCGAGGCTGGGCTCGAGTGGAGG

AGCCGCCGCGCG ← RP

Figure 3.3: Sequence of *SOX2* promoter region. The arrow indicates forward and reverse primers respectively.

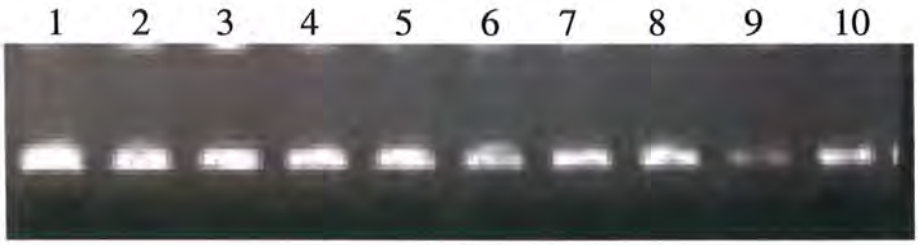

Figure 3.4: Electropherogram of ethidium bromide stained 2 % agarose gel of the PCR amplified product, showing samples 1 to 10. The serial number indicates the individual number as listed in the table 2.1.

Sequencing Results

As discussed previously, single nucleotide change (G→ C) observed in the sequence of samples at serial number; 1, 4, 5, and 10 (Figure: 3.5). Concentrating on the whole population we studied, it is concluded that this variation was only observed in the samples we collected from Pakhtoon origins, and the remaining samples possessed no variation at any nucleotide and their sequences were the same as shown through Ensemble.

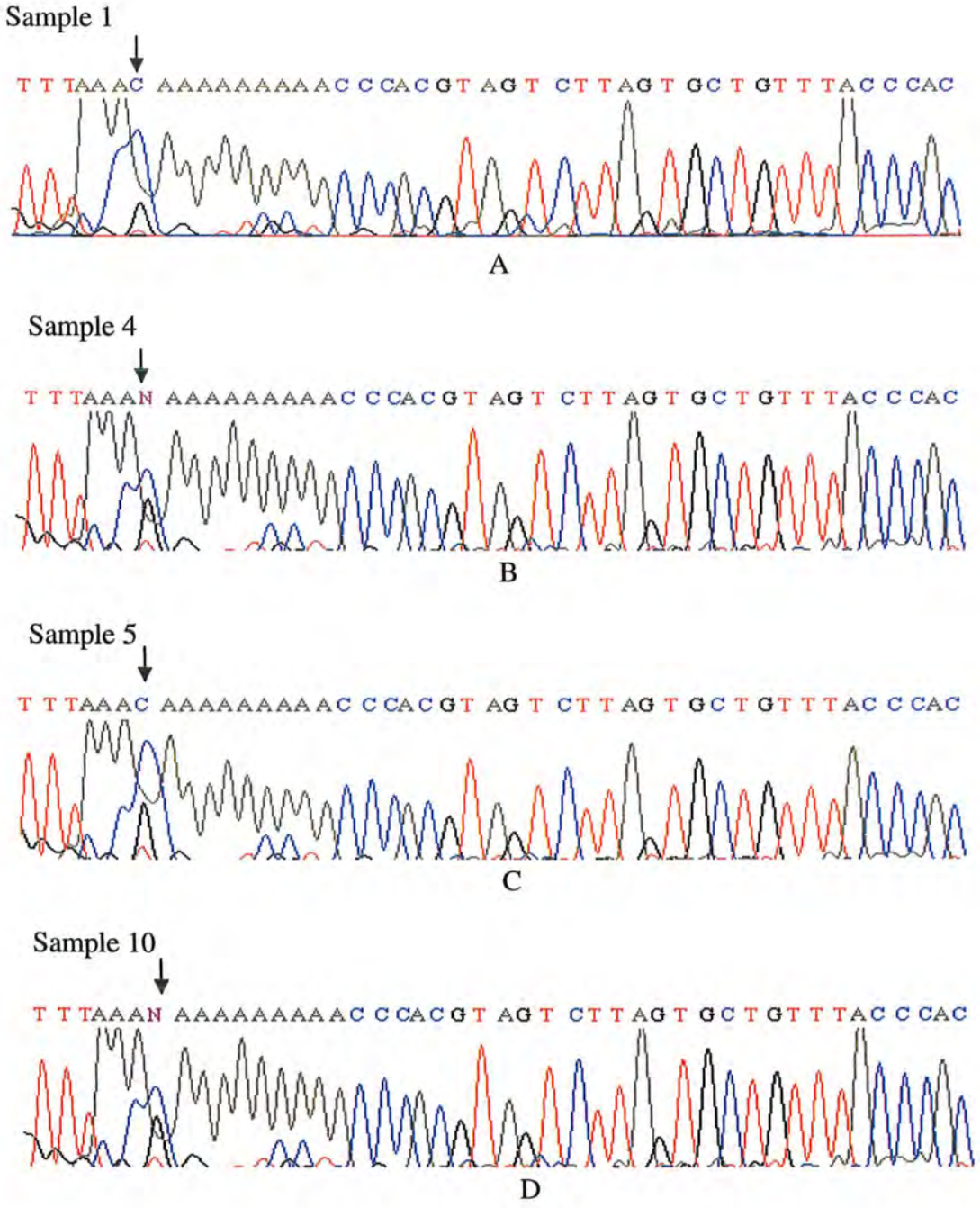

Figure 3.5: Sequence analysis of *Sox2* promoter in different individuals .The arrow shows variation at respective nucleotide of the up-stream promoter sequence; (A) Sequence analysis of sample 1 (B) Sequence analysis of sample 4 (C) Sequence analysis of sample 5 (D) Sequence analysis of sample 10

Figure 3.6: *Sox2* up-stream promoter sequence. The region highlighted as red shows the enzyme specificity for that region. The downward arrow indicates the exact nucleotide (at 741 bp up-stream region from the start site) 'G' replaced by 'C'.

Digestion with Restriction Enzyme (RE)

The variation showed at the nucleotide 'C' created a site for the restriction enzyme. The enzyme found for the respective site has the specificity shown in the figure 3.7. This site is also mentioned in the *Sox2* promoter sequence downloaded through Ensemble genome browser shown in the figure 3.6, highlighted as red. The arrow indicates G to C transition (G ⟶ C).

5....GTTT↓AAAC...3′

3′....CAAA↑TTTG....5′

Figure 3.7: Restriction enzyme (PmeI) sequence. Arrows indicates the specific digestive sites.

Other population of 60 individulas was also subjected for the current study and their PCR Products were treated with Restriction Enzyme (PmeI) according to the protocol discussed in chapter 2. The purpose of the enzyme treatment was to search for the variation existence in other groups of the general population. Many samples were treated with the respective enzyme to see the digested bands on the agarose gel if there is change in the same nucleotide exist where we found earlier.

After treatment with the RE, no digestion has been observed (Figure: 3.8).The RE activity was also tested by repeating the same process for those samples where we had already the variation occurred (sample 1, 4, 5, and 10), but again no digestion

occurred (Figure: 3.9). The process was repeated 4-5 times to confirm for the existence of variation in the *Sox2* up-stream region, but no digestion occurred.

Hence, it shows that the observed polymorphism in the sequence of sample 1, 4, 5, and 10 was due to the error in the sequencing as the enzyme didn't cleave the PCR product from the observed polymorphic samples.

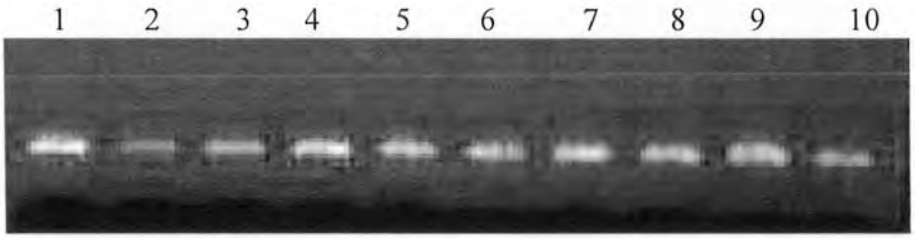

Figure 3.8: Electropherogram of ethidium bromide stained 2 % agarose gel, showing no activity of the restriction enzyme. Samples 1 to 10 are the PCR amplified products treated with restriction enzyme (PmeI).

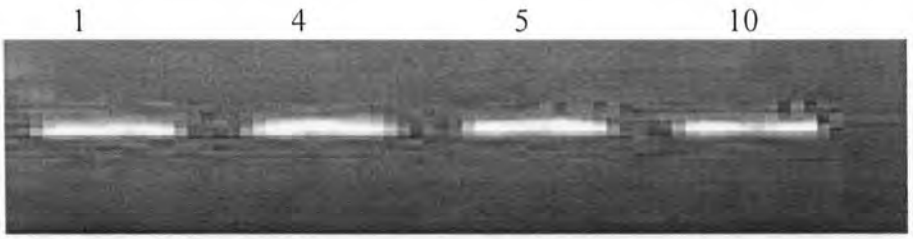

Figure 3.9: Electropherogram of ethidium bromide stained 2 % agarose gel, showing no activity of the restriction enzyme. Samples 1, 4, 5, and 10 are the PCR amplified products (of individuals having variation observed previously) treated with restriction enzyme (PmeI).

Discussion

Discussion

Human embryonic stem (ES) cells, derived from the inner cell mass of the mammalian blastocyst are pluripotent cells that can differentiate into multiple cell lineages and can give rise to the three germ layers (endoderm, mesoderm and ectoderm). ES cells have the capacity of self renewal and because of these characteristics, ES cells are ideal models for studying molecular mechanisms that determine cell fate, and have the potential to be utilized in replacement and regenerative therapy for treating a variety of human diseases.

Pluripotency of ES is maintained by a limited set of transcription factors, designated "core" pluripotency factors and when introduced in combination can re-program differentiated cells back to a pluripotent state (Okita *et al.*, 2007; Takahashi and Yamanaka, 2006; Wernig *et al.*, 2007; Yu et al., 2007). The key pluripotency factors include *Oct4*, *Sox2*, *Nanog*, and *Klf4* target genes and function in an extensive regulatory circuit that silences the expression of transcription factors required for differentiation and activates the expression of genes important for maintenance of pluripotency (Jiang *et al.*, 2008; Kim *et al.*, 2008; Loh *et al*,. 2006). Other important factors, including c-Myc, regulate a distinct set of genes and can enhance reprogramming of adult cells to the pluripotent state. The delicate balance between gene activation and repression may be regulated by the extent of promoter co-occupancy by the four different pluripotency factors. Silent promoters are generally bound by a single factor while actively transcribed promoters are simultaneously bound by multiple pluripotency factors (Kim *et al.*, 2008). The correlation between promoter occupancy by pluripotency factors and gene expression also coincides with specific chromatin modifications (Kim *et al.*, 2008).

In present study upstream promoter region of ESC pluripotency factor, *SOX2* gene was analyzed for the existence of any sequence variant (polymorphism) in general Pakistani population. Seventy normal individuals were included in study subject selected from different ethnic groups. Genomic DNA was extracted from blood samples collected from these individuals. The sequence variant was hunted by using specific primer set for the promoter region of *SOX2* gene. After PCR amplification, sequence analysis was carried out to find any sequence variation. Four Pakhtoon groups showed G to C transition in the upstream promoter region of *SOX2* while the

remaining members possessed normal sequence. PCR amplified DNA products in which sequence variant were detected were treated with specific restriction enzyme (PmeI) to confirm polymorphism. Agarose gel electrophoresis after 16 hours of enzyme treatment revealed that the region in which variation was detected is not digested by restriction enzyme, suggesting that the sequence variation was not a polymorphism but it would be due to any sort of error in sequencer reading.

SOX2 is a single-exon gene that lies in an intron of the SOX2OT (*SOX2* overlapping transcript) gene on chromosome 3q26.33 encoding a 317-amino acid protein (Fantes *et al.*, 2003; Stevanovic *et al.*, 1994). *Sox2* is one of the key transcription factors required in induced pluripotent stem cells (Zhao *et al.*, 2008). As forced *Oct4* expression induces pluripotency in *Sox2*-null cells, a group of researchers concluded that the primary role of *Sox2* in induced pluripotent stem cells is controlling *Oct4* expression, and they perpetuate their own expression when expressed concurrently (Masui *et al.*, 2007).

SOX2 coexpresses with *OCT4* starting very early in embryogenesis. These factors are found within the inner cell mass (ICM) of the blastocyst of the pre-implantation embryo. Expression of both factors persists within the epiblast, the tissue that differentiates into the embryo and germ cells after blastocyst implantation. *SOX2* and *OCT4* are also expressed in Embryonic Stem Cells (ES) which are typically, but not always, derived from the ICM of blastocysts. The importance of *SOX2* and *OCT4* as regulators of pluripotency has been dramatically illustrated by the demonstration that these factors together with c-Myc and *Klf4* or Nanog and LIN28 can induce the dedifferentiation of somatic cells into induced pluripotent stem cells (iPS) with many of the features of embryonic stem cells (Takahashi and Yamanaka, 2006; Wernig *et al.*, 2007; Yu *et al.*, 2007). The successful replacement of embryonic stem cells with induced pluripotent stem cells for scientific research and as candidates for clinical therapies will require an extensive knowledge of the roles played by *SOX2* and *OCT4* as gate-keepers of toti- and pluripotency.

SOX2 and *OCT4* bind DNA through their HMG and POU domains, respectively. *SOX2* and *OCT4* regulate the expression of each other which is essential for stabilization of pluripotency within cells, such as ES cells (Masui *et al.*, 2007). *SOX2* and *OCT4* regulate the expression of *Nanog*, a transcription factor that co-occupies many of the

same genes promoters (Rodda *et al.*, 2005). Nanog is also a very important early regulator of pluripotency. Together *SOX2*, *OCT4* and Nanog co-regulate a growing list of downstream target genes. Target genes include *YES1*, *FGF4*, *UTF1*, *Fbx15*, *Zic3* and *ZFP206*, but this is only a sampling of the hundreds of genes that are involved. The targets of *SOX2*, *OCT4* and *Nanog* have recently been identified using time course microarray and genome-wide immunoprecipition data (Sharov *et al.*, 2008).

Loss of function *SOX2* mutations have been linked to the rare disease microphthalmia syndrome type 3, small eye (MCOPS3) (Ragge *et al.*, 2005; Verma and Fitzpatrick, 2007).

Conclusion

In present study we suggested that *SOX2* is very important gene, especially involved in the regulation of embryonic stem cells pluripotency and play a key role in the developmental process during embryonic stages. The lack of polymorphism shows that sequence in the upstream promoter region of *SOX2* gene is highly conserved. Because of its more importance in the human embryos, the evolution has not allowed any variation in it.

References

References

Ambros V (2004). The functions of animal microRNAs. Nature 431: 350–355

Amit M, Carpenter MK, Inokuma MS, Chiu CP, Harris CP, Waknitz MA, Itskovitz-Eldor J, Thomson JA. (2000). Clonally derived human embryonic stem cells lines maintain pluripotency and proliferative potential for prolonged periods of culture. Dev Biol 227: 271–278

Amit M, Itskovitz-Eldor J (2002). Derivation and spontaneous differentiation of human embryonic stem cells. J Anat 200: 225–232

Aoi T, Yae K, Nakagawa M, Ichisaka T, Okita K, Takahashi K, Chiba T, Yamanaka S (2008). Generation of pluripotent stem cells from adult mouse liver and stomach cells. Science 321: 699–702

Armstrong L, Hughes O, Yung S, Hyslop L, Stewart R, Wappler I, Peters H, Walter T, Stojkovic P, Evans J, Stojkovic M, Lako M (2006). The role of PI3K/AKT, MAPK/ERK and NFkappabeta signalling in the maintenance of human embryonic stem cell pluripotency and viability highlighted by transcriptional profiling and functional analysis. Hum Mol Genet 15: 1894–1913

Assady S, Maor G, Amit M, Itskovitz-Eldor J, Skorecki KL, Tzukerman M (2001). Insulin production by human embryonic stem cells. Diabetes 50: 1691–1697

Azuara V, Perry P, Sauer S, Spivakov M, Jorgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merkenschlager M, Fisher AG (2006). Chromatin signatures of pluripotent cell lines. Nat Cell Biol 8: 532–538

Babaie Y, Herwig R, Greber B, Brink TC, Wruck W, Groth D, Lehrach H, Burdon T, Adjaye J (2007). Ana lysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. Stem Cells 25: 500–510

Beattie GM, Lopez AD, Bucay N, Hinton A, Firpo MT, King CC, Hayek A. (2005). Activin A maintains pluripotency of human embryonic stem cells in the absence of feeder layers. Stem Cells 23: 489–495

Berger J, Suzuki T, Senti KA, Stubbs J, Schaffner G (2001). Genetic mapping with SNP markers in Drosophila. Nat Genet 29: 475–481

Bernstein BE, Mikkelsen TS, XieX, 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

Bongso A, Fong CY, Ng SC (1994). Isolation and culture of inner cell mass cells from human blastocysts. Hum Reprod 9: 2110–2117

Boyer LA, Mathur D, Jaenisch R (2006). Molecular control of pluripotency. Curr Opin Genet Dev 16: 455–46

Brivanlou AH, Gage FH, Jaenisch R (2003). Stem cells. Setting standards for human embryonic stem cells. Science 300: 913–916

Burdon T, Stracey C, Chambers I, Nichols J, Smith A (1999). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. Dev Biol 210:30–43

Bylund M, Andersson E, Novitch BG, Muhr J (2003). Vertebrate neurogenesis is counteracted by Sox1-3 activity. Nature Neurosci 6: 1162-1168

Caplan AI (1994). The mesengenic process. Clin Plast Surg 21: 429–435

Carlone DL, Lee JH, Young SR, Dobrota E, Butler JS, Ruiz J, Skalnik DG (2005). Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. Mol Cell Biol 25: 4881–4891

Carpenter MK, Rosler E, Rao MS (2003). Characterization and differentiation of human embryonic stem cells. Cloning Stem Cells 5: 79–88

Chambers I, Colby D, Robertson M (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell 113: 643–655

Cheng LC, Tavazoie M, Doetsch F (2005). Stem cells: From epigenetics to microRNAs. Neuron 46: 363–367

Cowan CA, Klimanskaya I, McMahon J, Atienza J, Witmyer J, Zucker JP, Wang S, Morton CC, McMahon AP, Powers D, Melton DA (2004). Derivation of embryonic stem-cell Lines from human blastocysts. N Engl J Med 13:1353-1356

Dvorak P, Dvorakova D, Koskova S, Vodinska M, Najvirtova M, Krekac D, Hampl A (2005). Expression and potential role of fibroblast growth factor 2 and its receptors in human embryonic stem cells. Stem Cells 23: 1200–1211

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

Fantes J, Ragge NK, Lynch SA, McGill NI, Collin JR, Howard-Peebles PN, Hayward C, Vivian AJ, Williamson K, van Heyningen V, FitzPatrick DR (2003). Mutations in SOX2 cause anophthalmia. Nature Genet 33: 461-462

Fontes PA, Thomson AW (1999).Stem cell technology. BMJ 319: 319-1308

Ginis I, Luo Y, Miura T (2004). Differences between human and mouse embryonic stem cells. Dev Biol 269: 360–380

Griffiths-Jones S (2004). The microRNA Registry. Nucleic Acids Res 32: 109–111

Gronthos S, Brahim J, Li W, Fisher LW, Cherman N, Boyde A, DenBesten A, Robey P G, Shi S (2002). Stem cells properties of human dental pulp stem cells. J Dent Res 81: 531-535

Hanna J, Markoulaki S, Schorderet P, Carey BW, Beard C, Wernig M, Creyghton MP, Steine, EJ, Cassady, JP, Foreman, R (2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell 133: 250–264

Henderson JK, Draper JS, Baillie HS, Fishe Sl, Thomson JA, Moore H, Andrews PW (2002). Preimplantation human embryos and embryonic stem cells show comparable expression of stage-specific embryonic antigens. Stem Cells 20: 329–337

Hogan B, Tilly R (1977). In vitro culture and differentiation on normal mouse blastocysts. Nat 265: 626–629

Houbaviy HB, Murray MF, Sharp PA (2003). Embryonic stem cell-specific MicroRNAs. Dev Cell 5: 351–358

Ivanova NB, Dimos JT, Schaniel C (2002). A stem cell molecular signature. Science 298: 601–604

Jackson M, Krassowska A, Gilbert N, Chevassut T, Forrester L, Ansell J, Ramsahoye B (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol Cell Biol 24: 8862–8871

Jiang J, Chan YS, Loh YH (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol 10: 353–360

Jiang Y, Jahagirder BN, Reinhardt RL (2002). Pluripotency of mesenchymal stem cells derived from adult bone marrow. Nat 418: 41–49

Kanellopoulou C, Muljo SA, Kung AL, Ganesan S, Drapkin R, Jenuwein T, Livingston DM, Rajewsky K (2005). Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. Genes Dev 19: 489–501

Keiko A, Hirofumi S, Shin-ya Y, Norio N, Eihachiro K (2010). Role of SOX2 in maintaining pluripotency of human embryonic stem cells. Genes to Cells 15: 455–469

Kim J, Chu J, Shen X, Wang J, Orkin SH (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell 132: 1049–1061

Kim JB, Sebastiano V, Wu G, Arauzo-Bravo MJ, Sasse P, Gentile L, Ko K, Ruau D, Ehrich M, van den Boom D (2009). Oct4-induced pluripotency in adult neural stem cells. Cell 136: 411–419

Laga AC, Lai CY, Zhan Q, Huang SJ, Velazquez E F, Yang Q, Hsu MY, Murphy GF (2010). Expression of the embryonic stem cell transcription factor SOX2 in human Skin. Am J Pathol 176: 903-913

Lengner CJ, Welstead GG, Jaenisch R (2008). The pluripotency regulator Oct4: a role in somatic stem cells? Cell Cycle 7: 725–728

Lim JWE, Bodnar A (2002). Proteome analysis of conditioned medium from mouse embryonic fibroblast feeder layers which support the growth of human embryonic stem cells. Proteomics 2: 1187–1203

Liu NA, Min LU, Tian X, Han Z (2007). Molecular mechanisms involved in self-renewal and pluripotency of embryonic stem cells. J. Cell. Physiol 211: 279–286

Liu Y, Mark Li, Yiu M, Cheung PC, Sham, Michael K, Ng (2010). SKM-SNP: SNP markers detection method. Journal of Biomedical Informatics 43: 233–239

Loh YH, Wu Q, Chew JL, (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet 38: 431–440

Lopez-Schier H, St Johnston D (2001). Delta signaling from the germ line controls the proliferation and differentiation of the somatic follicle cells during Drosophila oogenesis. Genes Dev 15: 1393–1405

Machold R, Hayashi S, Rutlin M, Muzumdar MD, Nery S, Corbin JG, Gritli-Linde A, Dellovade T, Porter JA, Rubin LL, Dudek H, McMahon AP, Fishell G (2003). Sonic hedgehog is required for progenitor cell maintenance in telencephalic stem cell niches. Neuron 39: 937–950

Maherali N, Hochedlinger K (2008). Guidelines and Techniques for the Generation of Induced Pluripotent Stem Cells. Cell Stem Cell 3: 595–605

Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Yachechko R, Tchieu J, Jaenisch, R (2007). Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 1: 55–70

Martin GR, Evans MJ (1975). Differentiation of clonal lines of teratocarcinoma cells: formation of embryoid bodies in vitro. Proc Natl Acad Sci USA 72: 1441–1445

Masui S, Nakatake Y, Toyooka Y, Shimosato D, Yagi R, Takahashi K, Okochi H, Okuda A (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nature Cell Biology 9: 625–635

Michaelidis TM, Lie DC (2008). Wnt signaling and neural stem cells: caught in the Wnt web. Cell Tissue Res. 331: 193–210

Mitalipov S, Kuo HC, Byrne J, Clepper L, Meisner L, Johnson J, Zeier R, Wolf D (2006). Isolation and characterization of novel rhesus monkey embryonic stem cell lines. Stem Cells 24: 2177–2186

Moore JC, Atze K, Yeung PL, Toro-Ramos AJ, Camarillo C, Thompson K, Ricupero CL, Brenneman MA, Cohen RI, Hart RP (2010). Efficient, high-throughput transfection of human embryonic stem cells. Stem Cell Res Ther 6: 1-10

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

Otsubo T, Akiyama Y, Hashimoto Y, Shimada S, Goto K, Yuasa Y (2011). MicroRNA-126 Inhibits SOX2 expression and contributes to gastric carcinogenesis. PLoS ONE 6: 137-6617

Paling NR, Wheadon H, Bone HK, Welham MJ (2004). Regulation of embryonic stem cell self-renewal by phosphoinositide 3-kinase-dependent signaling. J Biol Chem 279: 48063–48070

Pankratz MT, Li XJ, Lavaute TM, Lyons EA, Chen X, Zhang SC (2007). Directed neural differentiation of human embryonic stem cells via an obligated primitive anterior stage. Stem Cells 25: 1511–1520

Papaioannou VE, McBurney MW, Gardner RL, Evans MJ (1975). Fate of teratocarcinoma cells injected into early mouse embryos. Nat 258: 70–73

Qi X, Li TG, Hao J, Hu J, Wang J, Simmons H, Miura S, Mishina Y, Zhao GQ (2004). BMP4 supports self-renewal of embryonic stem cells by inhibiting mitogen-activated protein kinase pathways. Proc Natl Acad Sci USA 101: 6027–6032

Ragge NK, Lorenz B, Schneider A, Bushby K, de Sanctis L, de Sanctis U, Salt A, Collin JR, Vivian AJ, Free SL, Thompson P, Williamson KA, Sisodiya SM, van Heyningen V, Fitzpatrick DR (2005). SOX2 anophthalmia syndrome. Am J Med Genet A 135: 1-7

Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA (2002). "Stemness": Transcriptional profiling of embryonic and adult stem cells. Science 298: 597–600

Rao M (2004). Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem cells. Dev Biol 275:269–286

Reubinoff BE, Pera MF, Fong CY, Trounson A, Bongso A (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. Nat Biotechnol 18: 399–404

Richards M, Fong CY, Chan WK (2002). Human feeders support prolonged undifferentiated growth of human inner cell masses and embryonic stem cells. Nat Biotechnol 20: 933–936

Rodda DJ, Chew JL, Lim LH, Loh YH, Wang B, Ng HH, Robson P (2005). Transcriptional regulation of nanog by OCT4 and SOX2. J Biol Chem 280: 24731-24737

Ruohola H, Bremer KA, Baker D, Swedlow JR, Jan LY (1991). Role of neurogenic genes in establishment of follicle cell fate and oocyte polarity during oogenesis in Drosophila. Cell 66: 433–449

Saha K, Jaenisch R (2009). Technical challenges in using human induced pluripotent stem cells to model disease. Cell Stem Cell 5: 584–595

Sato N, Meijer L, Skaltsounis L, Greengard P, Brivanlou AH (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med 10: 55–63

Sharov AA, Masui S, Sharova LV, Piao Y, Aiba K, Matoba R, Xin L, Niwa H, Ko MS (2008). Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. Genomics 9:269- 278

Solter D, Skreb N, Damjanov I (1970). Extrauterine growth of mouse egg- cylinders results in malignant teratoma. Nat 227: 503–504

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

Stevanovic M, Zuffardi O, Collignon J, Lovell-Badge R, Goodfellow P (1994). The cDNA sequence and chromosomal location of the human SOX2 gene. Mammalian Genome 5: 640-642

Stevens LC (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. Dev. Biol. 21: 364–382

Suh MR, Lee Y, Kim JY, Kim SK, Moon SH, Lee JY, Cha KY, Chung HM, Yoon HS, Moon SY, Kim VN, Kim KS (2004). Human embryonic stem cells express a unique set of microRNAs. Dev Biol 270: 488–498

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

Takahashi K, MurakamiM, Yamanaka S (2005). Role of the phosphoinositide 3-kinase pathway in mouse embryonic stem (ES) cells. Biochem Soc Trans 33: 1522–1525

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

Valdimarsdottir G, Mummery C (2005). Functions of the TGFbeta superfamily in human embryonic stem cells. APMIS 113: 773–789

Verma AS, Fitzpatrick DR (2007). Anophthalmia and microphthalmia. Orphanet J Rare Dis 2: 47-55

Wernig M, Meissner A, Foreman R (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature 448: 318–324

Wianny F, Bourillot PY, Dehay C (2011). E mbryonic stem cells in non-human primates: An overview of neural differentiation potential. Elsevier 1: 142–152

Wright NA (2000). Epithelial stem cell repertoire in the gut: Clues to the origin of cell lineages, proliferative units and cancer. Int J Exp Pathol 81: 117–143

Xu C, Inokuma MS, Denham J, Golds K, Kundu P, Gold JD, Carpenter MK (2001). Feeder-free growth of undifferentiated human embryonic stem cells. Nat Biotechnol 19: 971–974

Yamanaka, S (2007). Strategies and new developments in the generation of patient-specific pluripotent stem cells. Cell Stem Cell 1: 39–49

Ying QL, Nichols J, Chambers I, Smith A (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell 115: 281–292

Yu F, Yao H, Zhu P, Zhang X, Pan Q, Gong C, Huang Y, Hu X, Su F, Lieberman, J, and Song E (2007). Let-7 regulates self renewal and tumorigenicity of breast cancer cells. Cell 131:1109–1123

Yu J, Vodyanik MA, Smuga-Otto K (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science 318: 1917–1920

Zhao R, Daley GQ (2008). From fibroblasts to iPS cells: induced pluripotency by defined factors. J Cell Biochem 105: 949–55