The Transcriptomic Profiling Of Murine Embryonic Stem Cell-derived Mesodermal Cell Lineages And Functional Cardiomyocytes: An Insight Into Mesodermal Patterning

Inaugural-Dissertation

Zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Michael Xavier Doss.J

Aus

Paramakudi, TN, Indien.

Köln, Februar 2007

Referees/Berichterstatter:

Prof. Agapios Sachinidis

Prof. Jürgen Hescheler

Prof. Diethard Tautz

Date of oral examination : 25th April 2007.

Tag der mündlichen Prüfung : 25th April 2007.

Declaration

The research work presented herewith was carried out by me under the supervision of Prof. Agapios Sachinidis, Prof. Jürgen Hescheler and Prof. Diethard Tautz at the Institute of Neurophysiology, Medical Faculty, University of Cologne, Germany from November 2003 to February 2007.

Erklärung

Diese Arbeit wurde von November 2003 bis February 2007 in institut für Neurophyiologie der Medizinischen Fakultät der Universität zu Köln unter der Leitung von Prof. Agapios Sachinidis, Prof. Jürgen Hescheler und Prof. Diethard Tautz angefertigt.

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit- einschließlich Tabellen und Abbildungen-, die anderen Werke im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie- abgesehen von unten angegebenen beantragten Teilpublikationen- noch nicht veröffentlicht ist, sowie, dass ich eine Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Agapios Sachinidis , Prof.Dr. Hescheler und Prof. Diethard Tautz betreut worden.

Köln den Michael Xavier Doss.J

Acknowledgements

First and the foremost, I would like to express my sincere gratitude to my supervisor Prof. Agapios Sachinidis and our institute director Prof. Jürgen Hescheler for their whole hearted support that they conferred through out my stay in the laboratory. I would like to thank very much Prof. Sachinidis for having given me the opportunity to carry out my research in his laboratory and his guidance and support all the way through. I am very grateful to Prof. Diethard Tautz, Institute of genetics, for his advise and valuable suggestions during my doctoral thesis work. I would like to thank specially Prof. Helmut W. Klein, Institute of Biochemistry, for chairing my thesis committee during my thesis defense.

I am very grateful to Prof. Hescheler whose initial ideas and constant efforts made "FunGenES" project sanctioned by the EU consortium. I gained a lot of thought provoking new ideas and excellent challenges from the experts in the fields of embryonic stem cell research thanks to the FunGenES project.

I would like to thank Prof. Francis Stewart, Technical University, Dresden for having given me a chance to attend the workshop on BAC Recombineering and providing the reagents. I would like to thank also Prof. Norbert Hübner, MDC, Berlin for the workshop on high throughput Affymetrix microarray data analysis and Affymetrix hybridization.

I would like to thank Mrs. Rita Altenburg for her kind assistance all the time in carrying out successfully all the large-scale experiments in very organized ways and for all those warm and friendly discussions, which I had with her during this time.

I would like to thank Dr. Norbert Huebner and his team Dr.Herbert Schulz, Dr. Oliver Hummel and also Dr. Jaak Vilo from Egeen, Tartu, Estonia for their Bioinformatics support in analyzing the large scale microarray data generated from all my lab experiments.

I would like to thank our collaborators Dr. Margarete Odenthal and Dr.Claudia Wickenhauser, Institute of pathology.

I would like to thank Miss. Ruth Rottscheidt, Insitute of Genetics for her kind help in Real time PCR analysis.

I am also thankful to the project-granting agency “European Community” (6th Framework Programme, Thematic Priority: Life sciences, genomics and biotechnology for health, Contract No: FunGenES LSHG-CT-2003-503494).

I am also thankful to ARTTIC community for their well-planned organizations for the FunGenES summer schools and workshops.

I would like to thank all the people of Institute of neurophysiology for their helpful scientific discussions, especially, Dr. Johannes Winkler, Kurt Pfannkuch, Dr. Tomo Saric and Dr. Dimitry Spitkovsky. I would like to thank Dr. Marcel Halbach and Miss. Huamin Liang for their help in electrophysiology. I also thank my lab collegues Shiva potta, Abdul Rahman Akasaha and Kleomenis Dardoussis.

I especially thank Prof.Schneider for having taken the responsibilities and his advice for our individual’s lab safety.

I would like to thank Mrs. Wood for her kind administrative support during my stay. I also thank Mr. Frank Stassen for his excellent computer support, Mr. Metzner and his team for the technical assistance.

I am also thankful to my former lab colleagues Miss. Conny Gissel who optimized the ES cell culture conditions and for her valuable scientific discussions, Dr.Vitam Kodelja who helped me to be familiar with the gene expression atlas.

During my stay here, I came in contact with several fascinating people with whom I shared great companionship especially Christoph Koehler, Kristine Obernier, Vijay Shankar and Velu.

I also thank 3Gs from Gastroenterology lab - Gita, Gudrun and Gissel for their kind help in fluorescence microscopy.

Last but certainly not the least, I would like to let my family know how much they mean to me. Without their support and constant encouragement, I would not come successful to this far.

Table of Contents

Summary

The underlying basis of vertebrate mesodermal patterning by which the mesodermal cells become sequentially determined to more precisely defined cell fates during embryonic development remains largely unknown and is one of the classical problems in developmental biology. This demands an in-depth analysis of the endogenous signaling cascades and transcription factor networks. Till date, this fascinating puzzle is not yet solved since isolation of pure early stage mesodermal cells was not feasible due to practical difficulties and hence the functional genomics of the mesodermal cells has not yet been successful.

Recently, embryonic stem (ES) cells have been proven to be a promising tool to study the early embryonic development in a more physiological context *in vitro* because of their versatility over the other conventional systems. Furthermore, ES cells may revolutionize medicine by providing an unlimited renewable source of cells capable of replacing or repairing tissues that have been damaged in all degenerative diseases. Using ES cell model *in vitro*, for the first time pure populations of T Brachyury and BMP-2 expressing mesodermal lineages and mesodermal derived pure and functional cardiomyocytes have been isolated and their entire transcriptomes were profiled using microarrays. To isolate the pure lineage specific cells, stable transgenic ES clones transfected with a targeting vector construct in which the promoter of lineage specific marker (T, BMP-2 or αMHC) drives the expression of both puromycin resistance and enhanced green fluorescence bicistronically with the help of an IRES element were generated. The EBs made from these clones were treated at the onset of the particular transcript expression with puromycin to enrich the respective lineages.

Furthermore, analysis of the gene signatures specific to mesodermal cells and cardiomyocytes that defines their unique cellular and genetic identities contributing a major molecular insight into mesodermal patterning has been performed. In addition, a significant number of novel putative transcripts whose functions are unknown but expressed in specific lineages are enlisted. Few of these transcripts were functionally evaluated for their role in germ layer development and patterning. In addition, the purified populations of T Brachyury and BMP-2 mesodermal cells are capable of priming themselves to give rise to cardiomyocytes along with other mesodermal lineages. Interestingly, the BMP-2 positive cells contained predominantly neural crest stem cells (NCSCs) and their lineages namely smooth muscle cells, epithelial cells, melanocytes, and astrocytes along with cardiomyocytes. The transcriptome characterization of ES cell-derived αMHC cardiomyocytes clearly proves that these cells are exhibiting the properties more like the cardiomyocytes developed in *in vivo* conditions from the point of biological processes and hence are promising candidates for the future cardiac cell replacement therapy.

In overall, the part of the work presented here will better contribute substantially to the lineage specific transcriptomic atlas which will shed light in understanding the complex embryonic developmental process.

Zusammenfassung

Die grundlegenden Mechanismen des sogenannten “mesodermalen Patternings (Musters)”, durch welches sich mesodermale Zellen während der Embryonalentwicklung von Wirbeltieren nacheinander zu genau definierten Zelltypen entwickeln, sind weitgehend unerforscht. Für ein genaues Verständnis des “mesodermalen Musters” ist eine eingehende funktionelle Genomanalyse der endogenen Signalkaskaden und der beteiligten Transkriptionsfaktoren erforderlich. Bisher wurde das faszinierende Zusammenspiel dieser Faktoren nicht eingehend untersucht, da eine Isolierung von frühen mesodermalen Zellen aus experimentell-praktischen Gründen schwer durchführbar ist. Bisher existieren folglich diesbezüglich keine erfolgreichen Untersuchungen an mesodermalen Zellen.

Aufgrund ihrer *in-vitro*-Differenzierbarkeit können embryonale Stammzellen (ES-Zellen) als vielversprechendes Modell dienen, um frühe embryonale Entwicklungsprozesse in einem physiologischen Kontext zu studieren. Aufgrund dessen ist das ES-Zell-Modell herkömmlichen entwicklungsbiologischen *in vivo*-Modellen überlegen. Darüber hinaus können ES-Zellen die Medizin revolutionieren, indem sie als eine unbegrenzte Quelle erneuerbarer Zellen zur Verfügung stehen, die zur Reparatur oder als Ersatz für geschädigtes Gewebe eingesetzt werden können. Solche beschädigten Gewebe kommen in den meisten degenerativen Krankheiten vor.

Im Rahmen der vorliegenden Dissertation konnten aus ES-Zellen zum ersten Mal weltweit reine mesodermale T (Brachyury)- und BMP-2-positive Zellpopulationen sowie von mesodermalen Zellen abstammende, funktionelle Kardiomyozyten gewonnen und deren Transkriptom mittels DNA-Microarrays detailliert analysiert werden. Um diese reinen Zellpopulationen zu gewinnen, wurden ES-Zellen mit IRES-Vektoren stabil transfiziert, die gleichzeitig eine Puromycin-Resistenz-Kassette und eine Reporter-Kassette (grün fluoreszierendes Protein) unter der Kontrolle zellspezifischer Promotoren (T, BMP-2, αMHC) exprimieren. Diese transgenen Zelllinien wurden *in vitro* zu “Embryoid Bodies” (EBs) differenziert, aus denen nach Puromycinbehandlung die jeweiligen Zelltypen (T-, BMP-2-, und αMHC-positive Zellen) in reiner Form isoliert werden konnten

Die Analyse der Microarray-Expressionsdaten führte zur einer Identifikation zellspezifischer Genexpressionssignaturen für mesodermale Zellen sowie für Kardiomyozyten und somit zur Charakterisierung ihrer spezifischen zellulären und genetischen Identität. Diese Untersuchungen gewähren weiterhin einen Einblick in das“mesodermale Musters”. Darüber hinaus wurde eine große Anzahl von Transkripten ermittelt, deren Funktion bisher unbekannt ist. Einige dieser Transkripte wurden funktional bezüglich ihrer Rolle bei der Entwicklung von Keimblättern charakterisiert. Weiterhin konnte gezeigt werden, dass reine T Brachyury- und BMP-2-Zellen in der Lage sind, sich zu Kardiomyozyten und anderen gewebsspezifischen Zelltypen mesodermalen Ursprungs zu differenzieren. Interessanterweise wurden neben Kardiomyozyten zusätzlich Neuralleisten-Stammzellen, glatte Gefäßmuskelzellen, Epithelzellen, Melanozyten und Astrozyten nachgewiesen. Die Transkriptomanalyse der aus ES-Zellen abgeleiteten Kardiomyozyten zeigt, dass diese biologische Eigenschaften von *in vivo* Kardiomyozyten aufweisen und infolgedessen zukünftig für eine zelluläre Ersatztherapie von degenerativen Herzerkrankungen eingesetzt werden könnten.

Die vorliegende Arbeit wird zur Erstellung eines zellspezifischen Transkriptom-Atlas führen, der zu einem besseren Verständnis komplexer entwicklungsbiologischer Prozesse beitragen wird.

Abbreviations used

dpc	days postcoitum
EBs	Embryoid bodies.
μl	micro litre
ng	nanogram
d	day (s)
LB	Luria Bertani
IRES	Internal Ribosome Entry Site
LIF	Leukemia Inhibitory Factor
BMP-2	Bone morphogenetic protein-2
TGF	Transforming growth factor
αMHC	α-Myosin Heavy Chain (*Myh6*)
nm	Nanometer
μm	micrometer
FGF	Fibroblast growth factor
EGFP	Enhanced green fluorescent protein
Bsd	Blasticidin S
PCR	Polymerase Chain Reaction
RT-PCR	Reverse Transcriptase- Polymerase chain reaction
cDNA	Complimentary deoxy ribonucleic acid
DNA	Deoxy ribonucleic acid
RNA	Ribonucleic acid
siRNA	small interfering RNA
qPCR	Qunatitative PCR
FACS	Flourescence Activated Cell Sorting

1. Introduction

Introduction

The vertebrate early stage embryo consists of three primary germ layers-the ectoderm, mesoderm and endoderm, from which all tissues of the adult are derived later. The mesodermal germ layer plays a fundamental role in organizing the vertebrate axes, and give rise to, among other structures, the skeletal, muscular, and circulatory systems. Mesoderm is an inducible fate, that is, the mesoderm is formed by a mechanism called induction[1]. Induction refers to a process in which extracellular signals bring about a change from one cell fate to another in a particular group of cells. The inducing factors and signaling cascades that regulate the differentiation and patterning of the mesodermal germ layer have been the focus for intensive study for the last decades.

1.1 Early embryonic development in mice

Much of the earlier investigation about the germ layer formation and patterning came from the *in vivo* experiments involving early embryos. Hence it is worthwhile to start with the actual developmental processes in early embryonic development in mice to better understand the process of mesodermal patterning.

1.1.1 Pre-implantation embryo

Prior to fertilization, the egg (oocyte) enlarges, divides by meiosis, and matures in its ovarian follicle until it reaches metaphase II, a stage of meiotic division. At this point, the follicle releases the haploid oocyte into the oviduct, one of two tube-like structures that lead from the ovaries to the uterus. The union of the oocyte and sperm and subsequent fertilization take place in the ampulla of the oviduct. The fertilized egg (zygote) divides and develops as it travels along the oviduct to the uterus (Figure 1). In mice, this process takes about 4-5 days. The egg has a protective membrane, the zona pellucida, which stops it from implanting in the oviduct wall.

By the time it reaches the uterus, the egg has undergone many cell divisions to form a blastocyst, which hatches from the zona pellucida to implant into the uterine wall. In all mammals studied, the pre-implantation stages are characterized by a relatively synchronous doubling of cell numbers until the 8-cell stage followed by asynchronous cell divisions after compaction[2]. During compaction, the blastomeres become polarized and develop tight junctions with the other blastomeres. At the 8-16 cell stage the embryo enters the uterine environment, developing into a blastocyst, in which the first events of cellular differentiation are observed (Figure 1).

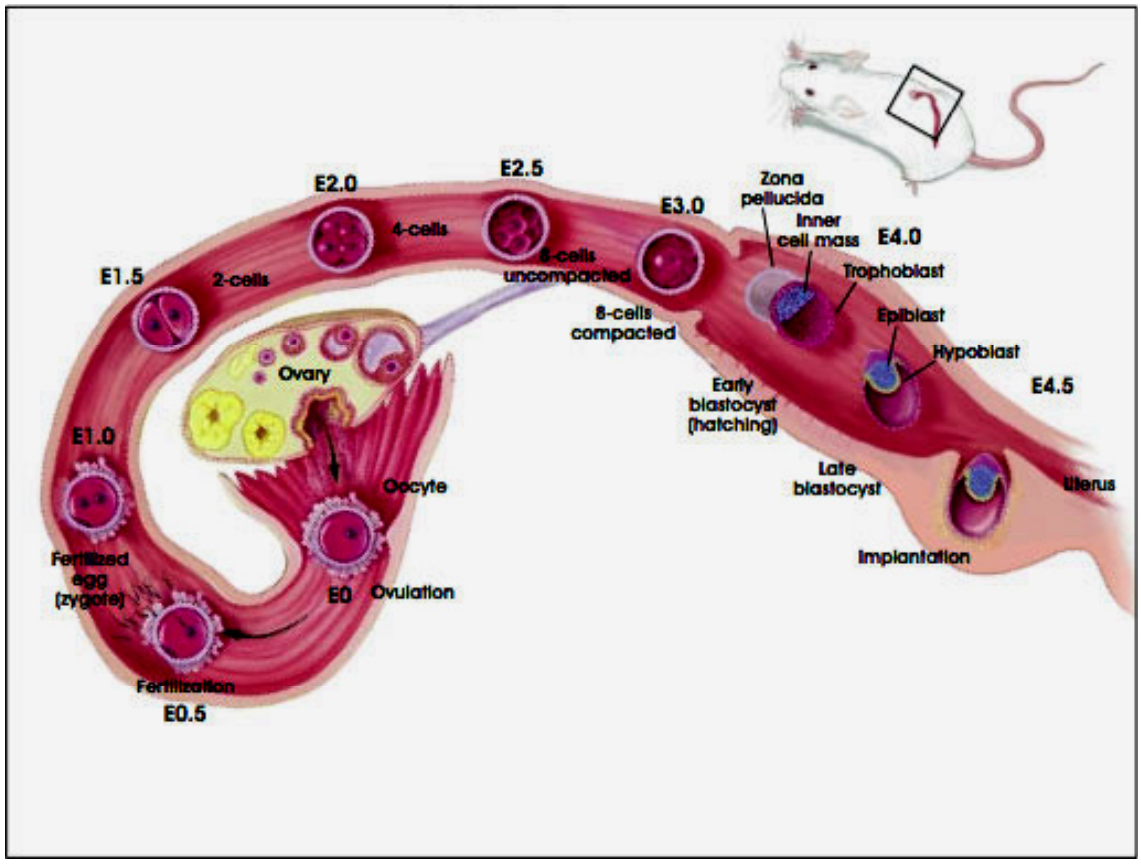


Figure 1. Development of the Preimplantation Blastocyst in Mice from Embryonic Day 0 (E0) Through Day 5 (E5.0). This picture was adopted from National Institutes of Health. Stem Cells: Scientific Progress and Future Research Directions. Stem Cell Information [online], http://stemcells.nih.gov/info/scireport (2001)).

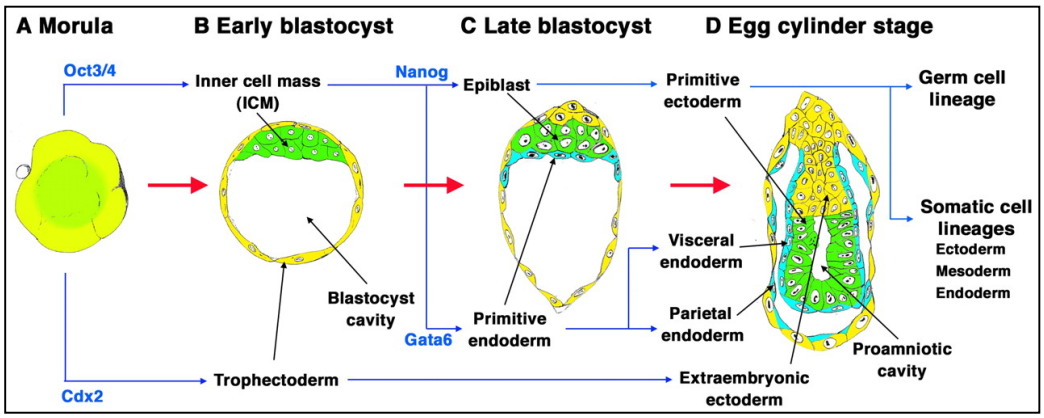


Figure 2. A schematic view of mouse pre-implantation development. (**A**) Pluripotent stem cells (green) are imaged in a morula as the inner cells, which (**B**) then form the inner cell mass (ICM) of the blastocyst. (**C**) After giving rise to the primitive endoderm on the surface of the ICM, pluripotent stem cells then form the epiblast and start to proliferate rapidly after implantation. (**D**) They then form the primitive ectoderm, a monolayer epithelium that has restricted pluripotency, which goes on to give rise to the germ cell lineage and to the somatic lineages of the embryo. Certain key transcription factors (blue) are required for the differentiation of the various embryonic lineages. Adopted from Niwa, H. Development 2007; 134:635-646.

1.1.2 Post Implantation embryo

Around the time of implantation, the mouse embryo is composed of three distinct cell types- the trophectoderm, the primitive endoderm, and the inner cell mass (ICM) (Figure 2, 3A and 3B)[3] . At approximately 4.5 to 5.0 dpc, implantation of the blastocyst into the uterus begins. Shortly after implantation, the ICM undergoes rapid proliferation and differentiates to form a pluripotent population known as primitive ectoderm. As the primitive ectoderm population expands, the innermost group of cells undergoes apoptosis, generating a cavity known as the proamniotic cavity. The remaining cells surrounding this cavity differentiate further to form a pseudostratified columnar epithelium, which results in the formation of a cup-like structure known as the epiblast[4]. Primitive ectoderm and epiblast cells differ from ICM cells in that they have upregulated Fgf5 expression, downregulated Rex expression and have lost their ability to generate chimeras following blastocyst injection[5-8]. Coinciding with development of the epiblast, primitive endoderm undergoes differentiation, giving rise to both visceral and parietal endoderm. Visceral endoderm forms a tissue that surrounds the epiblast, whereas parietal endoderm surrounds the extraembryonic tissues[9] (Figure 2 and Figure 3).

Gastrulation is the major milestone in early post implantation development. During gastrulation, pluripotent epiblast cells are allocated to the three primary germ layers of the embryo (ectoderm, mesoderm and definitive endoderm) and the extraembryonic mesoderm of the yolk sac and amnion. The primary germ layers are the progenitors of all tissue lineages. The specification of tissue lineage is accompanied by the restriction of the developmental potency and the activation of lineage specific gene expression. This process is strongly influenced by cellular interactions and signaling[10,11].

Although much remains unknown about how the embryonic axis is laid down in the mouse, it is now clear that reciprocal interactions between the extraembryonic and embryonic lineages establish and reinforce patterning of the embryo. At early post-implantation stages, the extraembryonic ectoderm appears to impart proximal–posterior identity to the adjacent proximal epiblast, whereas the distal visceral endoderm signals to the underlying epiblast to restrict posterior identity as it moves anteriorward. At gastrulation, the visceral endoderm is necessary for specifying anterior primitive streak derivatives, which, in turn, pattern the anterior epiblast. Polarity of these extraembryonic tissues can be traced back to the blastocyst stage, where asymmetry has been linked to the point of sperm entry at fertilization[2].

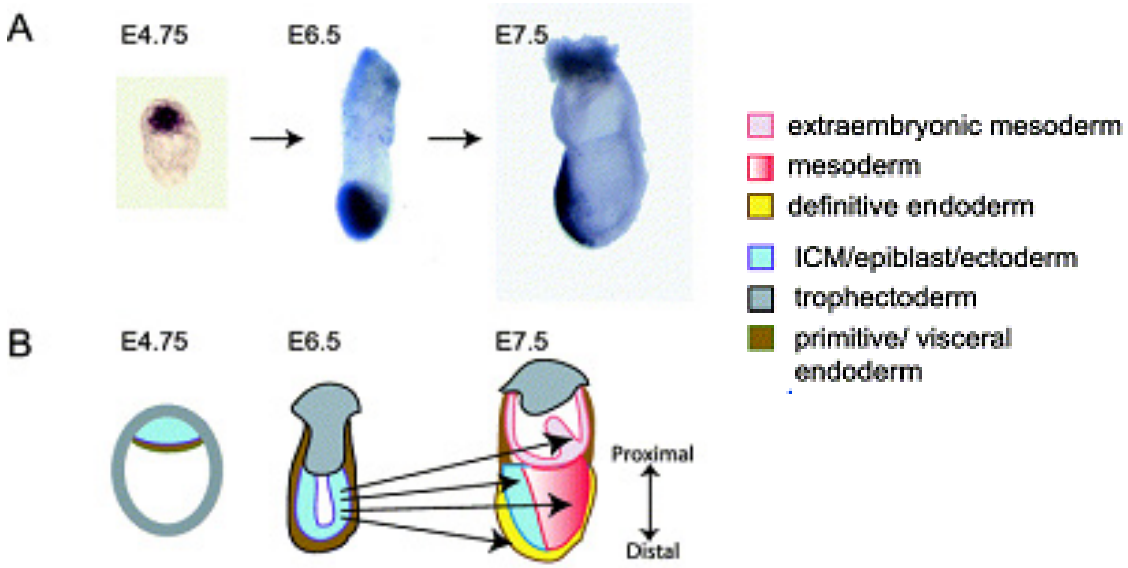


Figure 3. (A) Mouse embryo at peri-implantation (E4.75, *Pou5f1* expression in the ICM), early-streak (E6.5, *Otx2* expression in the epiblast), and early-bud (E7.5, *Cer1* expression in the anterior definitive endoderm) stages. (B) The differentiation of trophectoderm (grey), ICM (blue), and primitive endoderm (brown) of E4.75 embryo to the extraembryonic and embryonic tissues of E7.5 embryo (germ layer tissues are color-coded, see key). Taken from *D.A.F. Loebel et al. / Developmental Biology 264 (2003) 1–14.)*

1.2 Lineage specification

One of the distinguishing features of early mammalian development is the maintenance of a population of developmentally plastic, pluripotent stem cells, which give rise to all cells that constitute the mature organism. The earliest stage of lineage specification in the mouse embryo occurs at approximately 3.5 days post coitum (dpc) with formation of the blastocyst [12]. Around 10-20 cells, which appear to be developmentally equivalent, make up the inner cell mass (ICM), located at one end of the 3.5 dpc (days post coitum) blastocyst. Mammalian development entails the regulated proliferation of these cells and allocation of descendants to specific cell lineages following differentiation[13].

Around 4.0 days post coitum (dpc), ICM cells lining the blastocoelic cavity differentiate to extraembryonic primitive endoderm. Pluripotency is retained by internal ICM cells, referred to as the “epiblast”, now surrounded by extraembryonic endoderm and trophectoderm. At about the time embryo implants into the uterine wall these pluripotent cells begin to proliferate rapidly such that the 20-25 cells present at 4.5d.p.c expand to give rise to about 660 cells by 6.5dpc and 8060 cells by 7.5dpc[14]. Concurrent with pluripotent cell proliferation, primitive endoderm cells migrate along the pluripotent cell surface and differentiate into one of two cell types. Primitive endoderm cells that remain in contact with the pluripotent cells differentiate into visceral endoderm while those primitive endoderm cells

that migrate onto the blastocoelic surface of trophectoderm differentiate into parietal endoderm [9] (Figure 2).

Formation of the proamniotic cavity within the pluripotent cell population is first visible at 5.0dpc and is accompanied by reorganization of the cells into a pseudostratified columnar epithelial sheet. Concomitant with this process, the pluripotent cells become polarized, and undergo changes in gene expression and developmental potential suggesting that this morphologically distinct pluripotent cell population, termed primitive ectoderm, is derived from the ICM by differentiation. Primitive ectoderm provides the substrate for gastrulation, which is initiated at around 6.5 dpc. This process transforms the pluripotent monolayer into a multilayered embryo consisting of the 3 primary germ layers, mesoderm, endoderm and ectoderm. While coordinated regulation of pluripotent cell proliferation, differentiation and morphogenesis underpins development of the mammal, relatively little is known of the signaling mechanisms that regulate these processes at a cellular or molecular level. Accumulating evidence emerging from close analysis of gene expression and knock out phenotypes implicates visceral endoderm as a source of signals controlling primitive ectoderm formation, maintenance and differentiation[15].

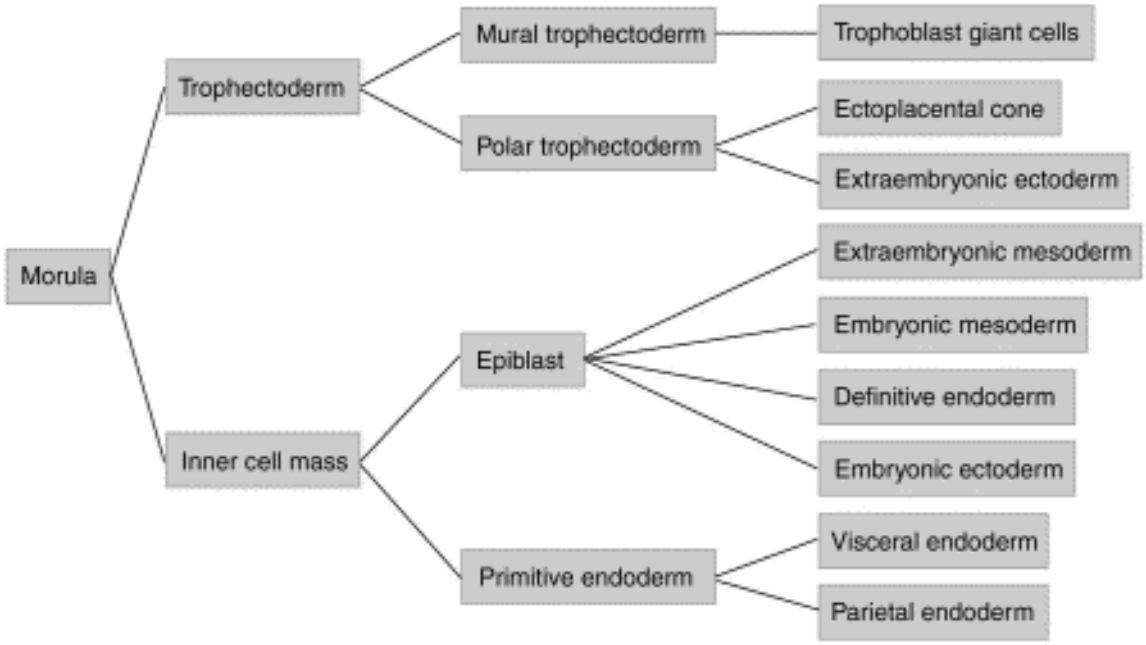


Figure 4. Cell lineage relationships in the early mouse embryo. Taken from Cindy C. Lu et al., Current Opinion in Genetics & Development Volume 11, Issue 4 , Pages 384-392.

The means by which cells are allocated to either embryonic or extra embryonic lineages in the blastocyst is not understood. However, the early establishment of cell fate may be regulated in part by the POU transcription factor *Pou5f1* (also known as *Oct4*). *Pou5f1* is expressed in all blastomeres at the four-cell stage but is downregulated in the trophectoderm by the blastocyst stage[16]. Its expression is maintained in the ICM, with highest protein concentrations accumulating in the primitive endoderm. *Pou5f1* is required for ICM formation, as *Pou5f1* mutant blastocysts consist only of trophectoderm[17]. Consistent with the mutant phenotype, recent studies have shown that embryonic stem (ES) cells lacking *Pou5f1* differentiate into trophectoderm, whereas cells expressing a single copy of *Pou5f1* maintain their pluripotent ES cell state. By contrast, a twofold increase in *Pou5f1* expression causes ES cells to adopt VE or mesodermal fates[18]. Thus, tight control of *Pou5f1* levels may be important for cell fate allocation in the blastocyst. It will be interesting to learn whether the initial polarity at fertilization or the timing of cell division in blastomeres is linked to expression levels of *Pou5f1*.Whereas *Pou5f1* mutant blastocysts lack an ICM, the novel gene *Taube nuss* (*Tbn*) seems to be required for survival or maintenance of the ICM[19]. *Tbn*-deficient embryos appear normal at the early blastocyst stage, but by late blastocyst stage consist only of trophoblast cells as a result of increased apoptosis in the ICM. Tetraploid aggregation experiments in which trophectoderm cells were wild type but ICM cells were mutant indicated that *Tbn* is required cell-autonomously for ICM survival. Another gene whose role in trophoblast development is the T-box gene *Eomesodermin* (*Eomes*), which is first expressed in the trophectoderm at the blastocyst stage[20]. Embryos lacking *Eomes* arrest at the blastocyst stage, and mutant trophectoderm dies in culture even in the presence of the trophectoderm proliferation factor Fgf4, suggesting that *Eomes* is required cell-autonomously for trophectoderm development.

Embryonic stem (ES) cells can be derived *in vitro* from the ICM of the peri-implantation embryo (Figure 5). Like the ICM and the early epiblast, ES cells are capable of differentiating into a wide range of fetal tissues in culture[21] or in chimeric embryos and can even form the entire fetus[22]. Trophectodermal stem (TS) cells, with the molecular properties of diploid trophoblast cells, may be isolated from the blastocyst or extra embryonic ectoderm of peri-implantation embryos. TS cells have a restricted differentiation potential: they can form giant cells in vitro [23,24] and have only been observed to contribute to trophectodermal cell lineages in vivo. Whereas ES cells can be induced to become TS cells[18], TS cells have not been shown to acquire the lineage potential of ES cells. As yet, there are no published reports of stem cells-derived from the third cell type, the primitive endoderm of the embryo, but this remains an exciting possibility (Figure 5).

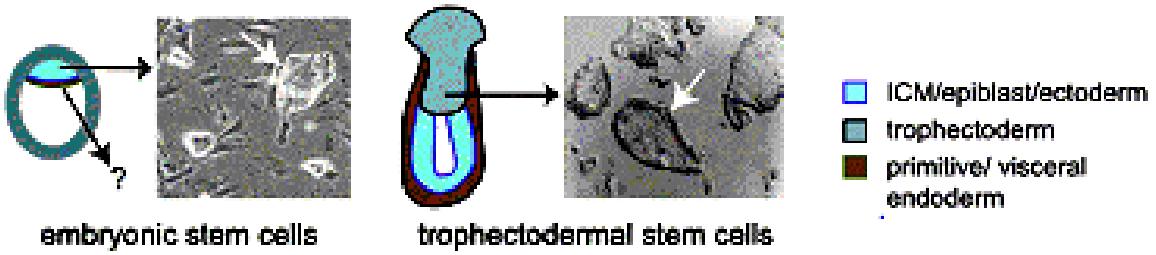


Figure 5. Colonies of embryonic and trophectodermal stem cells (white arrows) derived from the ICM and extra embryonic ectoderm, respectively. Taken from D.A.F. Loebel et al. / Developmental Biology 264 (2003) 1–14.

1.3 Mesoderm development

In mouse embryo, major decisions about lineage allocation and cell fate are made during gastrulation, between E6.25 –7.5[25]. Mesoderm is generated from the epiblast or embryonic ectoderm through the process of gastrulation that is initiated at approximately day 6.5 of gestation[10]. At the onset of gastrulation, the epiblast cells in the region that defines the posterior part of the embryo undergo an epithelial to mesenchymal transition and form a transient structure known as the primitive streak from which the mesoderm emerges. The newly formed mesoderm migrates away from the primitive streak, moves laterally and anteriorly and is patterned into various populations with distinct developmental fates. Brachyury is expressed in all nascent mesoderm and downregulated as these cells undergo patterning and specification into the derivative tissues including skeletal muscle, cardiac muscle and connective tissues in addition to blood and endothelium[26,27]. The first mesodermal cells to develop within the embryo contribute predominantly to the extra-embryonic tissues, giving rise to the hematopoietic and vascular cells of the yolk sac[28,29].

1.2.1 Major Molecular control of mesodermal differentiation

Wnt signaling

WNT3A signaling induces cells from the epiblast to adopt a paraxial mesodermal rather than a neuroectodermal fate after gastrulation[30] (for review[3]) . In *Wnt3a*$^{-/-}$ embryos, epiblast cells ingressing into the PS are diverted to a neuroectodermal fate instead of forming mesodermal cells. Consequently, ectopic structures resembling neural tubes are formed in place of the somites[30]. Regulation of genes acting downstream of WNT3A signaling requires the transcription factors *Lef1* and *Tcf1*. Mouse mutants lacking both *Lef1* and *Tcf1* develop ectopic neural tubes and are defective in paraxial mesoderm differentiation[31]. WNT3a signaling, via *Lef1/Tcf1,* maintains expression of the mesodermal transcription factor *T* in presumptive paraxial mesoderm[32]. Maintenance of *T* expression can be rescued by a constitutively active LEF1-β-catenin fusion protein[33]. The WNT signaling pathway interacts

with the FGF pathway to induce mesoderm differentiation (Figure 6, red area). *Fgfr1* is required for expression of *T* and *Tbx6*[34,35]. In chimeric embryos, ectopic neural tubes form entirely from *Fgfr1*$^{-/-}$ cells, which fail to fully traverse the PS and take on an ectodermal fate [30,36]. The phenotypes of *Fgf8* and *Fgfr1* mutant mice suggest a role in the migration of nascent mesoderm away from the streak, possibly by repression of E-cadherin[,35-38]. E-cadherin prevents the interaction of LEF/TCF proteins with the transcriptional co-activator β-catenin, disrupting WNT signaling. Sequestering E-cadherin rescues expression of a *T-lacZ* reporter, presumably due to the restoration of the WNT signaling pathway[35].

BMP signaling

BMP signaling is essential for the choice between ectodermal and mesodermal fates (for review [3]). BMP signaling is required to induce mesoderm formation *in vivo* (Figure 6, red area). Most *Bmp4*$^{-/-}$ embryos fail to gastrulate and form no mesoderm[39]. Loss of *Bmpr1* and *Bmpr2b* function also results in reduced epiblast proliferation and no formation of mesoderm [40,41]. The similarities in the phenotypes of *Bmp4, Bmpr1,* and *Bmpr2* mutants suggest that BMP4 signals predominantly through heterodimers of these receptors[40,41]. BMP signaling activates receptor regulated SMADs 1, 5,and 8, which interact with the common partner SMAD4 to regulate gene expression[42]. *Smad4* is required for epiblast proliferation, gastrulation, and mesoderm formation[43,44]. For correct patterning and differentiation of the embryo, the influence of BMP signaling must be modulated by antagonistic molecules. The BMP antagonists *Chordin* and *Noggin* are expressed in the mouse organizer and promote differentiation of ectodermal lineages (Figure 6, blue area). *Chordin* and *Noggin* have overlapping functions during embryogenesis and are required for anterior neural differentiation in the embryo[45].

Nodal signaling

Nodal, a member of the TGF-β superfamily is expressed throughout the epiblast in mice. Embryos mutant for *nodal* lack a proper primitive streak and show a severe reduction in mesoderm formation. Cofactors of Nodal belonging to the EGF-CFC family are essential for Nodal signaling and they have been found in *Xenopus*, fish and mouse. The mouse counterpart of these cofactors, *Cripto* is also expressed throughout the epiblast and null mutant mice fail to undergo the movements of the epiblast cell to the primitive streak and that of the anterior visceral endoderm to its proper location[46]. Further more, mutants for the Lefty/antivin proteins- a class of TGF-β molecules that antagonize Nodal in the mouse, zebrafish and *Xenopus* –display the opposite mutant phenotype to that of nodal. Mouse mutants exhibit an expanded primitive streak and excess mesoderm[47].

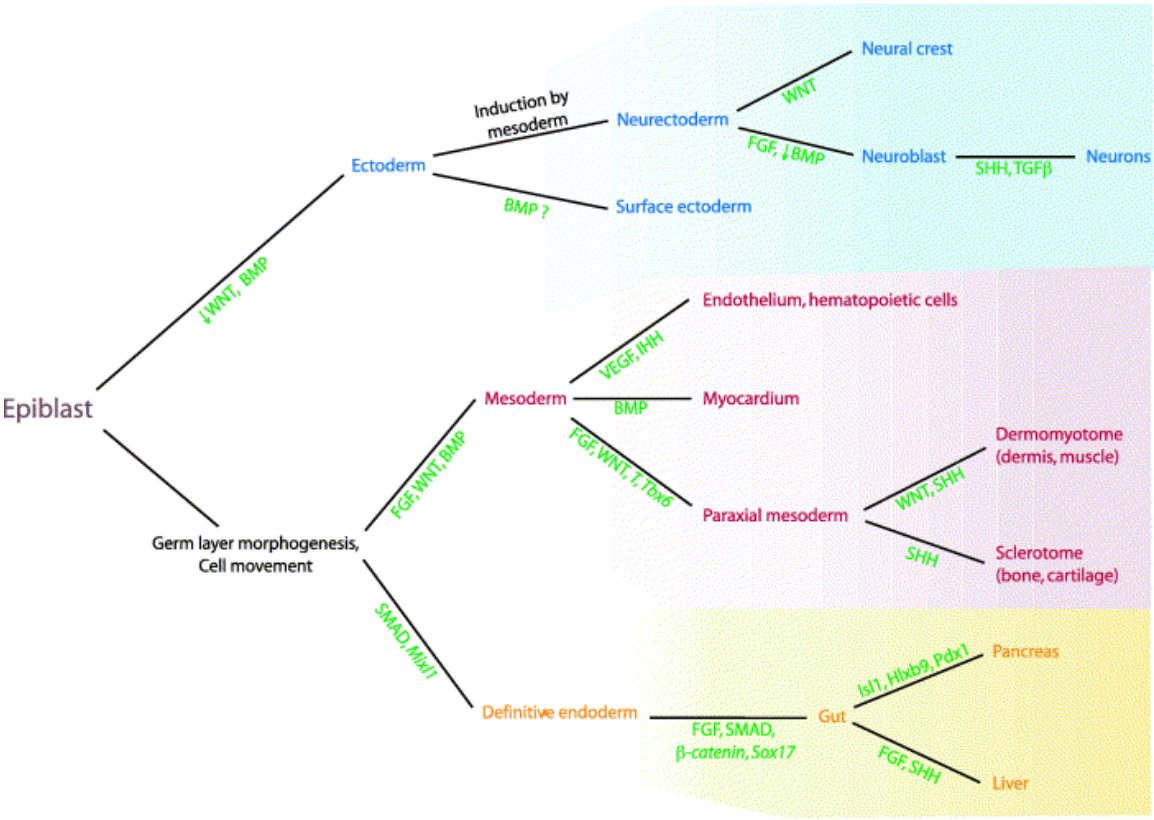


Figure 6. Specification and differentiation of germ layer derivatives in mouse embryos. Ectodermal (blue), mesodermal (red), and definitive endodermal (orange) lineages derived from the epiblast (brown) are shown, with the signaling pathways and transcription factors indicated in green. The question mark signifies the lack of direct evidence for involvement of BMP4 in surface ectoderm differentiation in mouse embryos. Taken from *D.A.F. Loebel et al. / Developmental Biology 264 (2003) 1–14.*

1.4 Cardiac development

The heart is the first organ to form during embryogenesis and its circulatory function is critical from early on for the viability of the mammalian embryo. The development of the vertebrate heart can be considered an additive process, in which additional layers of complexity have been added throughout the evolution of a simple structure (linear heart tube) in the form of modular elements (atria, ventricles, septa, and valves)[48,49]. Each modular element confers an added capacity to the vertebrate heart and can be identified as individual structures patterned in a precise manner. An understanding of the individual modular steps in cardiac morphogenesis is particularly relevant to congenital heart disease, which usually involves defects in specific structural components of the developing heart. Hence, the developmental mechanisms that orchestrate the formation and morphogenesis of this organ have received much attention among classical and molecular embryologists. The initiation of cardiac differentiation has been the topic of vigorous investigation. However, no single

transcription factor that is responsible for the differentiation of lateral plate mesoderm into cardiac cells has been identified[50].

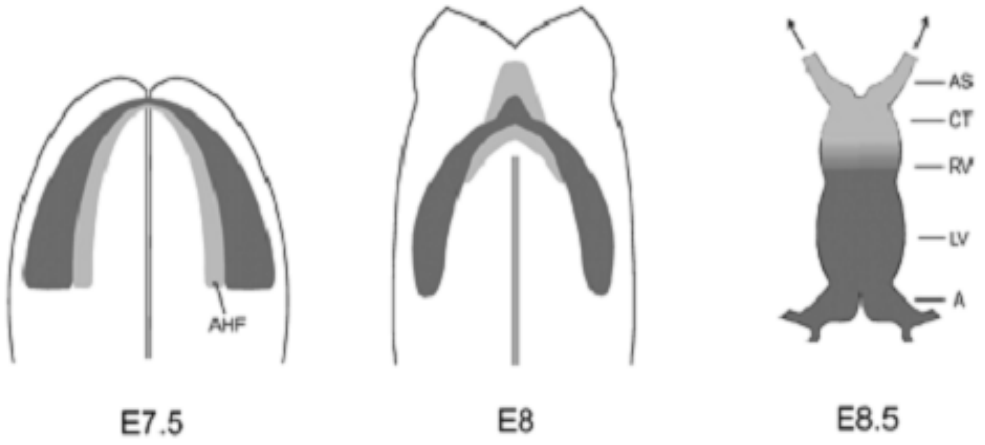


Figure 7. Stages of Mouse cardiogenesis at E7.5, E8, and E8.5. Major heart forming field is shown in black and the anterior (secondary) heart-forming field (AHF) in gray. AS indicates atria, sinus venosa; CT, conotruncus; RV, right ventricle; LV, left ventricle; and A, atria. Adopted from Zaffran and Frasch, Circ Res 2002; 91:457-469.

In mouse, the heart arises from cells in the anterior lateral plate mesoderm of the early embryo, where they are arranged in bilateral fields on either side of the prechordal plate and rostral notochord [51]. These fields include the precursors of both myocardial and endocardial cells, although there is apparently no common pool of bipotential precursors for these two heart cell lineages[52,53]. The cells of the cardiogenic mesoderm are brought to these positions by gastrulation movements that occur in close association with ingressing cells of the rostral endoderm. In mammalians and birds, the bilateral fields of the cardiogenic mesoderm merge at their anterior margins to form the so-called "cardiac crescent"[54] (Figure 7). More recent studies have identified a second type of heart field that is located more medially in the splanchnic mesoderm, directly adjacent to the cardiac crescent. The cells from this bilateral field, termed anterior heart-forming field, are fated to generate anterior heart structures of the outflow tracts (see review[55] (Fig7and 8B). The specification of cardiomyocytes occurs just before or during the formation of the cardiac crescent, and differentiation markers become expressed shortly thereafter (review [51]). The earliest steps of assembly of the heart tube are initiated by the convergence and fusion of the bilateral heart primordia along the midline. The cells of the anterior heart-forming field, after having migrated anteriorly, are added to the anterior end of the linear tube during and after this period. The resulting beating tubular heart is composed of an external myocardial and an internal endocardial layer and also possesses a polarity along the anteroposterior axis, in which the prospective tissues of the aortic sac, outflow tract (conotruncus), right ventricle, left ventricle, and atria are present in an anterior to posterior order along the tube (Figure 7). In all

vertebrates, the tubular heart undergoes a process known as rightward looping. The morphogenetic steps required to achieve looping are guided by molecular asymmetries that are established in and around the heart by the embryonic left/right axial pathway. Furthermore, in higher vertebrates, septal division of the chambers and formation of the valves, which involves endothelial cells, are essential steps leading to the formation of an integrated 4-chambered heart with separate venous (or inflow) and arterial (or outflow) poles. During the growth process of the cardiac epithelium another distinct cell lineage, the migrating cardiac neural crest cells, populate the heart through the outflow channel and contribute to the formation of the great vessels and outflow septum[51].

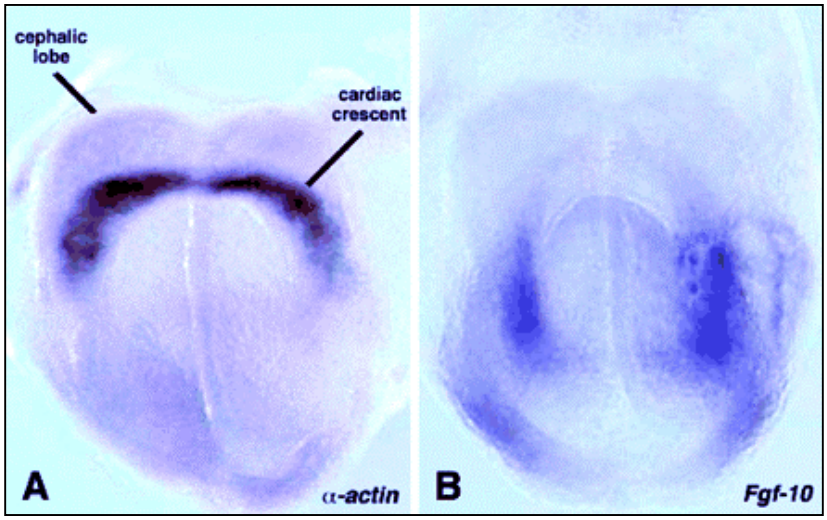


Figure 8. Early markers for the cardiac crescent (A) and anterior heart-forming field (B) in the mouse. Adopted from Zaffran and Frasch, Circ Res 2002; 91:457-469.

Transcription Factors involved during cardiomyogenesis

Heart formation requires the precise integration of cell type–specific gene expression and morphological development; both are intertwined in their regulation by transcription factors. Although many transcription factors have been described as regulators of cardiac-specific gene expression, the transcriptional regulation of cardiac morphogenesis is still not well explored. The most notable cardiac transcription factors and their roles in heart formation are summarized in Figure 9 and the transcriptional network in Figure 10.

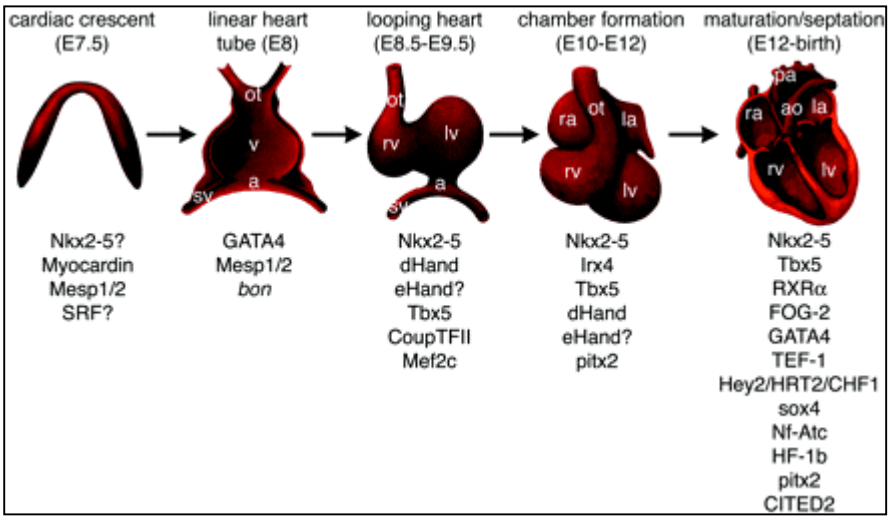


Figure 9. Summary of mouse heart development. Five major stages of heart development are shown: (1) cardiac crescent formation at embryonic day (E) 7.5; (2) formation of the linear heart tube at E8; (3) looping and the initiation of chamber morphogenesis at E8.5 to E9.5; (4) chamber formation; and (5) chamber maturation and septation and valve formation. The transcription factors involved or suspected of involvement in these processes are listed below each stage. ao indicates aorta; a, atrium; la, left atrium; lv, left ventricle; ra, right atrium; rv, right ventricle; ot, outflow tract; sv, sinus venosa; and pa, pulmonary artery. Adopted from Bruneau, B. G. Circ Res 2002;90:509-519.

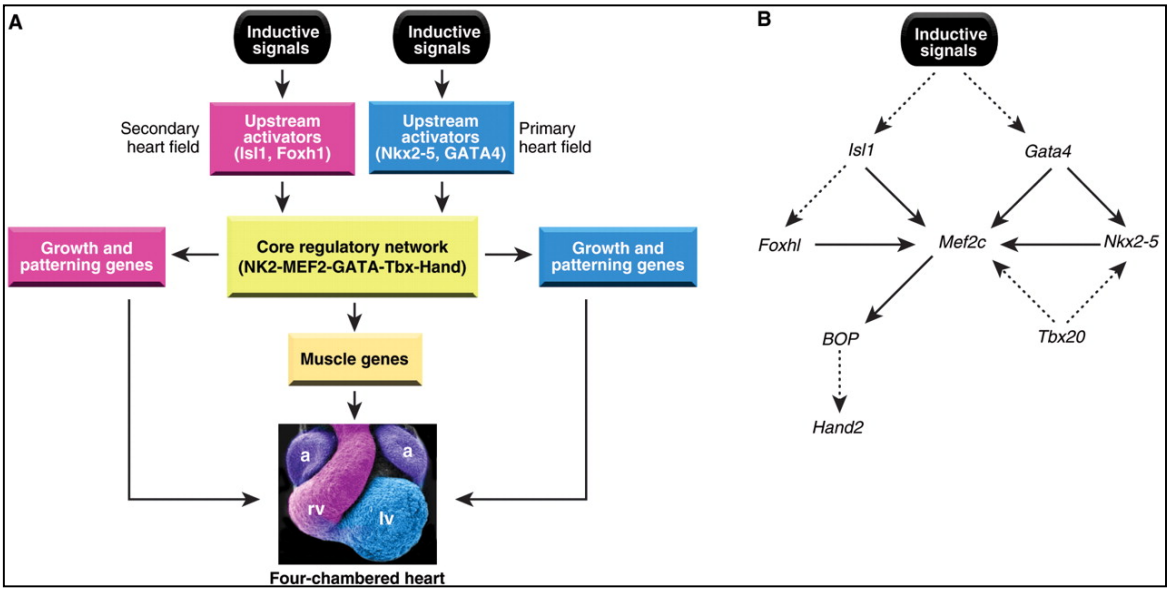


Figure 10. Schematic of transcriptional networks involved in mammalian heart development. (**A**) Inductive signals activate a set of upstream regulatory genes, encoding transcription factors, in the primary and secondary heart fields. The products of these genes activate the genes in the core cardiac network (NK2-MEF2-GATA-Tbx-Hand). Some components of the network, such as Nkx2-5, are also activated in the primary heart field in response to inductive signals. The core network genes cross- and autoregulate their expression and serve as the central regulatory network for the activation of muscle-specific genes and genes that control the growth and patterning of derivatives of the primary and secondary heart fields. The primary heart field gives rise to the left ventricle (lv) and portions of the atria (a), whereas the secondary heart field gives rise to the right ventricle (rv), portions of the atria, and the outflow

tract. A scanning electron micrograph of a mouse heart at embryonic day 14.5 is shown at the bottom. Derivatives of primary and secondary heart fields are shown in blue and pink, respectively. The atria, which are derived from the primary and secondary heart fields, are shown in purple. (**B**) Regulatory interactions among cardiac transcription factors in the secondary heart field. Solid lines indicate direct transcriptional connections, and dotted lines indicate connections not yet shown to be direct. [Adapted from Eric N. Olson, *Science* 29 Vol. 313. no. 5795, pp. 1922 - 1927]

Signaling Pathways During the Induction of Cardiogenic Mesoderm

BMP Signaling

The heart-inducing activity of BMP signaling has been confirmed by studies using inhibitors of BMP signaling, including the BMP inhibitor noggin, truncated versions of type I (tALK3) or type II (tBMPRII) BMP receptors, and inhibitory SMAD6[51]. These studies show that BMP signal transduction is indeed required within the cardiogenic mesoderm, and not only within the anterior endoderm, to promote cardiac differentiation. A second important conclusion from these data are that BMP signaling is required for the maintenance of *Nkx2-5, T-box* genes*, bHLH*-factor encoding *Hand* genes and *Gata* genes expression at stages during and after the fusion of the bilateral heart primordia, but apparently not for the initial activation of these genes in earlier stage embryos[56-58].

Fibroblast Growth Factor Signaling

Function of FGF signaling in cardiac induction has been obtained in the zebrafish system. Zebrafish FGF8 is expressed in the cardiogenic fields of the lateral plate, as well as in specific areas of the neural tube. *fgf8* (*acerebellar*) mutant embryos display strong heart defects with a particular loss of ventricular structures[59]. FGF signals not only act during early mesoderm development but are also required more directly for the induction of cardiogenic transcription factors during subsequent stages[51].

Notch Signaling

The direct target genes of Notch in heart development are not known. At least in certain contexts, the *Hrt* (*Hesr*/*Hey*) genes of the *Hairy/Enhancer of Split* family were proposed as candidates although there is presently no evidence that the differential expression of these genes along the anteroposterior axis of the heart tube is controlled via Notch signaling[60]. Nevertheless, the observations in *Xenopus*, the heart defects of hypomorphic *Notch2* mouse mutant embryos, and the heart abnormalities associated with *Jagged1* haploinsufficiencies in human Alagille syndrome patients indicate that Notch signaling is likely to be required during multiple processes in early vertebrate cardiogenesis[61].

1.5 Embryonic stem cells

-A versatile tool to study early embryonic development

The pluripotency and immortality of mouse embryonic stem ((ES) cells have made them attractive for basic studies of regenerative medicine, as well as for gaining molecular insight into cellular differentiation at early developmental stages, which are much more difficult to assess by *in vivo* approaches.

Due to their remarkable ability to recapitulate early embryonic development on their own, embryonic stem (ES) cells are revolutionizing the field of developmental biology as a potential tool to investigate the molecular mechanisms occurring during the process of differentiation from the early embryonic stage to the adult phenotype. ES cells harvested from the inner cell mass (ICM) of the early embryo can proliferate indefinitely in vitro culture without senescence in an undifferentiated state in the presence of leukemia inhibitory factor (LIF) or on the top of a layer of mitotically inactivated mouse embryonic fibroblasts (MEFs). In the absence of LIF, ES cells spontaneously differentiate into multicellular ES cell aggregates (ESCAs). These ESCAs, also called as embryoid bodies (EBs) represent a more physiological context in which the formation of germ layers-ectoderm, mesoderm and endoderm and their subsequent differentiation into several germ layer specific phenotypes can be studied more extensively than most other *in vitro* systems (Figure 11). The ES cell system offers several important advantages over the conventional studies in the early embryos such as the possibility to isolate large numbers of lineage specific cells in sufficient quantities at different developmental stages for molecular, biochemical, or functional studies using appropriate induction and selection strategies, access to the earliest stages of development to investigate the inducing and patterning properties of different soluble factors and relative easiness to genetically manipulate with loss or gain of function and to generate reporter cells lines to track the spatial and temporal expression of a particular gene of interest[12,62]. Thus ES cell differentiation system paves the platform for the discovery of new genes and serves as an elegant model for developmental biology as well as for its future potential in tissue replacement therapies.

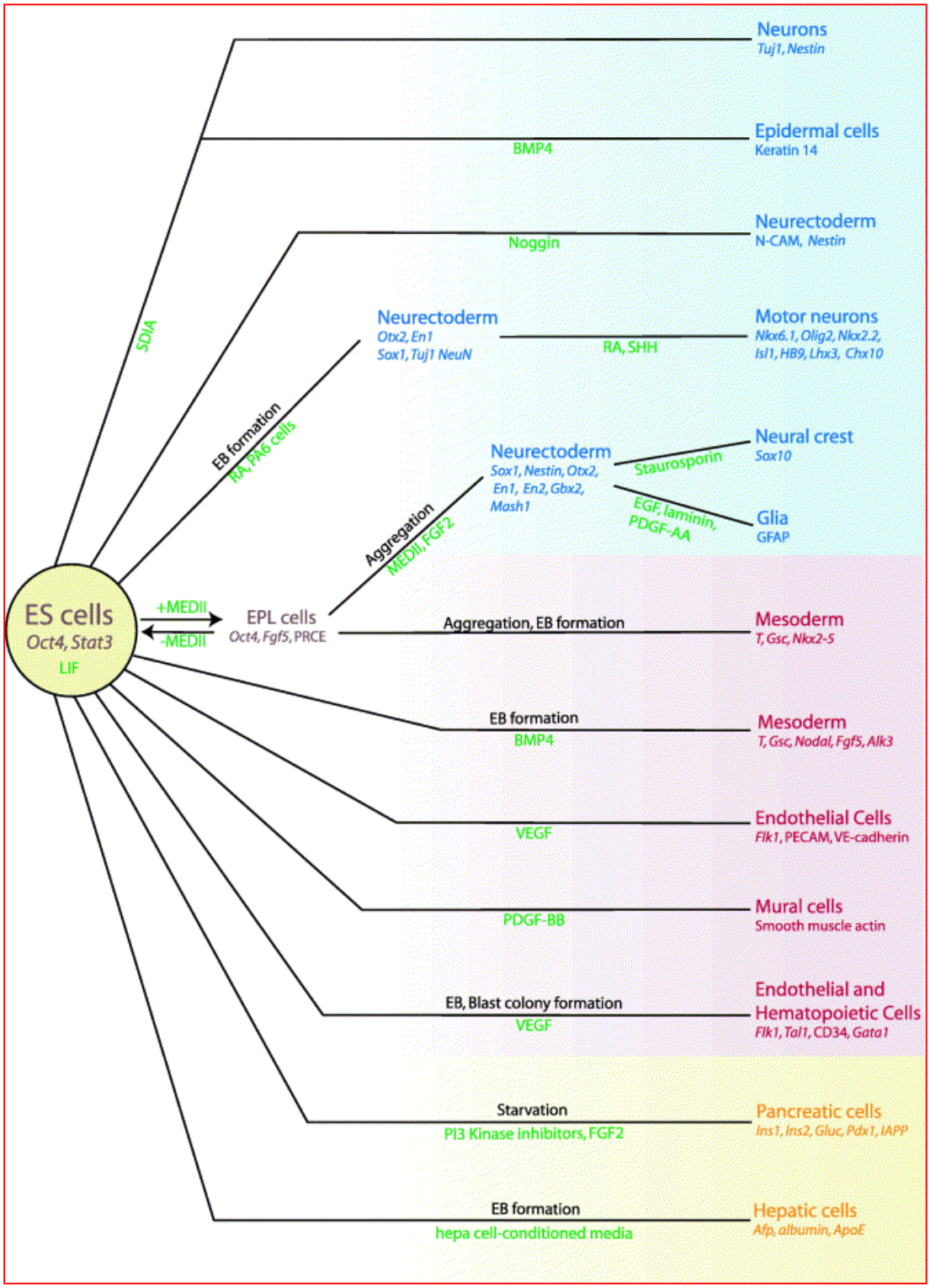


Figure 11. i*n vitro* differentiation of embryonic stem cells. Culture supplements are shown in green, culture conditions in black, and germ layer derivatives and tissue-specific markers in blue for ectodermal tissues; red for mesodermal tissues; orange for endoderm. EB, Embryoid body. Adopted from Developmental Biology Volume 264, Issue 1,1 December 2003, Pages 1-14.

2. Objective

During early embryonic development, the mesodermal cells become sequentially determined to more precisely defined cell fates and give rise to a diverse array of functionally important cell types, including the muscles, heart, vasculature, blood, kidney, gonads, dermis and cartilage. How the prospective mesodermal cells integrate the various signals they receive and how they resolve this information to regulate their morphogenetic behaviors and cell-fate decisions is largely unknown. This demands an in-depth analysis of the endogenous signaling cascades and transcription factor networks. Till date, isolation and transcriptome characterization of pure mesodermal cells have not been successful due to the practical difficulty in isolating the cells in pure form and in sufficient quantities from early stage embryos. This leaves a wider gap in our understanding about the mesodermal patterning. The present work is aimed to attempt to decipher the underlying mechanisms of mesodermal patterning during the differentiation of mesodermal cells into cardiomyocytes using ES cell model as an alternative approach and in particular to identify novel genes whose functions have not yet been annotated till date but are differentially expressed in mesodermal cells and cardiomyocytes.

The objective of this thesis work is to isolate ES cell-derived pure mesodermal cells and their lineages in order to characterize their unique transcriptomes by large scale microarray analysis and to functionally evaluate few of the novel transcripts, if any, for their role in development by “loss of function” and “gain of function” studies.

3. Materials and Methods

3.1 MATERIALS

ES cell line

CGR8 cell line (European Collection of Cell culture (ECACC), catalogue number: 95011018) was used. The germ-line competent cell line CGR8 was established from the inner cell mass of a 3.5-day male pre-implantation mouse embryo (Mus muculus, strain 129/ola). These cells retain the ability to participate in normal embryonic development. Differentiation of CGR8 cells is inhibited by the pleiotrophic cytokine Differentiation inhibiting Activity (DIA) which is identical to Leukemia Inhibitory Factor (LIF). Addition of DIA/LIF allows culture of CGR8 without the use of feeder layers. Karyotype: 40XY.

Consumables:

S.No	Consumables	Manufacturer
1	T25- tissue culture flask	Becton Dickinson
2	10 cm petri dishes (bacteriological)	Greiner
3	6 cm petri dishes (bacteriological)	Greiner,
4	6-well cell culture plates (	Becton Dickinson
5	24-well cell culture plates	Becton Dickinson
6	5 ml pipettes	Sarstedt
7	10 ml pipettes	Sarstedt
8	15 ml tubes	Falcon
9	50 ml tubes (F	Falcon
10	cryotubes	Nunc
11	cell counting chamber	Neubauer/ Merck
12	pasteur pipettes	6127-1722
1	Multipipette	Eppendorf
14	1000 µL micro molecular-pipette	Gilson (Amersham)
15	blue tips	Biolabs
16	Sterile Gene pulser Cuvette	Bio-Rad
17	Trypan Blue 0.4%	Gibco BRL
18	Neubauer Counting Chamber	Merck
19	Hoechst dye	Molecular Probes
20	ProLong Gold Mounting medium	Molecular Probes

Primers included in our study.

S.No	Gene	Forward Primer (5'→3')	Reverse Primer (5'→3')
1	1700108l22rik	gagaacatcagcgaggatgag	caacaggatacacacgcacac
2	Activin	cctagaaagcgggacctatga	tacacgggtgcagggtaaggc
3	Adipsin	tggtatgatgtgcagagtgtagtg	ggtaggatgacactcgggtatag
4	Afp	cca gaa cct gcc gag agt tgc	gcc ttc agg ttt gac gcc att
5	Aggrecan	agagctatttccacacgctacac	gtaggttctcactccagggaact
6	Alkaline Phosphatase	acaccttgactgtggttactgct	ggaatgtagttctgctcatggac
7	ANP	cag agt ggg aga ggc aag acc	cct ctg aga gac ggc agt g
8	Ap2 (Fabp4)	ctggaagcttgtctccagtga	tatgatgctcttcaccttcctgt
9	Bmp-2	tct tag acg gac tgc ggt ctc	cct gag tgc ctg cgg tac aga
10	Bmpr1a	ctagacaccagagccctactcaa	cctggtattcaagggtatgtcaa
11	Cbfa/Runx2	atccatccactccaccacgc	aagggtccactctggctttgg
12	Cd34	tatggaaaagcaccaatctgact	tctccgtgtaataagggtcttca
13	Cdx2	gatacatcaccatcaggaggaaa	caaggaggtcacaggactcaag
14	c-Fms	taagcaagatctggacaaagagc	gagtcattcatgatgtccctagc
15	C-Kit	aactccatgtggctaaagatgaa	ccagaaaggtataagtgcctcct
16	Collagen Ii/Col2a1	aggggtaccaggttctccatc	ctgctcatcgccgcggtccta
17	Collagen X	atgccttgttctcctcttactgga	ctttctgctgctaatgttcttgacc
18	Cyclophilin B	gcacaggaggaaagagcatc	cccacacagacagctgctta
19	Cytokeratin 20	cgtatagctgtggaagctgatct	ctatctccaggttctggtaggtg
20	Cytokeratin 7	agatcaagaccctcaacaacaaa	caacacaaactcattctcagcag
21	Cytokeratin 8	gatgcagaacatgagcattcata	ttcaatcttcttcacaaccacag
22	Desmin	aatggtacaagtccaaggtttca	agaaggtctggataggaaggttg
23	E-Cad	tatgatgaagaaggaggtggaga	aacaccaacagagagtcgtaagg
24	EGFP	agctgaccctgaagttcatctg	Tgatatagacgttgtggctgttg
25	Eomesodermin	agcttcaacataaacggactcaa	gaaggtctgagtcttggaaggtt
26	F4/80	aatgctaccttcacaatccttga	ctgccaagttaatggactcattc
27	Flk-1	ctgaactcaagatcctcatccac	gatacacttccttgatgccaaga
28	Flt-1	cagactcttgtcctcaactgcac	ggctctccacattgttgatctta
29	Fsp-1 (S100a4),	gtccaccttccacaaatactcag	ctcctggaagtcaacttcattgt
30	Gapdh	cagcaaggacactgagcaag	gggtgcagcgaactttattg
31	Gata4	ggtgcctccagcggtaactcc	tgattacgcggtgattatgtc-3´
32	Gene Model 397	gcaagagcagcctatctctga	gacttcaggttggtcttgtgg
33	GFAP	ggatttggagagaaaggttgaat	tgcaaacttagaccgataccact
34	Glut4	tccaactggacctgtaacttcat	caagttctgtactgggtttcacc
35	Goosecoid	ggctacaacagctacttctacgg	cttaaaccagacctccaccttct
36	Hand1	atcatcaccatcatcaccactc	gcgccctttaatcctcttct

Primers included in our study (continued)

S.No	Gene	Forward Primer (5'→3')	Reverse Primer (5'→3')
37	Lpl	agcagcaagatgtacctgaagac	ttctccctagcacagaagatgac
38	$\tilde{\alpha}$Actinin	cagagcctccttcaaccact	gagtcataatgcgggcaaat
39	Mash2 (Ascl2)	agagtacattcggaccctctctc	ccaccttactcagcttcttgttg
40	Cardiac Actin	tcgggacctcactgactacc	cgccatctcgttctcaaaat
41	Caridiac Troponin C	cagaaagggaagtctgagg	cgtaatggtctcacctgtgg
42	Mef2c	agcactgacatggataaggtgtt	ggtgagtgcataagaggagtcag
43	Mesp1	agaacctgaccaagatcgagac	tctagaagagccagcatgtcg
44	Mesp2	ctaggaacaagactggacactgg	aaagttcaggacagccactgag
45	MGC117846	tattcagcagctggtcctgtc	ttggttcgagaattgaagagc
46	MHC-ß	ccctcctcacatcttctcca	gcggcaataacagcaaaata
47	Myocardin	tcctggctcagaaagtgaca	cggttcttactgtcacccaaa
48	Myo-D	ctacccaaggtggagatcctg	gtggagatgcgctccactat
49	Myogenin	tgtgtaagaggaagtctgtgtcg	agcaaatgatctcctgggttg
50	Myostatin	agctcctaacatcagcaaagatg	ggtttgatgagtctcaggatttg
51	Nanog	gtggttgaagactagcaatggtc	tggtggagtcacagagtagttca
52	Nestin	gaggaagaagatgctgatgaaga	gccactgatatcaaaggtgtctc
53	NF-M	tagagcgcaaagattacctgaag	ttgacgttaaggagatcctggta
54	Nkx.2.5	ccaaagaccctcgggcggata	gcgcagctgtagccgggactg
55	Osteocalcin	aggagggcaataaggtagtgaac	gataccgtagatgcgtttgtagg
56	Osteocalcin Primers 2 (Bglap)	atgctactggacgctggagggt	gcggtcttcaagccatactggtc
57	Osteopontin (Spp1)	atgaatctgacgaatctcaccat	cttagactcaccgctcttcatgt
58	$P75^{NTR}$	attgctttcaagagatggaacag	ctactgtagaggttgccatcacc
59	Pax-7	agaaagaagaagatggcgagaag	gtagagtccggcagctggtag
60	PPAR-□	agaaccttctaactccctcatgg	cggcagttaagatcacacctatc
61	Pref- 1 (Dlk1)	gaaattctgcgaaatagacgttc	actcttgttgagctctttcatgg
62	Runx1	tacctgggatccatcacctc	gacggcagagtagggaactg
63	Shh	gga agg tga gga agt cgc tgt	ttg gcc atc tct gtg atg aac
64	ACTA2	ctattcaggctgtgctgtcc	ccaagtccagacgatgat
65	Sox9	cacggaacagactcacatctctc	aaggtctcaatgttggagatgac
66	Tyrp1	tccagaagcaacttcgattctac	tgttgtattgcctgttatgtcca
67	Vimentin	cttgaagctgctaactaccagga	Gaagtctcattgatcacctgtcc

Primers for T Brachyury Subcloning

5′ T-bra primer

- **GCT GAG TCT TCT CTG TCA TTT AAA TTA GGC TCT GTC TCA GTT TGC CAT TCA GAT CTT CAC AGC TTG TCT GTA AGC GGA TG(80 bp)**

3′ T-bra primer

- **CCT CCC GCC ACC CTC TCC ACC TTC CAG GAG TCT TGA CTC CCT ACC CAA CAA GAT CTG CTC TCC TGA GTA GGA CAA ATC (78 bp)**

Bgl II restriction enzyme sites are indicated in blue and forward and reverse primer sequences to prime minimal vector are indicated in red and underlined.

Cell culture Reagents

S.No	Medium	Catalogue Number	Manufacturer
1	Glasgows buffered minimal essential medium (GMEM)	21710082	Gibco BRL
2	Fetal bovine serum (FBS)	10500064	Invitrogen
3	Phosphate buffered saline (PBS) (w/o Ca^{2+}/Mg^{2+}	. 14190-094	Invitrogen
4	Phosphate buffered saline* (PBS*) (w Ca^{2+}/Mg^{2+})	14040-091	Invitrogen
5	Trypsin/EDTA solution (Ca^{2+}/Mg^{2+}-free) (	25300-054	Invitrogen,
6	Leukemia inhibitory factor (LIF	ESG1107	Chemicon
7	ß-2-mercaptoethanol (ß-ME)	31350-010	Invitrogen
8	Gelatine Type II	G-2500	
9	Di-methyl sulfoxide (DMSO)	D-8418	Sigma Alldrich
10	Lipofectamine™ 2000 Reagent	52758	Invitrogen
11	Puromycin dihydrochloride	P8833	Sigma
12	Neomycin (G418 Sulphate)	11811-031	Invitrogen
13	Blasticidin	R210-01	Invitrogen

Bacterial clones/Bacterial Artificial clones (BACs)

S.No	Bacterial clone	Catalogue/ Clone ID	Company
1	RP23-376B1	RPC123.C	Invitrogen, Germany
2	MGC117846	30462068	Geneservice,UK
3	1700108L22Rik	IRAVp968A0470D6	RZPD, Germany.
4	Gm397	IRAKp961F07111Q2	RZPD, Germany.

Instruments Used

Zeiss Axiovert 200 fluorescence microscope
FACScan (BD, Pharmingen)
Biorad Gene Pulser™

Plasmids used

S.No	Plasmid name	Catalogue name	Company/from
1	pIRES2 EGFP	632306	Clontech
2	pIRES Puro3	631619	Clontech
3	pCAGGS	Kind gift	Dr. Miyazaki, Tohoku University
4	pBMP-2-GL3	Kind gift	Dr. Nandini Ghosh-Choudhury, Texas,USA
5	αMHC	Kind gift	Dr. J. Robbins Cincinnati, OH)
6	RedET plasmids	BAC Subcloning kit	Genebridges,Germany.
7	pPyCAGIP	Kind gift	Stem cell Technologies
8	pPPGKIP	Kind gift	Stem cell Technologies
9	PMGD20neo	Kind gift	Dr. M Gassmann, CA, USA

Antibodies

S.No	**Antibodies**	**Company**
1	Anti-BMP-2	Santa Cruz
2	anti-alpha Actinin	Sigma
3	anti-Connexin43	Sigma,
4	Anti-mouseIgG1-AlexaFluor555	Molecular Probes.
5	and anti-rabbit-Ig-AlexaFluor647	Molecular Probes.

Molecular Biology Kits/Reagents

S.No	Kit	Catalogue number	Company
1	QIAquick Gel Extraction Kit	28704	Qiagen
2	Restriction enzymes		New England Biolabs /MBI Fermentas
3	DNA Ligase		MBI Fermentas
	QIAquick PCR Purification Kit	28104	QIAGEN
4	QIAGEN Plasmid Mini Kit	12123	QIAGEN
5	QIAGEN Plasmid Maxi Kit	12162	QIAGEN

Pre-designed On-TARGETplus SMARTpool siRNA Reagents from Dhramacon

S.No	Target gene	Catalogue no
1	Cycophilin B	D-001820-20-05
2	Non-Target	D-001810-01-05
3	MGC117846	L-073711-00-0010
4	1700108L22Rik	L-053323-00-0010
5	Gm397	L-064700-00-0010

Self prepared media and solutions

SOC Media

SOC Media was prepared to be used in DNA transformation by dissolving 20g of bacto-tryptone, 5g of bacto-yeast extract, and 0.5g of NaCl in 950mL water, then adding 10mL of 250mM KCl. The solution was titrated to a pH of 7.0 by using NaOH, then diluted to 1L using water. The new solution was sterilized by autoclaving and cooled before adding 20mL of sterile 1M glucose and 5mL of sterile 2M $MgCl_2$.

TB buffer

Mixed 10 mM HEPES pH 6.7, 15 mM $CaCl_2$, 250 mM KCl and adjusted the pH to 6.7 with KOH. Then added 55 mM $MnCl_2$ and filter sterilized the mixture using a 0.22-µm filter.

SOB^{++} medium

Added 20 g bacto-tryptone, 5 g yeast extract, 0.5 g NaCl, 0.186 KCl, and the volume made up with distilled water to 1 litre. Sterilized by autoclaving and after cooled, added filer sterilized 10 mM $MgCl_2$, 10 mM $MgSO_4$ just before.

LB broth

Added 20g bacto-tryptone, 5g yeast extract and 5g NaCl in distilled water and the volume made up to 1 liter. The pH was adjusted to 7.2 at 37^0C. Autoclaved for 30 minutes.

LB Agar plates

Just before autoclaving the LB broth, added 15% agar-agar and then autoclaved. Once cooled to luke warm, added appropriate antibiotics and poured into plates. Once solidified, stored them in dark at 4^0C

Antibiotics

antibiotics	Stock solutions	Working concentration
Ampicillin	100 mg/ml in water	100 µg/ml
Chloramphenicol	30 mg/ml in ethanol.	15 µg/ml for BACs 50 µg/ml for high-copy plasmids.
Kanamycin	50 mg/ml in water	50 µg/ml
Tetracycline HCl	10 mg/ml in 75% ethanol.	2 µg/ml for pSC101-BAD-gbaA, 10 µg/ml for high copy plasmids

ES cell culture medium:

GMEM supplemented with 10% heat inactivated FBS, 100Units/ml LIF and 50 µM ß-2-mercaptoethanol (ß-ME). FCS has to be critically checked for optimal results for every batch.

Differentiation **medium**

IMDM supplemented with 20% FBS, 100 µM β-mercaptoethanol, and 1x concentrated MEM.

Freezing medium

10% DMSO in 90% FBS:

Wash medium

GMEM supplemented with 5% FBS.

Gelatine solution

5 g type II gelatine in 500 ml PBS was solubilized by heating in a microwave for 3 –5 minutes till clear solution was obtained. Autoclaved this 1% gelatine solution 30 min and stored at 4°C.

3.2 METHODS

3.2.1 ES cell culture

Gelatine coating of dishes

Added 5 ml of 0.2% gelatine solution (prepare diluting the 1% gelatine solution 1:5) per T25 flask and placed it in the sterile incubator for at least 3 hours before use for the best coating.

Defreezing cells

Prepared a 15ml falcon tube with 10ml wash media pre-warmed at 37°C.Thawed rapidly the vial at 37°C water bath. Once thawed, transferred all the contents into the falcon tubes with 10ml wash media. Centrifuged at 1000rpm for 6 minutes. Resuspended the pellet in 5ml medium with LIF Removed the gelatine solution contained in the culture flask. Transferred the cell suspension from the tube into the gelatine-coated dishes. Placed it in the incubator.

Passaging

Passaged the cells when they approach 70-80% confluency. Prepared new 0.2% gelatine coated flasks or dishes one day before. Aspirated the medium with a sterile Pasteur pipette. Washed cells with 5ml 37°C warm PBS Added 2ml Trypsin/EDTA solution. Incubated cells for 2-5min at 37°C until they have lost the contact to the flask. Rocked the flask in between. After 5-10 minutes, transferred the cells with trypsin into a 15 ml falcon tube with 10 ml wash medium. Centrifuged at Room Temperature for 6 min at 1000rpm Removed the supernatant with a sterile Pasteur pipette Resuspended the cells in 1ml propagation medium with LIF. Dispensed 0.2 to 0.3 ml of this cell suspension into 0.2% gelatine coated dishes with 5ml culture medium with LIF.

Counting cells

Prepared a cell counting chamber as described by the manufacturer. Added the appropriate volume of the cell suspension. Counted the cells under a microscope and calculate the total cell number in your cell suspension as described below. Prepared a 1 x 10^6 cells/ml cell suspension finally.

- Cell titer (cells/ml) = number of counted cells x chamber factor (Neubauer chamber = chamber factor 10^4)
- Number of total cells = cell titer x volume of cell suspension

Freezing cells:

When cells were around 70% confluency, frozen them. Aspirated the medium with a sterile Pasteur pipette. Washed cells with 5ml 37°C warmed PBS. Added 2ml Trypsin/EDTA solution. Incubated cells for 2-5min at 37°C until they had lost the contact to the flask. Rocked the flask in between. After 5-10 minutes, transferred the cells with trypsin into a 15 falcon tube with 10 ml wash medium. Taken an aliquot for cell counting. Centrifuged at room temperature for 6 min at 1000rpm. During centrifugation, counted the cells. Removed the supernatant with a sterile Pasteur pipette. Resuspended the cells quickly and carefully in ice cold freezing medium (the

titer should be 1 x10^6 cells / ml transfer into vials (put 1.5ml suspension into one vial). Placed on ice and stored it for two days at –80°C. Then transferred to Liquid N_2. Thawed one control vial to check everything ok with the freezing procedure 24 hrs after transfer to Liquid N_2.

Default differentiation protocol-Hanging drop protocol for Generation of embryoid bodies (EBs)

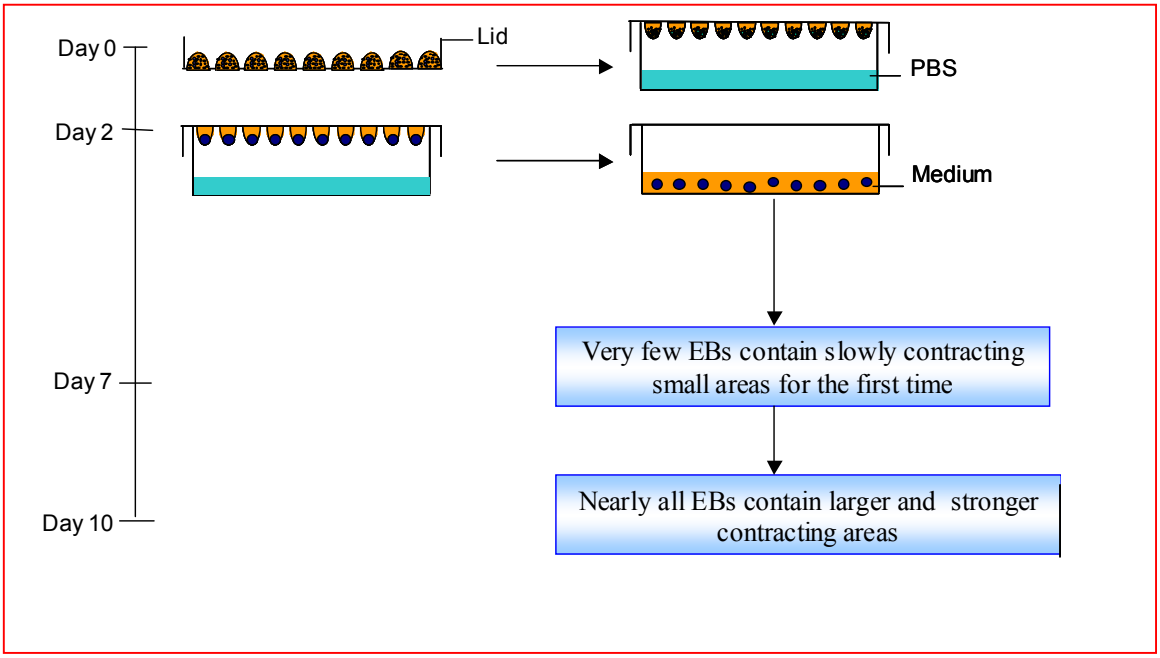


Figure 12. Schematic representation of the hanging drop protocol used for making EBs

Used a 70% confluent cell culture. Harvested the cells by trypsinisation and made viable cell counting with Trypan Blue dye exclusion method using Neubauer counting chamber. Cell suspension of 1X10^6 cells /ml was made initially. Then diluted 1:40 with EB differentiation medium to get the final titer of 2.5x10^4 cells per ml. Using a multi channel pipette, spotted 20µl of the cell suspension onto the inside of the upper lid of a10 cm bacteriological dishes with the lower lid containing 5ml PBS (Figure 12). Closed the dish carefully by inverting gently and incubated it for two days in the incubator for the cells spotted in droplets to form EBs. On day 2, the EBs were collected and transferred into a bacteriological dish containing 10ml of EB differentiation medium. Made medium change whenever the medium turns little yellow and made sure that the EBS don't starve of the medium. On 7^{th} –8^{th} days, first contractile areas will observed and almost all EBs will be contracting.

3.2.2 Basic Molecular Biological Methods.

DNA agarose gel electrophoresis

Agarose gel electrophoresis was performed to analyse the length of DNA fragments after restriction enzyme digests and polymerase chain reactions (PCR), as well as for the purification of PCR products and DNA fragments. DNA fragments of different molecular weight show different electrophoretic mobility in an agarose gel matrix. Optimal separation results were obtained using 0.5-2% gels in TAE buffer at 10 V/cm. Horizontal gel electrophoresis apparatus of different sizes were used. Before loading the gel, the DNA sample was mixed with 1/6 volume of the 6x DNA-loading buffer. For visualization of the

DNA fragments under UV-light, agarose gels were stained with 0.1µg/ml ethidium bromide. In order to define the size of the DNA fragments, DNA molecular standard markers were also loaded onto the gel.

BAC DNA isolation

Picked colonies and inoculated in 2 ml of LB culture containing the appropriate antibiotics. Incubated at 37°C over night with shaking at 1100 rpm. Next day, spun down the 2 ml over night cultures for 1 min at 11,000 rpm. Discarded the supernatant and resuspended the cell pellet in 200 µl buffer P1 with RNase (from QIAGEN DNA Maxi-preparation Kit). Added 200 µl of buffer P2 (Qiagen) and mixed by inverting the tube several times. Added 200 µl of buffer P3 (Qiagen) and mixed by inverting the tube several times. Spun down the white lysate at highest speed for 4 min. Transferred the clear supernatant into a new 1.5ml-Eppendorf tube and added 0.50 ml of 2-propanol. Mixed by inverting the tube and spin down the DNA at highest speed for 5 min. Discarded the supernatant and add 1 ml of 70% ethanol to rinse the pellet.10. Dried the pellet under the speed vacuum for 2 min or leave the tube open on the bench for 5 to 10 min until the DNA pellet is completely dried. The BAC DNA was resuspended in 10mM Tris-HCl, pH 8.0. Stored at 4^0C.

Plasmid and DNA isolation

QIAprep Spin Miniprep Kit and EndoFree Plasmid Maxi Kit were used for the isolation of plasmid DNA and followed exactly the manufacturers protocol.

Preparation of chemically competent DH5α E.coli

Inoculated 3 ml LB medium with the DH5α *E. coli* strain and incubated the culture overnight at 37°Cin an orbital shaker. The next day, added the overnight culture to 500 ml SOB^{++} medium and incubated the culture at 25-30°C until the absorbance at 600 nm reached approx. 0.5 (between 0.4 and 0.6). Chilled the culture for at least 10 min on ice. In the following steps, the cell suspension was kept on ice as much as possible. Centrifuged the cell suspension for 10 min at 4,500 rpm (Sorvall RCB4 rotor) or 6,000 rpm (Sorvall GSA rotor) at 4°C. Gently resuspended the pellet in 100 ml ice-cold TB buffer. Incubated the cell suspension on ice for 10 min. Centrifuged for 10 min at rpm (Sorvall SS-34 rotor) at 4°C. Gently resuspended the pellet in 18.6 ml ice-cold TB buffer and add 1.4 ml DMSO. Incubated the cell suspension on wet ice for at least 10 min. Aliquoted the cell suspension 200 µl per 1.5ml eppendorf tube. Shock-frozen the cell suspension in liquid nitrogen and stored the tubes at -80°. The transformation efficiency of the competent cells was determined using 5 ng of pure plasmid DNA and 200 µl competent cell suspension (see transformation protocol). This gave approx 1 x 10^8 transformants per µg plasmid DNA with DH5 α.

Transformation of chemically competent cells by heat shock

Transformation of DH5α *E.coli* with plasmid DNA was performed by incubating 200 µl of the competent cell suspension on ice after taken from the -80° storage followed by heat shock at 42^0C for 45 seconds and the subsequent incubation on in ice for 5 minutes. Then, 750ul SOC media was added and incubated at 37 in orbital shaker for 45 –60 minutes. Then the cells were plated onto the top of the agar plates containing the appropriate antibiotics.

3.2.3 Recombineering

Oligo design

Chosen 50 nucleotides immediately of the left of region to be subcloned from the BAC Ordered an oligonucleotide with this sequence at the 5' end and 3' end of this oligo included PCR primer sequence for amplification of the ColE1+amp template, given in *italics*

5'-(N)50 *TCACAGCTTGTCTGTAAGCGGATG* -3'. Chosen 50 nucleotides immediately right of the site to be subloned and transferred them into the ***reverse complement orientation***. Ordered an oligonucleotide with this sequence at the 5' end and the 3' end of this oligo included the 3' PCR primer sequence for the ColE1+amp template, given in *italics* 5'-(N)50 *GCTCTCCTGAGTAGGACAAATC* -3'. Included restriction sites between the 5' homology regions and the 3' PCR primer sequences.

PCR

The oligonucleotides were suspended in dH_2O at a final concentration of 25 pmol/µl. PCR reaction was set up in volume of 50 µl with 38.5 µl dH_2O, 5.0 µl 10 x PCR reaction buffer, 2.0 µl 5 mM dNTP, 1.0 µl upper oligonucleotide, 1.0 µl lower oligonucleotide, 2.0 µl ColE1+amp template and 0.5 µl Taq polymerase (5 U/µl). The following thermal conditions were used. 95^0C for 60 sec, 57^0C for 30 sec and 72° C 2min. 30 cycles. Checked 3 µl PCR products on a gel to ensure the PCR was successful. The size of the PCR product is around 2.7kb . Precipitated using 5 µl 3 M sodium acetate, pH 7.0, and 150 µl 100% ethanol. Mixed well and precipitate for 5 min at -80°C or 30 min at -20°C. Spun down the DNA at maximal speed for 5 min. Carefully washed the pellet once with 500µl 70% ethanol. Dried the pellet at 37°C using a heating block for 5 -10 min or vacuum dry for 2 min. Resuspended in 5 µl 10mM Tris-HCl, pH 8.0 (0.2 -0.5 µg/µl).

Electroporation of the Red/ET Plasmid pSC101- tetR-gbaA into the E.coli strain carrying the BAC with the gene of interest

Set up an overnight culture. Picked at least ten colonies carrying the BAC and inoculated them together in an Eppendorf tube containing 1.0 ml LB medium with appropriate antibiotics to select for your endogenous BAC. Punctured a hole in the lid for air. Incubated at 37ºC over night with shaking. Next day, set up an Eppendorf tube containing fresh 1.4 ml LB medium conditioned with the same antibiotics and inoculated with 30 µl of fresh overnight culture. Cultured for 2-3 hours at 37ºC, shaking at 1000 rpm. Centrifuged for 30 seconds at 11,000 rpm in a cooled Eppendorf benchtop centrifuge (at 2°C). Discarded the supernatant by quickly tipping out the supernatant twice, and place the tube on ice. Resuspended the pellet with 1 ml chilled dH2O, pipetting up and down three times to mix the suspension. Repeated the centrifugation and resuspended the cells again. Centrifuged and tip out the supernatant once more; 20 to 30 µl will be left in the tube with the pellet. Kept the tube on ice. Added 50ng the Red/ET recombination protein expression plasmid pSC101-tetR-gbaA and mixed gently. Kept the tube on ice. Fitting a larger pipette to the same tip, transferred up to 30 µl of cell suspension from the tube to the chilled electroporation cuvette. Electroporated at 1350 V, 10µF, 600 Ohms. with an Eppendorf® Electroporator 2510 using a 1 mm electroporation cuvette. Resuspended the electroporated cells in 1 ml LB medium without antibiotics and returned them to the Eppendorf tube. Incubated at 30ºC for 70 min, shaking at 1000 rpm. Using a small loop, plated 100 µl cells on LB agar plates containing tetracycline (25 µg/ml) plus the appropriate antibiotics for the BAC. Incubated the plates at 30°C overnight (or for at least 15 hours).

Subcloning

Once the cultures were incubated at 30^0C , the cells were induced with arabinose incubated for 1 hour with shaking at 1100rpm. Then the PCR product with the homology arms were electroporated as mentioned above. Then the cells were plated on LB agar plates prepared with appropriate antibiotics. The next day, the colonies were analyzed with restriction digestion for the subcloned construct.

3.2.4 Generation of ES cell targeting construct and clone generation

ES cell targeting construct

pIRES2 EGFP was purchased from Clontech, Germany. Human cytomegalovirus (CMV) immediate early promoter and enhancer were removed by a double digestion with *Nhe I* and *Ase I* and then subsequently blunt end ligated to get pIRES2 EGFP ϕ CMV. Puromycin cDNA flanked on its both ends by BamH1 restriction sites which was PCR amplified from pIRESPuro3 by *Pfu* DNA polymerase (Promega, Germany was inserted at the *Sma I* site of pIRES2 EGFP φ CMV construct to get pPuro IRES2 EGFP

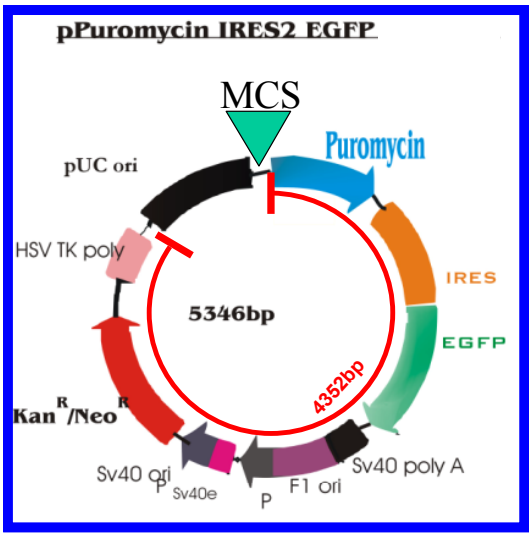


Figure 13. Plasmid map of pPuro IRES2 EGFP. This vector was constructed to derive the expression of puromycin N acetyl transferase to confer puromycin resistance and EGFP bicistronically with use of an IRES element. The multiple cloning sites just upstream of puromycin resistance cassette can be used to clone the promoter of interest to generate a reporter construct for a gene of interest. The plasmid encodes for neomycin resistance for selecting the transformants after electroporation and is the under the control of a ubiquitous promoter (SV40 promoter).

Clone Generation

30-40 μg plasmid was linearised in 100 μl reaction with appropriate enzyme(s). The enzyme is then heat inactivated at 80^0 C in a PCR machine or phenol chloroform extracted if the enzyme is resistant to heat inactivation at 80^0 C. Then the DNA was added to 10^7 cells in the Gene pulser electroporation cuvette and incubated on ice for 20 minutes before the electroporation. Then, the cells were electroporated with 500μFD and 0.24 Volts to get a time constant of around 9.8 for optimal results. The cuvette with the cells was place on ice and incubated for the next 10 minutes. The electroprated cells were divided equally to seed 3 gelatine-coated petriplate (100m diameter) with 10 ml media. At the end of 48 hours, antibiotic selection was applied to obtain stable transfectants either with neomycin at 400ug/ml or puromycin at 3ug/ml. When the negative background was cleared, around 9-11 days post transfection, the colonies were picked with aid of a microscope and propagated in the normal way

Generation of βactin clone expressing Puromycin resistance and EGFP constitutively

1.7kb βactin promoter excised from pCAGGS (kindly provided by Dr. Miyazaki, Tohoku University) with EcoRI and SalI was blunt end ligated to EcoRI digested, klenowed and 5' dephosphorylated pPuro IRES2 EGFP to get pβactin^p Puro IRES2 EGFP. This was then linearised with *SacI* and electroporated in CGR8 ES cell line and the neomycin resistant clones were picked, propagated and analyzed for the functional integrity of the targeting vector.

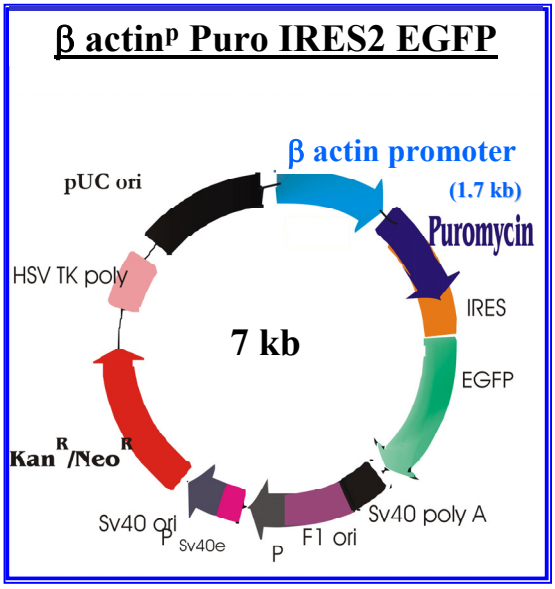


Figure 14. Plasmid map of pβactin^{p}Puro IRES2 EGFP

3.2.5 Derivation of T-brachyury lineages

T-brachyury Promoter Cloning by Recombineering

5kb promoter region upstream of the translation initiation site of the T-brachyury gene was subcloned from a BAC into the ES cell targeting construct pPuroIRES2 EGFP by a recombineering approach as follows (Figure 15). Sequences of 50 bases flanking on either side of the promoter region to be subcloned were verified by sequencing. Each 50b sequence was added to the 5' end of the forward/reverse primers with appropriate restriction enzyme recognition sites in between (Figure 16). These primers were then used to amplify the linear minimal vector containing ColE1 origin and Ampicillin resistance cassette. The linear amplified product was electroporated into the *E.Coli* harboring the BAC and induced to express Red ET proteins. Due to the homologous recombination events mediated by the Red E/T proteins and the 50bp homology regions, the minimal vector containing the promoter was positively selected with ampicillin. Then the minimal vector was digested with the appropriate restriction enzymes to release the T-brachyury promoter, which is then ligated in pPuroIRES2 EGFP to get pT-bra^{p} PuroIRES2 EGFP. This was then linearised and electroporated in CGR8 ES cell line and the neomycin resistant clones were picked, propagated and analyzed for the functional integrity of the targeting vector.

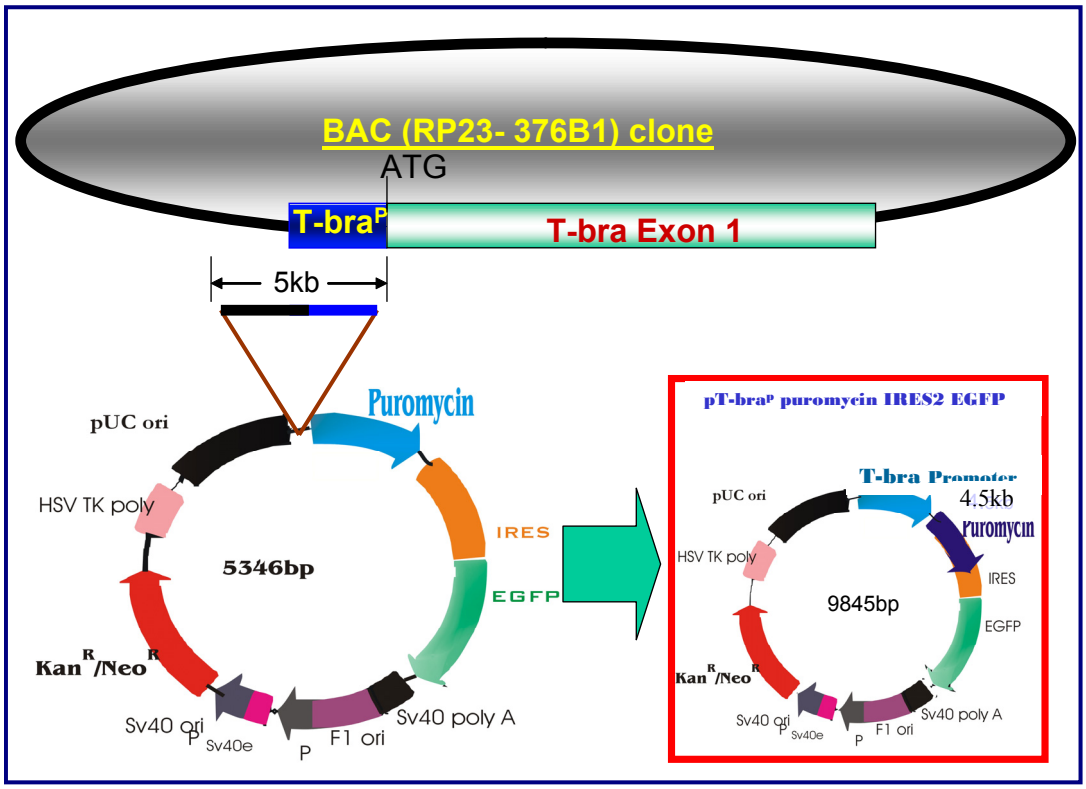


Figure 15. The overall view of the BAC subcloning by recombineering approach to clone T Brachyury promoter in the reporter vector approach.

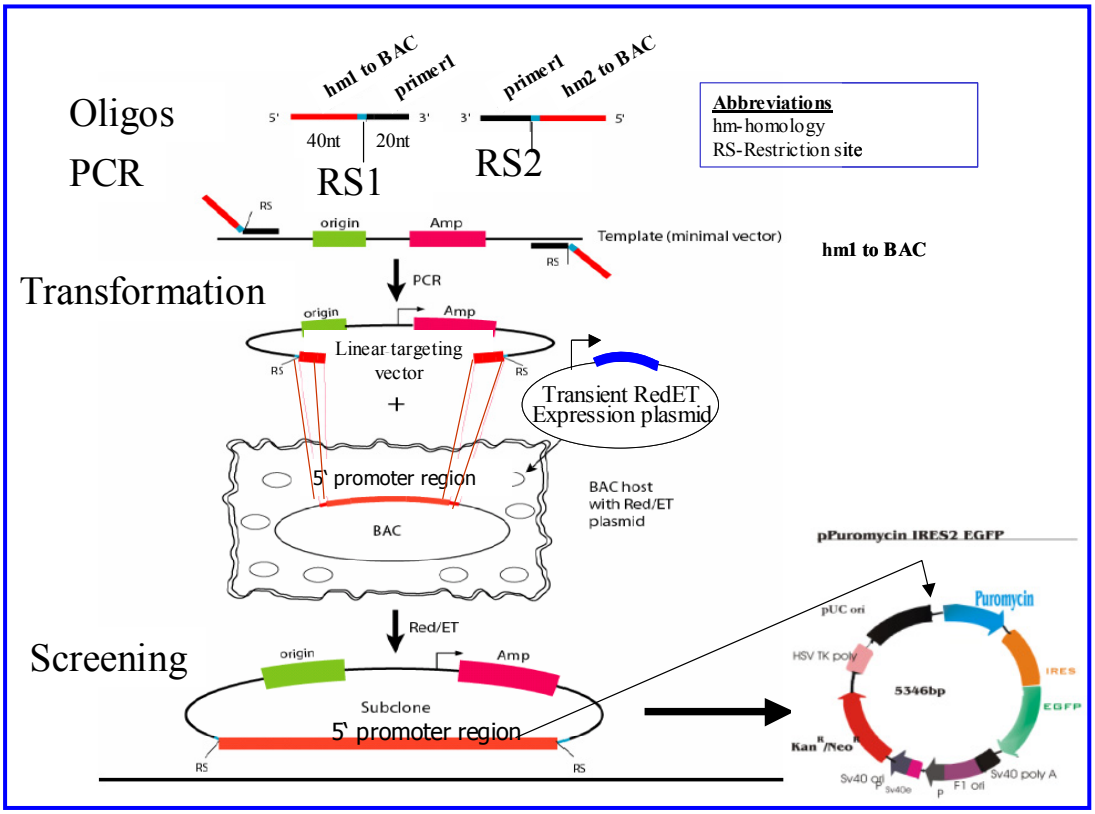


Figure 16. Schematic representation of recombineering process involved in sub cloning the T brachyury promoter from BAC.

Generation of T brachyruy reporter construct and reporter ES cell clones.

The 4.5 kb T Brachyury contained in the minimal vector was excised with *Bgl II* and ligated to *Bgl II* digested and terminal 5' phosphate removed (by calf intestinal phosphatase) pPuro IRES2 EGFP to generate pT Bra^{+} Puro IRES2 EGFP. The plasmids cloned in the right orientation were screened by restriction digestion with other enzymes and one among was checked by sequencing before electroporation. . This construct was electroporated in CGR8 with 500µF and 240V in a Bio-Rad Gene Pulser^{TM}. The transfected clones were selected by 400 µg/ml neomycin and after selection, the clones were maintained with 200µg/ml neomycin.

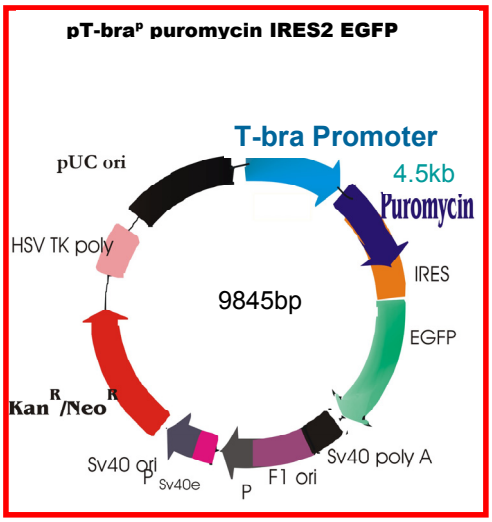


Figure 17. Plasmid map of T Brachyury reporter ES cell targeting construct

3.2.6 Derivation of BMP-2^{+} lineages

Generation of BMP-2 reporter cell line

A 2.9 kb BMP-2 promoter fragment with both proximal and distal transcription start sites, excised with Hind III and sac I digestion from pBMP-2-GL3 (kindly provided by Nandini Ghosh-Choudhury,USA) were blunt end ligated to EcoRI digested and klenowed pPuro IRES2 EGFP to generate pBMP-2^{p} Puro IRES2 EGFP. This BMP-2 reporter construct drives the expression of both puromycin resistance and EGFP under the control of BMP-2 promoter by the use of IRES sequence. This construct was electroporated in CGR8 with 500µF and 240V in a Bio-Rad Gene Pulser^{TM}. The transfected clones were selected by 400 µg/ml neomycin and after selection, the clones were maintained with 200µg/ml neomycin. During the EB generation, the neomycin selection was discontinued.

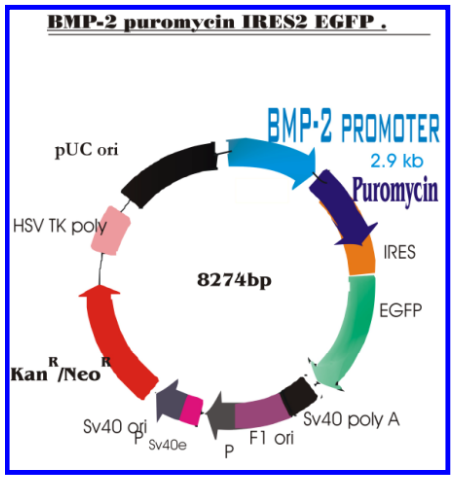


Figure 18. Plasmid map of BMP-2 reporter ES cell targeting construct

Isolation of BMP-2$^+$ cells

To induce differentiation, ES cell suspension of 1.6x10^4 cells/ml was made in IMDM supplemented with 20% FCS, 1% none-essential amino acids (vol/vol), 2 mM L-glutamine and 100 µM β-ME. Each bacteriological dish containing 25 ml of this ES cell suspension was placed in a 37^o C incubator with 5% CO_2 for 2 days. Then the 2-days old EBs were transferred into a 0.2% gelatine coated 10 cm tissue culture dish. On day 4, the EBs were treated with 3µg/ml puromycin for subsequent 3 days with every day medium change with fresh puromycin. On day 7, the puromycin resistant BMP-2$^+$ cells were used for the experiments.

3.2.7 Derivation of αMHC lineages

Generation of αMHC reporter construct and αMHC ES clones.

The 5.5 kb αMHC promoter construct was a kind gift from Dr. J. Robbins (Cincinnati, OH). The promoter was klenowed on one end and the other end digested with *Sac*II was inserted with one end *Eco*RI digested and klenowed and other end digested with *Sac*I fragment of pPuro IRES2 EGFP to generate p αMHC^p Puro IRES2 EGFP. This αMHC^p reporter construct drives the expression of both puromycin resistance and EGFP under the control of αMHC promoter by the use of IRES sequence. This construct was electroporated in CGR8 with 500µF and 240V in a Bio-Rad Gene Pulser^{TM}. The transfected clones were selected by 400 µg/ml neomycin and after selection, the clones were maintained with 200 µg/ml neomycin. During EB generation, the neomycin selection was discontinued.

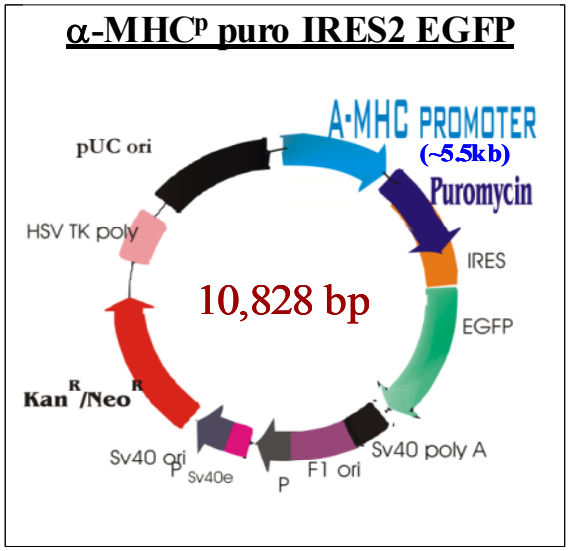


Figure 19 Plasmid map ofαMHC reporter ES cell targeting construct

Isolation of αMHC$^+$ cardiomyocytes

To induce differentiation, the hanging drop protocol was used. Briefly, an ES cell suspension of 2.5x10^4 cells per ml titer was prepared in IMDM supplemented with 20 % FCS, 1 % non-essential amino acids (vol/vol), 2 mM L-glutamine and 100 µM β-ME. 20 µl of this ES cell suspension was spotted on the inside of the upper lid of a 10cm bacteriological dish and then covered over its bottom dish containing 5 ml PBS. On day 2, the formed multicellular aggregates (EBs) were transferred to suspension in a new dish with 20 ml IMDM supplemented with 20% FCS, 1% non-essential amino acids (vol/vol), 2 mM L-glutamine and 100 µM β-ME. On day 8, contracting clusters were observed. At this stage, cultures were either treated with puromycin (4µg/ml) or left without treatment for 7 days with a medium change on every alternative day. On day 15, the beating clusters were trypsinised and used for FACS analysis or used for RNA extraction.

3.2.8 Patch clamp and sharp electrode electrophysiological studies

AP recordings were performed in single αMHC cardiomyocytes and in multicellular αMHC$^+$ cell aggregates at days 11-16. Single cardiomyocytes were obtained by dissociating puromycin-purified αMHC$^+$ cell aggregates using a protocol described before and incubated in DMEM + 20% FCS at 37°C and 5% CO_2 for 24-36 hours prior to the measurements. APs of spontaneously beating single αMHC$^+$ cardiomyocytes were measured by means of the whole-cell current-clamp technique using an EPC-9 amplifier and the PULSE software package (Heka Elektronik, Lambrecht, Germany). Patch clamp electrodes had a resistance of 3-4 MΩ when filled with intracellular solution containing (in mM) 50 KCl, 80 KAspartate, 1 $MgCl_2$, 3 MgATP, 10 EGTA and 10 HEPES, pH 7.4.

Spontaneous electrical activity of cardiomyocytes within plated □MHC$^+$ cell aggregates was assessed by conventional microelectrode recordings. The resistance of the sharp electrodes, which were filled with 3 M KCl, was 20-50 MΩ. Signals were acquired by a SEC-10LX amplifier (npi electronic, Tamm, Germany) connected to the PULSE software via the interface of the EPC-9.All experiments were performed at 37°C in standard extracellular solution containing (in mM) 140 NaCl, 5.4 KCl, 1.8 $CaCl_2$, 1 $MgCl_2$, 10 HEPES and 10 glucose or NaCl 136, KCl 5.4, NaH_2PO_4 0.33, $MgCl_2$ 1, glucose 10, hepes 5, $CaCl_2$ 1.8 (pH 7.4 adjusted with NaOH). All electrophysiological data were analyzed off-line with a custom made analysis software (kindly provided by Philipp Sasse).

3.2.9 Immunohistochemistry

One day prior to the sample processing, 50,000 cells per each well of Lab-Tek Permanox slide Chambers (Nalge Nunc International, USA) were seeded and cultured in the absence of puromycin. After 24 hours, the samples were fixed with 4% Paraformaldehyde or with –20 ^{0}C cooled methanol-acetone (1:1) solution, permeabilised with 0.1%Triton X-100, and labeled with the appropriate primary antibody in the optimal dilutions followed by labeling with the respective secondary antibodies. The specificity of the antibodies has been tested using the appropriate tissues and isotype control antibodies.

3.2.10 Sudan Red and Alizarin Red S stainings

The paraformaldehyde fixed samples were treated with Alizarin Red S for 5 minutes and then excess Alizarin Red was washed off by dipping in Aceteone: Xylene (1:1), followed by wash in Xylene. Then stained for haemtoxylin for 10 minutes and the excess dye was washed off in running water.

3.2.11 Semi-quantitative RT-PCR

Design of the RT-PCR primers

For the design of the primers, the online software “Primer3” was used. (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Care is taken to choose the exon spanning primers to avoid the genomic DNA amplification. Primers with 23 bases and melting temperature around 60^0 C which give rise to amplicon sizes from 150-400 base pairs were preferred. Most of the Oligos were ordered from Metabion (Martinsried, Germany) and the rest from MWG (Ebersberg, Germany).

cDNA synthesis and PCR

Total RNA was extracted using RNeasy Mini Kit (Qiagen,Hilden,Germany) with on-Column DNase I (Qiagen) digestion according to the manufacturer’s instructions. 5μg total

RNA was reverse transcribed using SuperScript II Reverse transcriptase (Invitrogen) with random primers according to the manufacturer's recommended protocol. PCR amplification was done with REDTaq ReadyMix (Sigma) with 0.4μM each primer. GAPDH was used as an internal control. The following conditions were used. An initial denaturation at 95^0C for 2 minutes, followed by 22-35 cycles of 30 sec denaturation at 95^0C, 30 sec annealing at 60^0C and 60 seconds of elongation at 72^0. A final extension at 72^0C for 5 minutes was included. Electrophorectic separation of PCR products was carried out on 2% agarose gels with 0.001% Ethidium Bromide.

3.2.12 FACS analysis

Single cell suspension was prepared by trypsinisation. Cell clumps were removed by passing through cell strainer cap of round bottom tube from Falcon® (BD, Germany). Propidium Iodide staining (Sigma) was included to exclude dead cells. Acquisition of 10,000 live (PI negative) cells was made with FACScan (BD Biosciences) and the data analysis was done with CellQuest software (Becton Dickinson). The respective wild type EBs on the same day as the sample EBs were used as the control.

3.2.13 Affymetrix Hybridization

Total RNA was extracted from undifferentiated ES cells and EBs using the RNeasy total RNA isolation kit (Qiagen GmBH, Hilden, Germany). The preparation quality was assessed by agarose-formaldehyde gel electrophoresis. Three independent total RNA preparations, each 15 µg from the enriched specific lineages and the mixed cell population and from the undifferentiated ES cells were labeled with Target Labeling and Control Reagent package (Affymetrix) as described in the manufacturer's instructions. Briefly, double-stranded cDNA was synthesized using the one-cycle cDNA synthesis module. Biotinylated cRNA was synthesized with the IVT labeling kit and cleaned up using the sample cleanup module. After fragmenting of the cRNA for target preparation using the standard Affymetrix protocol, 15 µg fragmented cRNA was hybridized for 16 h at 45 °C to Mouse Genome 430 2.0 arrays which carry probes representing 45101 probe sets. Following hybridization arrays were washed and stained with streptavidin-phycoerythrin in the Affymetrix Fluidics Station 400 and further scanned using the Hewlett-Packard GeneArray Scanner G2500A. The image data were analyzed with GCOS 1.4 using Affymetrix default analysis settings and global scaling as normalization method. After RMA normalization [63] three pair-wise comparisons have been performed using the Student t-test (unpaired, assuming unequal variances). A Student t-test p-value < 10^{-2} and a fold change > 2 were used to identify and restrict the number of differentially expressed genes. For differentiation of treatment and developmental aspects hierarchical clustering of probe sets differential expressed between the three conditions in the F-test comparison (p-value cut off 10^{-7}; n = 617 probe sets) has been performed. The cluster analysis has been done using cluster version 2.11 [64] applying mean-centering and normalization of genes and arrays before average linkage clustering with uncentered correlation.

3.2.14 Affymetrix Analysis

DAVID functional annotation tool was used to classify the transcripts according to the Gene Ontologies. (http://david.abcc.ncifcrf.gov/summary.jsp). Hierarchical clustering was performed with Gene Cluster (Eisen et al. (1998) PNAS 95:14863).

3.2.15 Quantitative Real time PCR

Validation of the Affymetrix data was performed by quantitative real time PCR analysis with the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA). 100ng RNA was reverse transcribed with ThermoScript™ Reverse

Transcriptase (Invitrogen). Then Real time PCR was perfomed in triplicates for every sample using TaqMan Gene Expression Assays from Applied Biosystems (Foster City, CA). The Gene Expression Assays included for the validation were *Brachyury (T)* (Mm00436877_m1), *Sox17* (Mm00488363_m1), *αMHC* (Mm00440354_m1), *BMP-2* (Mm01340178_m1), *GAPDH*(Mm99999915_g1), and *Nanog* (Mm02019550_s1). Averaged C_t values of each qRT-PCR reaction from the target gene were normalised with the average C_t values of the housekeeping gene, *GAPDH,* that ran in the same reaction plate to get ΔC_t value . The fold change was calculated by using the formula, fold-change= $2^{-(\Delta C_t \text{ gene1} - \Delta C_t \text{ gene2})}$. ΔC_t of the gene in the sample in which it is expressed lowest is taken as ΔCt gene2 to calculate the fold change using the above formula. The resulting fold change is expressed as percentage of the maximum.

3.2.16 Episomal overexpression

For episomal overexpression studies, the original pMGD20 (a kind gift from Dr. M Gassmann) was linearised with *PvuI*, klenowed and blunt end ligated to EcoRI digested, klenowed pEF-1/Bsd vector (purchase from Invitrogen) to get pMGD20 Blasticidin. This plasmid encodes the polyoma large T antigen required for the episomal maintenance of the vector pPyCAGIP. T Brachyury clones were transfected with linearised pMGD20 Bsd with *NotI.* The novel transcripts cDNA clones were purchased from RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH, Germany and geneservices, UK. The cDNAs were cloned in pPyCAGIP and the resulting overexpression plasmid was used to transfect the T Brachyury ES cells with lipofectamine 2000 transfection which are already transfected with pMGD20 Bsd. Stable clones were generated with puromycin selection and used for the experiments and the expression of the Large T antigen was checked by semi-quantitative PCR as shown in figure 20A.

3.2.17 siRNA Knock down

Optimization of siRNA delivery with Lipofectamine

The optimal amount of lipofectamine and optimal concentration of siRNA was determined using the siControl Reagent from Dharmacon. Positive control siRNA labeled with Rhodamine and targeted against Cycophilin B was used for the optimization. $1x10^5$CGR8 cells were plated 24 well dishes one day before transfection and were transfected with gradient amounts of Lipofectamine from 0.5μl to 1.5μl with different concentrations of siRNA for determining the optimal conditions. When increased the Lipofectamine amounts more than 1.5 μl, toxicity was observed to a greater extent. As shown in the figure 20B, 1.0μl Lipofectamine 2000and 100nm siRNA was found to be optimal, when performed in 24 well dishes. When transfected in 6 well plates, the amounts were linearly scaled up (4 fold) as recommended by Dharmacon.

A

B

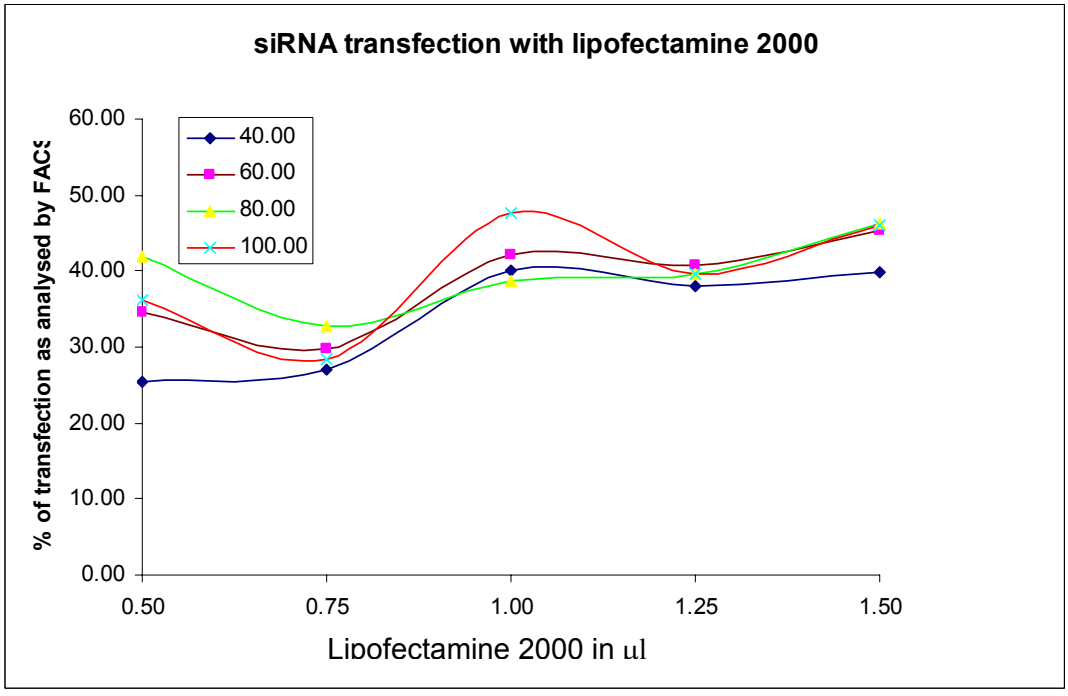


Figure 20. **A.,** Verification of the clones for their Large T antigen expression for episomal over expression studies. **B.,** Optimization of siRNA delivery in CGR8 ES cells with Lipofectamine 2000 for knockdown studies.

4.Results and Discussion

4.1 EXPERIMENTAL DESIGN

For the isolation of mesodermal lineages and cardiomyocytes, the promoters of the following candidate genes were made use of to generate the respective stable ES cell lines with reporter (enhanced green fluorescent protein) and selection marker (puromycin resistance) under the control of the respective promoters.

1. *T-brachyury* ---Isolation of nascent mesodermal cells
2. *Bmp2* ---Isolation of late mesodermal cells
3. *αMyosin heavy chain* ---Isolation of cardiomyocytes.

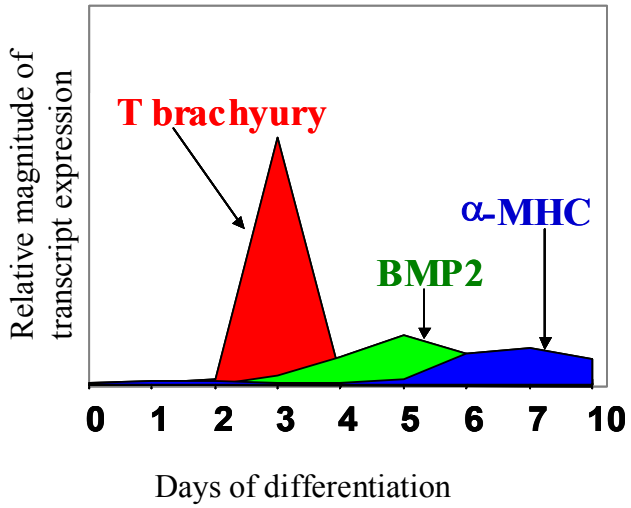


Figure 21 Expression patterns of the T Brachyury, BMP-2 and αMHC transcripts used for isolation of nascent mesoderm, late mesoderm and cardiomyocytes respectively during the course of differentiation, as obtained from microarray analysis.

Once the pure populations of mesodermal and cardiac cells were obtained, transcriptomic profiling of specific lineage cells were performed using microarrays. In addition, phenotypic analysis of the isolated cells was analyzed by immunohistochemistry and semi-quantitative PCR. Using bioinformatics approach, all the differentially expressed genes were analyzed. In the latter part, functional evaluation of the novel transcripts highly expressed in mesodermal cells but whose functions are still not yet annotated was done by using loss of function (siRNA mediated knock down) and gain of function (episomal over-expression) studies.

4.2 Transcriptomic profiling of wild type ES cells and differentiating EBs

ES cells are pluripotent and can give rise to any type of cells depending on the culture conditions. ES cells when maintained in the presence of adequate amount of LIF or cultivated on mouse embryonic fibroblasts remain in undifferentiated state and represents the inner cell mass of the blastula in the embryonic context.

When applied hanging drop protocol or placed in suspension in the absence of LIF, ES cells differentiate into 3 dimensional multicellular aggregates called Embryoid bodies. The mesoderm is established on day 3, the time point at which T Brachyury expression peaks high. The first contracting cardiomyocytes are observed on day 7 onwards. For the selective lineage of any mesodermal cells, either the formation of more mesodermal cells should be favoured at the expense of other germ layers or the differentiation of mesoderm into the lineage of choice should be enhanced at the expense of other mesodermal lineages. Till date, the mechanisms of mesoderm development and other germ layers development remain still elusive.

In order to better understand the molecular mechanism of germ layer development and their lineage commitment, a large scale transcriptomic profiling of undifferentiated embryonic stem cells and the differentiating EBs at daily intervals over a period of 10days, the latest time point at which the onset of cardiomyogenesis is complete, was performed as a first step towards characterizing the pure mesodermal and cardiac population. The morphologies of the embryonic stem cells and the differentiating EBs are shown in the Figure 22. We used feeder layer free embryonic stem cell line CGR8 and “Hanging drop protocol” as outlined in section (3.2.6) to obtain more reliable and reproducible experimental results. The differentiation of embryonic stem cells is stage specific. Establishment of ectoderm precedes mesoderm formation and occurs between day 0-3. Establishment of mesoderm and definitive endoderm occurs 2 days onward and 3days onwards respectively. Thus the germ layer development from ES cells mimics the embryonic development sequentially (Figure 23).

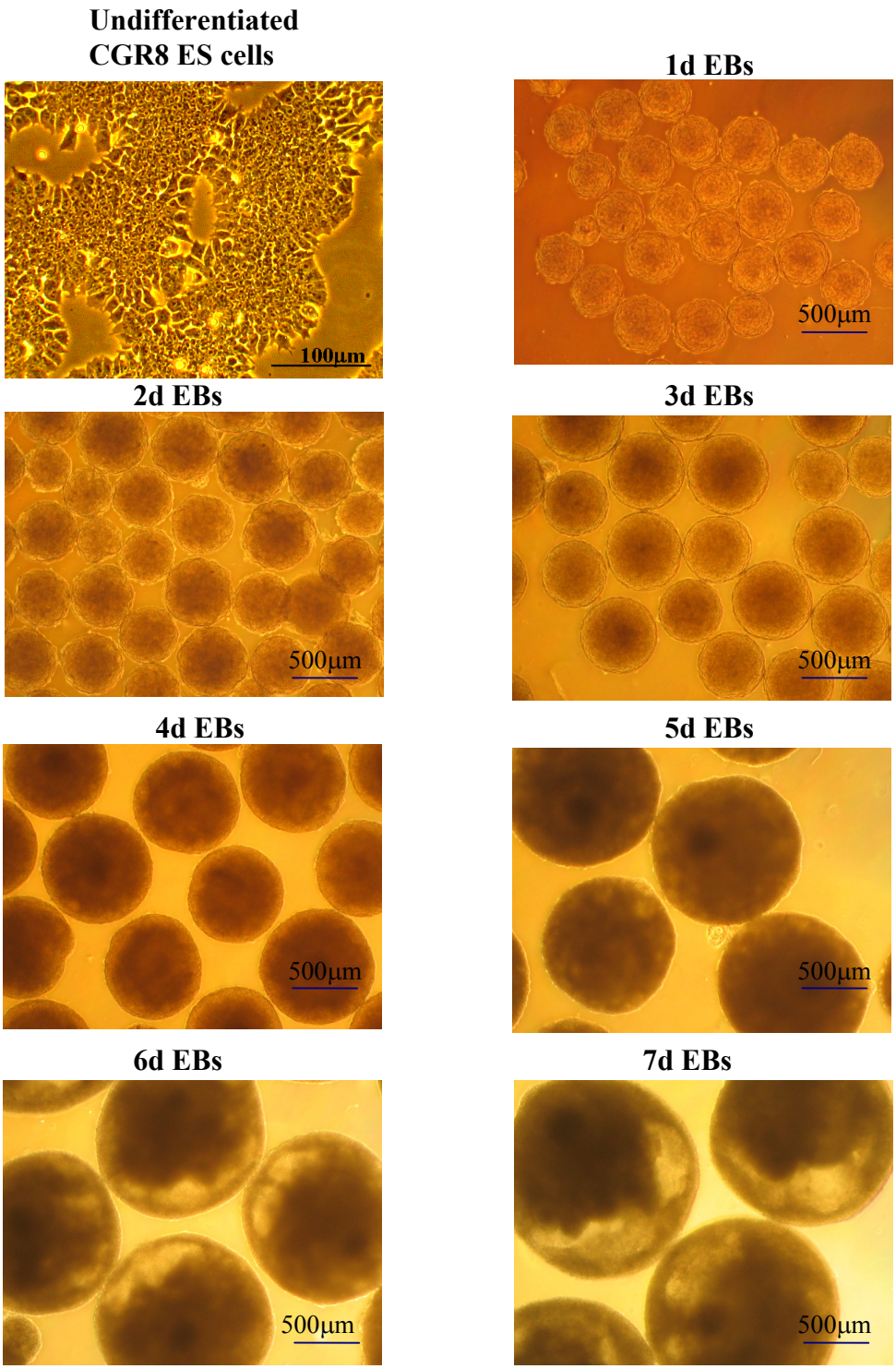


Figure 22: Morphologies of the differentiating EBs on daily intervals. Cavitation occurs in the same way as in the in vivo embryonic development, starting from day 5 at which time point BMP-2 is expressed at its peak and is postulated to be the initiator of cavitation by its apoptotic property.

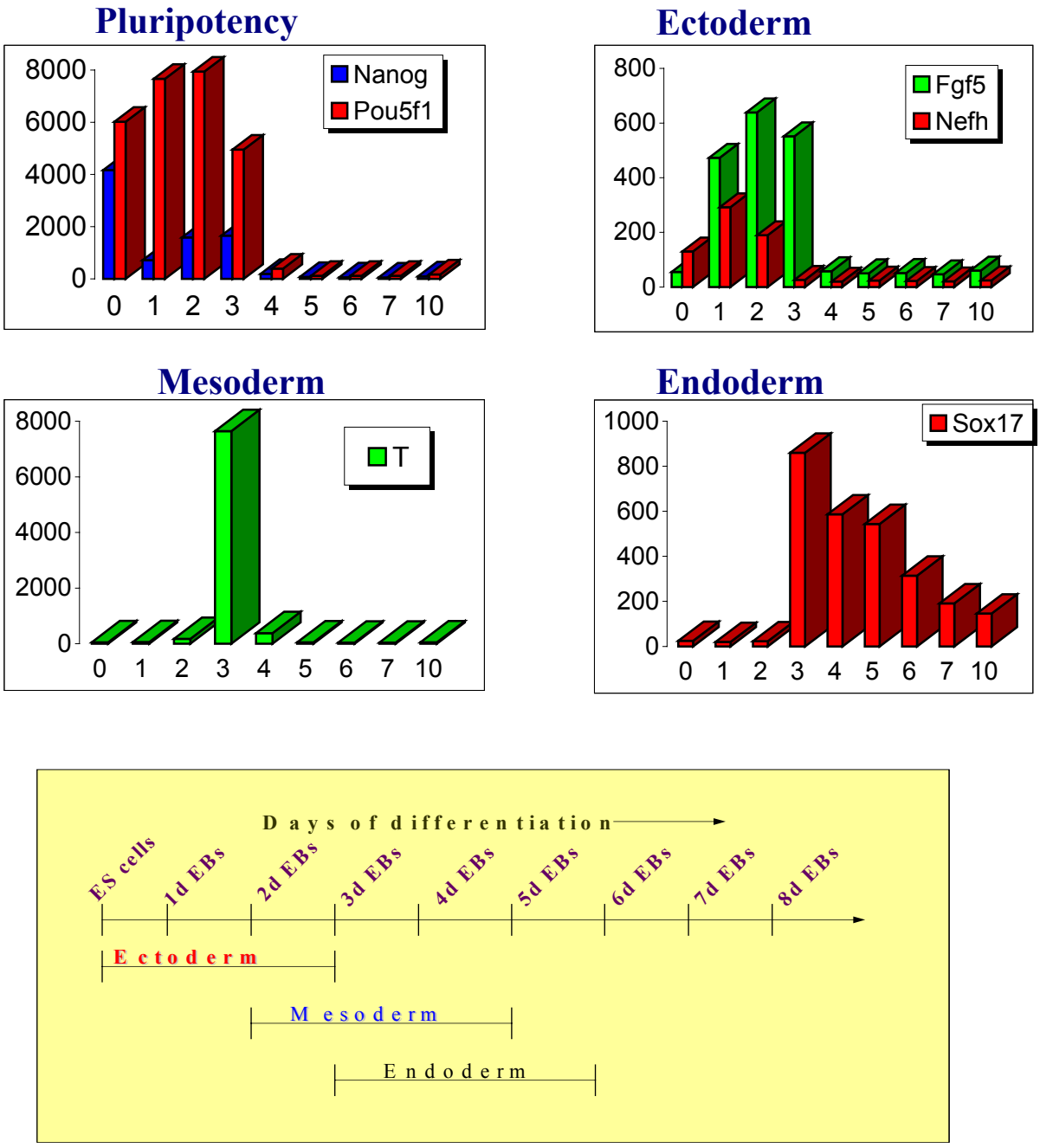


Figure 23 Stages of germ layer formation during embryonic stem cell differentiation. *Nanog* and *Pou5f1* are the markers for pluripotent stem cells, *Fgf5* , marker for primitive ectoderm, Neurofilament heavy(*Nefh*), marker for ectodermal lineages, *T Brachyury*, a pan-marker for nascent mesoderm and *Sox17*, pan marker for definitive endoderm. The values on Y axis represent the relative magnitude of expression.

4.2.1 Validation of the Affymetrix data by Real Time PCR Quantitation

The affymetrix data obtained from the ES cells and the differentiating EBs were verified by quantitative real time PCR analysis of the pluripotent and germ layer specific genes –*Nanog, T- Brachyury, Bmp-2, Neurofilament H, Sox 17* and α- myosin heavy chain. As seen in the figure 24, the Real time PCR data correlated very well with the affymetrix data.

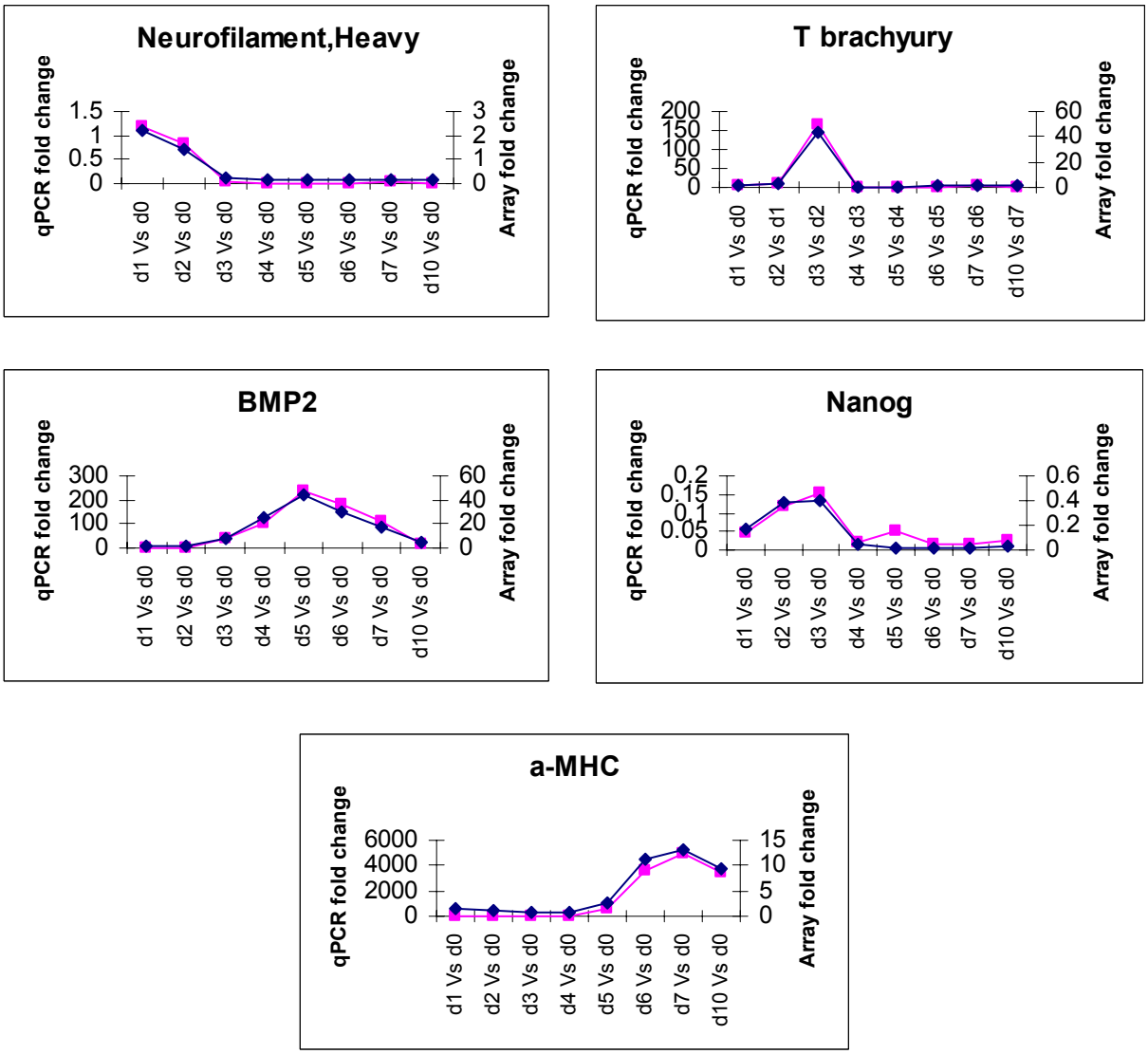


Figure 24 Validation of the affymetrix results by qPCR. The fold change was calculated by using the formula, fold-change = $2^{-(\Delta C_t \text{ gene1} - \Delta C_t \text{ gene2})}$. ΔC_t of the gene in the sample in which it is expressed lowest is taken as ΔCt gene2 to calculate the fold change using the above formula. The pink line represent results from quantitative PCR and the blue line from affymetrix data.

4.3 CHARACTERIZATION OF T-BRACHYURY^{+} MESODERMAL POPULATION

In the mouse embryo, mesoderm is generated from the epiblast or embryonic ectoderm through the process of gastrulation that is initiated at approximately day 6.5 of gestation (for review see [10]). At the onset of gastrulation, the epiblast cells in the region that defines the posterior part of the embryo undergo an epithelial to mesenchymal transition and form a transient structure known as the primitive streak from which the mesoderm emerges. The newly formed mesoderm migrates away from the primitive streak, moves laterally and anteriorly and is patterned into various populations with distinct developmental fates. Brachyury is expressed in all nascent mesoderm and downregulated as these cells undergo patterning and specification into the derivative tissues including skeletal muscle, cardiac muscle and connective tissues in addition to blood and endothelium[27,65]. In the ES/EB model system, mesoderm, as defined by expression of the T-box gene Brachyury, is induced within 48 hours of the onset of differentiation and persists until day 4 (Figure 25). To visualize and to isolate these short lived nascent mesodermal cells which are at the very beginning of patterning, we generated transgenic ES cell clones transfected with a plasmid construct in which T Brachyury promoter (4.5kb) drives the expression of both EGFP and puromycin resistance cassettes bicistronically with the help of an IRES element.

4.3.1 Generation and analysis of the T-brachyury ES cell clone

Stable T-Brachyury clones were generated by electroporation with linearised pTbra^{p} IRES2 EGFP construct with neomycin selection. Embryoid Bodies, when made from the T-brachyury ES cell clone by hanging drop protocol and analyzed at regular intervals, showed the transient expression of T-brachyury transcripts as shown by RT-PCR in figure 25a. FACS analysis of the EBs showed the same pattern of EGFP expression as the transcript expression (Figure 25D). The ES cells and EBs from this transgenic clone behaved faithfully with the wild type ES and EBs. The T-brachyury expression is distributed uniformly throughout the EBs as evident from the EGFP expression in every EB (Figure 25C).

4.3.2 Isolation and Transcriptomic Profiling T-brachyury^{+} population

On day 0, EBs were made from T-brachyury ES (CGR8) cell line by hanging drop protocol. On second day, the hanging drops EBs were transferred into suspension. On third day, the EBs were treated or untreated with puromycin at a concentration of 5ug/ml for 2 days. On 5th day, the medium was changed and added medium with or without puromycin at a concentration of 2μg/ml for 1 day. On day 6th , RNA was isolated from EBs treated or untreated with puromycin for affymetrix analysis. Application of puromycin led to the enrichment of the T-brachyury cells up to 79% (Figure 26). The puromycin resistant T-

brachyury population along with the undifferentiated ES cells from the T-brachyury ES cell clone and the control 6day EBs without puromycin treatment were included for transcriptomic profiling.

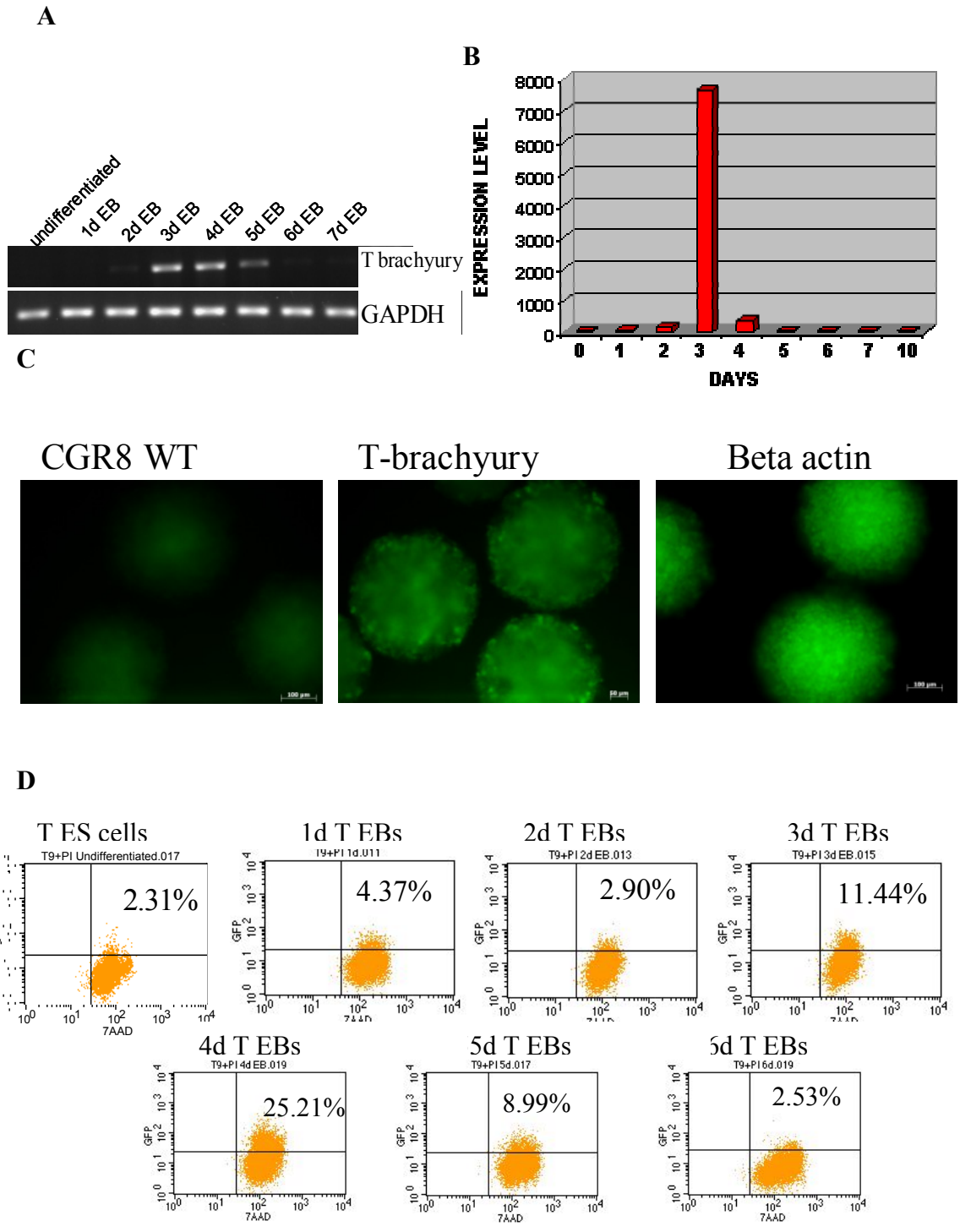


Figure 25 A) Semi quantitative RT-PCR analysis of T Brachyury transcript expression in ES cells and differentiating wild type EBs. B). Affymetrix analysis of the T Brachyury transcript expression in ES cells and in differentiating wild type EBs. C) Fluorescence microscopic pictures of 4-day-old EBs from CGR8 wild type, T Brachyury clone, and β actin clone. The latter expressed GFP ubiquitously. The intensity of green fluorescence emitted by Brachyury EBs very low compared to the strong (βactin) promoter driven EGFP. D) FACS analysis of fluorescence emission by Brachyury ES cells and EBs during the course of differentiation, as an index of T-brachyury expression.

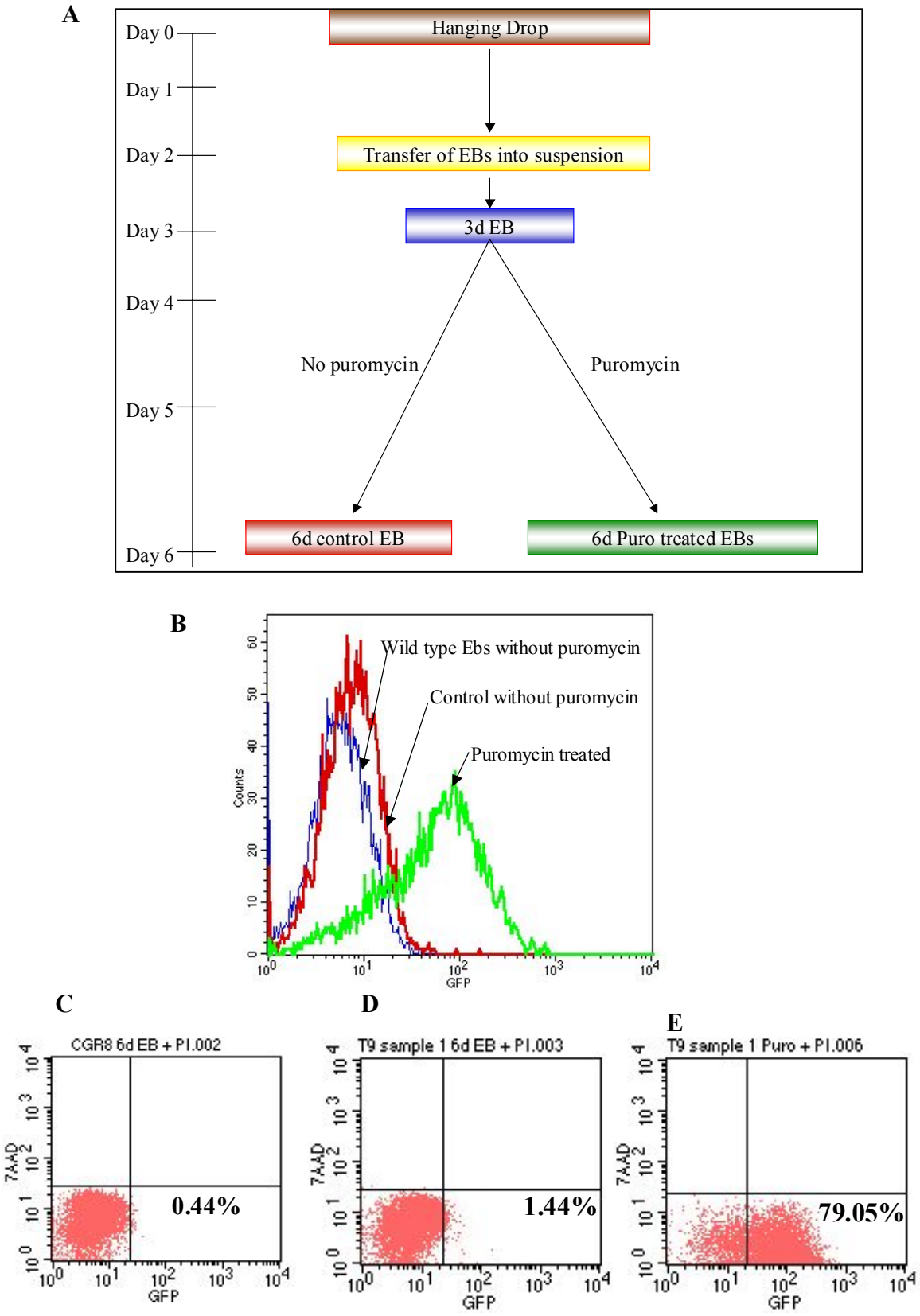


Figure 26 A) Schematic outline of the protocol for the enrichment of T Brachyury population by puromycin selection. B, C, D and E). The purity of the T-brachyury population obtained by the protocol (A) as assessed by FACS analysis.

4.3.3 Validation of Affymetrix data by quantitative real time PCR

The affymetrix data obtained from the ES cells, puromycin treated 6day EBs (T-brachyury^{+} cell population and puromycin untreated 6 day EBs) were verified by quantitative real time PCR analysis of the pluripotent and germ layer specific genes –Nanog, T-Brachyury, BMP-2, Neurofilament H, Sox 17 and α- myosin heavy chain. As seen in the figure 27, the Real time PCR data correlated very well with the affymetrix data.

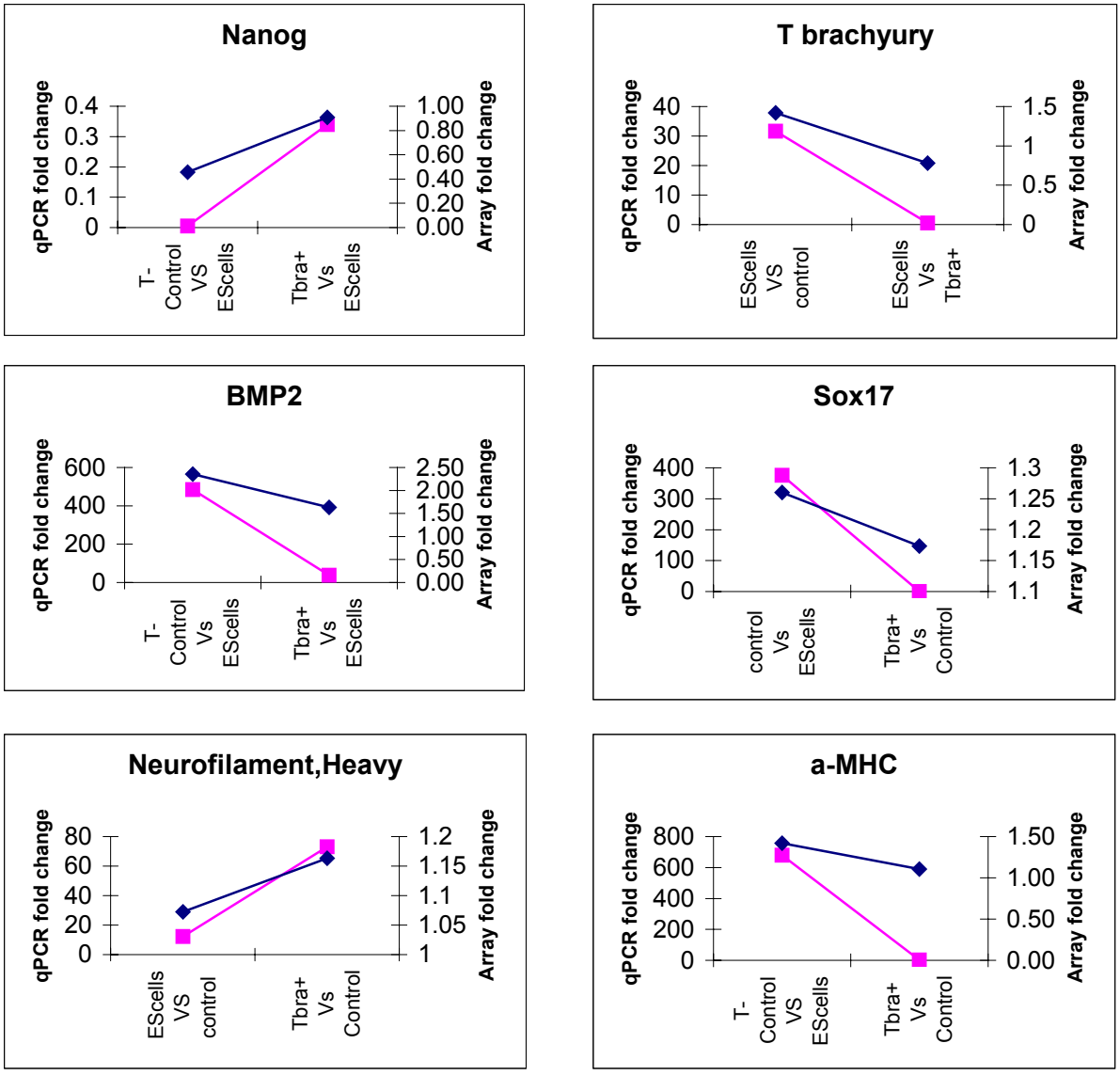


Figure 27. Validation of the affymetrix results by qPCR. The fold change was calculated by using the formula, fold-change = $2^{-(\Delta C_t \text{ gene1} - \Delta C_t \text{ gene2})}$. ΔC_t of the gene in the sample in which it is expressed lowest is taken as ΔCt gene2 to calculate the fold change using the above formula. The pink line represent results from quantitative PCR and the blue line from affymetrix data.

4.3.4 Affymetrix analysis of the transcriptome of T-brachyury$^+$ population

To rule out the differentially expressed transcripts are specific to T-brachyury cells and are not due to the puromycin application, transgenic β actin clones were generated in the same way like the T-brachyury clones except that the βactin promoter in the place of T-brachyury was used to serve as puromycin treatment controls. There were 6 unique transcripts with student t-test value of $p<0.01$ differentially expressed between puromycin treated 7day β actin EBs and untreated β actin EBs (as discussed for BMP-2 transcriptome analysis in the next chapter) and were excluded from the analysis. Thereafter the transcriptomic analysis was performed as follows.

4.3.4.1 Major pathways / biological processes differentially regulated in T Brachyury cells

Overall, there are 5616 unique transcripts differentially expressed (2 fold up or 2 fold down with an ANOVA p value cut off of less than 0.01) compared to either the T Brachyury ES cells or 6day old control EBs (i.e., the intersection of all differentially expressed genes in T Brachyury population in comparison to at least one of either the ES cells or 6day EBs. Among these transcripts, 711 transcripts are participating in pathways and 723 genes are implicated in embryonic development as shown in figure 28. Strikingly 95 transcripts participate in MAPK signaling pathway, 54 transcripts in Wnt Signaling and 41 transcripts in TGFβ signaling pathways (Figure 28). The differential expression of these transcripts in each pathway is summarized in figures 29,30and 31. Most of the rest transcripts take part in metabolism and a significant number of them in cell cycle, p53 signaling pathway, focal adhesion and axon guidance as shown in annexure 1.

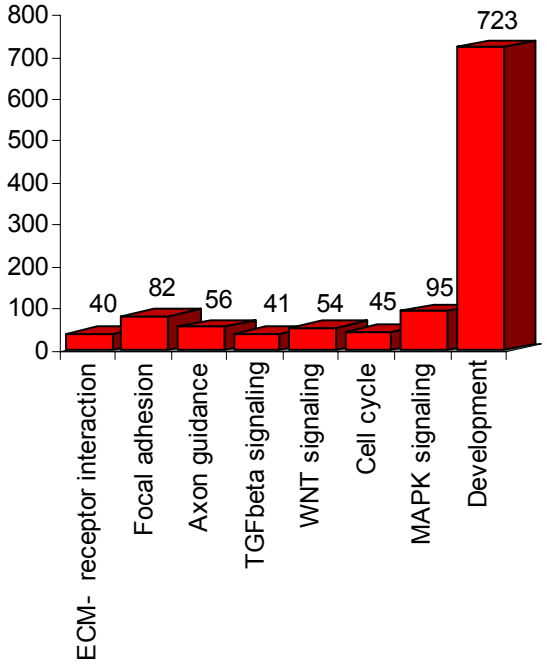


Figure 28. An overview of the number of transcripts differentially expressed in T brachyury population according to different Gene Ontologies(GOs), as obtained from DAVID analysis.

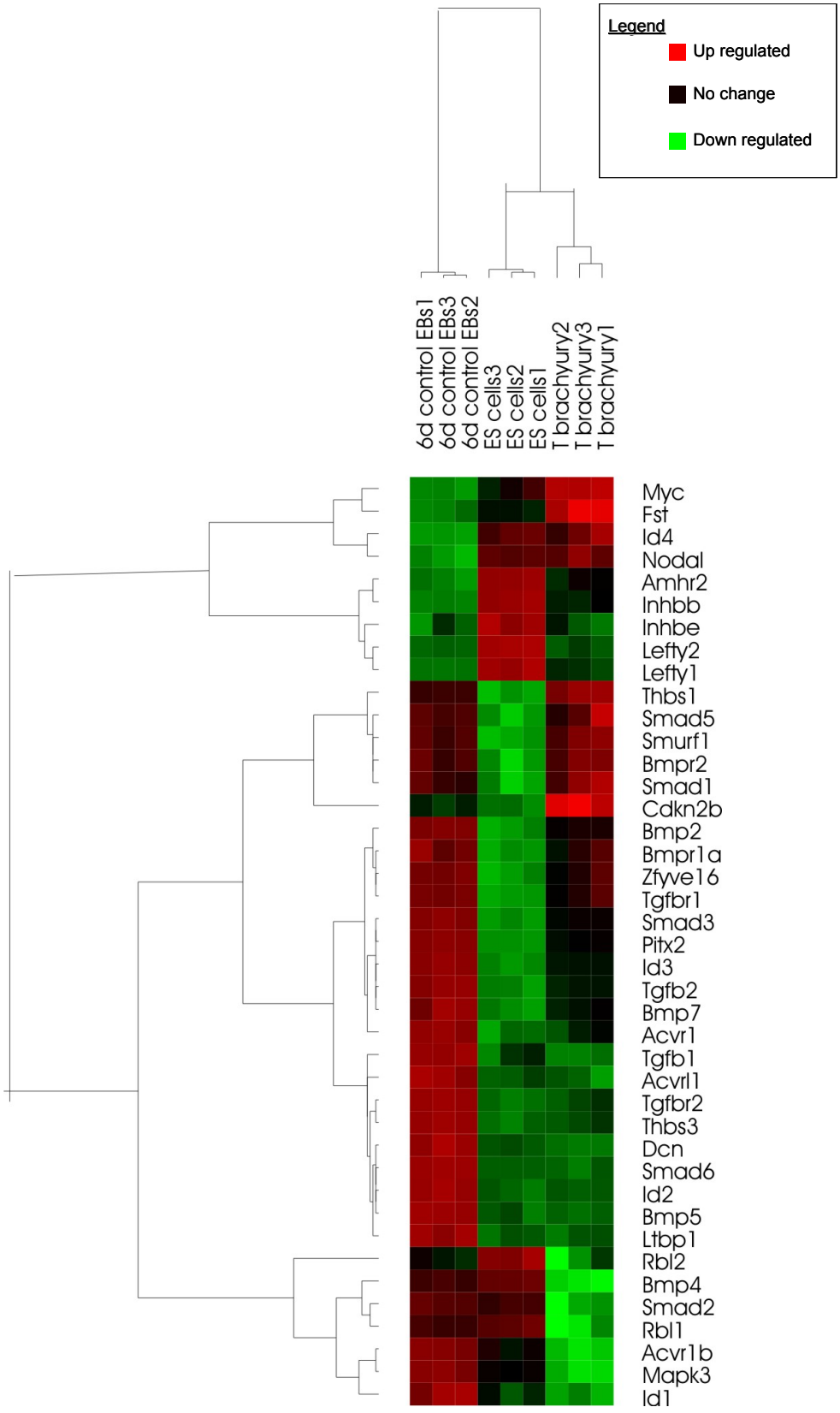


Figure 29. Hierarchical clustering of the transcripts differentially expressed in T-brachyury^{+} cells that participate in TGF β signaling pathway

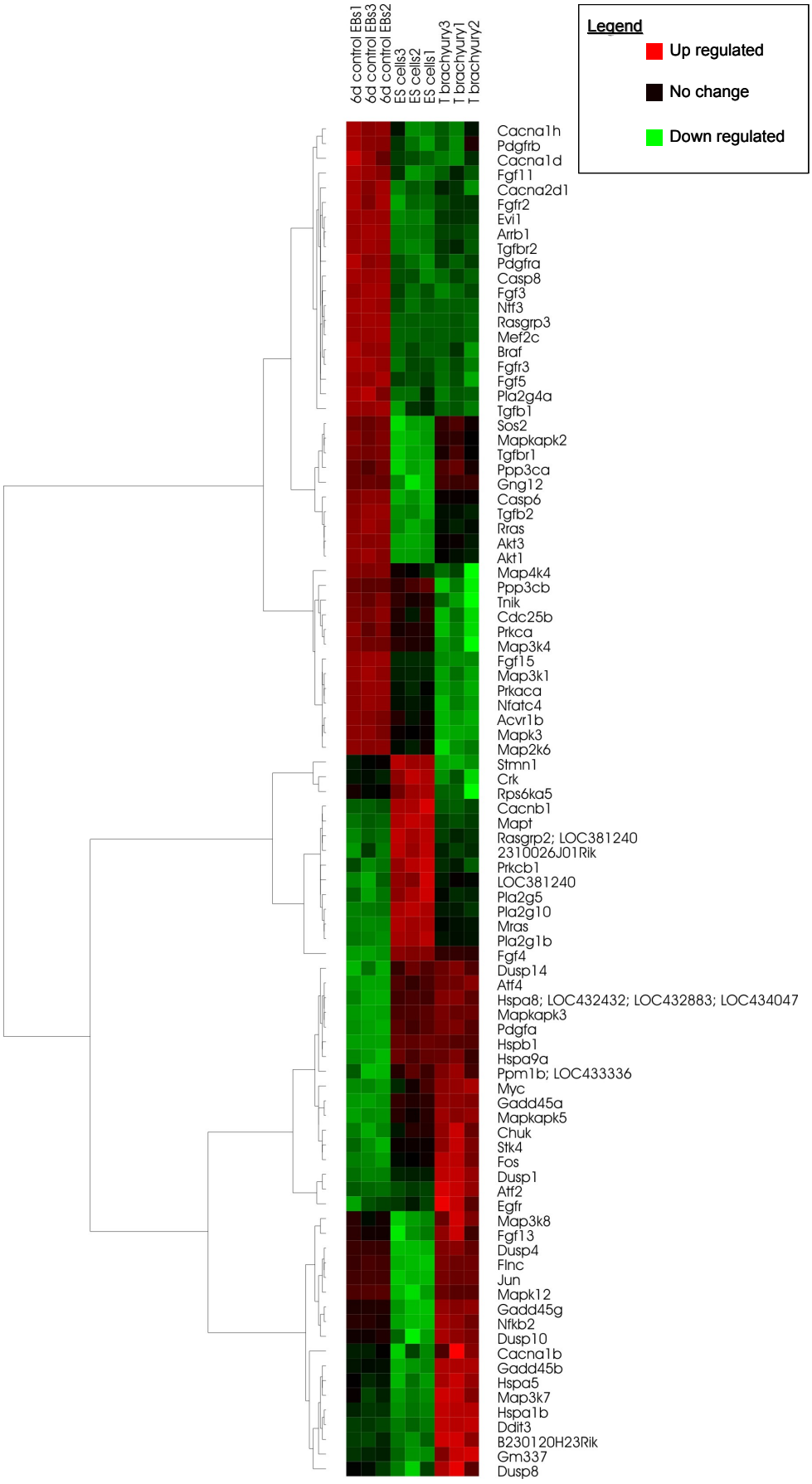


Figure 30. Hierarchical clustering of the transcripts differentially expressed in T brachyury^{+} cells that participate in MAPK signaling pathway

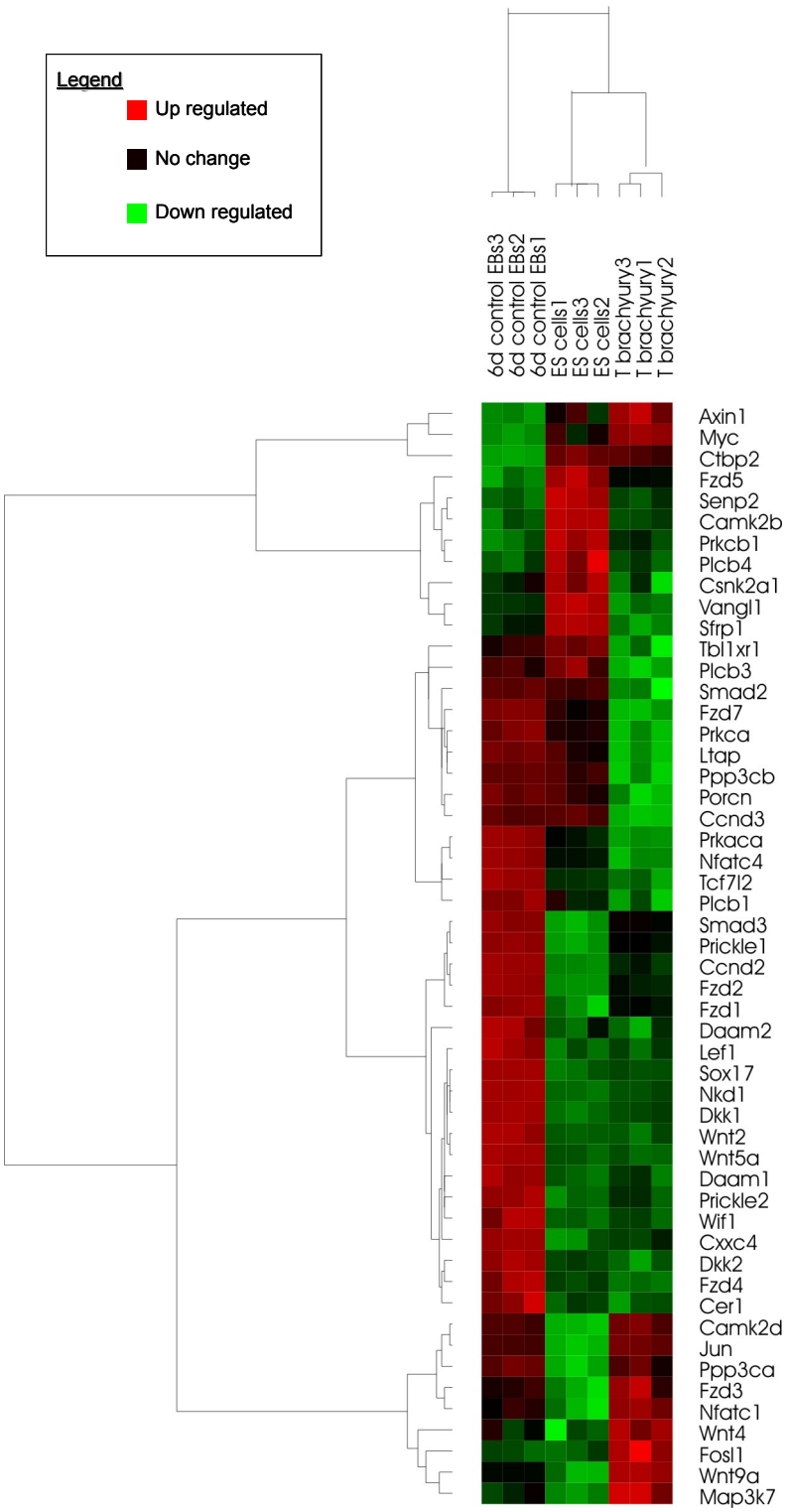


Figure 31. Hierarchical clustering of the transcripts differentially expressed in T brachyury^{+} cells that participate in Wnt signaling pathway

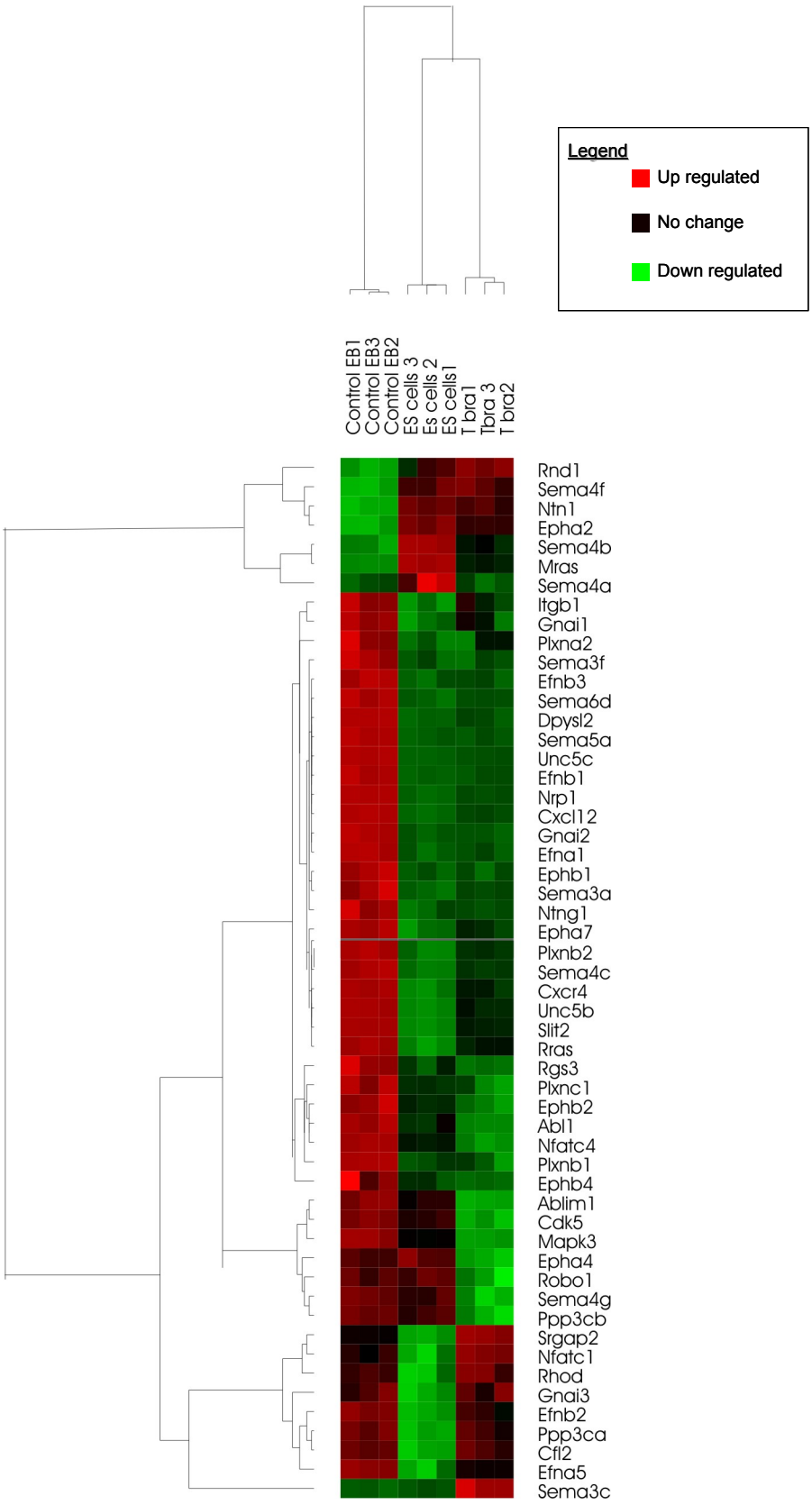


Figure 32. Hierarchical clustering of the transcripts differentially expressed in T Brachyury cells that involve in axon guidance.

This shows that the T Brachyury population has several dynamically ongoing functional signaling cascades, which are implicated to play a crucial role in early embryonic patterning. The transcripts involved in axon guidance are almost down regulated indicating the transition of primitive ectoderm into mesoderm and development of neuronal phenotypes in the control but not in T Brachyury population (Figure 32).

4.3.4.2 Classification of differentially expressed transcripts in T Brachyury in comparison to ES cells and 6day control EBs

T Brachyury population is a transient one during differentiation under normal conditions and needs co-operation from other 2 germ layers for the further lineage commitment. When they were kept in isolation from the other germ layers, their interaction with the other germ layers is eliminated and hence they slow down in terms of the differentiation and remain in differentiation state between the starting ES cell state and the 6d control EBs which, are far advanced in terms of differentiation due their intact interaction with the other 2 germ layers. Hence characterization of differentially expressed transcriptome of T Brachyury population needs to be compared with both the T ES cells and 6d Control EBs. Thus the analysis will need to be addressed as follows.

1. *Upregulated in T Brachyury$^+$ cells than both ES ells and 6day control EBs*
    i. Uniquely upregulated in T Brachyury cells (Not Differentially expressed in ES cells Vs 6day EBs)
    ii. Not Uniquely expressed in T Brachyury cells (Differentially expressed in ES Vs 6d EBs)
2. *Downregulated in T Brachyury$^+$ than both ES ells and 6day control EBs*
    i. Uniquely downregulated in T Brachyury cells (Not Differentially expressed in ES cells Vs 6day EBs)
    ii. Not Uniquely expressed in T Brachyury cells (Differentially expressed in ES versus 6d EBs)
3. *Upregulated in T Brachyury$^+$ than ES cells but downregulated than 6d control EBs*
4. *Upregulated in T Brachyury$^+$ than 6d control EBs but downregulated than ES cells*

Analysis of upregulated transcripts in T Brachyury$^+$ cells in comparison to both ES cells and 6day EBs.

There are 427 transcripts upregulated in T Brachyury cells in comparison to both ES cells and 6day old control EBs. Among them, there are 342 transcripts, which do not show any differential expression among ES cells and 6day EBs. The other 85 transcripts even though are upregulated compared to both ES cells and 6day EBs, they in turn show differential expression among ES cells and EBs. The distribution of the transcripts according to Gene Ontology "Biological Process" as analyzed with David functional annotation tool is presented in the Figure 33. The 393 transcripts upregulated uniquely in T Brachyury cells but not in ES VS EBs are the bona fide transcriptomic signatures of T Brachyury$^+$ mesodermal cells. There are 42 T Brachyury cell specific transcripts involved in development according DAVID analysis and are enlisted in table 1. The unique T Brachyury cell specific transcription factors and the transcripts involved in signaling and development are enlisted in tables 2 and table 3. The rest transcripts are represented by more than one Gene ontologies and are shown as heat map in annexure 2. Among the transcripts upregulated in T Brachyury cells that in turn showed differential expression in ES cells Vs EBs, there are 18 transcripts involved in developmental processes and are listed in the annexure 3.

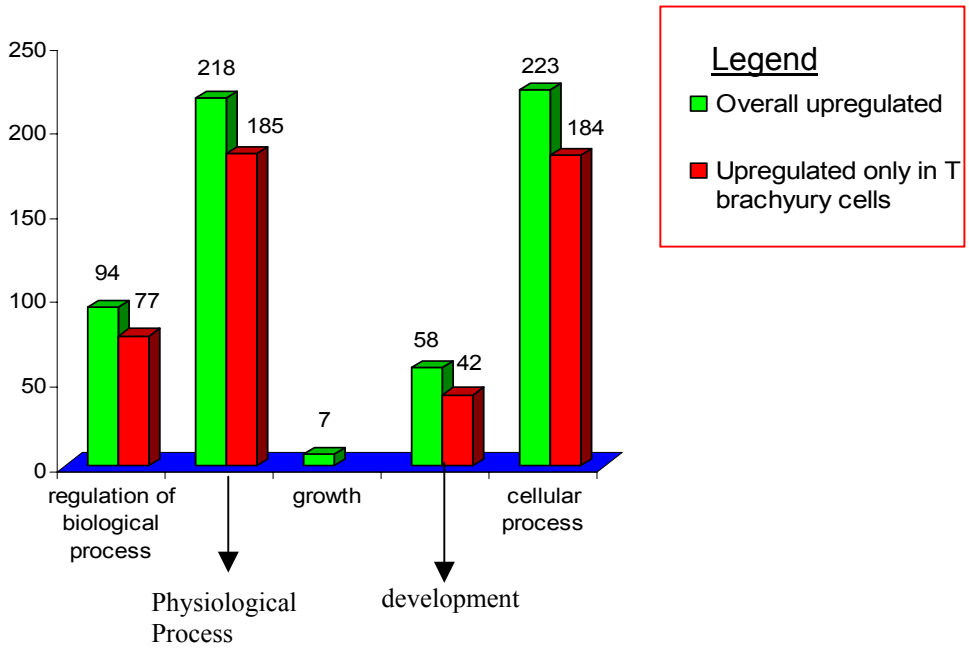


Figure 33. Transcripts upregulated more than both ES cells and control EBs according to the biological processes. A) All transcripts which are upregulated in T Brachyury cells than both ES cells and 6 day control EBs. B) The T Brachyury cell specific unique transcriptome that does not show differential expression between ES cells Vs 6 day EBs but show up-regulation in T Brachyury cells than both ES cells and EBs.

Table. 1a The development related transcripts specifically upregulated in T Brachyury population. The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

Affy_IDs	NCBI ID	Gene Name	ES cells	6d Control EB	T Brachyury+	ANOVA p Value
1416192_at	NM_025898	N-ethylmaleimide sensitive fusion protein attachment protein alpha	1.17	1.00	2.40	1.24E-06
1416953_at	NM_010217	connective tissue growth factor	1.36	1.00	4.65	4.32E-07
1417311_at	NM_024223	cysteine rich protein 2	1.00	1.69	4.18	2.22E-08
1417850_at	NM_009029	retinoblastoma 1	1.00	1.29	3.62	1.93E-06
1418289_at	NM_016701	nestin	1.00	1.15	2.32	6.18E-03
1418357_at	NM_008241	forkhead box G1	1.00	1.93	8.03	5.47E-05
1418901_at	NM_009883	CCAAT/enhancer binding protein (C/EBP), beta	1.78	1.00	6.42	1.99E-06
1419204_at	NM_007865	delta-like 1 (Drosophila)	1.61	1.00	6.70	3.22E-06
1419340_at	NM_031260	Moloney leukemia virus 10-like 1	1.51	1.00	3.48	6.07E-05
1419440_at	NM_021447	tripartite motif-containing 54	1.74	1.00	4.16	2.83E-05
1420696_at	NM_013657	sema domain, immunoglobulin domain (Ig),secreted, (semaphorin) 3C	1.05	1.00	3.01	5.71E-06
1420964_at	NM_007930	ectodermal-neural cortex 1	1.63	1.00	3.65	1.74E-05
1421207_at	NM_008501	leukemia inhibitory factor	1.44	1.00	4.45	8.26E-06
1421273_at	NM_080843	suppressor of cytokine signaling 4	1.00	1.04	2.82	1.13E-04
1421365_at	NM_008046	follistatin	1.52	1.00	3.25	8.25E-07
1421375_a_at	NM_011313	S100 calcium binding protein A6 (calcyclin)	1.37	1.00	6.84	5.79E-06
1422033_a_at	NM_053007	ciliary neurotrophic factor	1.00	1.01	4.40	2.08E-06
1422054_a_at	NM_011386	SKI-like	1.00	1.15	3.53	6.12E-05
1422959_s_at	NM_030743	zinc finger protein 313	1.01	1.00	2.26	6.25E-05
1423315_at	NM_133234	Bcl-2 binding component 3	1.00	1.22	3.21	3.75E-06
1424208_at	NM_008965	prostaglandin E receptor 4 (subtype EP4)	1.00	1.06	5.88	2.14E-07

Table.1b. The transcripts specifically upregulated in T Brachyury population which play a role in development (continued). The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

Affy_IDs	NCBI ID	Gene Name	ES cells	6d Control EBs	T Brachyury+	ANOVA p Value
1426430_at	NM_010588	jagged 2	1.01	1.00	2.93	2.71E-07
1427420_at	NM_183248	NK6 transcription factor related, locus 2 (Drosophila)	1.22	1.00	2.69	4.11E-06
1427735_a_at	NM_009606	actin, alpha 1, skeletal muscle	1.00	1.39	8.80	5.81E-07
1428319_at	NM_026131	PDZ and LIM domain 7	1.00	1.54	3.62	1.49E-06
1429974_at	NM_023814	T-box18	1.63	1.00	3.54	4.54E-03
1431320_a_at	NM_010864	myosin Va	1.00	1.59	3.40	3.65E-04
1432155_at	NM_028459	Wiskott-Aldrich syndrome-like (human)	1.00	1.14	2.38	1.44E-03
1437302_at	NM_007420	adrenergic receptor, beta 2	1.00	1.70	3.81	1.63E-03
1439016_x_at	NM_011468	small proline-rich protein 2A	1.00	1.04	2.66	9.65E-06
1440085_at	NM_175540	ectodysplasin A2 isoform receptor	1.22	1.00	4.46	5.23E-06
1440626_at	NM_008275	Homeo box D13 (Hoxd13), mRNA	1.30	1.00	3.07	2.06E-05
1444151_at	NM_010487	ELAV (embryonic lethal, abnormal vision, Drosophila)-like 3	1.25	1.00	2.88	1.80E-03
1448325_at	NM_008654	myeloid differentiation primary response gene 116	1.26	1.00	7.86	3.16E-09
1449391_at	NM_009554	zinc finger protein 37	1.00	1.00	2.02	2.07E-03
1450722_at	NM_016714	nucleoporin 50	1.00	1.13	2.74	4.16E-08
1450957_a_at	NM_011018	sequestosome 1	1.00	1.03	3.24	6.24E-06
1450971_at	NM_008655	growth arrest and DNA-damage-inducible 45 beta	1.00	1.96	8.00	4.74E-07
1451596_a_at	NM_011451	sphingosine kinase 1	1.00	1.06	3.08	6.46E-05
1452113_a_at	NM_008999	RAB23, member RAS oncogene family	1.35	1.00	2.76	1.62E-07

Table. 2a The transcription factors that are specifically upregulated in T Brachyury population. The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

Affy_IDs	NCBI ID	Gene Name	ES cells	6d Control EBs	T Brachyury+	ANOVA p Value
1416019_at	NM_026106	down-regulator of transcription 1	1.66	1.00	3.81	4.15E-05
1417487_at	NM_010235	fos-like antigen 1	1.01	1.00	2.26	1.04E-04
1417516_at	NM_007837	DNA-damage inducible transcript 3	1.00	1.29	8.27	1.08E-08
1417850_at	NM_009029	retinoblastoma 1	1.00	1.29	3.62	1.93E-06
1418323_at	NM_010193	feminization 1 homolog b (C. elegans)	1.04	1.00	3.31	3.58E-08
1418357_at	NM_008241	forkhead box G1	1.00	1.93	8.03	5.47E-05
1418616_at	NM_010757	v-maf musculoaponeurotic fibrosarcoma oncogene family, protein K	1.00	1.95	6.20	2.44E-06
1418901_at	NM_009883	CCAAT/enhancer binding protein (C/EBP), beta	1.78	1.00	6.42	1.99E-06
1419191_at	NM_010434	homeodomain interacting protein kinase 3	1.00	1.12	2.37	1.77E-04
1419558_at	NM_008575	transformed mouse 3T3 cell double minute 4	1.10	1.00	2.32	8.35E-04
1419641_at	NM_011221	purine rich element binding protein B	1.00	1.49	4.07	5.94E-06
1420433_at	NM_028958	TATA box binding protein -associated factor	1.94	1.00	4.89	2.93E-05
1421021_at	NM_001005605	AE binding protein 2	1.00	1.08	2.70	1.42E-04
1421867_at	NM_008173	nuclear receptor subfamily 3, group C, member 1	1.50	1.00	3.75	2.65E-04
1422033_a_at	NM_053007	ciliary neurotrophic factor	1.00	1.01	4.40	2.08E-06
1422893_at	NM_019460	Scm-like with four mbt domains 1	1.65	1.00	3.54	8.31E-06
1425356_at	XM_484882	zinc finger protein 142	1.09	1.00	2.29	3.06E-04
1426298_at	NM_010574	Iroquois related homeobox 2 (Drosophila)	1.00	1.11	3.22	7.26E-06
1427420_at	NM_183248	NK6 transcription factor related, locus 2 (Drosophila)	1.22	1.00	2.69	4.11E-06
1427559_a_at	NM_001025093	activating transcription factor 2	1.16	1.00	2.63	6.40E-06
1429217_at	NM_028298	zinc finger protein 655	1.10	1.00	2.40	5.69E-06
1429974_at	NM_023814	T-box18	1.63	1.00	3.54	4.54E-03
1432155_at	NM_028459	Wiskott-Aldrich syndrome-like (human)	1.00	1.14	2.38	1.44E-03
1433634_at	XM_284454	interferon regulatory factor 2 binding protein 2	1.00	1.76	4.29	7.96E-06
1434419_s_at	NM_001003898	TAR DNA binding protein	1.27	1.00	3.16	1.32E-05

Table. 2b The transcription factors that are specifically upregulated in T Brachyury population (continued). The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

Affy_IDs	NCBI ID	Gene Name	ES cells	6d Control EBsl	T Brachyury+	ANOVA p Value
1435197_at	NM_008900	POU domain, class 3, transcription factor 3 (Pou3f3), mRNA	1.00	1.07	2.84	2.96E-05
1435307_at	NM_175455	RIKEN cDNA 6430502M16 gene	1.35	1.00	26.79	2.92E-07
1437302_at	NM_007420	adrenergic receptor, beta 2	1.00	1.70	3.81	1.63E-03
1437788_at	NM_031183	trans-acting transcription factor 6	1.00	1.28	3.02	1.24E-04
1438725_at	XM_109726	thyroid hormone receptor associated protein 1	1.00	1.56	3.79	6.65E-05
1440085_at	NM_175540	ectodysplasin A2 isoform receptor	1.22	1.00	4.46	5.23E-06
1440626_at	NM_008275	Homeo box D13 (Hoxd13), mRNA	1.30	1.00	3.07	2.06E-05
1440855_at	NM_029705	Machado-Joseph disease homolog (human)	1.34	1.00	2.75	1.68E-05
1441935_at	NM_023472	ankyrin repeat, family A (RFXANK-like), 2	1.00	1.54	3.84	2.68E-04
1444712_at	XM_282974	PREDICTED: gene model 739 [Mus musculus], mRNA sequence	1.21	1.00	3.42	2.17E-04
1447930_at	XM_484142	bromodomain adjacent to zinc finger domain 1A	1.00	1.89	5.28	4.09E-05
1448022_at	---	---	1.30	1.00	10.18	2.40E-06
1449152_at	NM_007670	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	1.00	1.34	3.18	9.33E-06
1449391_at	NM_009554	zinc finger protein 37	1.00	1.00	2.02	2.07E-03
1450481_at	NM_008651	myeloblastosis oncogene-like 1	1.00	1.02	5.19	1.51E-05
1450957_a_at	NM_011018	sequestosome 1	1.00	1.03	3.24	6.24E-06
1451639_at	NM_009884	CCAAT/enhancer binding protein (C/EBP), gamma	1.12	1.00	2.36	6.30E-07
1455925_at	NM_029947	PR domain containing 8	1.00	1.16	2.98	6.81E-04
1457240_at	NM_130860	Cyclin-dependent kinase 9 (CDC2-related kinase) (Cdk9), mRNA	1.03	1.00	2.08	2.87E-04

Table.3 The transcripts specifically upregulated in T Brachyury population that encode for proteins involved in Signaling. The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

Affy_IDs	NCBI ID	Gene Name	ES cells	6d Control EB	T Brachyury+	ANOVA p Value
1416458_at	NM_007477	ADP-ribosylation factor 2	1.00	1.21	3.46	2.39E-06
1417516_at	NM_007837	DNA-damage inducible transcript 3	1.00	1.29	8.27	1.08E-08
1418105_at	NM_019675	stathmin-like 4	1.00	1.14	7.57	1.17E-06
1418819_at	NM_026011	ADP-ribosylation factor-like 10C	1.00	1.39	3.05	6.07E-04
1418943_at	NM_023057	RIKEN cDNA B230120H23 gene	1.00	1.16	2.50	3.93E-05
1421207_at	NM_008501	leukemia inhibitory factor	1.44	1.00	4.45	8.26E-06
1421273_at	NM_080843	suppressor of cytokine signaling 4	1.00	1.04	2.82	1.13E-04
1421867_at	NM_008173	nuclear receptor subfamily 3, group C, member 1	1.50	1.00	3.75	2.65E-04
1422033_a_at	NM_053007	ciliary neurotrophic factor	1.00	1.01	4.40	2.08E-06
1422638_s_at	NM_018750	Ras association (RalGDS/AF-6) domain family 5	1.00	1.45	2.92	9.07E-05
1423047_at	NM_023764	toll interacting protein	1.00	1.50	3.23	2.54E-05
1424208_at	NM_008965	prostaglandin E receptor 4 (subtype EP4)	1.00	1.06	5.88	2.14E-07
1424332_at	NM_139154	Rab40c, member RAS oncogene family	1.08	1.00	2.76	1.72E-04
1425420_s_at	NM_015771	large tumor suppressor 2	1.00	1.35	3.47	1.77E-04
1427679_at	XM_618738	large tumor supressor	1.00	1.18	2.80	2.63E-04
1429444_at	XM_485698	RAS-like, family 11, member A	1.27	1.00	3.93	1.26E-05
1435014_at	NM_175122	RAB39B, member RAS oncogene family	1.00	1.26	2.92	1.95E-04
1435432_at	NM_178119	centaurin, gamma 2	1.00	1.61	3.35	1.09E-06
1436192_at	XM_130646	ADP-ribosylation factor guanine nucleotide-exchange factor 2	1.30	1.00	3.15	2.12E-06
1437074_at	NM_024225	sorting nexin 5	1.83	1.00	4.16	2.58E-03
1437302_at	NM_007420	adrenergic receptor, beta 2	1.00	1.70	3.81	1.63E-03
1437741_at	NM_024454	RAB21, member RAS oncogene family	1.00	1.35	3.30	2.53E-06
1438097_at	NM_011227	RAB20, member RAS oncogene family	1.00	1.08	2.91	2.36E-05
1440085_at	NM_175540	ectodysplasin A2 isoform receptor	1.22	1.00	4.46	5.23E-06
1450957_a_at	NM_011018	sequestosome 1	1.00	1.03	3.24	6.24E-06
1450971_at	NM_008655	growth arrest and DNA-damage-inducible 45 beta	1.00	1.96	8.00	4.74E-07
1451596_a_at	NM_011451	sphingosine kinase 1	1.00	1.06	3.08	6.46E-05
1452113_a_at	NM_008999	RAB23, member RAS oncogene family	1.35	1.00	2.76	1.62E-07
1455399_at	XM_110525	connector enhancer of kinase suppressor of Ras 1	1.00	1.58	3.51	4.66E-07
1458595_at	NM_175515	PDZ domain containing 6	1.00	1.79	3.67	4.04E-05

Table.4 Devemopment related transcripts upregulated in T Brachyury cells than both ES cells and EBs but in turn show differential expression among ES cells and 6 day control EBs. The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

NCBI ID	Gene Name	ES cells	6d Control EB	T Brachyury+
NM_010029	DEAD (Asp-Glu-Ala-Asp) box polypeptide 4	18.81	1.00	92.42
NM_011580	thrombospondin 1	1.00	14.95	30.90
NM_011817	growth arrest and DNA-damage-inducible 45 gamma	1.00	7.53	26.97
NM_013657	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	1.00	2.03	24.21
NM_139298	wingless-type MMTV integration site 9A	1.00	2.96	15.99
NM_009255	serine (or cysteine) peptidase inhibitor, clade E, member 2	1.00	5.90	14.33
NM_011526	transgelin	1.00	6.55	14.00
NM_009964	crystallin, alpha B	1.00	5.74	13.16
NM_009573	zinc finger protein of the cerebellum 1	1.00	4.78	11.86
NM_008655	growth arrest and DNA-damage-inducible 45 beta	1.00	2.44	11.38
NM_013749	tumor necrosis factor receptor superfamily, member 12a	1.00	2.07	11.37
NM_176930	RIKEN cDNA C130076O07 gene	1.00	4.47	10.74
NM_008343	insulin-like growth factor binding protein 3	1.00	3.36	9.74
NM_011674	UDP galactosyltransferase 8A	4.22	1.00	9.39
NM_010118	early growth response 2	1.00	2.80	7.38
NM_013562	interferon-related developmental regulator 1	2.60	1.00	7.28
NM_010516	cysteine rich protein 61	1.00	2.48	5.31
NM_011536	T-box 4	1.00	2.03	4.67

2. Downregulated in T Brachyury than both ES cells and 6day control EBs

There are 411 unique transcripts downregulated in T Brachyury cells than both ES cells and 6day control EBs. Of them, there are 338 transcripts uniquely downregulated in T Brachyury cells (these do not show any differential expression among ES cells and 6day EBs). Among them, the developmental related transcripts and transcription factors are listed in the annexures 4 and 5. Apart from that, 73 transcripts are downregulated in T Brachyury cells compared to both ES cells and 6day EBs, which in turn show differential expression among ES cells and EBs and some of them are GTPase regulators and 9 others are related to development as listed in annexure 6A and B.

3. Upregulated in T Brachyury^{+} cells than ES cells but downregulated than 6day old control EBs

256 unique transcripts were found to be upregulated in T Brachyury cells compared to ES cells but down regulated in comparison to 6day control EBs. Among them, 5 transcripts are involved in TGF β signaling pathway and they are - MAD homolog 3, Paired-like homeodomain transcription factor 2, Transforming growth factorβ2, *BMP-2* and inhibitor of DNA binding 3. Interestingly the Id protein and Pitx2 are upregulated to direct mesoderm specification. There are 6 transcripts which play a role in Wnt signaling pathway are MAD homolog 3, Prickle like 1 (DROSOPHILA), SRY-BOX containing gene 17 (Sox17), Frizzled homolog 1, Cyclin D2 and Frizzled homolog 2. Sox17 is early marker for endoderm. Apparently, the transcript is present in the T-brachyury population indicating the existence of mesendoderm as already reported by several groups. Kubo and his colleagues used brachyury (T) as a marker for primitive streak mesoderm and showed that T-GFP-expressing cells in ES cell cultures could give rise to cells expressing endodermal markers[66]. Among the 256 transcripts, 45 transcripts are developmentally important genes as shown in the annexure 7.

4. The transcripts upregulated in T Brachyury cells than 6d control EBs but downregulated than ES cells.

There are 225 unique transcripts, which are upregulated in T Brachyury^{+} cells compared to 6day EBs but downregulated in comparison to ES cells. The development related transcripts and transcription factors are shown in annexures 8 and 9. It is interesting to note that the transcripts Oct4 and Fgf4 are upregulated in T Brachyury cells compared to the control EBs.

4.4 CHARACTERIZATION OF BMP-2$^+$ MESODERMAL POPULATION

BMP-2, the late mesodermal marker and a member of TGF-beta superfamily, plays a crucial role in early embryonic patterning as evidenced from gene-ablation studies[67,68]. Its expression immediately follows the transient expression of T-Brachyury in the nascent mesoderm. It is expressed normally in lateral plate mesoderm and extra embryonic mesoderm[67,69]. BMP-2$^+$ mesodermal cells at this stage denote a subset of mesoderm, important being the lateral plate cardiogenic mesoderm[25]. Administration of soluble BMP-2 to chick embryo explant cultures induces full cardiac differentiation in stage 5 to 7 anterior medial mesoderm, a tissue that is normally not cardiogenic[70].

Since BMP-2 is a cardiogenic factor as well as expressed in the cardiogenic mesoderm, it is highly imperative to investigate the molecular nature and phenotype of the mesodermal cells expressing BMP-2 during the early stages of development in the context of cardiomyogenesis. Also, It has been well documented that BMP-2 is a potent apoptotic inducer and a potent neurotrophic factor depending upon the target cells in a concentration gradient dependant manner, mostly by its paracrine mode of action[71-73]. Thus, BMP-2 plays a pivotal role not only during cardiomyogenesis but also during other early embryonic patterning and lineage specification. Till date, the molecular nature and phenotype of the mesodermal cells expressing BMP-2 during the early stages of development have not yet been characterized leaving a black box in understanding their molecular interactions with the target cells in the vicinity to conclude their role during early embryonic patterning and lineage determination. This is partly due to the pleiotrophic effects of BMP-2 and mainly due to the practical difficulty in isolating pure early stage BMP-2 expressing cells in sufficient quantities during early embryonic development *in vivo*.

Extensive investigation of the *in vitro* ES cells-based developmental model in the last 2 decades has proven it to be an invaluable tool to unravel the molecular enigma during the embryonic development in particular to better understand the mechanisms by which lineage decisions are made during early embryogenesis[12]. Recapitulation of early embryonic development, access to the early embryonic development, fine control over the process of differentiation, relative easiness to harvest the experimental material in sufficient quantity and quality in a very short time and to investigate molecular developmental events applying reporter gene manipulations make *the in vitro* ES-based model superior over the *in vivo* animal models[12,62].

To address the molecular nature and behavior of the BMP-2$^+$ mesodermal cells during their differentiation into their lineage specific somatic cells, we first established an ES cell-derived transgenic Bone morphogenetic protein-2 (BMP-2)-expressing lineage for

developmental studies *in vitro*. In order to identify all the possible signal transduction pathways and biological processes characteristic of the BMP-2$^+$ cells, we performed large-scale expression studies using Affymetrix expression microarrays. Our study on the phenotypic identification of the ES cell-derived BMP-2 lineage specific cells shows that the early BMP-2 population contained a heterogeneous population of predominantly neural crest stem cells and their lineages-smooth muscle cells (SMC), epithelial like cells, astrocytes and melanocytes. When early BMP-2$^+$ population was further cultured under certain conditions contained cardiomyocytes, macrophages and osteoblasts. Interestingly, these are the cell phenotypes which need BMP-2 for their phenotypic induction and demonstrates the presence of a multi-lineage progenitor phenotype resembling neural crest stem cells. More over, identification of the key signal transduction pathways induced or repressed in BMP-2$^+$ cells explains the observed potential of BMP-2 in modulating early embryonic development, in particular the mesodermal patterning.

4.4.1.Generation and analysis of the BMP-2$^+$ ES cell clone

The transgenic BMP-2$^+$ ES cell line was generated with the linearised pBMP-2-Puro IRES2 EGFP construct by stable transfection. The transfected clones are neomycin resistant. The ES cells do not express BMP-2 in the undifferentiated state but express BMP-2 during the course of differentiation at a later stage as shown in the RT-PCR data in figure 34A, when the ES cells were induced to undergo differentiation by the hanging drop protocol. Normally, the BMP-2 transcript expression starts to show up on day 2, gradually increases in expression level until its peak expression on day 5 after which the transcript expression is at the basal level even after 10days of differentiation. The ES cells and EBs from the BMP-2 clone faithfully behaved as the wild type ES cells and EBs (figure 34A and B). During the course of differentiation (by hanging drop protocol), the EGFP fluorescence increases significantly after 2 days and initially the EGFP expressing cells found to be scattered within the 2-days to 4-days old EBs (Figure 34C). As the differentiation continues, the EGFP fluorescence peaks on day 5 and the EGFP expressing cells are localized to a particular region in every EBs as shown in figure 34C.

4.4.2. Isolation and transcriptomic Profiling of BMP-2$^+$ cell population

Isolation and further characterization of the BMP-2$^+$, puromycin-resistant cells have been optimized according to the protocol described in Figure 35A. Briefly, single cell suspension of BMP-2$^+$ ES cells was seeded in bacteriological dishes for two days to form 2-day old EBs. Then, the 2-day old EBs were transferred into gelatine coated tissue culture dishes and cultured for further 2 days. Thereafter, plated EBs were treated with 3µg/ml puromycin for the subsequent 3 days. After trypsinisation of puromycin-resistant 7-day old

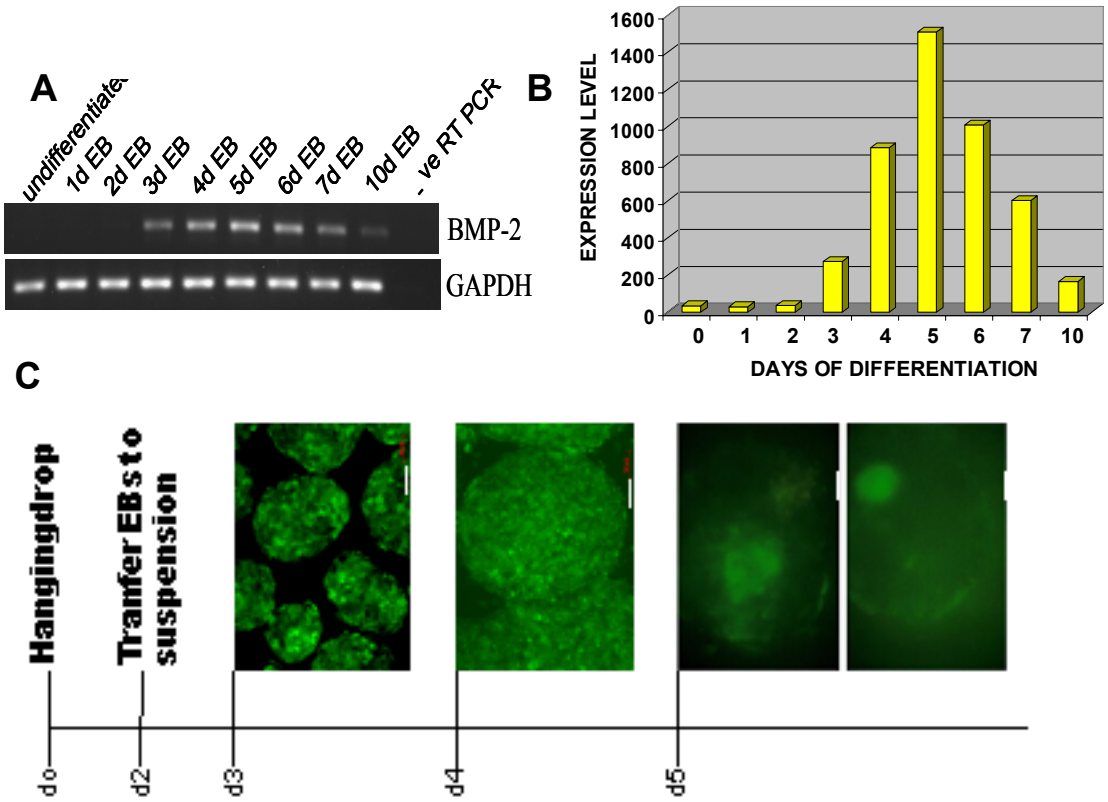


Figure 34 A., Analysis of the BMP-2 transcript expression by RT-PCR from EBs derived from BMP-2 ES clones, B., Differential expression pattern of BMP-2 transcript in differentiating EBs as obtained from the transcriptomic profiling by Affymetrix. C., Expression of the EGFP during differentiation of the BMP-2 ES cells induced by the conventional hanging drop protocol. Scale bar represents 50µm.

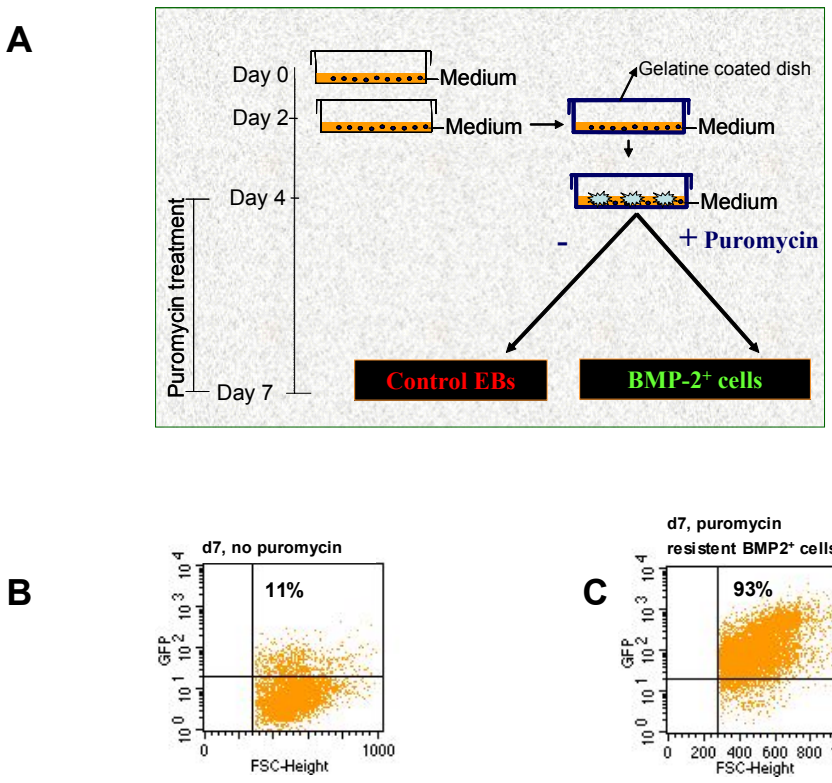


Figure 35 A, Protocol for isolation of puromycin resistant BMP-2$^+$ cells after treating the plated 4-days old EBs with 3µg/ml puromycin for 3 days. **B**, **C**, FACS analysis of the trypsinized untreated control and puromycin resistant BMP-2$^+$ cells.

BMP-2$^+$ cells, FACS analysis has been performed. As demonstrated in Figures 35.B and 35.C, after 3 days of puromycin treatment, EGFP fluorescing and puromycin resistant BMP-2$^+$ cells (hereafter called BMP-2$^+$ cells) accounted for 93% of the cells whereas the control EBs without puromycin treatment (hereafter called control EBs (=7-day old EBs)) contained only 11% of BMP-2$^+$ cells. This resulted in an enrichment of nearly 8.5 fold of the BMP-2$^+$ cells.

As demonstrated in figure 36A, plating of the BMP-2$^+$ cells in gelatine coated tissue culture dishes for the subsequent 3 days in the presence of puromycin results in a bright EGFP-positive BMP-2$^+$ cell population. Furthermore, the BMP-2 protein has been detected by immunostaining using BMP-2-specific antibodies (Figure 36B and 36C). The undifferentiated BMP-2 ES cells were included as negative control (Figure 36D). As demonstrated, BMP-2 is only detected in the cytosol and specifically in vesicles of the BMP-2$^+$ cells. Plating and culture of the BMP-2+ cells in the presence of puromycin for 8 more days resulted in a heterogeneous population/colonies of cells with different morphologies as shown in Figure 37.

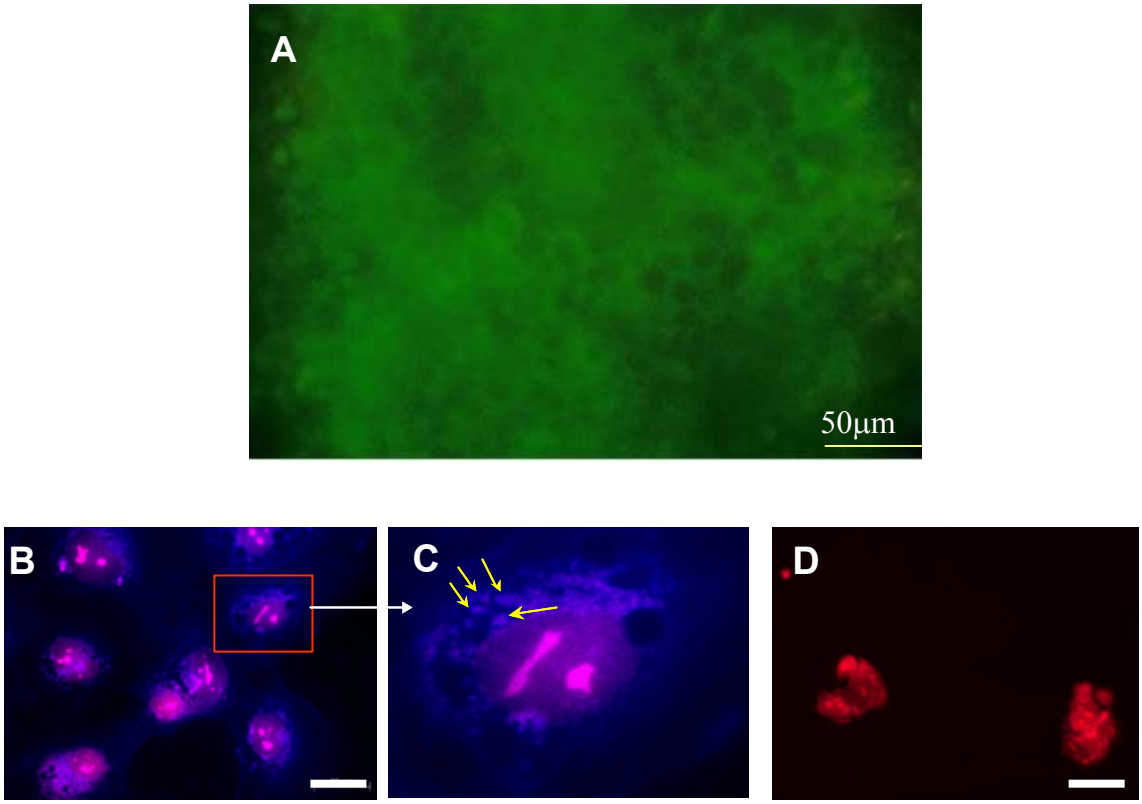


Figure 36 A, BMP-2$^+$, 3 days after plating in gelatine coated tissue culture dishes in the presence of 3µg/ml puromycin. Scale bar represents 50µm. **B, C, D,** Detection of BMP-2 in BMP-2$^+$ cells by immunohistochemistry staining. Stainings were done after the BMP-2$^+$ cells were trypsinized and plated on microscopic slides for 24 hours. **D**, ES cells were used as negative control. Scale bar represents 20µm.

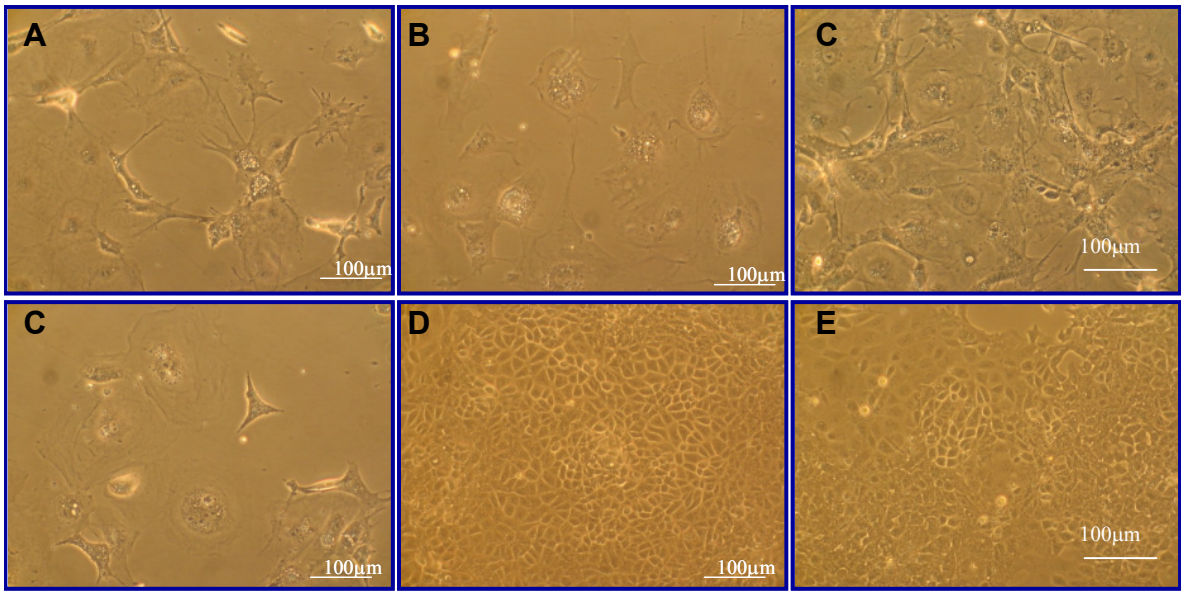


Figure 37 Morphology of the cells-derived from BMP-2$^+$ cells after culturing for 8 days in the presence of 3µg/ml puromycin

4.4.3 Phenotypic charecterisation of BMP-2$^+$ cell lineages

Expression of genes in the BMP-2$^+$ cells associated with plasticity, mesodermal and neural crest stem cell phenotypes

The early BMP-2 lineage cells are still in the state of plasticity

BMP-2$^+$ cells significantly upregulate Oct4 and Nanog transcripts expression compared to the 7-days old control EBs in which several somatic cell types are developed (Figure 38A) but at a level lower than ES cells. This implicates that there are some population of BMP-2$^+$ cells with multi-lineage progenitor phenotypes, which are still in the state of certain plasticity and which can be induced to give rise to different cell fates depending upon the stimuli they receive. This is further confirmed due to the upregulated expression of Leukemia Inhibitory Factor (LIF) in BMP-2$^+$ population compared to the 7-days old control EBs. Interestingly, the transcripts of Activin, Nodal and Cripto are also upregulated in BMP-2$^+$ population compared to the 7-days old control EBs. Recently, it has been demonstrated that TGF-β/Activin/Nodal signaling pathway is necessary for the maintenance of pluripotency in embryonic stem cells[74]. Therefore, the increased level of the pluripotency associated gene markers-Oct4 and Nanog might be partly explained by the increased expression of LIF, Activin and Nodal observed in the BMP-2$^+$ cells (figure. 38).

The BMP-2$^+$ cells exhibit mesodermal characteristics

The expression levels of the mesodermal markers in BMP-2$^+$ cells show increased levels of expression of nodal, activin, eomesodermin, cripto, and Mesp2 whereas the expression of T-bra and Mesp1 was lower in the BMP-2$^+$ cells compared to the control 7-days old EBs (Figure 38). In this context, it has been shown that Activin and Nodal which are members of transforming growth factor beta (TGF-β) superfamily play a pivotal role in inducing and patterning mesoderm and endoderm, and in regulating neurogenesis and left-right axis asymmetry (for review see[75]). Nodal genes have been identified in numerous vertebrate species and are expressed in specific cell types and tissues during embryonic development[75]. Moreover, Nodal null mouse mutants lack mesoderm and overexpression of Nodal in mouse ES cells leads to up-regulation of mesodermal and endodermal cell markers emphasizing the key role of Nodal for mesoderm formation[76]. Also, it is repeatedly shown that Activin is involved in the mesodermal pattering during *Xenopus* embryo development[77]. Cripto is the founder member of the EGF-CFC family initially related to the epidermal growth factor (EGF). Cripto is expressed in tumor tissues and studies in the mouse have established an essential role of cripto in the formation of precardiac mesoderm and differentiation to fucnctional cardiomyocytes[78]

The T-box gene *Eomesodermin (Eomes)* is required for mesoderm formation and the morphogenetic movements of gastrulation. It has been shown in the mouse that lack of *Eomes* abrogates the formation of embryonic and extraembryonic mesoderm[20]. In this context, it has been shown that *Eomes* is specifically required for the directed movement of cells from the epiblast into the streak in response to mesoderm induction[20]. Interestingly, we found a dramatically low level of T-Brachyury expression in BMP-2$^+$ cells compared to the 7-days old control EBs, indicating that the either BMP-2$^+$ cells represent a subset of mesodermal population, the mesoderm formation is nearly complete at the time of BMP-2 expression whereby T-Brachyury expression is downregulated or the necessary signals from the non-BMP-2 population to induce stronger T-brachyury expression are eliminated due to puromycin selection. T-Brachyury is an essential gene for the mesoderm formation as demonstrated in the mouse[65]. Mesp1 and Mesp2 are basic helix-loop-helix (bHLH) transcription factors that are co-expressed in the anterior presomitic mesoderm (PSM) just prior to somite formation[79] in the mouse embryo. Furthermore, it has been shown that Mesp1 has a significant role in the epithelialization of somitic mesoderm and it has been assumed that Mesp2 is responsible for the rostro-caudal patterning process itself in the anterior PSM[79]. Recently, it has been shown that Mesp1 is expressed in almost all of the precursors of the cardiovascular system in the mouse, including the endothelium, endocardium, myocardium, and epicardium[80]. Thus, the in vitro derived BMP-2 lineage cells exhibit more mesodermal characteristics. This conclusion is further supported by the

derivation of most of the mesodermal tissues and complete absence of endodermal phenotypes from these BMP-2 cells in the later stages as described below.

BMP-2$^+$ cells lack endodermal phenotypes

Interestingly, transcripts of α-feto protein (AFP) and Sonic Hedgehog (Shh) were not detectable in BMP-2$^+$ cells (figure 38A). AFP represents a marker for the endoderm-derived hepatocytes. The expression pattern of Sonic Hedgehog studied in several species indicates that Shh is essential for the endoderm-derived organ development such as foregut development, gut development, gastrointestinal duodenal and pancreas deveolpment[81]. These results show a dramatically reduced or more likely the complete absence of the endodermal lineage specific cells.

The BMP-2$^+$ cell lineage contains neural crest stem cells and their derivatives

The BMP-2$^+$cells contained enriched expression of ectodermal markers-Neurofilament H (NF-H,) and M (NF-M). But it is evident that the BMP-2 cells shared more mesodermal characteristics as described in the previous sections. Further investigation of this contrary situation led to the conclusion that these may be more likely due to neural crest stem cells, as reported in several observations where these cells share a more ectodermal and a less mesodermal characteristics[82-85]. Expression of the neural crest stem cells (NCSCs) specific p75NTR[86] and Nestin transcripts at higher levels compared to the 7-days old control EBs (Figure 39A) confirmed the presence of enriched NCSCs in BMP-2$^+$ cells. In addition, expression of smooth muscle α actin (SMA), astrocytes specific Glial fibrillary acidic protein (GFAP) and melanocytes specific Tyrosine phosphates 1 (Tyrp1) in BMP-2$^+$ cells as shown in figure 39A confirmed the presence of NCSC lineages as well in the BMP-2$^+$ lineage. Expression of p75NTR in BMP-2$^+$ cells was further confirmed by immunohistochemical staining with an antibody against p75^{NTR} on BMP-2$^+$ cells, one day after plating in the absence of puromycin. In addition, presence of glial cells has been confirmed by immunohistochemical staining with an antibody against Glial fibrillary acidic protein (GFAP), on BMP-2$^+$ cells, one day after plating in the absence of puromycin (Figure 39C.b). Notably, it has been reported that the differentiation of NCSCs into their lineage fates is mainly dependant on the presence of BMP-2 in the required threshold level and also the availability of the other factors in combination[83]. In this context, it is well demonstrated that exogenous addition of recombinant BMP-2 to cultured NCSCs isolated from chicken explants of cranial and trunk dorsal neural folds from stage 8/9 embryos resulted in the differentiation of NCSC into glial, melanocytes and smooth muscle cells[83]. It is surprising to note that the BMP-2 local concentration increased by BMP-2+cells

themselves are able to induce the differentiation of NCSCs into their lineages without the exogenous the addition of BMP-2.

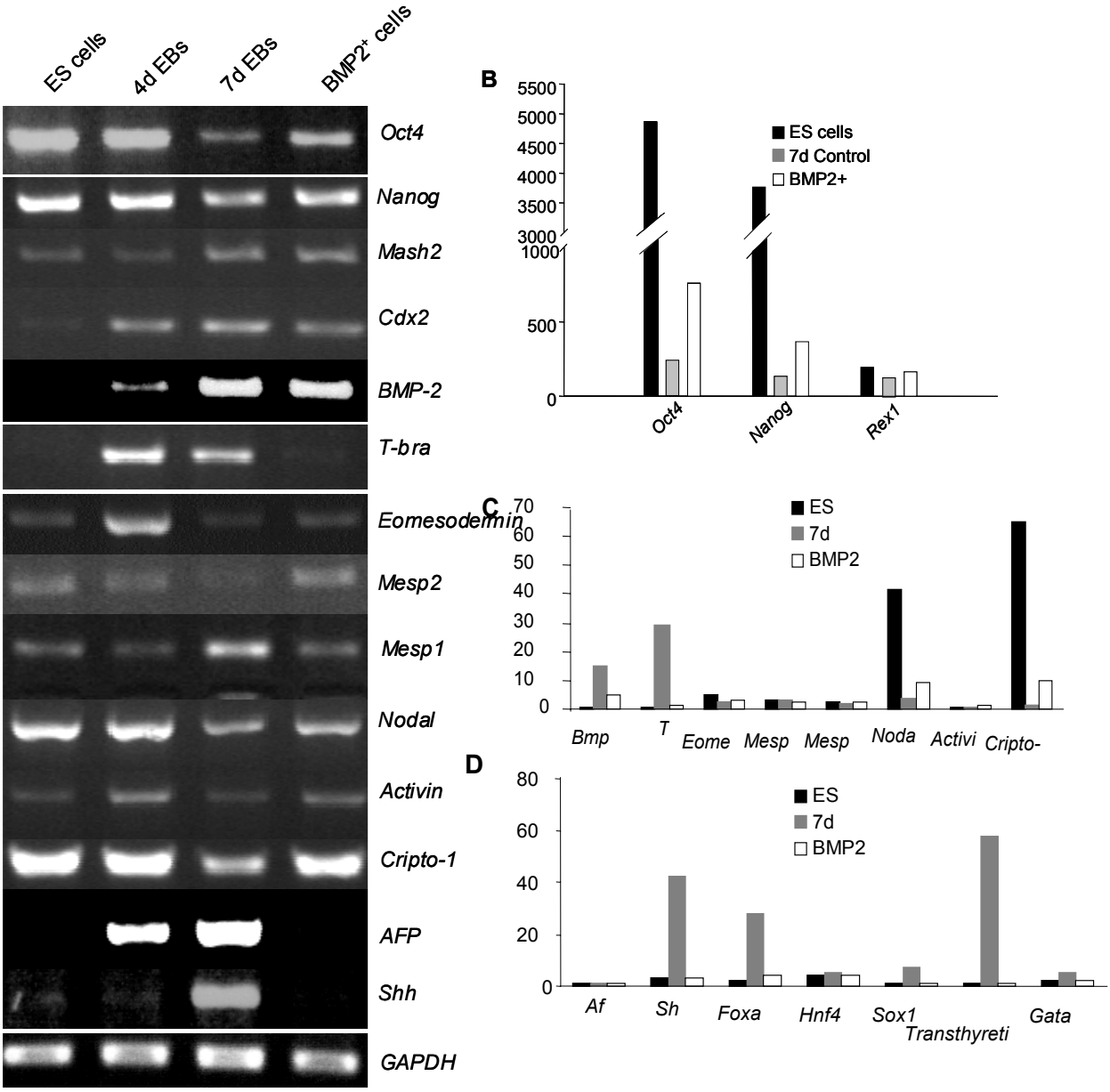


Figure 38. Expression of pluripotent, trophectodermal, mesodermal, and endodermal gene markers in BMP-2$^+$ cells. **A,** RT-PCR analysis for the representative genes, **B-D**, Relative fluorescence signal units as obtained from Affymetrix profiling for the genes presented in A and also other representative genes.

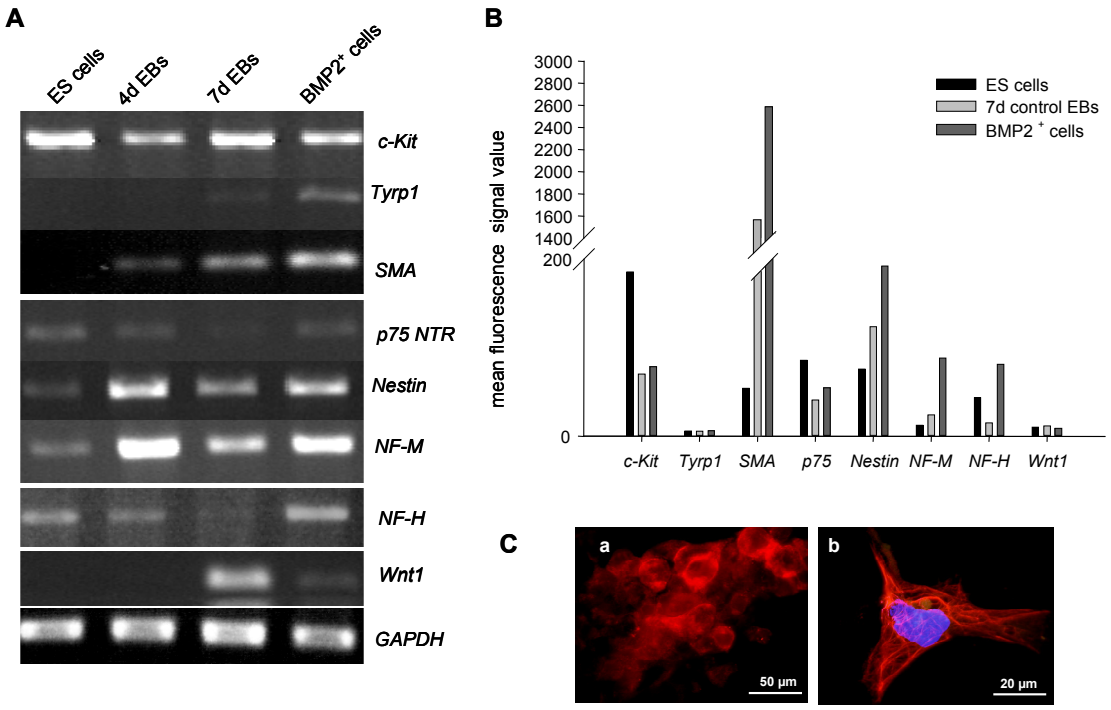


Figure 39. A-, RT-PCR analysis for the representative genes associated with neural crest stem cells, B. Relative fluorescence signal units as obtained from Affymetrix profiling for the genes presented in A. **C**, Detection of p75 (B) and GFAP in BMP-2$^+$ cells labeled by immunohistochemistry. Immunuostainings with anti-p75 (**Ca**) and anti–GFAP **(Cb)** to show the presence of neural crest stem cells and the astrocytes, respectively, in BMP-2$^+$ cells one day after plating (7+1 days in total).

During vertebrate embryonic development, as the notochord induces the transformation of surface ectoderm to neuroectoderm, a multipotential middle cell layer develops with characteristics of both cell types. These cells called, the neural crest cells migrate dorsolaterlally to form the neural crest, a flattened irregular mass between the surface ectoderm and neuroectoderm. This layer of cells separates into right and left portions and then migrate to various locations within the embryo to give rise to most structures of the peripheral nervous system (PNS) such as Schwann and glia cells of the autonomic and enteric nervous systems, endocrine cells such as the adrenal medulla and C-cells of the thyroid) as well as non-neural tissues such as pigmented pigment cells of the skin and internal organs, smooth muscle musculature of the cardiac outflow tract and great vessels, pericytes, craniofacial bones, cartilage and connective tissues[87,88]. Wnt1 and c-kit are well known mediators for melanocytes differentiation of NSCS[89,90]and are found to be expressed in the BMP-2$^+$ population. The tyrosine kinase, c-kit, has been found in the cell membranes of hematopoietic stem cells, primordial germ cells and presumptive subepidermal melanocytes[91] Intriguingly, wnt-1 and BMP-2 act synergistically to suppress

differentiation and to maintain NCSC marker expression and multipotency by combinatorial Wnt1/BMP-2 signalling[92]

The presence of Wnt1 transcripts present in BMP-2$^+$ population may be interpreted in 2 ways. 1) The Wnt1 may be involved in the maintenance of neural crest stem cells in their pluripotency state in combination with BMP-2. 2) Wnt1 may be involved in driving the differentiation of NCSCs into melanocytes. Both possible roles by wnt1 cannot be ruled out in BMP-2 population since there are proliferating NCSC on one side (as evident from increase in cell number when subjected to immunohistochemistry), and the occurrence of melanocytes in the same culture on the other hand. The local BMP-2 and/or Wnt1 gradient may drive the neural crest stem cells to produce either smooth muscle cells, pericytes or to melanocytes or to remaining their pluripotent state, respectively.

This is the first study that enables getting ES cells-derived neural crest stem cells and their derivatives all at the same time selectively via BMP-2 promoter based lineage selection approach. The co-expression of Wnt1 and BMP-2 in the same culture indicates the existence of both an environment to keep the neural crest stem cells in stemness and also the stage of ongoing differentiation of NSCS to form melanocytes. Thus, the study on these BMP-2 expressing cells during early differentiation of ES cells will pave the way for the better understanding of the neural crest stem cells and their differentiation into their lineages. The BMP-2$^+$ cells-derived from the ES cells may serve as an ideal model for the neural crest stem biology in the future since the neural crest stem cells and their derivatives can be selectively enriched by the BMP-2 promoter driven lineage selection approach. In addition, it provides a valuable system where the enriched neural crest stem cells prime themselves to differentiate into their cell specific lineages, since the enriched NSCS secrete BMP-2 and causes a BMP-2 gradient negating the supply of exogenous BMP-2 addition. Noggin, an inhibitor of BMP-2 can be used to keep the neural crest stem cells in their plasticity. Once the fine-tuning of BMP-2 derived neural crest stem cells to keep in their plasticity is made, then they will be the potential candidates for the cell replacement therapy, since they can mould into any tissue depending upon the local environment prevailing within the tissue in which they are injected.

BMP-2 cell contain predominantly smooth muscle cells

Figure 40A shows that culturing of the BMP-2$^+$ cells for 8 days in the presence of puromycin resulted in the development of a heterogeneous cell population of cells with different morphologies. Figure 37B indicates cells with smooth muscle-like morphology. The immunohistochemical staining of smooth muscle cells has been performed with an

antibody against smooth muscle α actin, 1 day after plating the BMP-2+ cells (7+1 days in total), 8 days after plating with puromycin (7+8 days in total), 18 days after plating with puromycin (7+18 days in total) and 18 days after plating without puromycin (7+18 days in total) (Figure 40Ca-d). Detection of Smooth muscle cells even 18 days after puromycin treatment indicates that SMC express BMP-2 and therefore they survive the puromycin treatment [93].

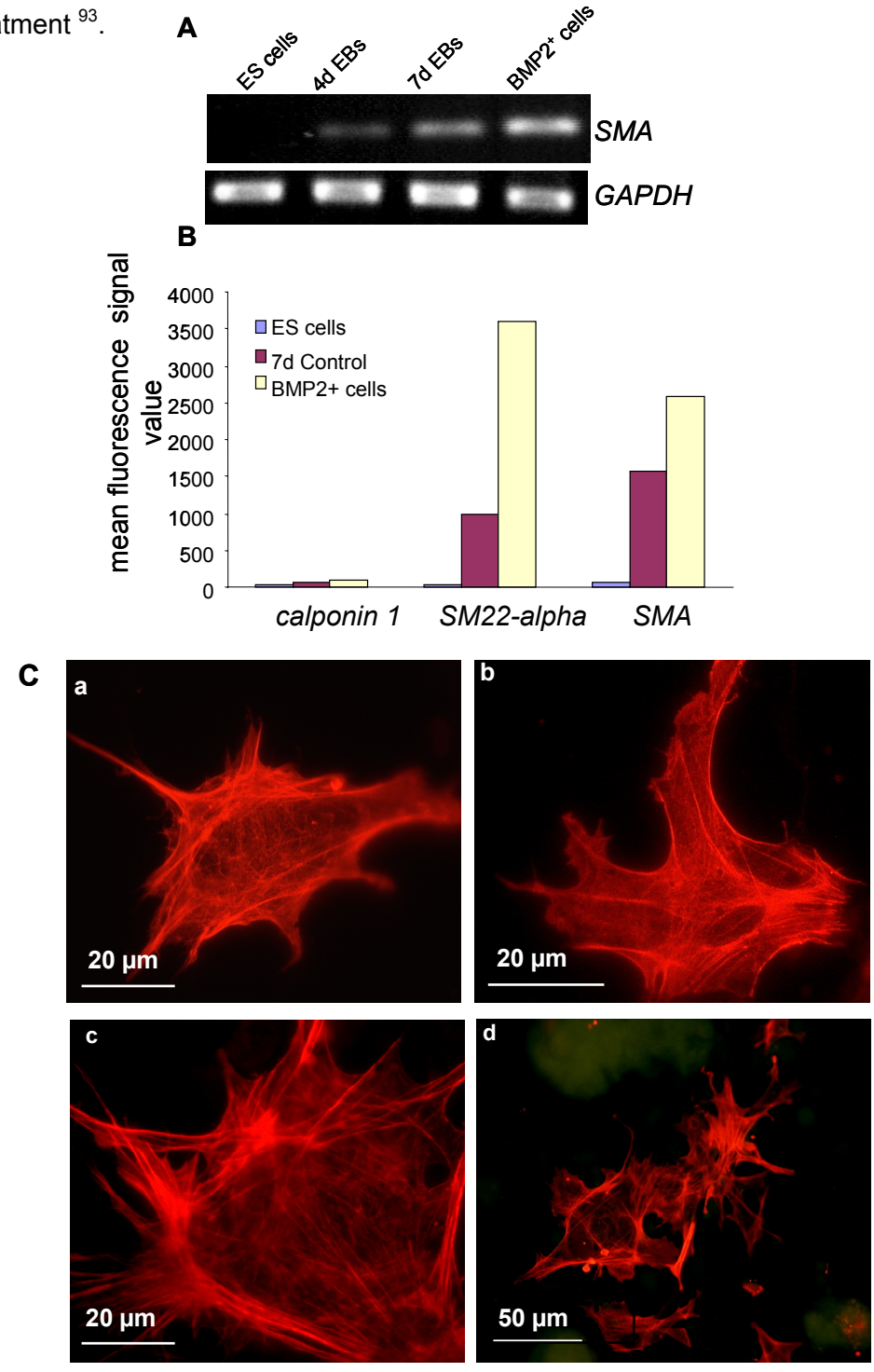


Figure 40. Detection of smooth muscle cells after differentiation of the BMP-2$^+$ cells. A, Expression of SMA in the BMP-2$^+$ detected by semi-quantitative RT-PCR method. B, Diagram indicates the microarray expression levels of various smooth muscle specific genes in BMP-2$^+$ cells. **C(a-d),** detection of smooth muscle cells, 1 day after plating the BMP-2+ cells (7+1 days in total) **(Ca)**, 8 days after plating with puromycin (7+8 days in

total) **(Cb)**, 18 days after plating with puromycin (7+18 days in total) (**Cc**) and 18 days after plating without puromycin (7+18 days in total) **(Cd)**.

BMP-2$^+$ cells give rise to cardiomyocytes after downregulation

The expression level of the cardiac marker genes NKx2.5 (early), MLC-2a, α-cardiac actin and Mef2c is reduced in the BMP-2$^+$ cells compared to the control 7-days old EBs (Figure 41A) suggesting that cardiomyogenesis is repressed in the BMP-2$^+$ cells. Accordingly, also Mesp1 was found to be repressed (Figure 38A). Culturing of the BMP-2$^+$ cells even for further 28 days (35 days in total) in the presence or absence of puromycin did not result in cardiac cell beating clusters (data not shown).

Preparation of hanging drops EBs applying the differentiation protocol (Figure 41B) as previously described[94] from a mixture of BMP-2$^+$ cells with wild type ES cells in a ratio 1:1, 1:2, 1:4, 10:1, 50:1, 1:10 and 1:50, respectively, did not augment/delay the onset of contractile activity in comparison to the control observed on day 12. Also, there were not any significant differences in terms of magnitude of the intensity of the contractility (data not shown). This corresponds to the observation where recombinant BMP-2 added during the differentiation of ES cells, no enhancement of cardiomyogenesis has been observed[95]. In order to investigate the capacity of the BMP-2$^+$ cells to differentiate in cardiac beating cells, secondary EBs were made from these BMP-2$^+$ cells by hanging drop protocol and the differentiation process was observed in comparison to the control EBs. The secondary EBs when made from the BMP-2$^+$ population were contracting on 11th day like the secondary control EBs made from the CGR8 wild type EBs without puromycin and secondary EBs made from the β–actin CGR8 clone transfected with β-actin promoter driven puromycin resistance and EGFP expression cassettes treated in the same way as the secondary BMP-2$^+$ EBs with puromycin. Notably, the intensity of the contracting areas in the BMP-2$^+$ EBs was significantly stronger, compared to the both controls. The whole EB was contracting in a jellyfish like fashion where as in the controls (Fig. 41C). The contracting areas persisted for a longer time when compared to the controls for more than a week. In order to demonstrate that an increased beating activity of the cardiomyocytes generated from the BMP-2$^+$-cells correlates with an increased expression of cardiomyocyte gene markers, we investigated the expression of α-cardiac actin and αMHC in EBs at day 11 secondary EBs. As clearly indicated, an increased expressed level of both cardiac specific genes was observed in the EBs generated from BMP-2$^+$ cells (Figure 41D) compared to both the control secondary EBs.

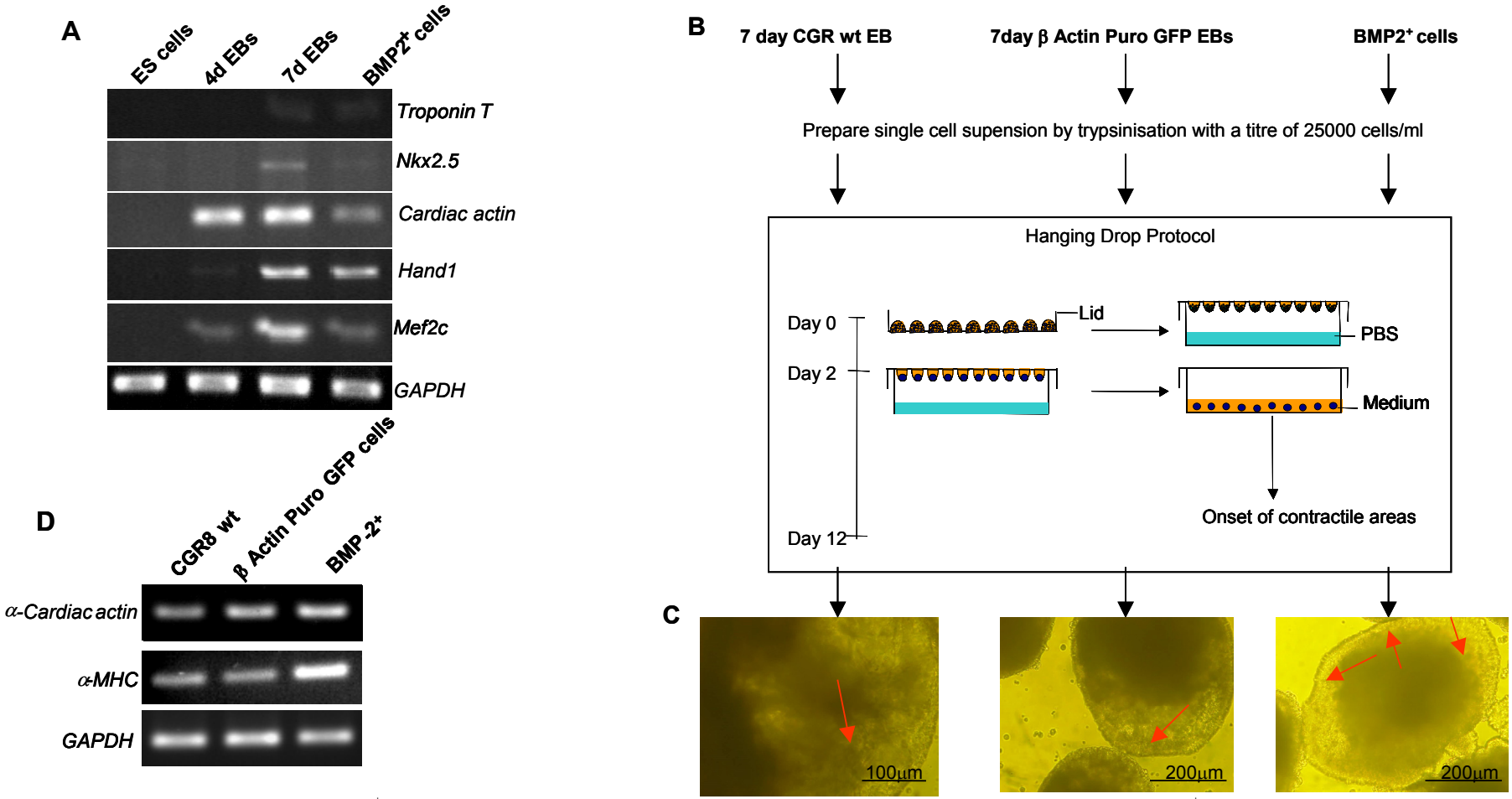


Figure 41. Differentiation of the BMP2$^+$ cells to cardiac cells. **A**, RT-PCR analysis of the cardiac markers in BMP2$^+$ cells and other controls. **B**. Schematic outline of the protocol used to derive cardiomyocytes from BMP2$^+$ cells. **C**, the morphology of the contracting EBs. The red arrows indicate the contracile area(s) in that EB and **D**, RT-PCR analysis of the representative cardiac markers in 12 day secondary EBs derived by hanging drop protocol from single cell suspension obtained from 7 day old primary EBs generated by the hanging drop protocol from CGR8 wild type EBs without puromycin treatment, β-actin puro EGFP EBs and BMP2 EBs with puromycin treatment.

Thus, BMP-2$^+$ cells have the capacity to develop into cardiomyocytes. Interestingly, the cardiomyogenic potential apparently seems to be regulated by the BMP-2$^+$ lineage cells only. These findings suggest that the BMP-2$^+$ cells are primed to become cardiomyocytes independently of the other BMP-2$^+$ negative lineage cells. But the secondary EBs maintained with puromycin and the contractile secondary EBs when treated with puromycin did not contain cardiomyocytes. So, it is imperative to note that the cardiomyocytes don't express BMP-2 but descend from the BMP-2$^+$ mesodermal cells. At the same time, the cardiomyocytes developed from the BMP-2$^+$ cells are dependant on paracrine action of BMP-2 secreted by the neighboring cells of BMP-2 lineage cells[96].

Vascular and haematopoietic cell gene markers from BMP-2 population

Expression of E-cadherin, Flk1, Flt1, Pecam1 and Runx1 shows the presence of vascular endothelial progenitors in the BMP-2$^+$ population (Figure 42A). As demonstrated in figure 42B, BMP-2$^+$ cells are able to differentiate into cells with an epithelial/endothelial morphology after culturing for 8 days in the presence of puromycin. Detection of epithelial-like cells has been confirmed by immunohistochemistry. As indicated in figure 42Ca-d epithelial-like cells were detected even after 11 days after culturing in the presence of puromycin.

Expression of c-kit and CD34 and Runx1 indicates the presence of hematopoietic stem cells in BMP-2 population (Figure 42A). *Ihh* that is implicated in haematopoiesis and vasculogenesis[97] also is expressed in the BMP-2$^+$ cells but the expression level is lower than that of the control cells. Flt-1 also known as vascular endothelial growth factor receptor 1 (VEGFR-1) is a high-affinity tyrosine kinase receptor for VEGF and is normally expressed only in vascular endothelial cells. However, Flt-1 transcript was recently found to be expressed in human peripheral blood monocytes[98]. Monocytes are known to differentiate into a variety of cell types such as osteoclasts in bone, Kupffer cells in liver, dendritic cells in the immune system, and mature macrophages in a number of tissues[98] . Expression of Flt-1 and c-fms in BMP-2 population indicates the possible monocyte-macrophage lineage existence in the BMP-2$^+$ cell population.

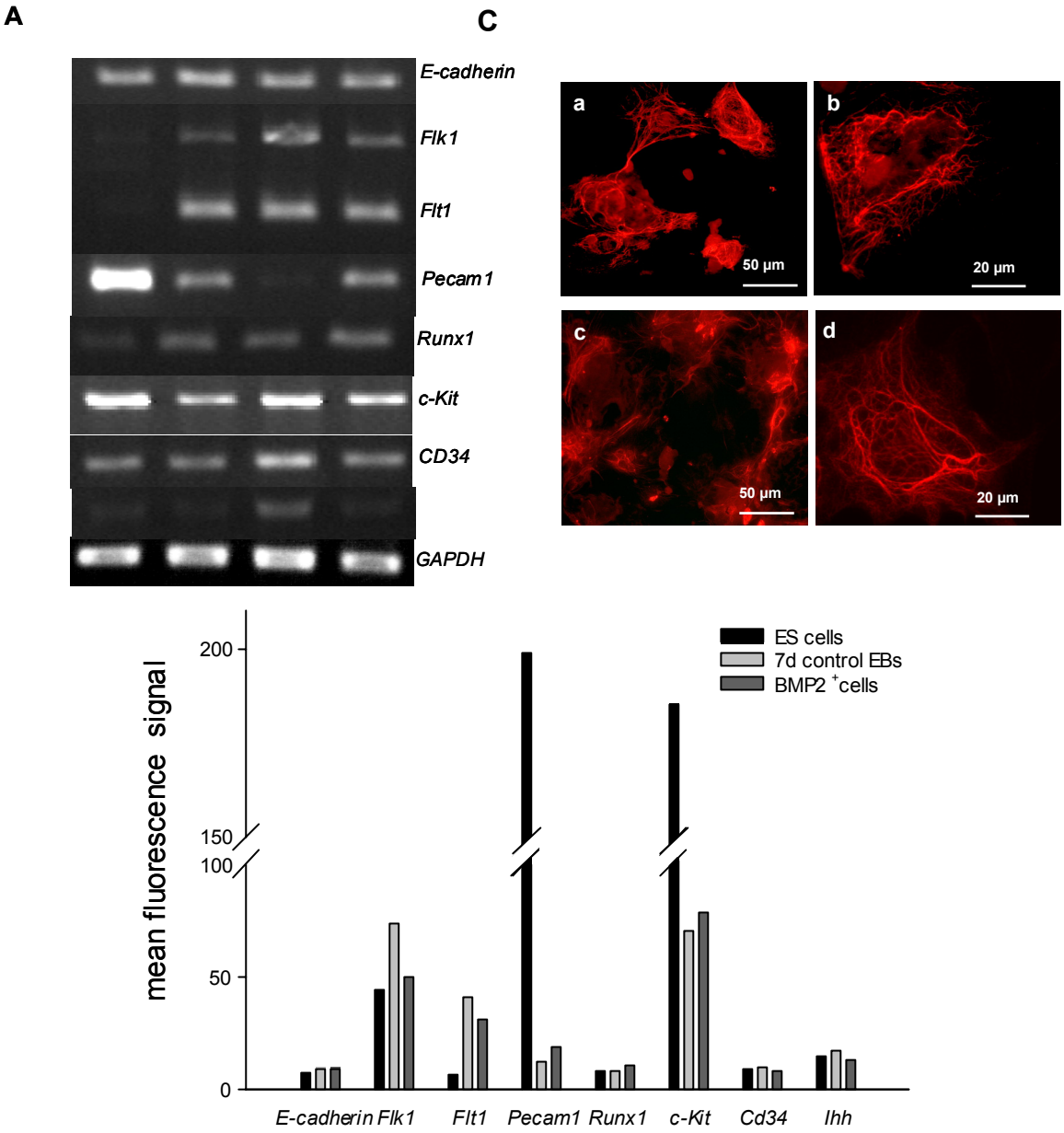
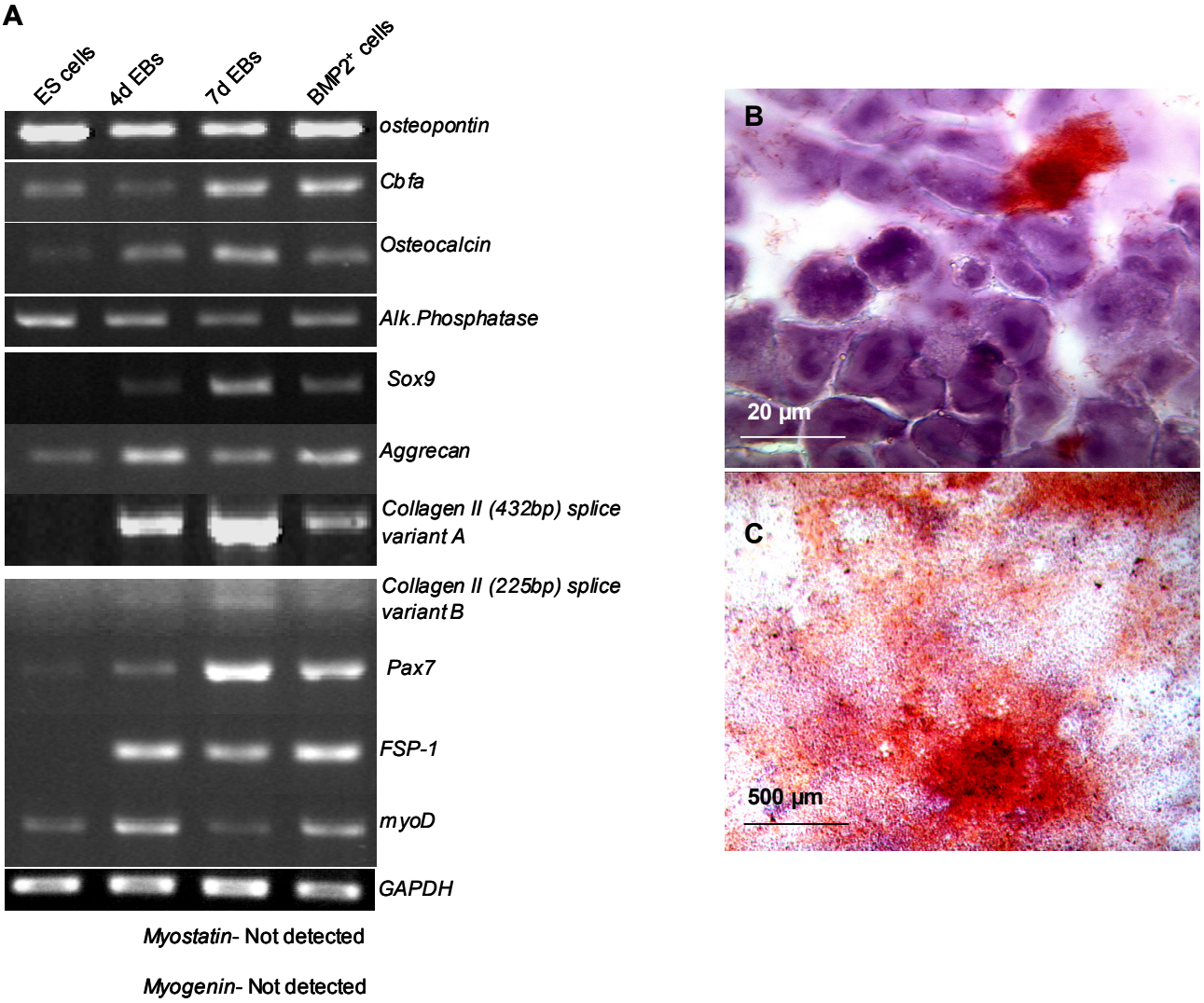


Figure 42 **A**nalysis of the vascular and haematopoietic cell gene markers in BMP-2^{+} cell population**. A**, RT-PCR analysis of the BMP-2^{+} cells. **B**, Relative fluorescence signal units as obtained from Affymetrix profiling for the genes presented in A. **C**, Immunuostainings with anti-pan cytokeratin over the period of time to show the presence of epithelial like cells. **Ca**, one day after plating the BMP-2+ cells (7+1 days in total). **Cb**, 8 days after plating with puromycin (7+8 days in total). **Cc**, 18 days after plating with puromycin (7+18 days in total). **Cd**, 11 days after plating without puromycin (7+11 days in total).

BMP-2$^+$ cells give rise to osteoblasts and express gene markers for satellite cells and fibroblasts

Expression of osteopontin, Cbfa, osteocalcin and alkaline phosphatase in BMP-2$^+$ cells showed the possible occurrence of osteoblasts (Figure 43A). When BMP-2$^+$ cells plated and maintained with the differentiation medium and puromycin showed Alizarin red positive clusters after 18days (Figure 43B) indicating the first appearance of osteoblasts. After 35 days of culturing in the presence of puromycin, brightly stained with Alizarin red were observed confirming the occurrence of osteoblasts in culture (Figure 43C). These results show that the puromycin resistant BMP-2$^+$ cells per se are cable of differentiating into osteoblasts. It may be also possible that the osteoblasts might be developed from the BMP-2$^+$ derived NCSC cells in the presence of an increased BMP-2 concentration for longer periods of time[99].

Expression of the transcription factor Sox9, the proteoglycan aggrecan, collagen II (with 2 alternate splicing forms) was detected in the BMP-2$^+$ population indicating the existence of chondrocytes[100](Figure 43A) . Sox9 is a key transcriptional factor for chondrocytic differentiation of mesenchymal cells via chondrocytes-specific enhancer of the pro alpha1 (II) collagen[101]. Expression of Pax-7 and myoD shows the presence of satellite cells[102,103]. Satellite cells represent a distinct lineage of myogenic progenitors responsible for maintenance of skeletal muscle[104]. MyoD is only expressed when satellite cells are activated to proliferate and differentiate into primary myoblasts, which will in turn differentiate into cells of the myofibres of skeletal muscles[103]. An elevated level of MyoD in the BMP-2$^+$ population indicates either the proliferation or the differentiation of satellite cells into primary myoblasts in BMP-2 culture. But the expression of myogenin and myostatin was not detected in the BMP-2$^+$ cell population suggesting the absence of skeletal muscle cells by RT-PCR even after 40 cycles. Interestingly, it has been reported that the satellite cells can spontaneously differentiate into adipocytes in an alternative mesenchymal pathway[105]. Expression of FSP-1 shows fibroblast phenotype occurring in the BMP-2$^+$ cell population. FSP-1 has been implicated in the epithelial-mesenchymal transition and is a representative marker for mesenchymal cells[106] .



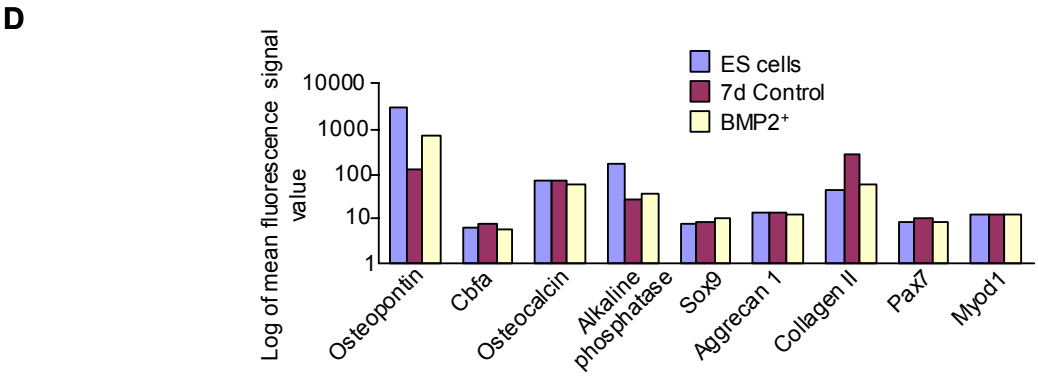


Figure 43. analysis of osteoblast, chondrocyte and myocyte specific markers in BMP-2^{+} cells. **A**, RT-PCR analysis in BMP-2^{+} cells. **B, C**, Alizarin stanings on 18th day after plating the BMP-2+ cells (7+18 days in total) and on 28 days after plating (7+28, 35 days in total). **D**, Relative fluorescence signal units as obtained from Affymetrix profiling for the genes presented in A.

BMP-2 cell lineage contains monocytes but not mature adipocytes

The RT-PCR analysis of the BMP-2$^+$ population showed the expression of the pre-adipocyte gene marker pref-1[107], the pan-adipocyte markers PPARγ[108] aP2, Glut4[109] and LPL as well as the pan- macrophage marker c-fms[107,108,110](Figure 44A and D). Although, the markers PPARγ, aP2 and Glut4 are unique to adipocytes, expression of these genes is also found to be expressed in macrophages. Similarly, trophoblast lineage also expresses c-fms[107]. Notably, the cells-derived from the BMP-2$^+$ cell population after 11 days of cultivation in the presence of puromycin were not found to be positive for either Oil Red O or anti F4/80 stainings eliminating the occurrence of mature adipocytes and macrophages in the later stages (data not shown). Also, no expression of adipsin[111], a marker for mature adipocytes, could be detected demonstrating the absence of the mature adipocytes in the BMP-2$^+$ cells. Thus, the BMP-2$^+$ cell population contains only the pre-adipocytes expressing PPARγ, aP2, Glut4 and Pref-1 but not adipsin. Apparently, other factors secreted by BMP-2 negative cells are required for maturation of preadipocytes to adipocytes.

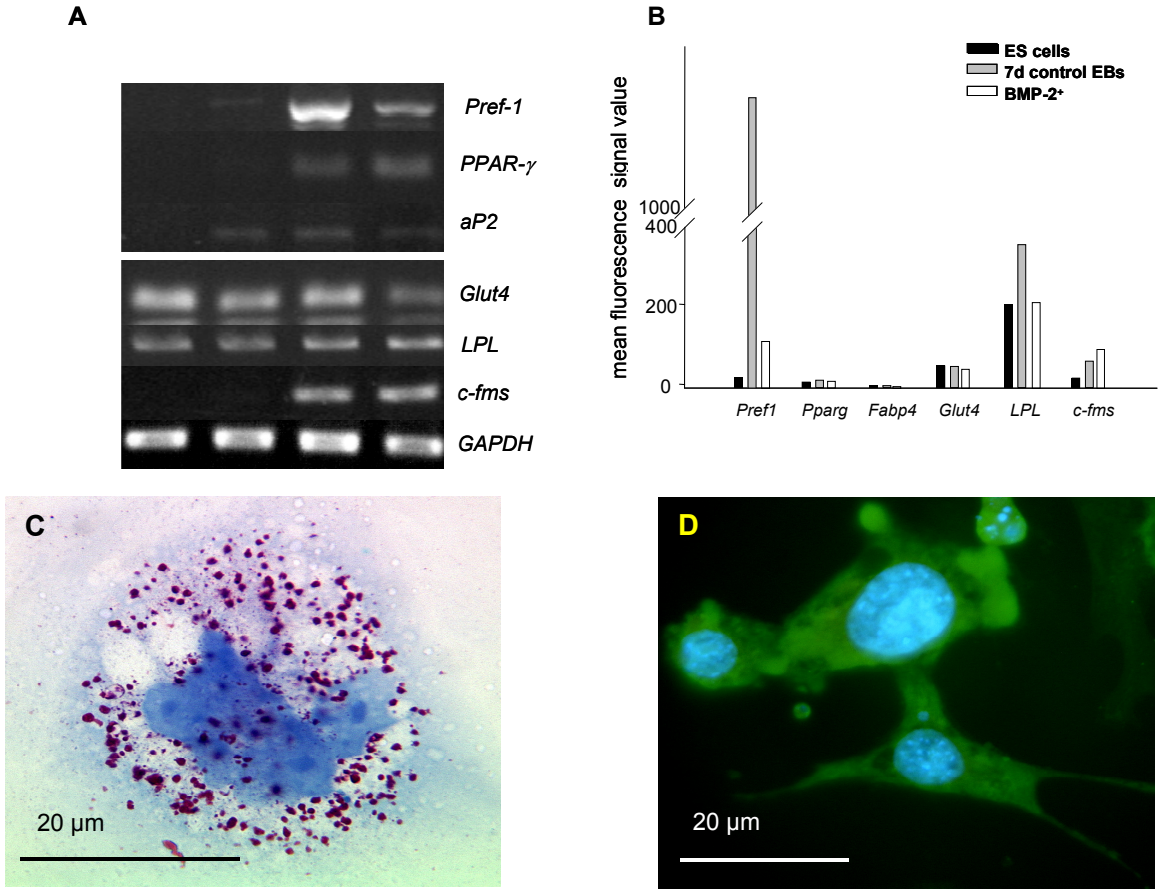


Figure 44. **A**, RT-PCR analysis of adipocyte specific markers. **B**, Relative fluorescence signal units as obtained from Affymetrix profiling for the genes presented in A. **C**, Sudan red staining on 11 days old puromycin untreated culture after plating the BMP-2$^+$ cells (7+11, 18 days in total). **D**, F4/80 Immunuostaining on 18 days old puromycin untreated culture after plating BMP-2$^+$ cells (7+18, 25 days in total).

Notably, after 11-days of BMP-2$^+$ cells culturing in the absence of puromycin, there were cells stained with Sudan Red and F4/80, a marker for macrophages as shown in figure 43B and 43C. These findings suggest that macrophages are derived from differentiated BMP-2$^+$ cells in which BMP-2 expression is downregulated at a later time.

4.4.4 Selective enrichment of different BMP-2$^+$cell phenotypes with puromycin

When the different day EBs were treated with puromycin for 3 days and when the subsequent puromycin treatment was either discontinued, or maintained all the time or maintained in 3 days alternating intervals as shown schematically in figure 45A, different BMP-2 lineages are enriched depending on the starting day of puromycin treatment and the subsequent puromycin selections. Cardiomyocytes were enriched when the 6day EBs were treated with puromycin for 3 days and discontinued the treatment after that. On 19th, when analyzed with semi-quantitative PCR, the cardiomyocytes are enriched when compared to the other BMP-2 lineage phenotypes. When the wild type CGR8 EBs were treated in the same manner, there were no viable cells even after 19 days, implying that the 3 days of puromycin at 3μg/ml concentration is very effective in selecting BMP-2 lineage cells with highest purity. The enrichment of the particular BMP-2 lineage at the respective conditions is color coded with a blue box in figure 46. It is worthwhile to note that in the all the conditions, the enrichment of a particular lineage was found but in addition to other phenotypes. This approach did not select only one phenotype at the whole expense of the others.

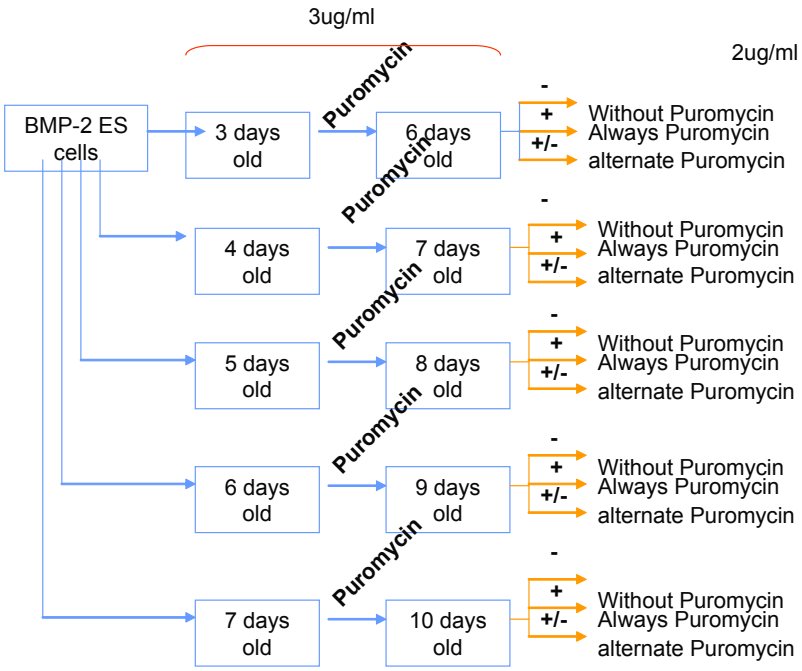


Figure 45. Schematic overview of puromycin treatment on different stage EBs and subsequent puromycin treatment conditions. All the time the first puromycin treatment was 3ug/ml for 3 days and later either the puromycin at a concentration of 2ug/ml was added to the culture or discontinued completely or added in regular 3 days intervals. Finally on day 19, the RNA was isolated from every sample and analyzed for the enrichment of the BMP-2 lineage cells.

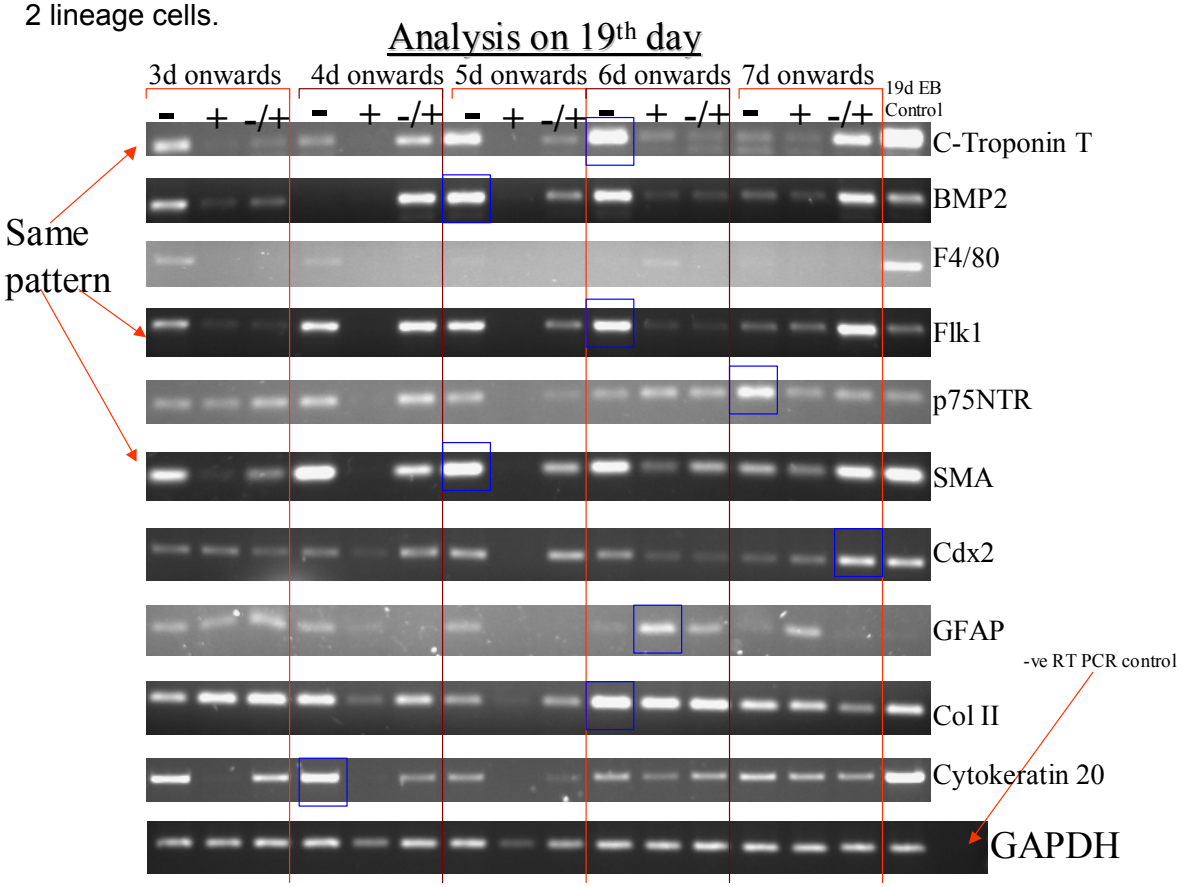


Figure 46. RT-PCR analysis of the BMP-2 mesodermal lineage marker genes on 19th day after the different day EBs were treated with puromycin for 3 days and then either discontinued or continued all the time or in alternative intervals of 3 days.

4.4.5 Validation of Affymetrix data by quantitative real time PCR

The affymetrix data obtained from the ES cells from the BMP-2 clone, puromycin treated 7day EBs (BMP-2+ population and puromycin untreated 7 day EBs) were verified by quantitative real time PCR analysis of the pluripotent and germ layer specific genes – Nanog, T- Brachyury, Bmp-2, Neurofilament H, Sox 17 and a- myosin heavy chain. As seen in the figure 47, the Real time PCR data correlated very well with the affymetrix data.

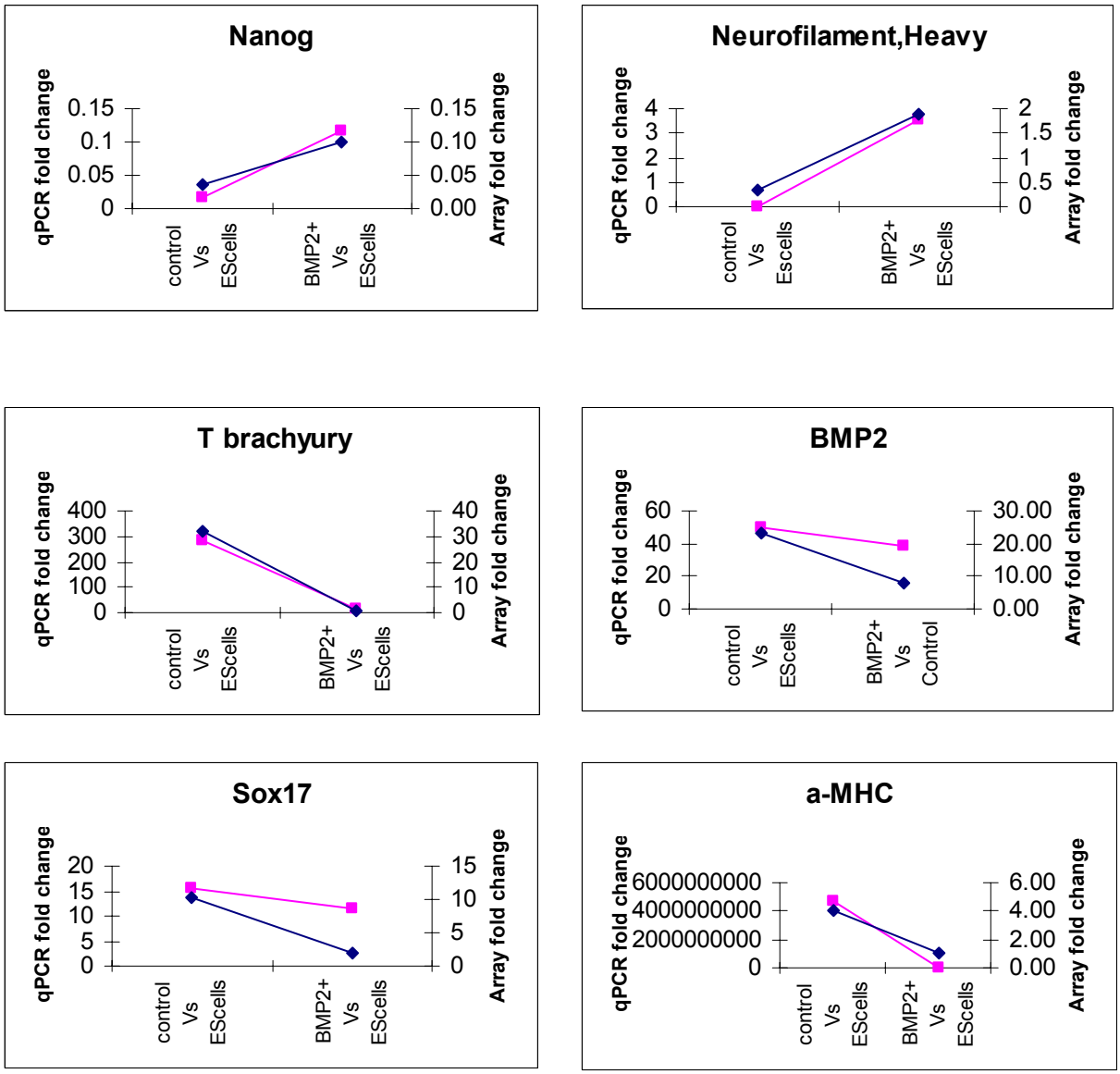


Figure 47. Validation of the Affymetrix data by Quantitative real-time PCR analysis of 5 representative genes. The fold change was calculated by using the formula, fold-change = $2^{-(\Delta C_t \text{ gene1} - \Delta C_t \text{ gene2})}$. ΔC_t of the gene in the sample in which it is expressed lowest is taken as ΔCt gene2 to calculate the fold change using the above formula. The pink line represent results from quantitative PCR and the blue line from affymetrix data.

4.4.6 Transcriptomic analysis of the BMP-2$^+$ cell population

To rule out the differentially expressed transcripts are specific to BMP-2$^+$ cell and are not due to the puromycin application, transgenic β actin clones were generated in the same way like the BMP-2 clones except that the β actin promoter instead of the BMP-2 promoter was used to serve as puromycin treatment controls. Under the same experimental conditions as for the derivation of BMP-2$^+$ cells, on day 4 the EBs were either treated with puromycin or untreated for the subsequent 3 days. On day 7, the RNA samples were prepared from the βactin ES cells, 7 day old puromycin treated β-actin EBs and untreated control β-actin EBs. The samples were affymetrix profiled in the same way as for the BMP-2 samples.

As the first step, the differentially expressed genes due to puromycin application were analyzed. There were 6 unique transcripts with student t-test value of $p<0.01$ differentially expressed between puromycin treated 7day β actin EBs and untreated β-actin EBs and were excluded from the BMP-2 transcriptome analysis.

4.4.6 Major pathways / biological processes differentially regulated in BMP-2$^+$ cells

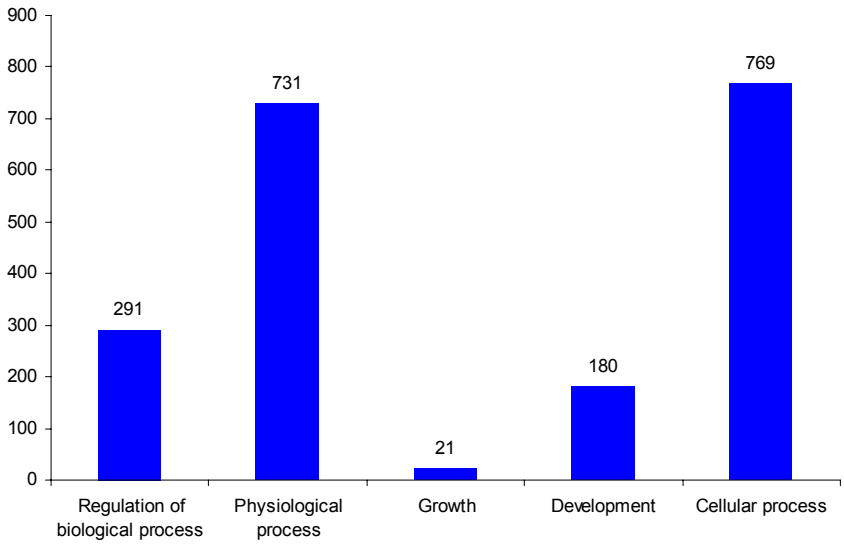


Figure 48. An overview of the number of transcripts differentially expressed in BMP-2$^+$ population according to different Gene Ontologies (GOs), as analyzed with DAVID functional annotation tool.

Overall, there are 1682 unique transcripts differentially expressed (2 fold up or 2 fold down with an ANOVA p value cut off of less than 0.01) compared to either the BMP-2 ES cells or 7day old control EBs. Among them, 28 transcripts participate in MAPK

signaling pathway, 19 transcripts in Wnt Signaling and 13 transcripts in TGFβ signaling pathways. 19 transcripts take part in cell cycle. This shows that the BMP-2 population also has several dynamically ongoing functional signaling cascades, which are implicated to play a crucial role in early embryonic patterning. Overall, among the 1682 transcripts differentially expressed in BMP-2$^+$ cells in comparison to BMP-2 ES cells and 7 day control EBs, 206 genes are implicated in embryonic development.

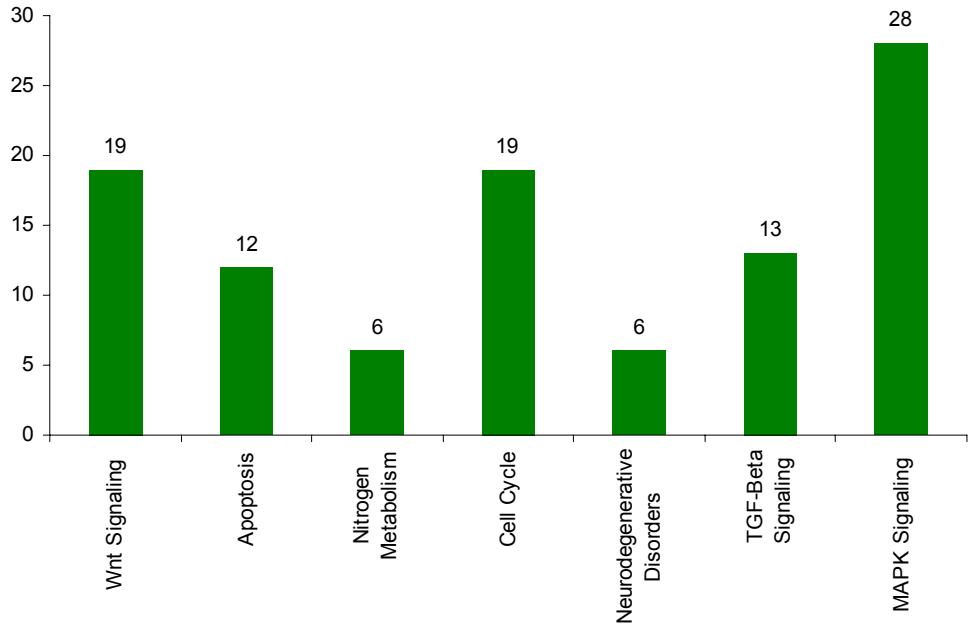


Figure 49. An overview of the number of transcripts differentially expressed in BMP-2$^+$ cell population involved in pathways, as obtained from DAVID analysis.

4.4.7 Classification of differentially expressed transcripts in BMP-2$^+$ cell in comparison to ES cells and 6day control EBs.

BMP-2$^+$ cell population represents a subset of mesoderm unlike the pan mesodermal marker, T Brachyury. Like the T Brachyury cells, BMP-2 cells also needs co-operation from other 2 germ layers for the further lineage commitment during differentiation under normal conditions. When they were kept in isolation, their interaction with the other germ layers is eliminated and hence they slow down in terms of the differentiation and remain in differentiation state between the starting ES cell state and the 6day control EBs which, are far advanced in terms of differentiation due their intact interaction with the other 2 germ layers. Hence characterization of differentially expressed transcriptome of BMP-2 population needs to be compared with both the BMP-2 ES cells and 6day Control EBs in the same way as performed for the T Brachyury transcriptome analysis. Thus the analysis will need to be addressed as follows

1. Transcripts upregulated in BMP-2$^+$ cells than both ES ells and 7day control EBs

    i. Uniquely upregulated in BMP-2$^+$ cells (Not differentially expressed among ES cells and 7day old control EBs)

    ii. Not Uniquely expressed in BMP-2$^+$ cells (Differentially expressed among ES cells and 7day old control EBs)

2. Transcripts downregulated in BMP-2$^+$ cells than both ES ells and 7day control EBs

    i. Uniquely downregulated in BMP-2$^+$ cells (Not differentially expressed among ES cells and 7day old control EBs)

    ii. Not Uniquely expressed in BMP-2$^+$ cells (Differentially expressed among ES cells and 7day old control EBs)

3. Transcripts upregulated in BMP-2$^+$ cells than ES cells but downregulated than 7day control EBs

4. Transcripts upregulated in BMP-2$^+$ cells than 7 day control EBs but downregulated than ES cells

Analysis of upregulated transcripts in BMP-2$^+$ cells in comparison to both ES cells and 7day EBs.

There are 271 transcripts upregulated in BMP-2$^+$ cells than both ES cells and 7day control EBs. Among them, there are 209 transcripts, which do not show any differential expression among ES cells and 7day EBs. The other 62 transcripts even though are upregulated than both ES cells and 7day EBs, they in turn show differential expression among ES cells and EBs. The 209 transcripts upregulated uniquely in BMP-2 cells but not differentially expressed among ES cells and EBs are the bona-fide transcriptomic signatures of BMP-2$^+$ mesodermal cells and these transcripts over-represent transcription factors and apoptosis, when analyzed with DAVID gene functional tool. The unique BMP-2 cell specific transcription factors are enlisted in table 5 and transcripts involved in apoptosis in table 6. . The transcripts upregulated in BMP-2 cells which in turn shows differential expression among ES cells and EBs did not over-represent any biological processes other than the normal physiological processes.

Table.5 List of transcription factors upregulated specifically in BMP-2$^+$ cells. The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

Unigene ID	Gene Name	ES cells	7day Control EBs	BMP-2$^+$
Mm.303534	Down-regulator of transcription 1	1.00	1.08	2.26
Mm.100273	Junction-mediating and regulatory protein	1.00	1.81	8.35
Mm.130752	ash1 (absent, small, or homeotic)-like (Drosophila)	1.07	1.00	2.44
Mm.27663	transcription elongation factor B (SIII), polypeptide 3	1.10	1.00	2.31
Mm.2444	Myelocytomatosis oncogene	1.32	1.00	2.67
Mm.196325	RIKEN cDNA A730098D12 gene	1.00	1.00	2.05
Mm.254233	ELK4, member of ETS oncogene family	1.02	1.00	2.79
Mm.209903	Activating transcription factor 2	1.00	1.32	3.30
Mm.156727	Vcell division cycle 73	1.00	1.17	2.89
---	Cleavage and polyadenylation specific factor 6	1.00	1.48	3.22
Mm.194486	PHD finger protein 3	1.41	1.00	3.14
Mm.206555	zinc finger protein 655	1.00	1.00	2.52
Mm.28646	RIKEN cDNA 6430596G11 gene	1.00	1.33	3.44
Mm.130883	Metadherin	1.04	1.00	3.57
Mm.132238	Adult male aorta and vein cDNA	1.31	1.00	2.72
Mm.267116	RAR-related orphan receptor alpha	1.00	1.65	5.07
Mm.218423	RIO kinase 2 (yeast)	1.25	1.00	2.66
Mm.379292	zinc finger, ZZ domain containing 3	1.00	1.33	3.54
Mm.128165	AT hook containing transciption factor 1	1.48	1.00	3.93
Mm.189270	Ectodysplasin A2 isoform receptor	1.00	1.02	3.15
Mm.271914	Machado-Joseph disease homolog (human)	1.00	1.38	3.21
Mm.325827	G1 to S phase transition 1	1.00	1.01	2.44
Mm.40828	Sequestosome 1	1.46	1.00	3.86
Mm.273090	CCAAT/enhancer binding protein (C/EBP), gamma	1.06	1.00	2.47

Table 6.List of transcription factors upregulated specifically in BMP-2 cells playing their role in apoptosis. The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

Unigene ID	Gene Name	ES cells	7d Control EBs	BMP-2+
Mm.271878	phorbol-12-myristate-13-acetate-induced protein 1	1.00	1.39	3.62
Mm.100273	junction-mediating and regulatory protein	1.00	1.81	8.35
Mm.7660	Bcl-2 binding component 3	1.00	1.80	3.70
Mm.2444	myelocytomatosis oncogene	1.32	1.00	2.67
Mm.29816	programmed cell death 6 interacting protein	1.00	1.27	2.65
Mm.7087	rabaptin, RAB GTPase binding effector protein 1	1.00	1.22	2.84
Mm.297905	DNA fragmentation factor, beta subunit	1.00	1.19	5.93
Mm.40828	sequestosome 1	1.46	1.00	3.86

Downregulated in BMP-2$^+$ cells than both ES cells and 7 day control EBs

The down-regulated transcripts are categorized according to the biological processes using DAVID functional annotation tool and is shown in annexure 10. The development related transcripts are listed in annexure 11.

Upregulated in BMP-2$^+$ cells than ES cells but downregulated than 7day control EBs

54 unique transcripts were found to be upregulated in BMP-2$^+$ cells compared to ES cells but down regulated in comparison to 7d control EBs. Among the 54 transcripts, 9 transcripts are developmentally important genes as shown in the annexure 12A.

The transcripts upregulated in BMP-2$^+$ cells than 7 day control EBs but downregulated than ES cells.

There are 119 unique transcripts that are upregulated in BMP-2$^+$ cells compared to 7day EBs cells but downregulated in comparison to ES cells and development related transcripts are shown in annexure 12B. It is interesting to note that the nanog transcript is upregulated in BMP-2$^+$ cells compared to the control EBs.

4.5 CHARACTERIZATION OF αMHC$^+$ CARDIOMYOCYTES

4.5.1.Generation and analysis of the αMHC$^+$ ES cell clone

Embryoid Bodies, when made from the αMHC$^+$ ES cell clone by hanging drop protocol and analyzed at regular intervals, showed the expression of αMHC$^+$ transcripts as shown by RT-PCR in figure 50. The EGFP expressing cells were first detectable in 7-8 days old EBs and the EGFP fluorescence increases significantly after 7 days. The ES cells and EBs from this transgenic clone behaved faithfully with the wild type ES and EBs.

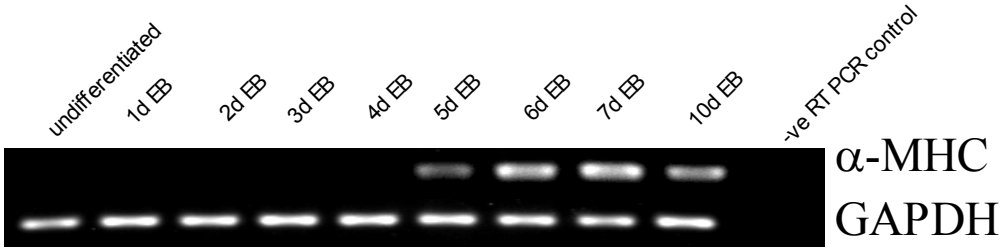


Figure 50. Expression pattern of αMHC transcript during the course of differentiation**.**

4.5.2.Isolation and transcriptomic profiling of αMHC$^+$ cardiomyocytes

The EBs were generated by hanging drop protocol. On day 7, the first contracting clusters started to appear. On the next day (day 8), the EBs were treated with or without puromycin at a concentration of 4μg/ml for the subsequent 7 days (Figure 51A). On day 15, the EBs untreated and puromycin treated were harvested by trypsinisation and analyzed with FACS to determine the purity of the cardiomyocytes in the puromycin treated cultures along with the control EBs. As shown in figure 51B, the puromycin treated cultures contained an enriched population of cardiomyocytes up to 91.15% compared to 16.72% of the control EBs. Thus an enrichment of the pure cardiomyocytes by 5.5 fold compared the control EBs was obtained by this approach. RNA was isolated from these cardiomyocytes for the transcriptomic profiling by Affymetrix.

The αMHC$^+$ cardiomyocytes thus derived from the ES cells were analyzed by immunohistochemistry. They showed a multi-angular (Fig. 52a and b), more rectangular (Figure 52c and d) and a tri-angular morphology (Figure 52e and f). Detection of cardiac α-actinin by immunocytochemistry (Figure 52b, d and f) clearly indicated the Z-disc specific protein and the characteristic striations of sarcomeric structures of the cardiac cells. The gap junction protein Connexin43 is highly expressed in the heart and was detected by immunocytochemistry (Figure 52g and h). Connexin43 is distributed in the cytosol and in the outer membranes in the cell border regions (Figure 52g and h, middle panel).

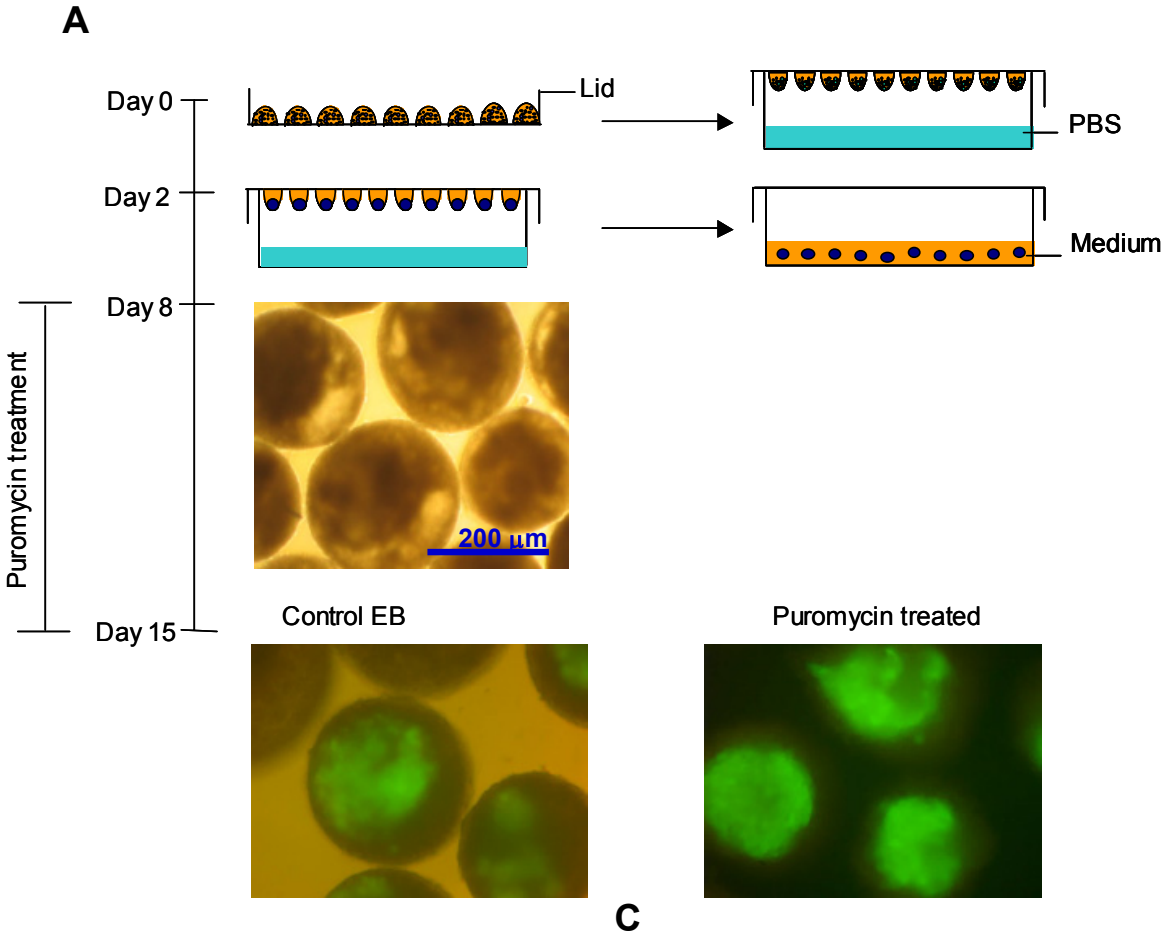


Figure 51. Enrichment of αMHC$^+$ cells isolated from the αMHC+ ES cell lineage after puromycin treatment. **A,** progressive purification of αMHC+ cardiac cell aggregates after treatment of the 8-days old EBs with 4µg/ml puromycin for 7 days. Puromycin containing medium was refreshed every second day. **B, C,** Cells from 15-days old EBs and 15-days old puromycin purified αMHC$^+$ aggregates were dissociated by trypsinisation and the purity of the αMHC$^+$ cells in the 15-days old EBs (C) and in the 15-days old αMHC$^+$ aggregates (D) was examined by FACS analysis.

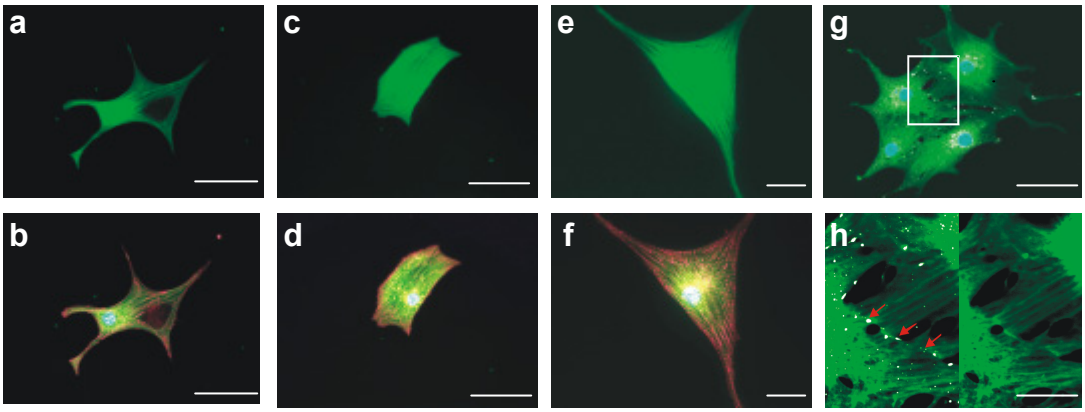


Figure 52. Characterization of the ES cell-derived cardiomyocytes by immunocytochemistry. αMHC$^+$ cardiomyocytes were dissociated with collagenase B and plated on fibronectin coated cover slips. (a, c, e) EGFP expression of single αMHC$^+$ cells with different morphology. (b, d, f) Detection of α-cardiac actinin and Connexin43 (g, h) was performed using anti-cardiac Actinin (1:400) and anti-Connexin43 1:400. Secondary detection was performed with anti-mouseIgG1-AlexaFluor555 and anti-rabbit-Ig-AlexaFluor647. Hoechst dye was used to stain nuclei. Bars in (a-f) represent 50 and bars in (g) 20 and (h) 7.5 µm.

4.5.3 Electrophysiological characterization of the cardiomyocytes

Functional characterization of the αMHC$^+$ cardiac cells has been performed by measurements of their typical spontaneous action potentials (APs). Spontaneous APs were measured in single αMHC$^+$ cardiomyocytes (n=32) as well as in multicellular αMHC$^+$ aggregates (n=24). All APs possessed characteristic parameters of cardiac APs. The minimal diastolic potential was -60.2±1.1 mV, membrane potentials normally showed a diastolic depolarization leading to a spontaneous AP frequency of 125.9±8.0 /min. The maximal upstroke velocity was 22.9±2.2 V/s, pointing to a contribution of voltage-activated Na^+ currents, which was confirmed by voltage-clamp measurements (data not shown). APD_{90}, APD_{50} and APD_{20} (AP duration from maximum to 90%, 50% and 20% repolarization) were 96.4±4.2 ms, 71.1±3.9 ms and 41.3±2.6 ms. APs had a variety of morphologies, including pacemaker-, atrial- and ventricular-like APs (Figure 53). In most cases, however, morphological properties did not match any type of specific differentiation. These unspecified APs mostly possessed a plateau phase, but had a much shorter APD_{90} than ventricular APs, which are characterized by a long APD_{90} of about 200 ms [112].

To characterize the hormonal regulation of αMHC cardiomyocytes, carbachol (Cch, an agonist of m-cholinoceptors), and isoproterenol (ISO, an agonist of β_1-adrenoceptors), were applied (Figure 53B and C). 1 µM Cch decreased the AP frequency significantly to 44.8±7.5 % of the control values (n=20; the frequency under control conditions was

determined for each recording and set to 100; Figure. 53D). 1 µM ISO evoked a significant increase of the frequency to 238.5±23.7 % of the control values (n=19; Fig. 2E). Intracellular recordings of spontaneous APs revealed typical cardiac AP parameters and morphologies confirming the cardiac differentiation and functionality of puromycin-selected αMHC$^+$ cells. The muscarinic and adrenergic regulation of the AP frequency, which is known for ES cell-derived cardiomyocytes [113] as well as for native murine cardiomyocytes at early developmental stages [114], further supports a physiological cardiac differentiation of αMHC cells. As described previously [115,116], APs at the intermediate developmental stage showed diastolic depolarizations and diverse shapes. APs with a distinct plateau phase were frequent but considered to be unspecific rather than ventricular-like in the majority of cases, since the APD_{90} was much shorter than reported for early stage murine ventricular cardiomyocytes as well as for murine ES cell-derived ventricular-like cardiomyocytes [117]. Since most APs had unspecific morphological properties, a general classification into pacemaker-like, atrial-like and ventricular-like APs could not be made, which accords well with previous findings from intermediate stage ES cell-derived cardiomyocytes [118,119]. Only few APs possessed typical morphological features of the respective differentiation types.

It has been reported recently that, in ES cell-derived cardiomyocytes expressing GFP under control of the αMHC promoter, green fluorescence is restricted to pacemaker- and atrial-like cells [120]. Since we found puromycin-purified αMHC$^+$ cardiomyocytes with a ventricular-like AP morphology in few cases, our data suggest that αMHC expression is not completely absent in ES cell-derived ventricular-like cardiomyocytes. This apparent discrepancy might arise from the complex stage-dependent expression pattern described for αMHC in murine embryoid bodies [121] and murine embryonic ventricles [122], as a different developmental stage of ES cell-derived cardiomyocytes was investigated in the present study (15-day old cardiomyocytes) as compared to the study by Kolossov et al. (9-day to 11-day old cardiomyocytes) [123].

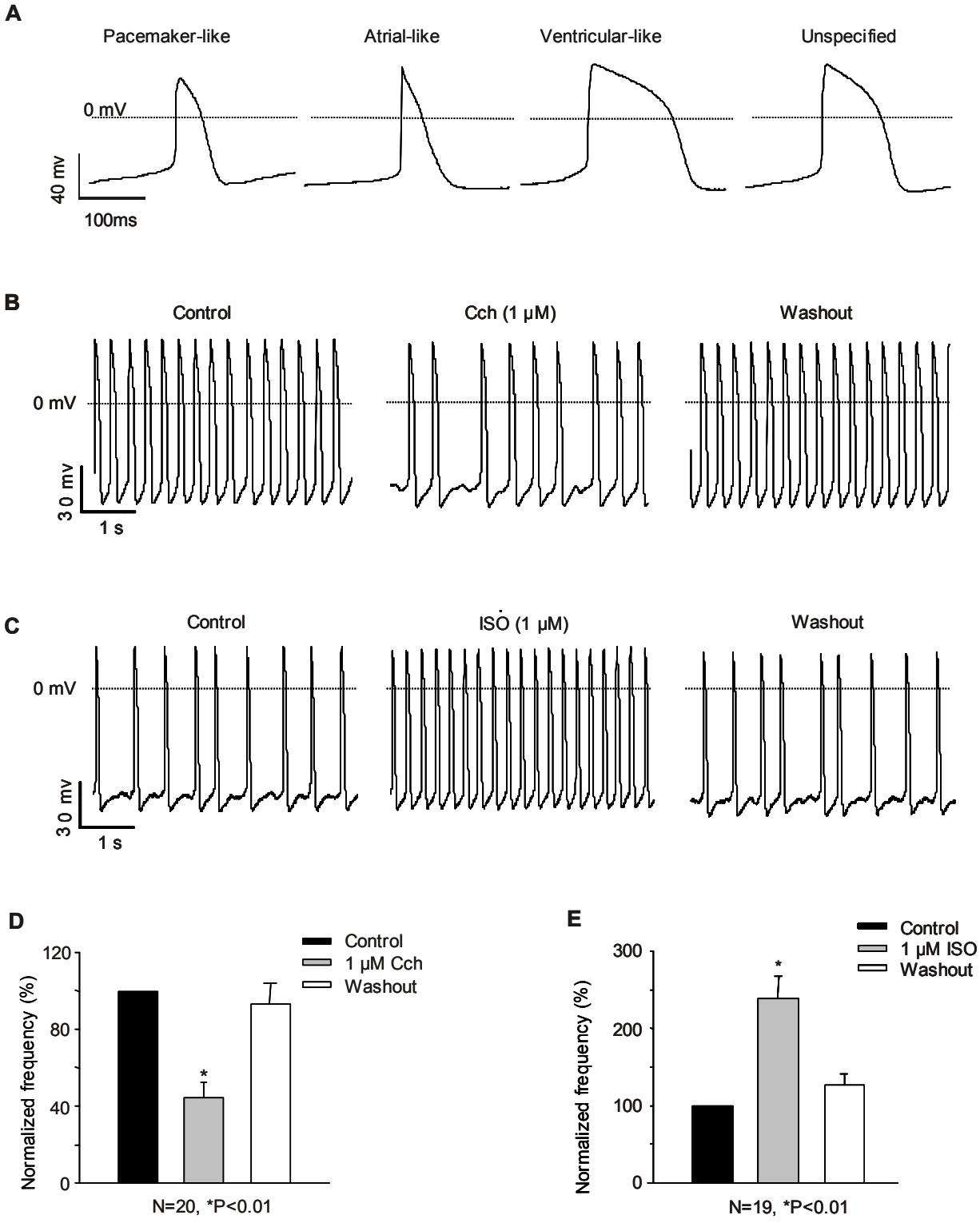


Figure 53. Electrophysiological characterisation of αMHC+ cells. **A,** Characteristic cardiac APs of puromycin purified αMHC cells. Most APs had a typical cardiac AP morphology but could not be further specified. Only few APs showed typical features of pacemaker-like atrial-like or ventricular-like APs. The minimal diastolic potential was –60.2±1.1 mV. The maximal upstroke velocity was 22.9±2.2 V/s. APD_{90}, APD_{50} and APD_{20} were 96.4±4.2 ms, 71.1±3.9 ms and 41.3±2.6 ms. **B, C,** Representative recordings showing the effect of Cch (1 µM) (B) and ISO (1 µM) (B) on the spontaneous AP frequency. **C, D,** Statistical analysis of the effects of Cch (1 µM) and ISO (1 µM), respectively, on the spontaneous AP frequency. Cch caused a decrease while ISO increased the spontaneous AP frequency.

4.5.4.Validation of Affymetrix data by quantitative real time PCR

The affymetrix data obtained from the ES cells, 15 days old puromycin treated and untreated αMHC EBs were verified by quantitative real time PCR analysis of the pluripotent and germ layer specific genes –Nanog, T- Brachyury, BMP-2, Neurofilament H, Sox 17 and a- myosin heavy chain. As seen in the figure 54, the Real time PCR data correlated very well with the affymetrix data. Additionally, the cardiomyocyte specific markers were analyzed in these samples by semi-quantitative PCR (Figure 55).

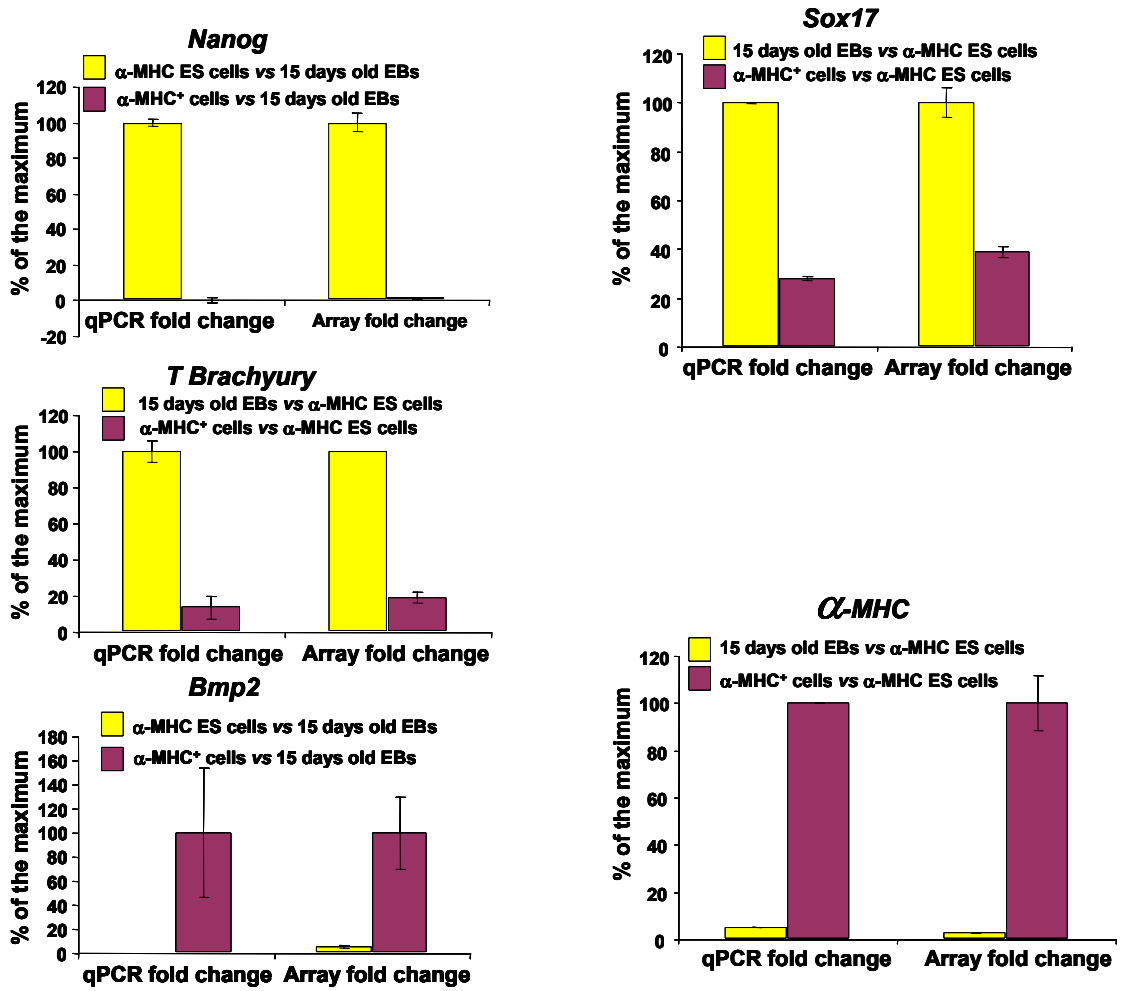
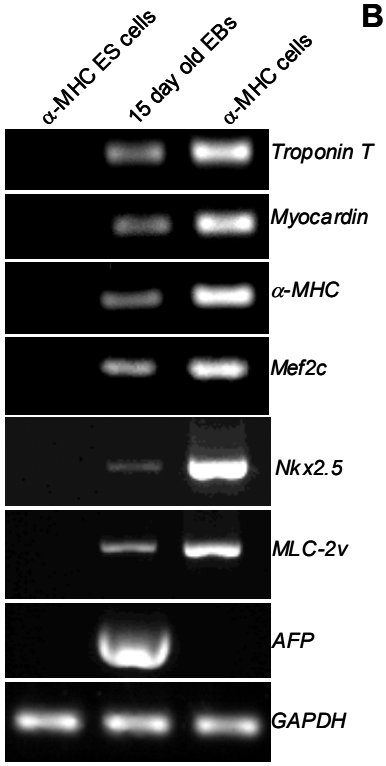


Figure 54. Validation of the Affymetrix data by Quantitative real-time PCR analysis of 5 representative genes. The fold change was calculated by using the formula, fold-change = $2^{-(\Delta C_t \text{ gene1} - \Delta C_t \text{ gene2})}$. ΔC_t of the gene in the sample in which it is expressed lowest is taken as ΔCt gene2 to calculate the fold change using the above formula. The resulting fold change is expressed as percentage of the maximum (=100%).Values are expressed as mean ± SD.

A



B

Relative mean fluorescence signal values as obtained from the Affymetrix microarrays

Gene	α-MHC ES cells	15 days old EBs	α-MHC$^+$ cells
Troponin T	9	4142	12873
Myocardin	12	142	1359
α-MHC	36	3814	15372
Mef2c	23	834	1889
Nkx2-5	49	101	787
MLC-2v	47	3642	16050
AFP	12	5120	14
Neurofilament-M	12	17	12

Figure 55. RT-PCR analysis of cardiomyocyte-specific markers and an endodermal marker (AFP). **B**, Relative mean fluorescence signal for the expression level of the genes obtained from the Affymetrix data.

4.5.5.Transcripomic analysis of αMHC$^+$ cardiomyocytes

To rule out the differentially expressed transcripts are specific to cardiomyocytes and are not due to the puromycin application, under the same experimental conditions as for the derivation of cardiomyocytes, on day 8 the EBs were either treated with puromycin or untreated for the subsequent 7 days. On day 15, the RNA samples were prepared from the β actin ES cells, 15 day old puromycin treated β actin EBs and untreated control βactin EBs. The samples were affymetrix profiled in the same way as for the αMHC samples.

As the first step, the differentially expressed genes due to puromycin treatment were analyzed. There were 235 unique transcripts with student t-test value of $p<0.01$ among the transcripts differentially expressed among puromycin treated β actin EBs and untreated control βactin EBs and these transcripts were excluded from the analysis of the αMHC transcriptome.

Overall, there are 7700 unique transcripts differentially expressed in αMHC$^+$ cardiomyocytes (2 fold up or 2 fold down with an ANOVA p value cut off of 0.01) compared to either the αMHC ES cells or 6day old control EBs. Among these transcripts, 1315 transcripts with ANOVA p value <0.01 are participating in pathways as shown in the annexure 13 and this comprises 17.08% of the 7700 unique transcripts differentially expressed in αMHC$^+$ cardiomyocytes. It is not surprise to see a significant number of transcripts differentially expressed that are participating in oxidative phosphorylation. This process is essential for the cardiomyocytes to derive the energy for their contraction. Strikingly 128 transcripts participate in MAPK signaling pathway, 68 transcripts in Wnt Signaling and 52 transcripts in TGFβ signaling pathways as shown in the figures 56,58-62. Among the 7700 transcripts differentially expressed in αMHC$^+$ cardiomyocytes in comparison to αMHC ES cells and 15 day αMHC control EBs, 873 genes are implicated in embryonic development as shown in Figure 57.

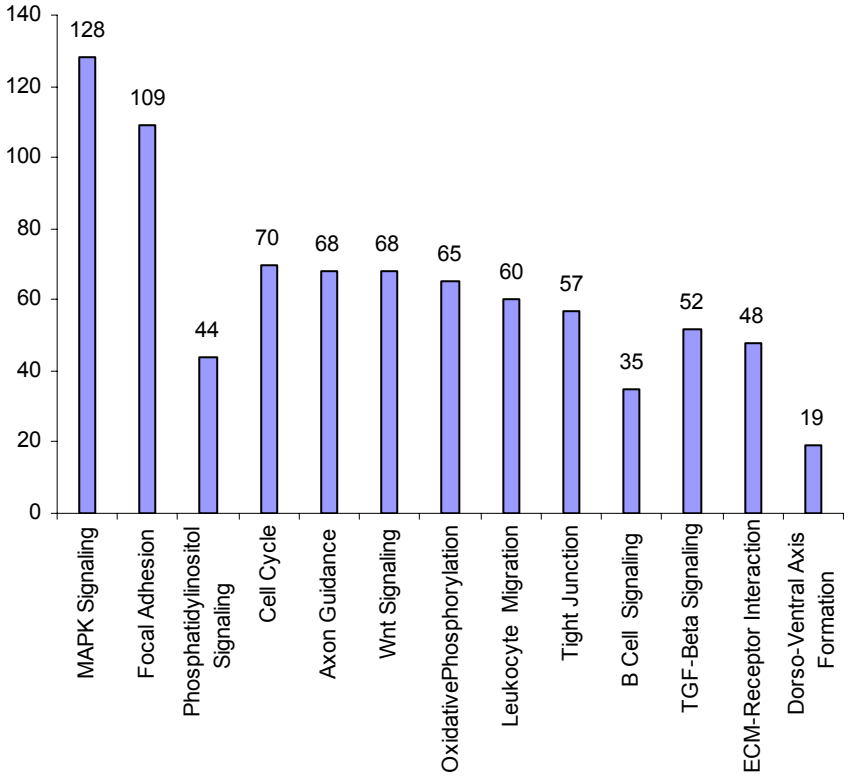


Figure 56. List of the transcripts that are differentially expressed in cardiomyocytes and are involved in pathways in comparison to ES cells, 15day old EBs or both. The count represents the relative ratio of the transcripts belonging to the particular pathway to the total transcripts differentially expressed in cardiomyocytes.

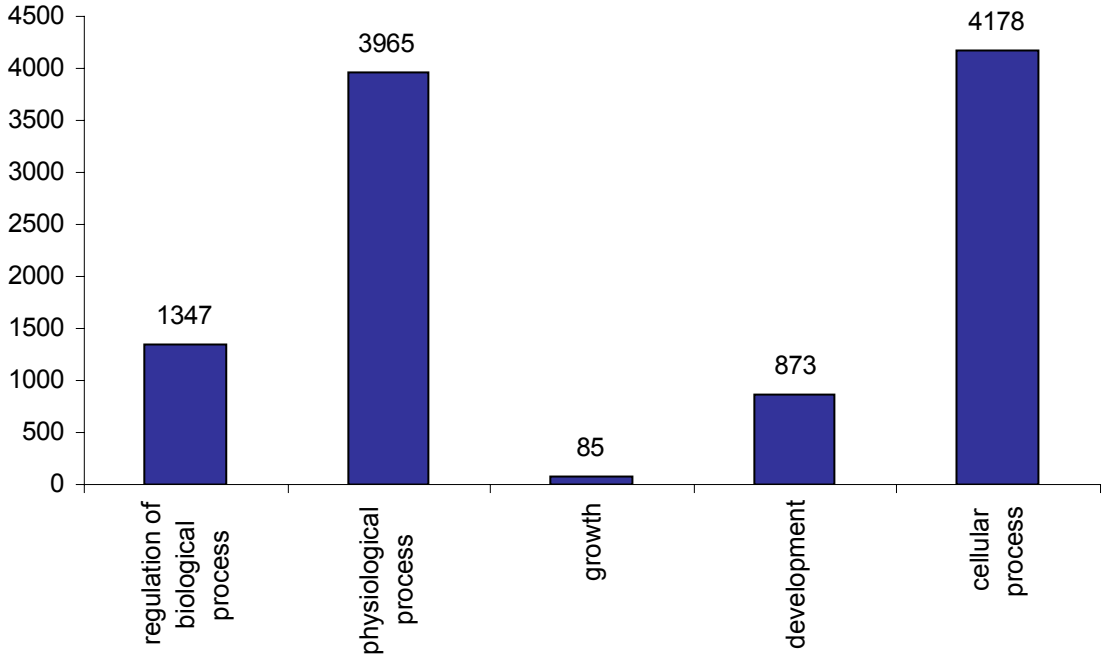


Figure.57 Schematic outline of the biological processes overepresented by the differentially expressed transcripts in cardiomyocytes in comparison to ES cells and 15day old EBs.

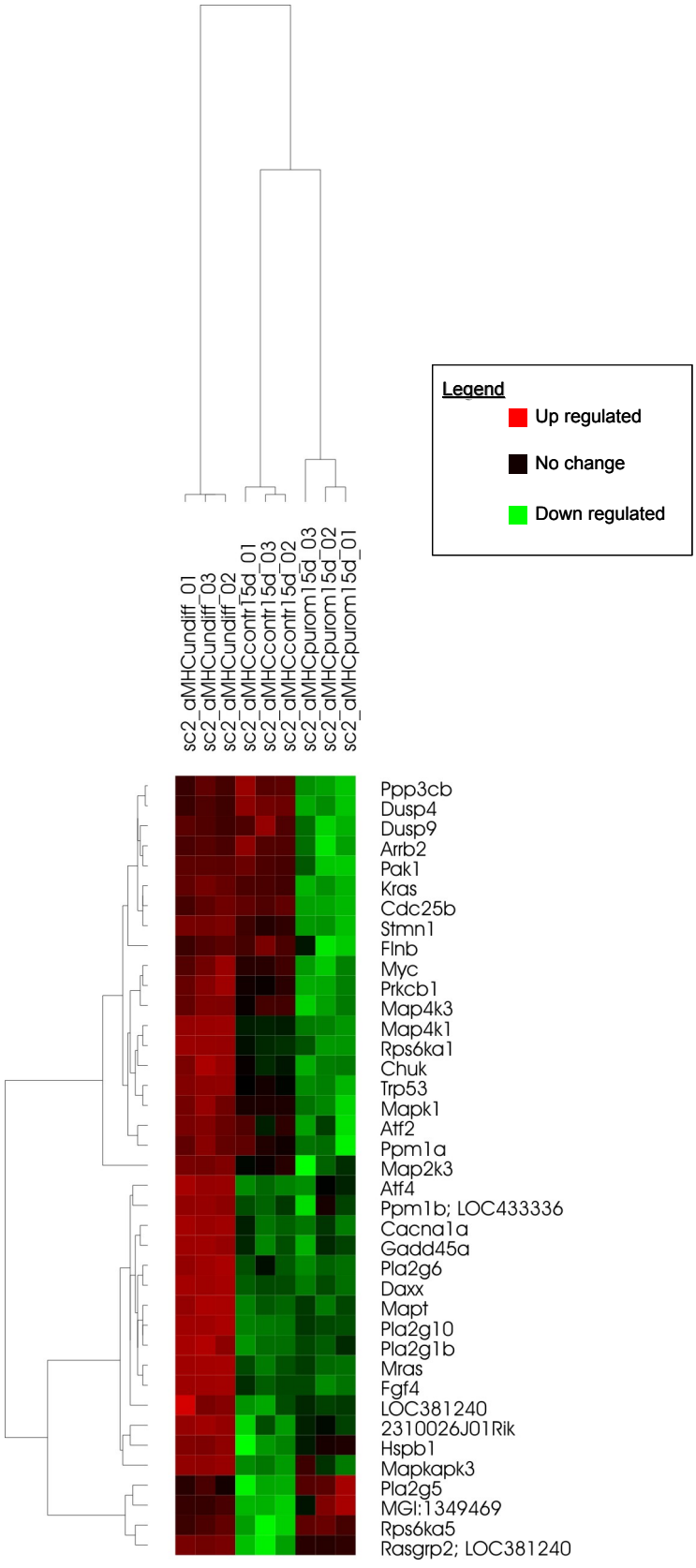


Figure 58. Hierarchical clustering of the transcripts differentially expressed in αMHC^{+} cells that participate in MAPK signaling pathway.

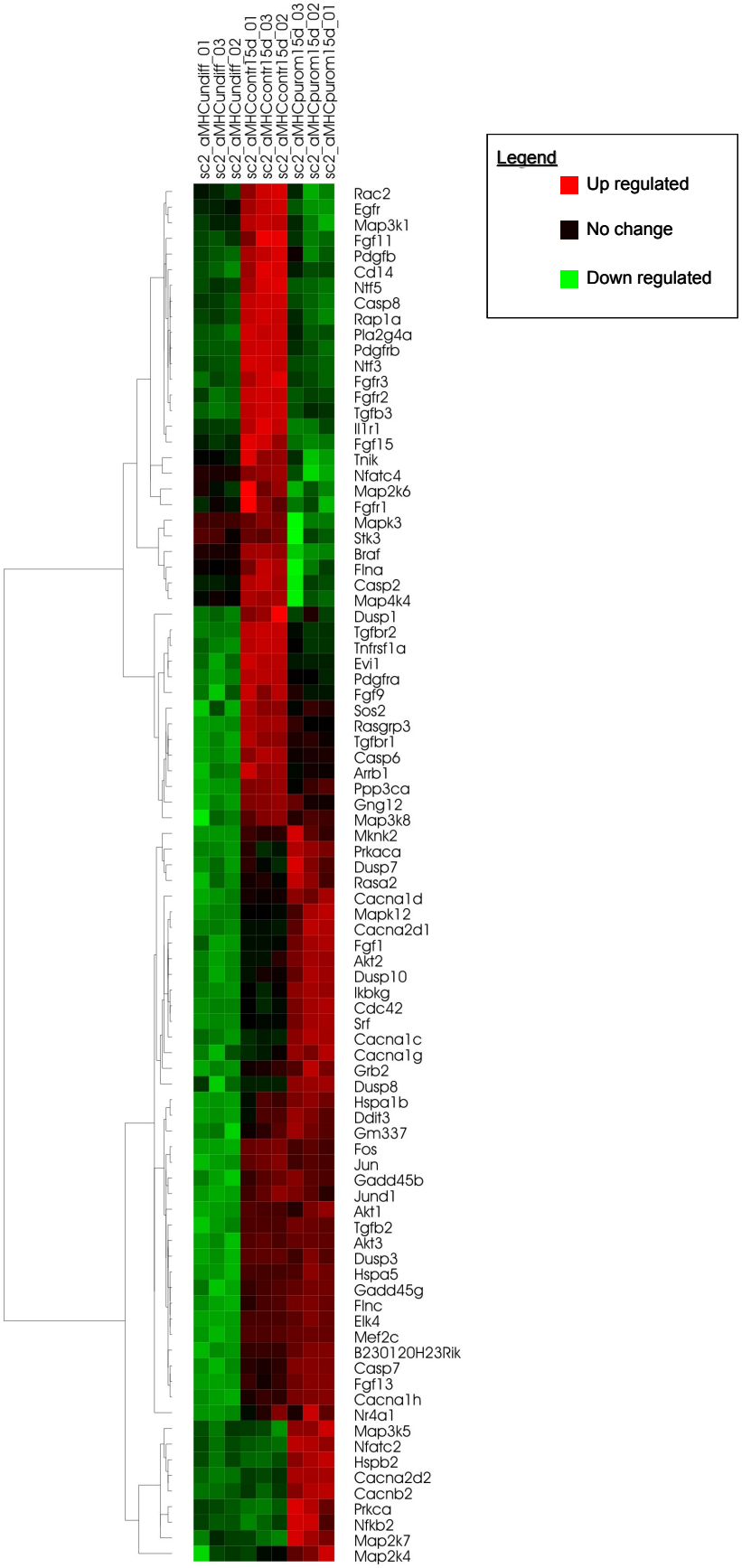


Figure 58 (continued). Hierarchical clustering of the transcripts differentially expressed in αMHC$^+$ cells that participate in MAPK signaling pathway (continued).

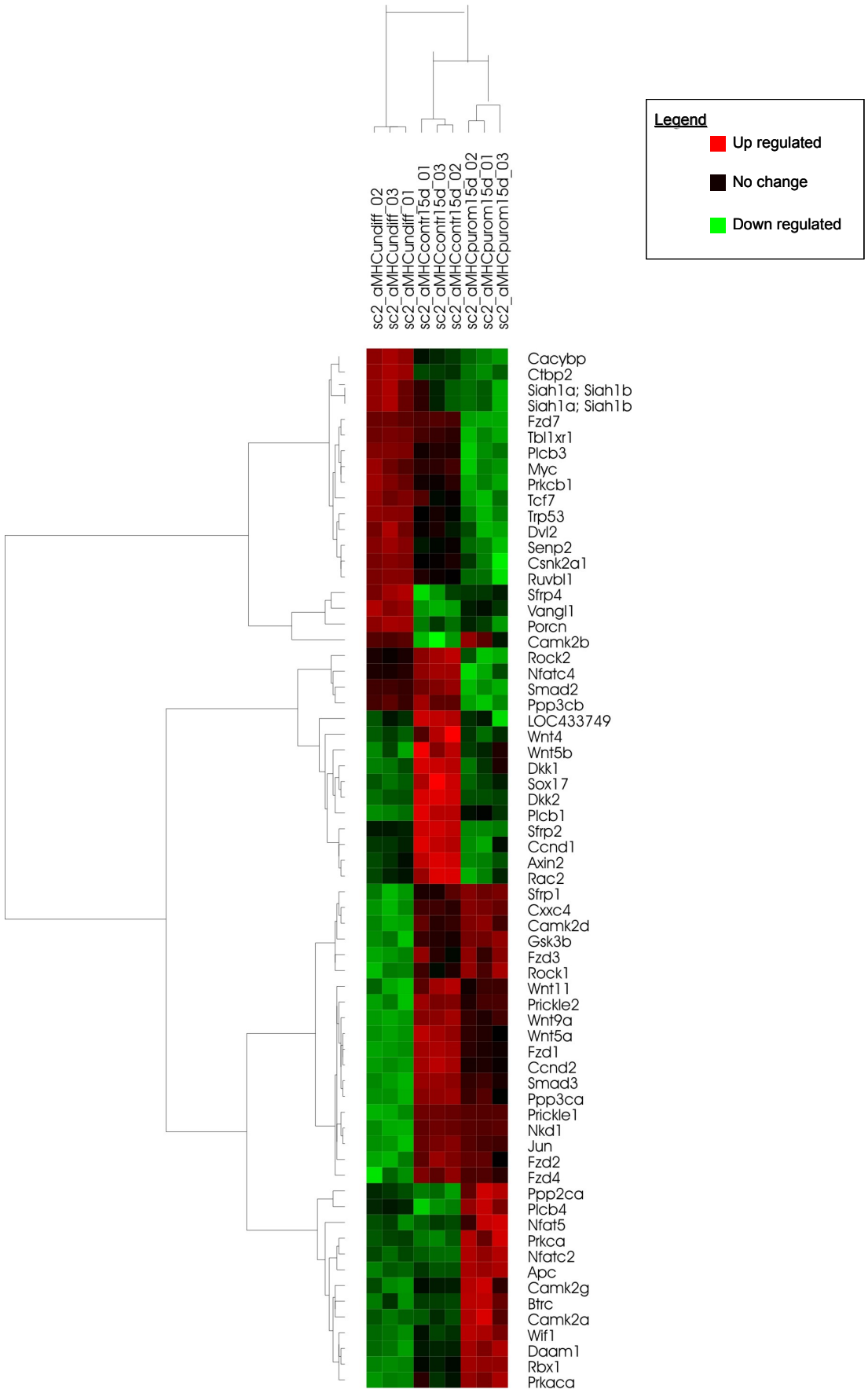


Figure 59.Hierarchical clustering of the transcripts differentially expressed in αMHC$^+$ cells that participate in Wnt signaling pathway

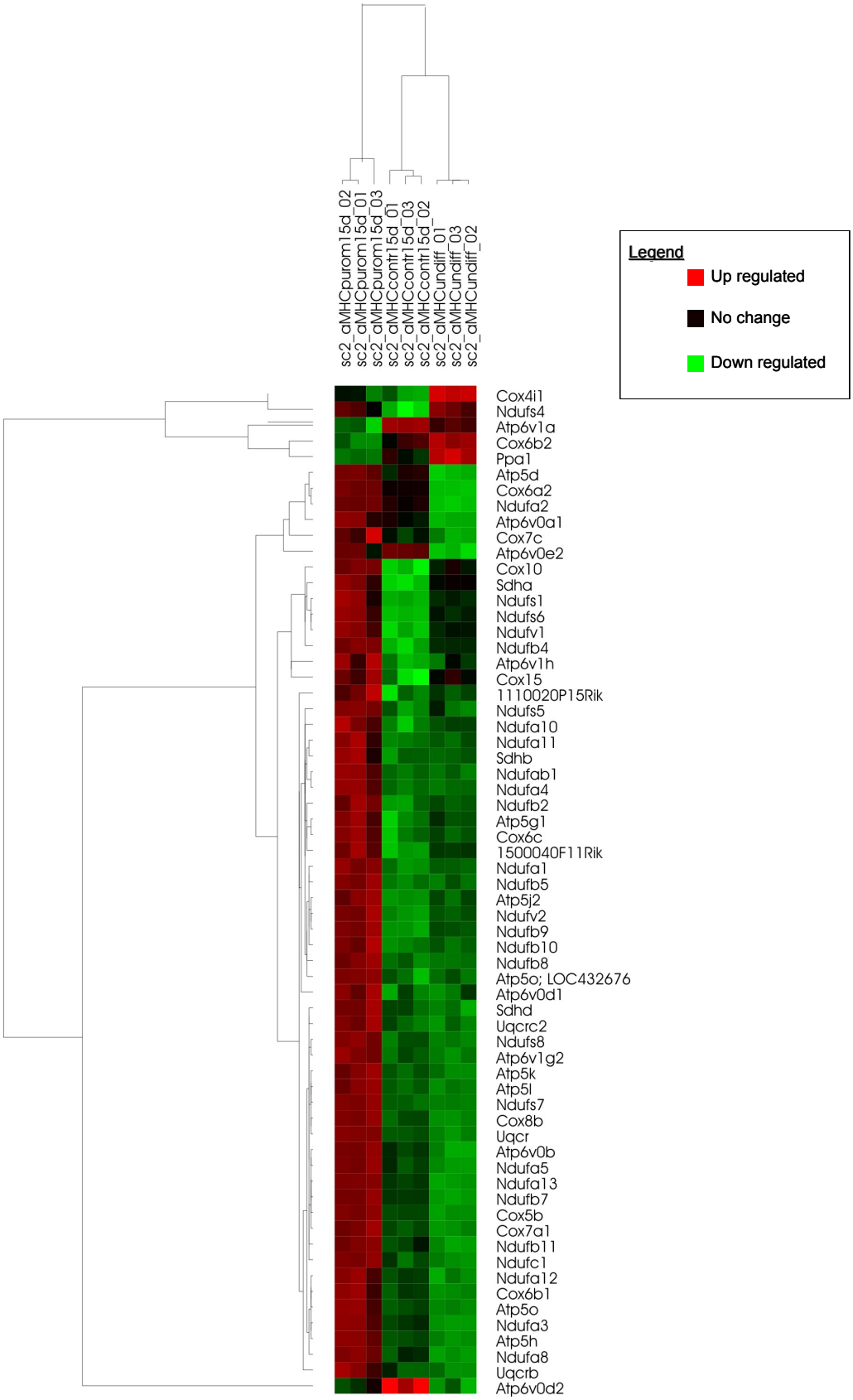


Figure 60.Hierarchical clustering of the transcripts differentially expressed in αMHC$^+$ cells that participate in oxidative phosphorylation.

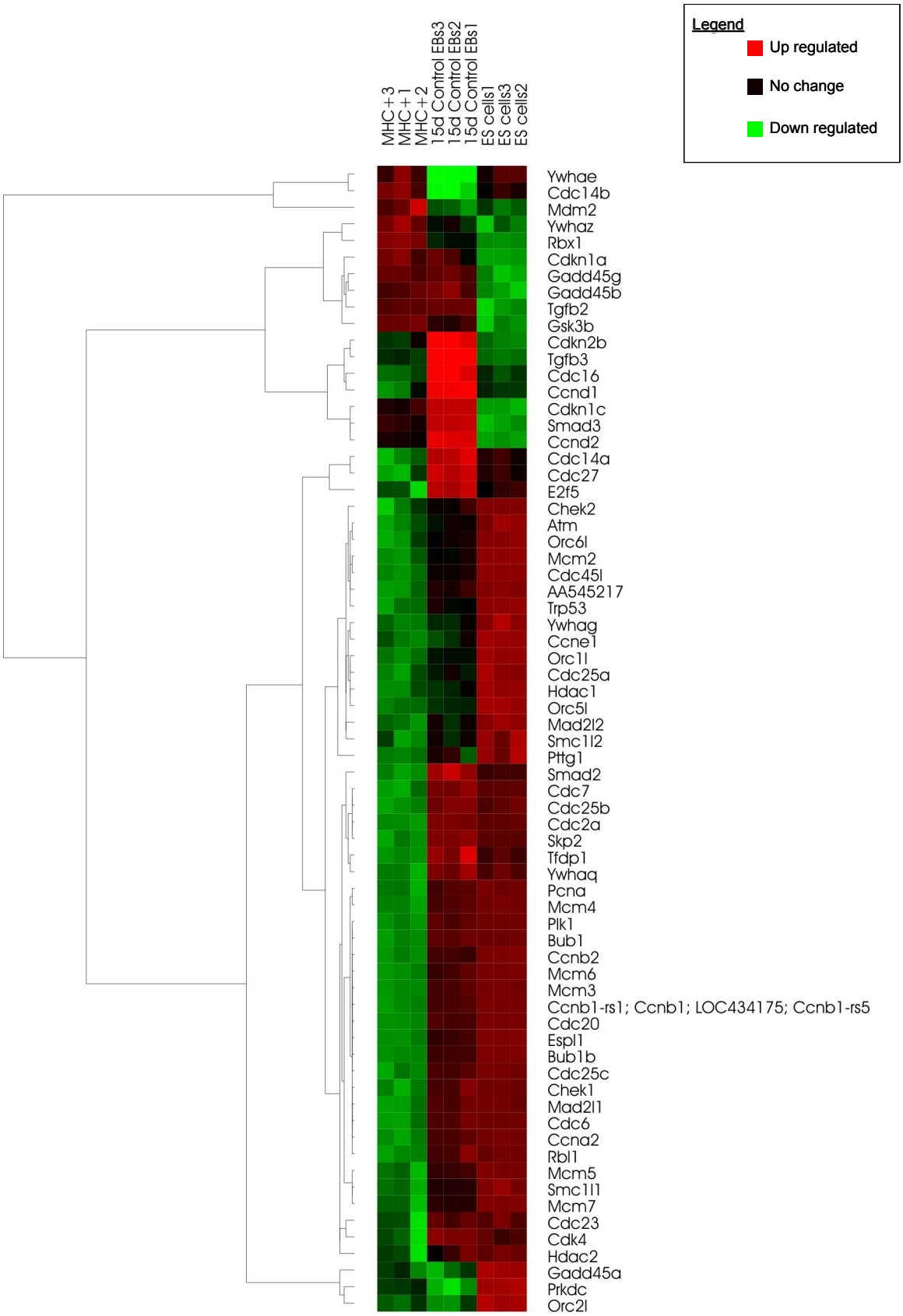


Figure 61. Hierarchical clustering of the transcripts differentially expressed in αMHC$^+$ cells that participate in cell cycle.

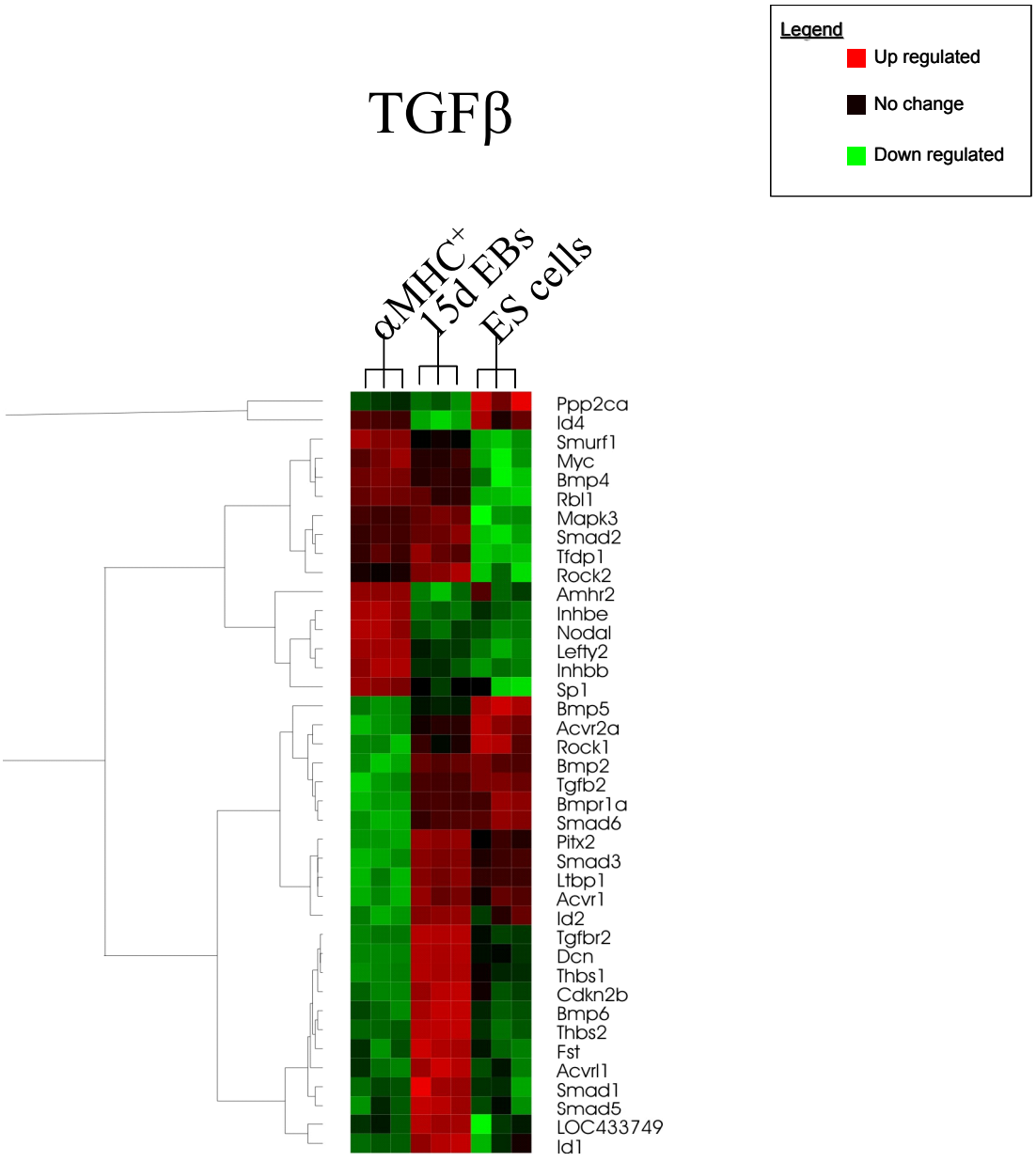


Figure 62. Hierarchical clustering of the transcripts differentially expressed in αMHC$^+$ cells that participate in TGFβ signaling.

Classification of differentially expressed transcripts in αMHC$^+$ cardiomyocytes in comparison to ES cells and 15day control EBs.

To analyze the transcriptome of αMHC$^+$ cardiomyocytes, the comparison of αMHC$^+$ cardiomyocytes with ES cells and 15 day control EBs were made in the same way as analyzed for T Brachyury$^+$ and BMP-2$^+$cells transcriptomes.

1. Transcripts upregulated in αMHC$^+$ cardiomyocytes than both ES ells and 15 day control EBs
    i. Uniquely upregulated in αMHC$^+$ cardiomyocytes (Not differentially expressed among ES cells and 15-days old EBs)
    ii. Not Uniquely expressed in αMHC$^+$ cardiomyocytes (Differentially expressed among ES cells and 15-days old EBs)
2. Transcripts downregulated in αMHC$^+$ cardiomyocytes than both ES ells and 15 day control EBs
    i. Uniquely downregulated in αMHC$^+$ cardiomyocytes (Not differentially expressed among ES cells and 15-days old EBs)
    ii. Not Uniquely expressed in αMHC$^+$ cardiomyocytes (Differentially expressed among ES cells and 15-days old EBs)
3. Transcripts upregulated in αMHC$^+$ cardiomyocytes than ES cells but downregulated than 15 day control EBs
4. Transcripts upregulated in αMHC$^+$ cardiomyocytes than 15 day control EBs but downregulated than ES cells

Analysis of upregulated transcripts in αMHC$^+$ cardiomyocytes in comparison to both ES cells and 15-day-old EBs

There are 1319 transcripts upregulated in αMHC$^+$ cardiomyocytes than both ES cells and 15 day control EBs. Among them, there are 927 transcripts, which do not show any differential expression among ES cells and 15day old control EBs. The other 392 transcripts even though are upregulated than both ES cells and 15 day old EBs, they in turn show differential expression among ES cells and EBs. The distribution of the transcripts according to Gene Ontology “Biological Process” by David Bioinformatics resource analysis is presented in the figure 63. The 927 transcripts upregulated uniquely in αMHC$^+$ cardiomyocytes but not among ES cells and EBs are the bonafide transcriptomic signatures of αMHC$^+$ cardiomyocytes. As seen from the figure 63B, the transcripts of cardiomyocyte transcriptome are involved in the cardiac specific physiological processes like ATP synthesis and oxidative phosphorylation. There are 81 αMHC$^+$ cardiomyocytes specific transcripts involved in development according DAVID analysis and are enlisted in table 8. Among the transcripts upregulated in αMHC$^+$ cardiomyocytes which in turn show differential

expression among ES cells and EBs, the 18 transcripts involved in developmental processes are listed in the annexure 14. 8 transcripts are involved in MAPK signaling pathway and 10 transcripts in calcium signaling pathway (annexure 15).

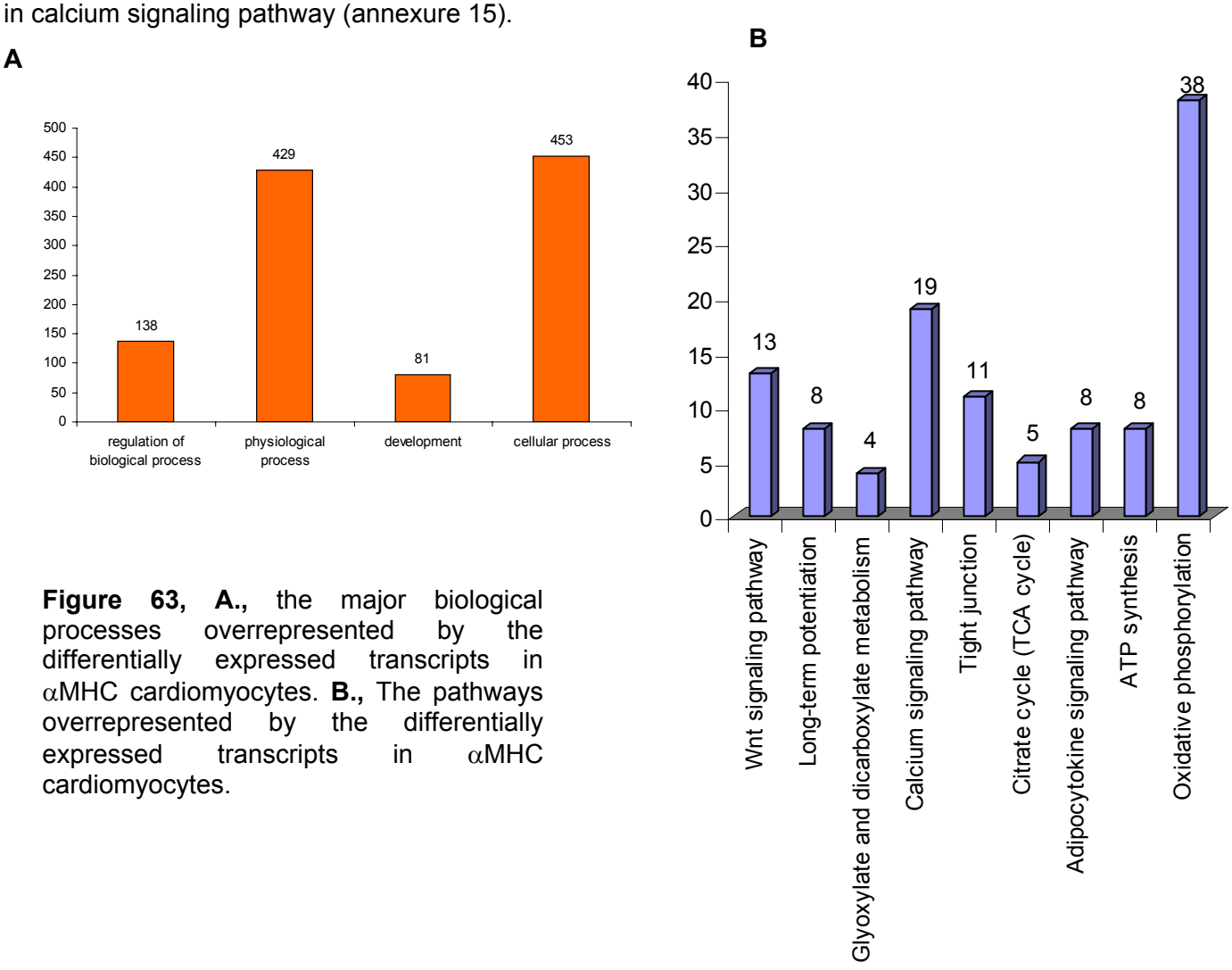


Figure 63, A., the major biological processes overrepresented by the differentially expressed transcripts in αMHC cardiomyocytes. **B.,** The pathways overrepresented by the differentially expressed transcripts in αMHC cardiomyocytes.

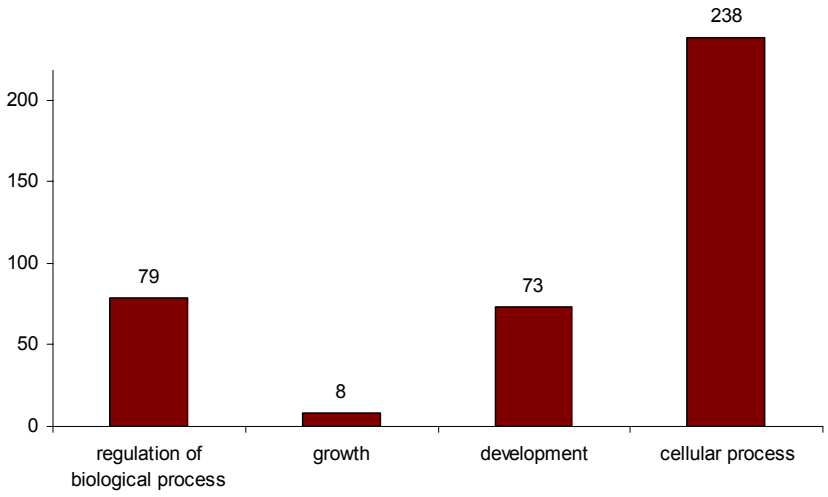


Figure 64, the major biological processes overrepresented by the unique αMHC cardiomyocytes specific transcriptome upregulated in comparison to ES cells and 15 day cardiomyocytes.

Table 8a. Development related transcripts specifically expressed in αMHC$^+$ cardiomyocytes.
The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

NCBI ID	Gene Name	ES cells	15day control EBs	α–MHC$^+$
NM_008664	myomesin 2	1.00	1.33	41.89
NM_011915	Wnt inhibitory factor 1	1.00	1.93	17.60
XM_136682	PREDICTED: v-erb-a erythroblastic leukemia viral oncogene homolog 4 [Mus musculus]	1.00	1.13	14.89
NM_021447	tripartite motif-containing 54	1.18	1.00	14.40
NM_010332	endothelin receptor type A	1.00	1.88	11.11
NM_008904	peroxisome proliferative activated receptor, gamma, coactivator 1 alpha	1.00	1.59	9.95
NM_145602	N-myc downstream regulated gene 4	1.00	1.83	9.95
NM_031260	Moloney leukemia virus 10-like 1	1.00	1.28	9.85
NM_019684	serine/threonine kinase 23	1.00	1.39	9.60
NM_007419	adrenergic receptor, beta 1	1.00	1.76	8.73
NM_011486	signal transducer and activator of transcription 3	1.00	1.95	6.96
NM_008092	Transcription factor GATA-4	1.00	1.80	6.73
NM_145136	myocardin	1.00	1.19	6.26
NM_010564	inhibin alpha	1.00	1.90	6.04
XM_487682	neuralized-like 2 (Drosophila)	1.00	1.81	5.90
NM_007956	estrogen receptor 1 (alpha)	1.00	1.09	5.79
NM_025282	Myocyte enhancer factor 2C	1.00	1.32	5.74
NM_021881	Quaking (Qk), mRNA	1.13	1.00	5.65
NM_009895	cytokine inducible SH2-containing protein	1.00	1.87	5.58
NM_007826	DACH protein (Dach)	1.00	1.45	5.22
NM_015763	lipin 1	1.00	1.54	5.08
NM_013883	sex comb on midleg homolog 1	1.00	1.76	5.08
NM_024226	reticulon 4	1.14	1.00	5.00
NM_011540	titin-cap	1.00	1.13	4.91
NM_054085	alpha-kinase 3	1.98	1.00	4.68
NM_024441	heat shock protein 2	1.00	1.06	4.62
NM_031158	ankyrin 1, erythroid	1.09	1.00	4.61
NM_009670	ankyrin 3, epithelial	1.07	1.00	4.58
NM_008434	KCNQ1b mRNA for Q1-type potassium channel spliced variant	1.77	1.00	4.49
NM_020496	T-box 20	1.00	1.22	4.47

Table 8b. Development related transcripts specifically expressed in αMHC$^+$ cardiomyocytes (continued). The fold change is expressed relative to the lowest value of expression that is normalized to 1.

NCBI ID	Gene Name	ES cells	15day control EBs	α–MHC$^+$
NM_026514	CDC42 effector protein (Rho GTPase binding) 3	1.00	1.65	4.16
NM_007555	bone morphogenetic protein 5	1.00	1.58	4.05
NM_001003817	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)	1.00	1.95	3.94
NM_001013580	Par-3 (partitioning defective 3) homolog (C. elegans) (Pard3), transcript variant 3, mRNA	1.11	1.00	3.74
NM_008937	prospero-related homeobox 1	1.00	1.08	3.67
---	antisense Igf2r RNA	1.00	1.54	3.44
XM_204015	MKIAA0458 protein	1.03	1.00	3.40
---	B6-derived CD11 +ve dendritic cells cDNA	1.00	1.11	3.40
NM_178740	SLIT and NTRK-like family, member 4	1.00	1.48	3.35
NM_172739	Premature mRNA for mKIAA1722 protein	1.41	1.00	3.34
NM_010789	Myeloid ecotropic viral integration site 1, mRNA	1.00	1.62	3.34
NM_145584	spondin 1, (f-spondin) extracellular matrix protein	1.00	1.17	3.33
NM_021527	McKusick-Kaufman syndrome protein	1.00	1.63	3.31
NM_009953	corticotropin releasing hormone receptor 2	1.05	1.00	3.29
NM_019411	Protein phosphatase 2a, catalytic subunit, alpha isoform, mRNA	1.17	1.00	3.28
NM_020295	limb region 1	1.06	1.00	3.22
NM_016798	PDZ and LIM domain 3 (Pdlim3), mRNA	1.00	1.09	3.20
NM_022433	sirtuin 3 (silent mating type information regulation 2, homolog) 3 (S. cerevisiae)	1.22	1.00	3.14
NM_175473	Fras1 protein	1.00	1.53	3.14
NM_010875	neural cell adhesion molecule 1	1.00	1.09	3.09
NM_026163	Plakophilin 2 (Pkp2), mRNA	1.00	1.08	2.99
NM_010017	dystroglycan 1	1.00	1.43	2.94
NM_009367	Transforming growth factor, beta 2, mRNA	1.00	1.11	2.94
NM_009274	serine/arginine-rich protein specific kinase 2	1.22	1.00	2.88
NM_139001	chondroitin sulfate proteoglycan 4	1.00	1.40	2.85
NM_007935	enhancer of polycomb homolog 1 (Drosophila)	1.00	1.09	2.84
NM_009762	SET and MYND domain containing 1	1.18	1.00	2.78
NM_009824	core-binding factor, runt domain, alpha subunit 2, translocated to, 3 homolog (human)	1.00	1.27	2.78
NM_008398	integrin alpha 7	1.03	1.00	2.72
NM_007462	Adenomatosis polyposis coli (Apc), mRNA	1.00	1.19	2.71

Table 8c. Development related transcripts specifically expressed in αMHC$^+$ cardiomyocytes (continued). The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

NCBI ID	Gene Name	ES cells	15day control EBs	α–MHC$^+$
XM_619767	Similar to hypothetical protein FLJ39502	1.00	1.00	2.61
NM_172868	paralemmin 2	1.01	1.00	2.60
NM_008844	phosphatidylinositol-4-phosphate 5-kinase, type 1 gamma	1.00	1.23	2.60
NM_008854	protein kinase, cAMP dependent, catalytic, alpha	1.00	1.29	2.58
NM_022305	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 1 (B4galt1), mRNA	1.00	1.21	2.48
NM_008601	microphthalmia-associated transcription factor	1.22	1.00	2.46
NM_027326	Myeloid/lymphoid or mixed lineage-leukemia translocation to 3 homolog (Drosophila) (Mllt3), transcript variant 1, mRNA	1.00	1.10	2.46
NM_183428	erythrocyte protein band 4.1	1.00	1.07	2.46
NM_172416	osteopetrosis associated transmembrane protein 1	1.13	1.00	2.42
NM_173182	MKIAA4164 protein	1.13	1.00	2.38
NM_007428	angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	1.09	1.00	2.32
NM_146104	anterior pharynx defective 1a homolog (C. elegans)	1.00	1.13	2.28
NM_023505	glutaredoxin 2 (thioltransferase)	1.00	1.08	2.27
XM_125745	integrin beta 1 binding protein 3	1.00	1.06	2.24
NM_024200	mitofusin 1	1.01	1.00	2.22
NM_019408	nuclear factor of kappa light polypeptide gene enhancer in B-cells 2, p49/p100	1.00	1.00	2.07
---	Transcribed locus	1.00	1.01	2.04

Downregulated Transcripts in αMHC⁺ cardiomyocytes than both ES cells and 6day control EBs

There are 1444 unique transcripts downregulated in αMHC⁺ cardiomyocytes than both ES cells and 15 day control EBs. Of them, there are 939 transcripts uniquely downregulated in αMHC⁺ cardiomyocytes cells which do not show any differential expression among ES cells and 15 day EBs. Those transcripts categorized according to the biological process they mediate are shown in figure 65. Apart from that, 494 transcripts are downregulated than both ES cells and 6d EBs but in turn show differential expression among ES cells. The developmental related transcripts are listed in annexure 16.

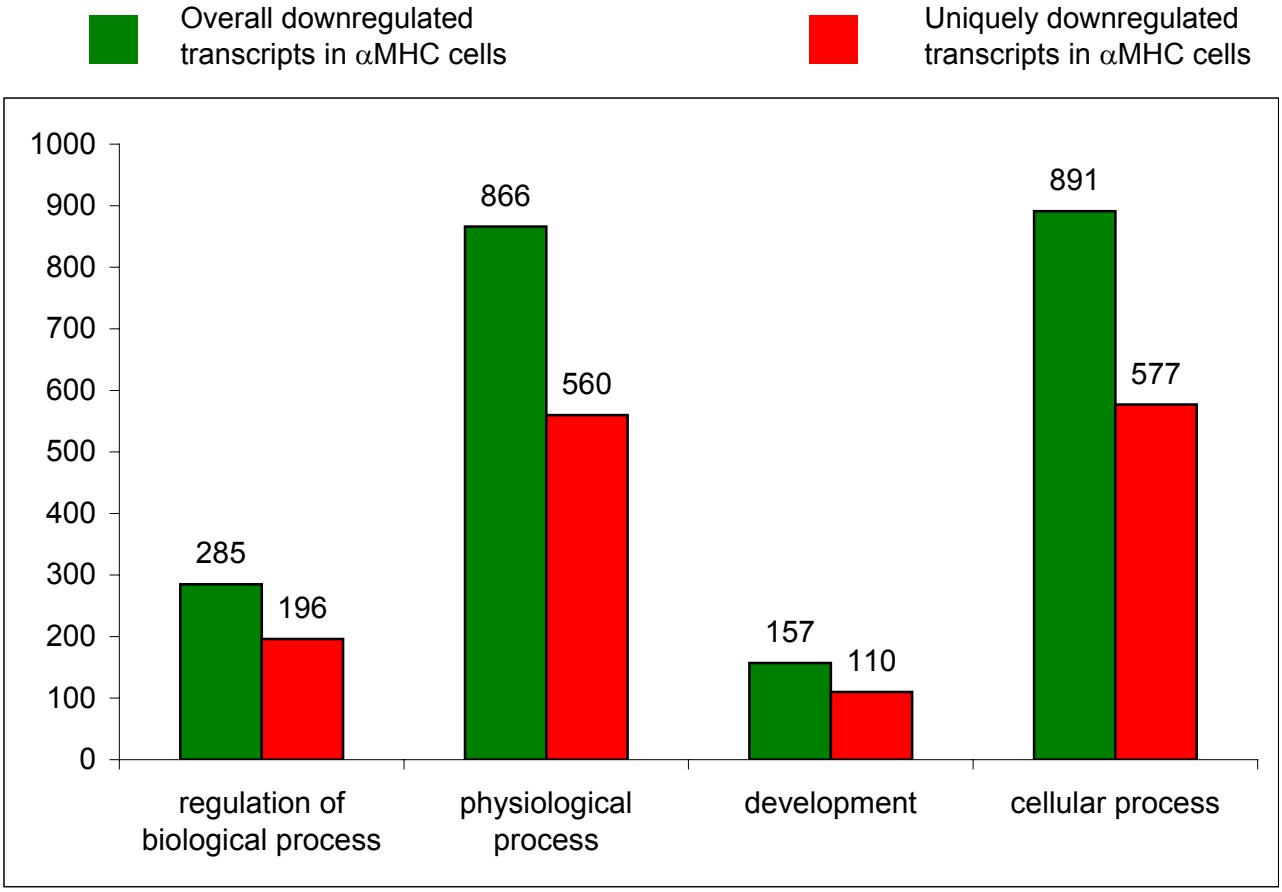


Figure 65. The major biological processes overrepresented by the transcripts downregulated than both ES cells and 15day EBs (green bar) and unique αMHC⁺ cardiomyocytes specific transcriptome downregulated in comparison to ES cells and 15-day cardiomyocytes (in red bar).

Transcripts upregulated in αMHC cardiomyocytes than ES cells but downregulated than 15 day control EBs

243 unique transcripts were found to be upregulated in αMHC cardiomyocytes compared to ES cells but down regulated in comparison to 15day control EBs. Among the 243 transcripts, 56 transcripts are developmentally important genes as shown in the annexure 17.

The transcripts upregulated in αMHC cardiomyocytes than 15 day control EBs but downregulated than ES cells.

There are 22 unique transcripts that are upregulated in αMHC^{+} cells compared to 15-day EBs cells but downregulated in comparison to ES cells (annexure 18).

4.6 FUNCTIONAL EVALUATION OF NOVEL TRANSCRIPTS

4.6.1 Experimental approach

Three novel candidate genes whose functions have not yet been annotated but highly enriched in $BMP\text{-}2^{+}$ cell lineages were chosen after assessing the pattern of transcript expression in differentiating EBs. Since the differentiating EBs can not be transfected efficiently with siRNA with non-viral delivery methods due to their multicellular 3 dimensional aggregations, the transcripts that are expressed even in the undifferentiated pluripotent state were specifically chosen. These transcripts were functionally evaluated by a systematic “loss of or gain of function“ studies using siRNA knockdown approaches and episomal overexpression systems.

The T Brachyury ES cell clone were transfected with siRNAs targeted to each of these novel transcripts. At 48 hours post transfection, the transfected ES cells were used for making “hanging drop” EBs and followed up. RNA were isolated at 24 hours post transfection, and just before making hanging drop EBs (48 hours post transfection) and then at regular intervals at days 3, 5 and 7, the time points at which T Brachyury, BMP-2 and αMHC peak their transcripts expression respectively (Figure 66) and on day 9. The RNA isolated were analyzed for magnitude of expression by the germ layer representative and other genes, *Nefh* (ectodermal marker), *T brachyury* (mesodermal marker), *BMP-2* (mesodermal marker), *Sox17* (endodermal marker), *Nanog (*pluripotent ES cells marker) and *αMHC* (cardiomyocyte marker, actual gene name is *Myh6*). siRNA transfected ES cells were monitored for any morphological changes associated with any differentiated cell types. On day 4, T Brachyury driven EGFP was quantified by FACS analysis as an index of mesodermal cells formed within the siRNA transfected EBs.

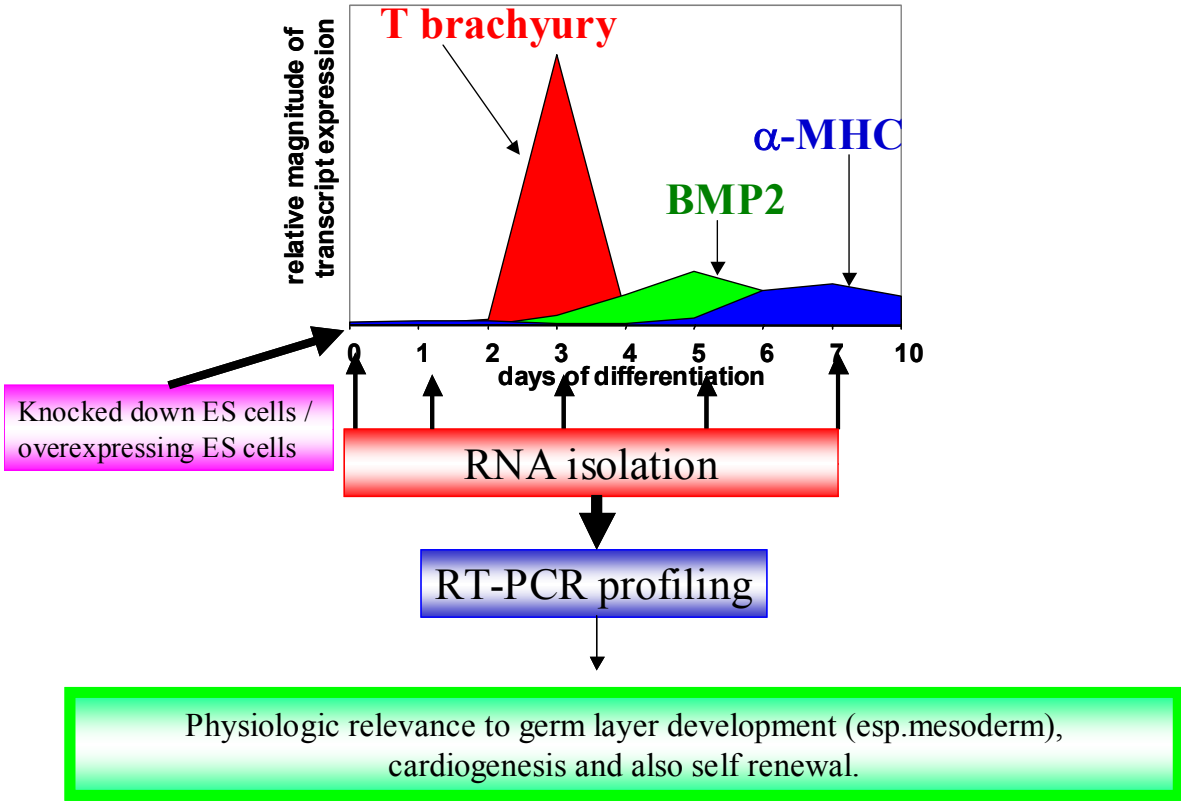


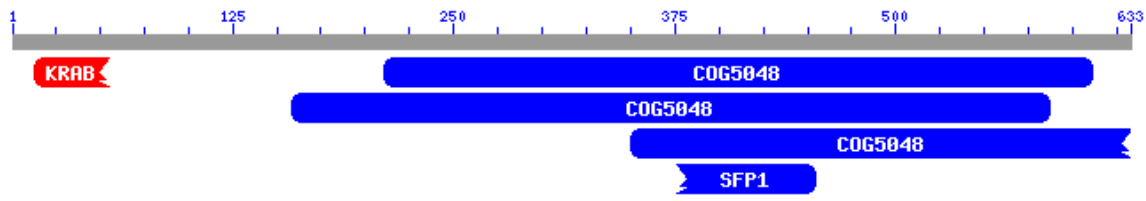
Figure 66 Overview of the experimental approach to functionally evaluate the novel transcripts using ES cells with siRNA knock down ("loss of function") and "gain of function" approaches.

4.6.2 The candidate genes- overview and results

1. MGC117846 (Similar to zinc finger protein 665)

This transcript is a putative transcription factor (for NCBI details, please see annexure 19) . It contains the single conserved domain each of krueppel associated box and SFP1 (Putative transcriptional repressor regulating G2/M transition) and multiple conserved domains of Zn-finger as shown in figure 67. During the early embryonic development, this transcript expression is developmentally regulated as seen in figure 70. In the ES cell/EBs system, this transcript is expressed at highest level in the undifferentiated ES cells and as its expression start to decline, the BMP-2 transcript start to be expressed (Figure 68). The fold changes in comparison to ES cells and the respective control EBs as obtained from the transcriptomic profiling of T Brachyury and BMP-2 lineages cells are shown in the figure 69. So, being a transcription factor expressed in pre-implantation stage and mid-gestation embryos and its downregulation precedes BMP-2 transcript expression, it may be likely that this transcript may have an influence on BMP-2 expression since this transcript is highly expressed in ES cell-derived BMP-2^{+} lineage cells. Hence this candidate was chosen for the functional evaluation to check whether this transcript has any inhibitory effect

on BMP-2 expression. The nucleotide and protein sequence information are shown in the



KRAB = krueppel associated box;
COG5048 = FOG: Zn-finger [General function prediction only].
SFP1 =Putative transcriptional repressor regulating G2/M transition

Figure 67 Conservation domains contained in the MGC117846.

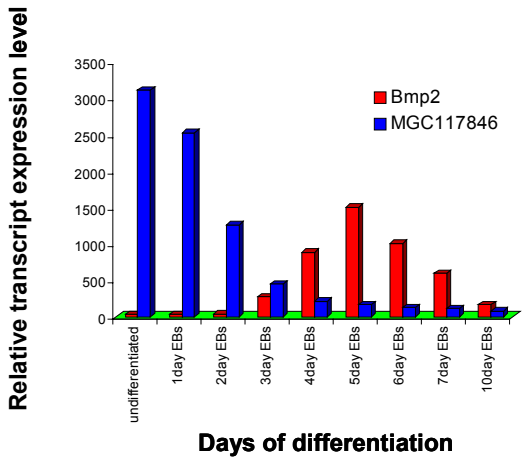


Figure 68 Expression of MGC117846 transcript during the course of differentiation.

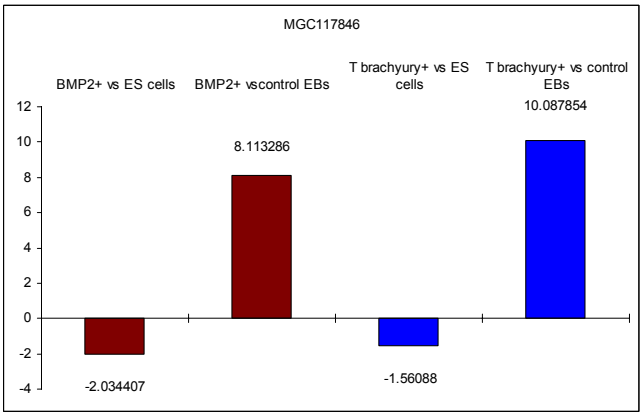


Figure 69. Fold change of the *MGC117846* transcript in pure T Brachyury and BMP2 mesodermal cells in comparison to ES cells and the control same day EBs as obtained from transcriptomic profiling.

A

Breakdown by Developmental Stage

	Mm.270762	
oocyte	0	0 / 20028
unfertilized ovum	0	0 / 20660
zygote	0	0 / 27827
pre-implantation embryo	19	3 / 150475
post implantation embryo	0	0 / 43247
mid-gestation embryo	45	8 / 176010
late gestation embryo	0	0 / 88290
fetus	3	2 / 608632
neonate	0	0 / 106970
juvenile	3	1 / 299617
adult	0	0 / 1026000

B

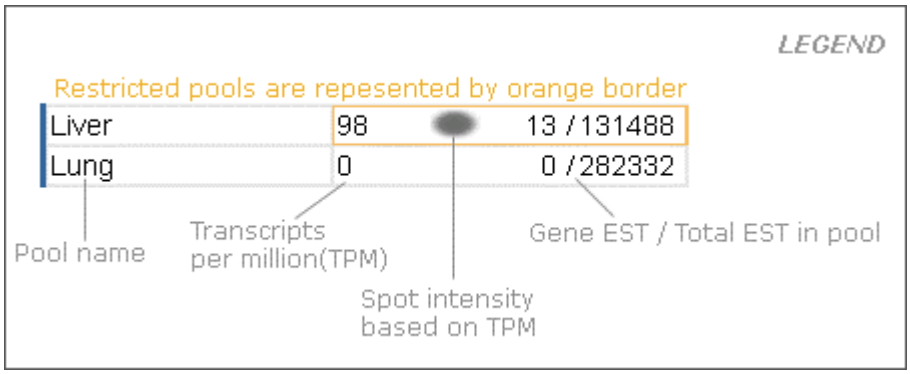


Figure 70. A, Developmental stage specific expression of *MGC117846* transcript. B, Legend.

siRNA knockdown:

Knockdown by siRNA targeted against MGC117846 did not cause any significant effect as shown in the figure 71. There were no morphological differences observed or growth rate.

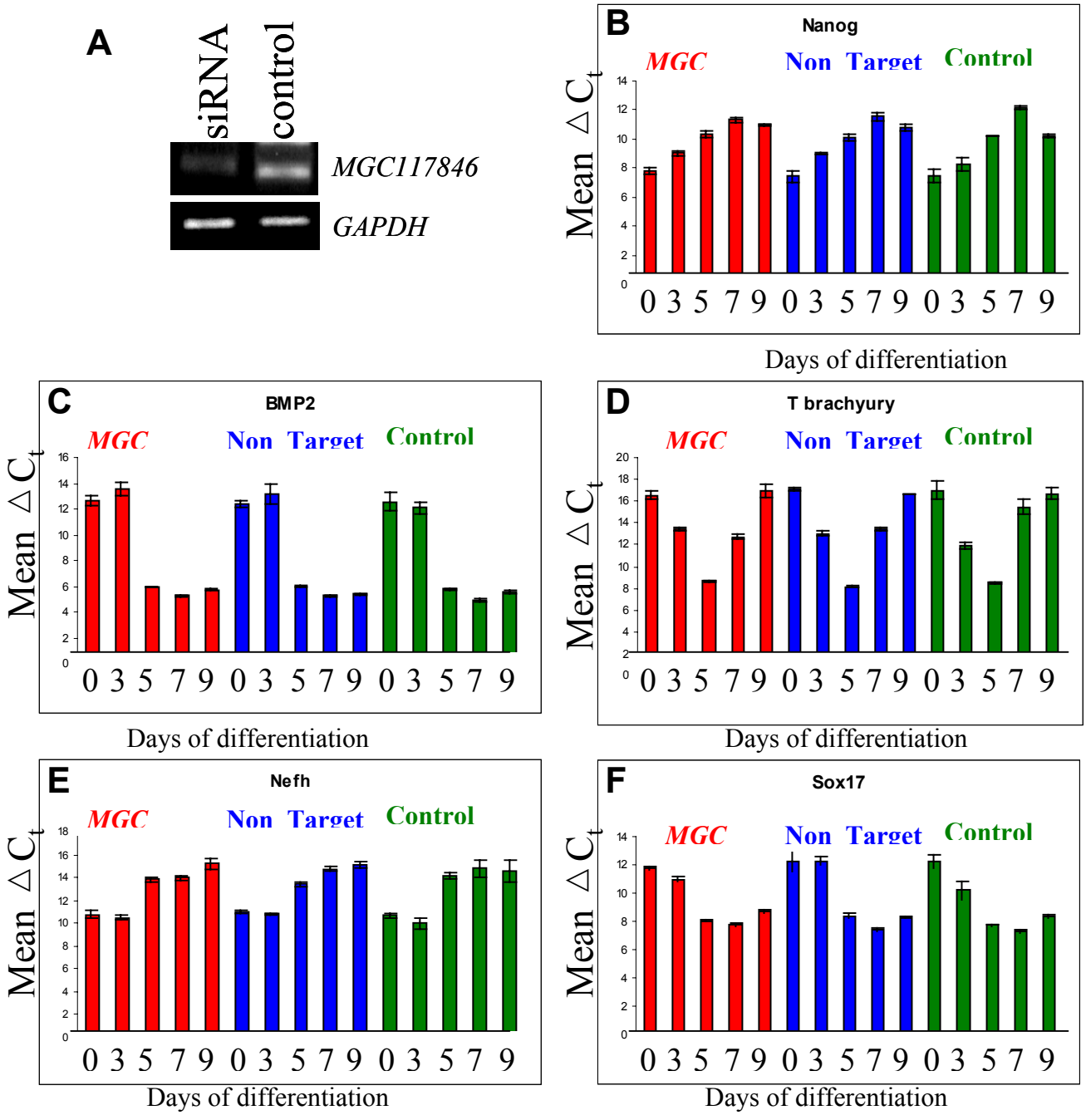


Figure 71. A, RT-PCR analysis of siRNA mediated knockdown of *MGC117846* in ES cells 24 hours post siRNA transfection. (B-F) Quantitative PCR analysis of the key developmental markers in the siRNA transfected samples targeted against MGC117846 and the non-target genes and the control samples that were not treated with siRNA or Lipofectamine 2000.

1700108L22Rik

This is RNA binding protein that is developmentally regulated in early embryonic stages as shown in figure 74. The NCBI details for this gene are presented in annexure 20. This transcript is enriched in pre-implantation embryos. The function of this gene has not yet been reported even though it is guessed from the LA domain contained in this gene that this codes for a protein which binds to RNA. Generally La protein plays a role in the transcription of RNA polymerase III. It is most probably a transcription termination factor. Binds to the 3' termini of virtually all-nascent polymerase III transcripts. It is associated with precursor forms of RNA polymerase III transcripts including tRNA and 4.5S, 5S, 7S, and 7-2 RNAs (By similarity). Since this transcript is temporally and spatially regulated during development and also upregulated in the T brachyury and BMP-2 cell lineages, this gene was chosen assuming that the rest of the protein apart from the La domain may be playing any role especially modulating the BMP-2 transcription as shown in the Figure It is interesting to note pattern of this transcript in comparison to BMP-2 transcripts that the downregulation of this transcript is coupled with the up regulation of BMP-2 transcript.

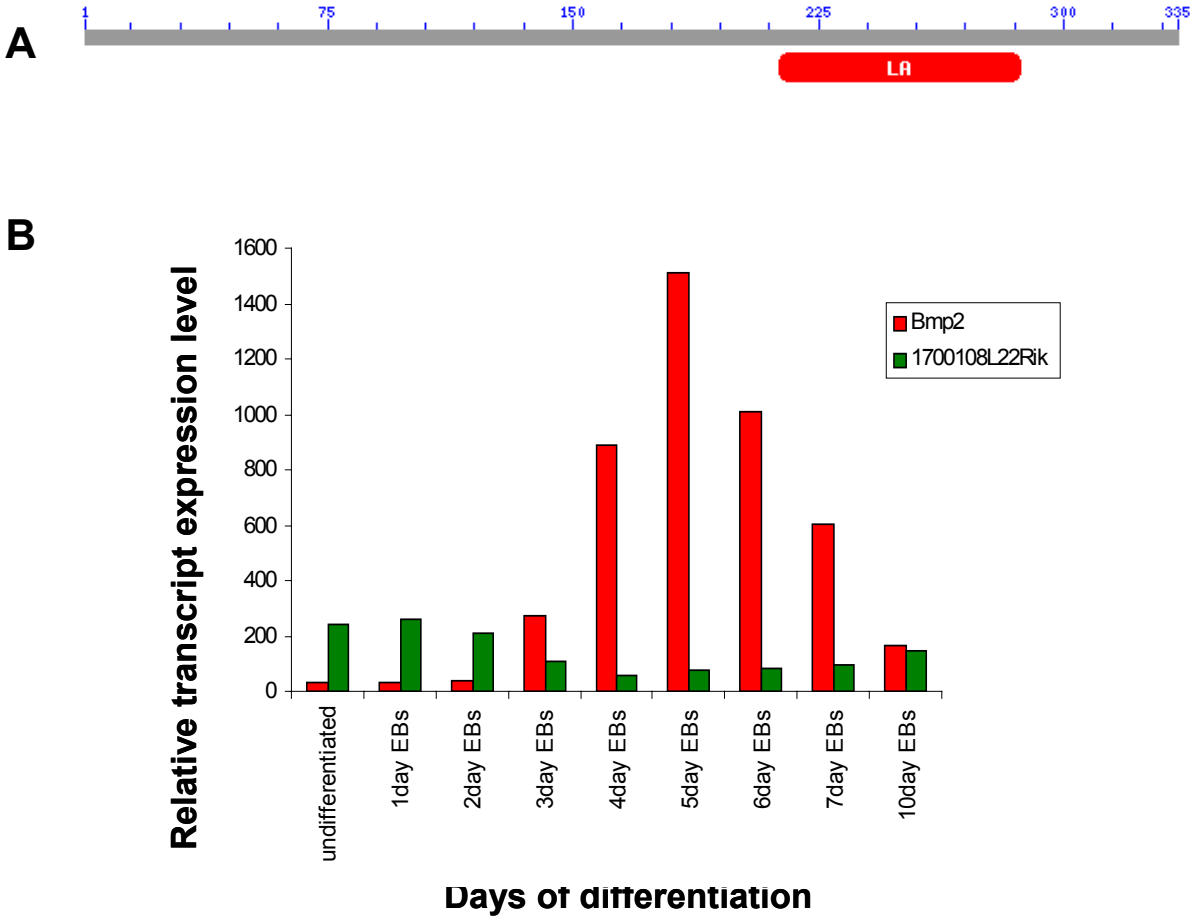


Figure 72 A. Conserved LA domain contained in the *1700108L22Rik.* B., Expression of *1700108L22Rik* transcript during the course of differentiation.

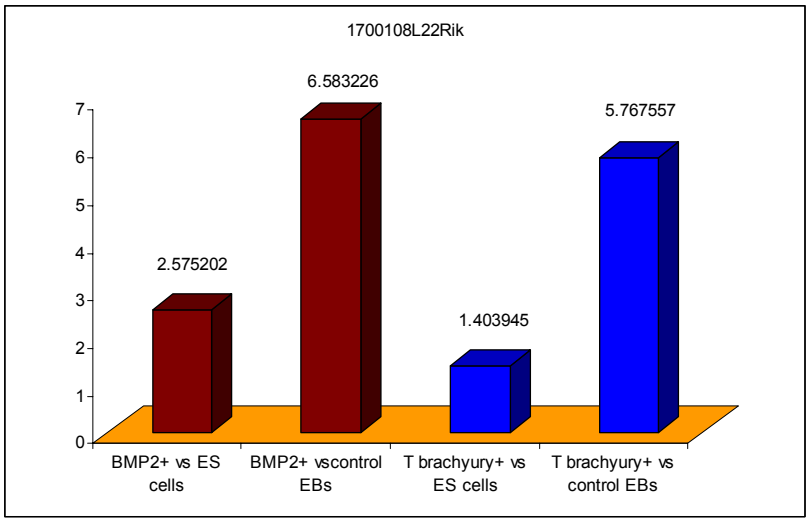


Figure 73 Fold change of the *1700108L22Rik* transcript in pure T Brachyury and BMP2 mesodermal cells in comparison to ES cells and the control same day EBs as obtained from transcriptomic profiling.

A

Breakdown by Developmental Stage

	Mm.46782	
oocyte	49	1 / 20028
unfertilized ovum	0	0 / 20660
zygote	0	0 / 27827
pre-implantation embryo	79	12 / 150475
post implantation embryo	0	0 / 43247
mid-gestation embryo	11	2 / 176010
late gestation embryo	0	0 / 88290
fetus	6	4 / 608632
neonate	9	1 / 106970
juvenile	30	9 / 299617
adult	42	44 / 1026000

B

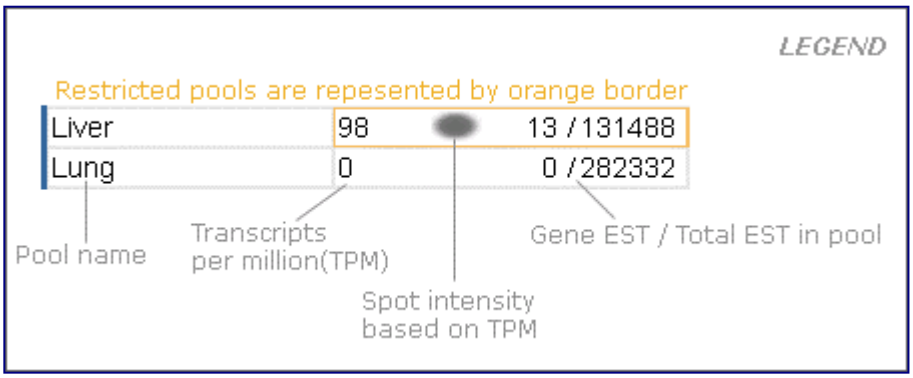


Figure 74. A., Developmental stage specific expression of *1700108L22Rik* transcript. **B.**, Legend.

siRNA knockdown of 1700108L22Rik

Knockdown by siRNA targeted against *1700108L22Rik* did cause a down regulation of BMP-2 transcripts on day 3. Apart from that, it did not cause any significant effect as shown in the figure 75. There were no morphological differences observed. The down regulation of the BMP-2 transcript by the knockdown of the 1700108L22Rik shows a positive interaction between these 2 transcripts. But in comparison to the time kinetics in differentiating EBs, there is reciprocal interaction may be inferred. This needs to be addressed further since there are several modes of auto-regulatory mechanisms existing. But there were no effects when over-expressed. So, this has to be addressed this contrary situation.

Episomal overexpression of 1700108L22Rik

Full Length cDNA Clone IRAVp968A0470D encoding the *1700108L22Rik* was purchased from RZPD, Germany and the *1700108L22Rik* cDNA was subcloned into the pPyCAGIP over expression plasmid. The resulting plasmid was transfected into T Brachyury ES cell clone and generated the overexpression clone. Hanging drop EBs made from the EBs were analyzed in terms of the lineage specific transcript levels including BMP-2. There was no significant effect was observed as shown in figure 76.

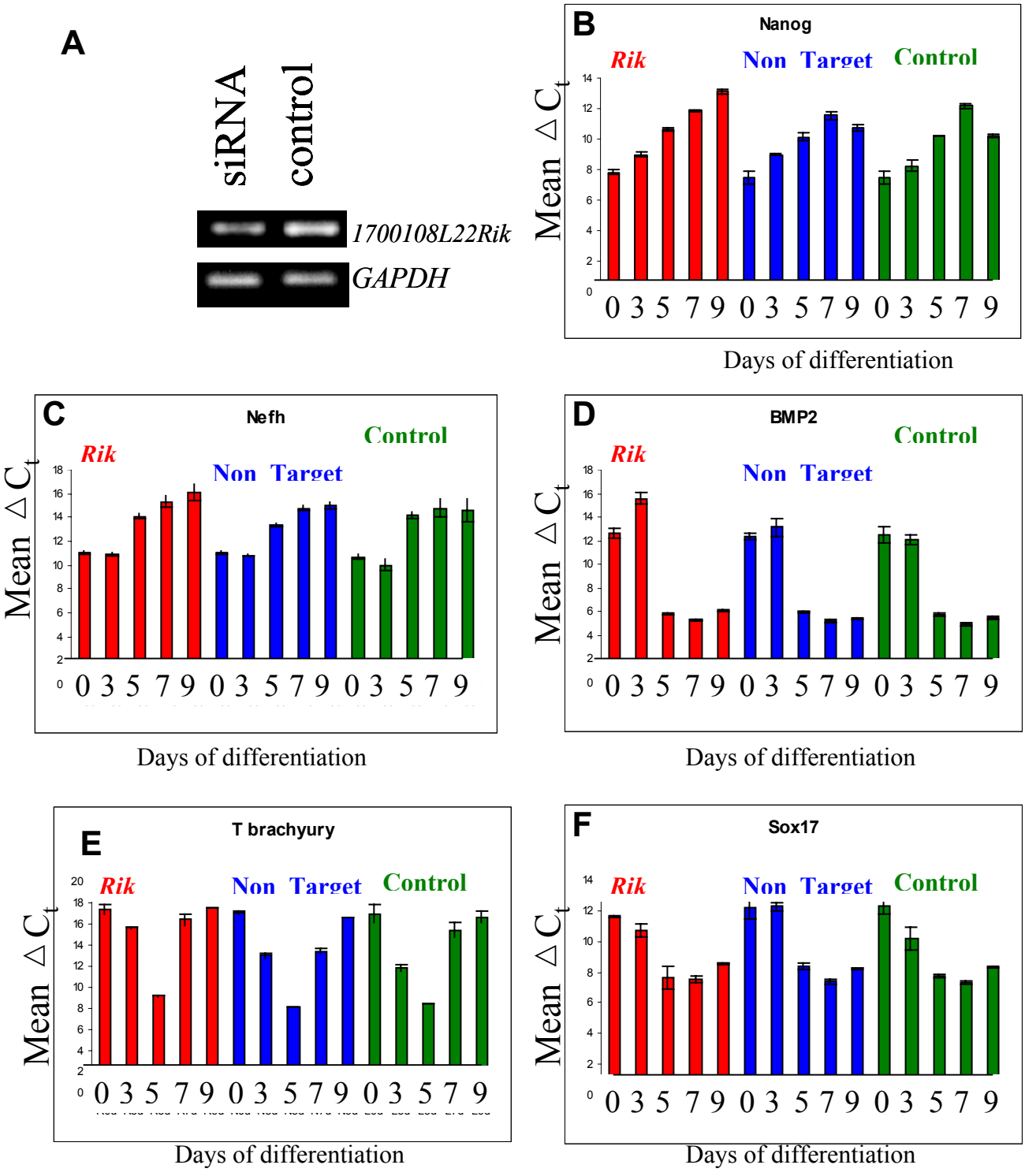


Figure 75. A., RT-PCR analysis of siRNA mediated knockdown of *1700108L22Rik* in ES cells 24 hours post siRNA transfection. (B-F) Quantitative PCR analysis of the key developmental markers in the siRNA transfected samples targeted against *1700108L22Rik* and the Non-target genes and the control samples that were not treated with siRNA or Lipofectamine 2000.

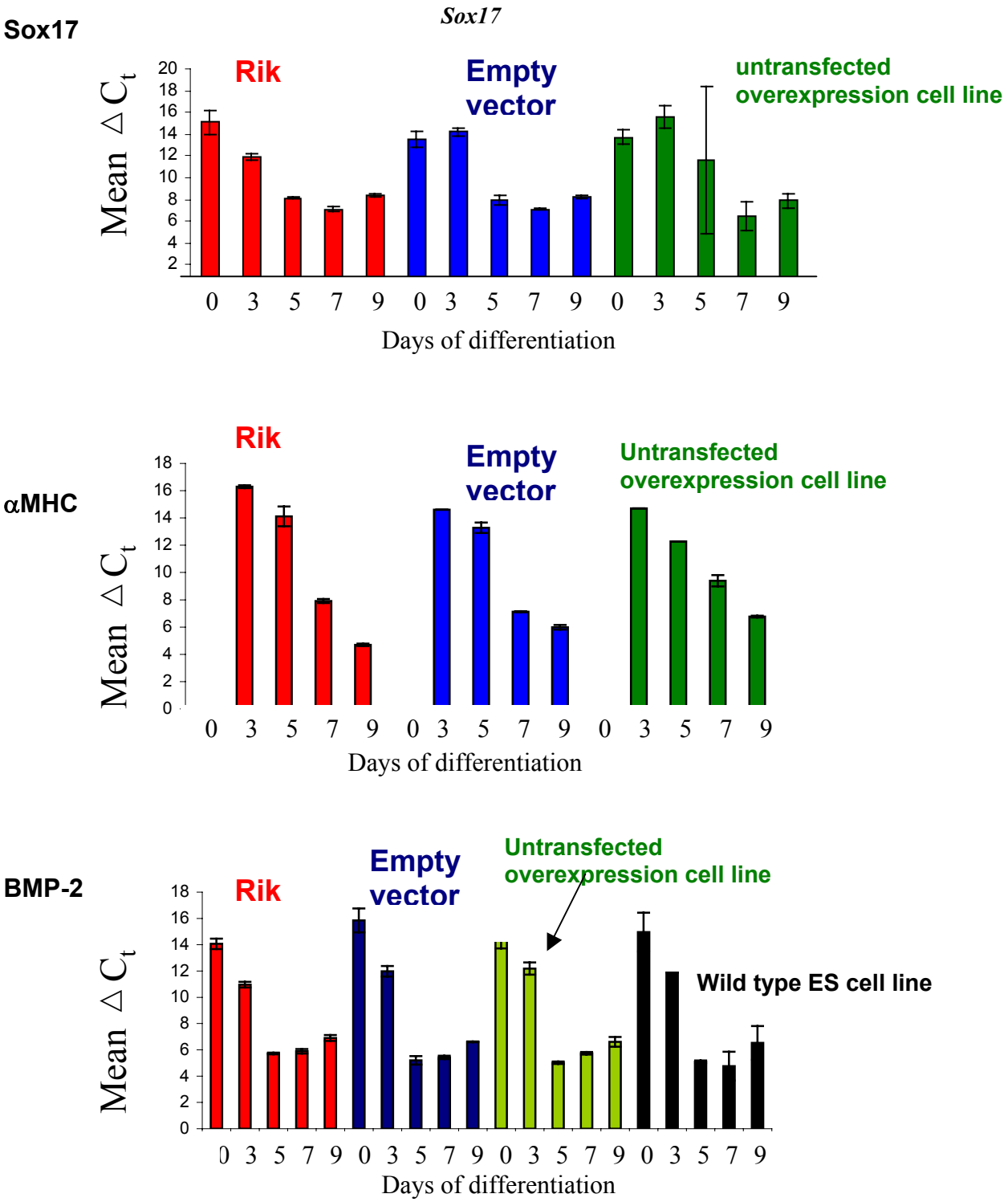


Figure 76. Quantitative PCR analysis of the key developmental markers in the episomally 1700108L22Rik overexpressing differentiating EBs.

Gm397

It is a novel transcript predicted by electronical models as shown in annexure 21. It is a very interesting gene since its expression is uniquely restricted to pre-implantation embryo and pluripotent undifferentiated ES cells (Figure 77 And 78). This gene was selected under the assumption that it might be involved in the maintenance of ES cell pluripotency. It is still a hypothetical protein, even then the transcripts are expressed and its mRNA sequence information is known. The protein has one leucine rich domain and shows zinc ion binding sites and electronically annotated to be a transcription factor. But in the development context, the function of this gene is still unknown even though they are enriched in the ES cells. This transcript is highly enriched in T Brachyury and BMP-2 lineage cells, unlike Pou5f1 *(Oct4),* the pluripotent associated marker.

A

B

Breakdown by Developmental Stage

Mm.131747

oocyte	0	0 / 20033
unfertilized ovum	0	0 / 20726
zygote	0	0 / 29187
pre-implantation embryo	26	4 / 151052
post implantation embryo	0	0 / 44710
mid-gestation embryo	0	0 / 182484
late gestation embryo	0	0 / 88308
fetus	0	0 / 630414
neonate	0	0 / 111715
juvenile	0	0 / 292578
adult	0	0 / 1060679

C

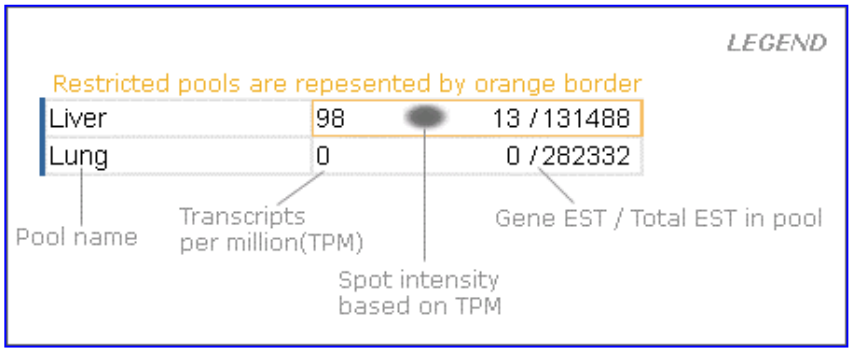


Figure 77. A., Conserved domains contained in the *1700108L22Rik* **B.,** Developmental stage specific expression of *1700108L22Rik* transcript. **C.**, Legend for B.

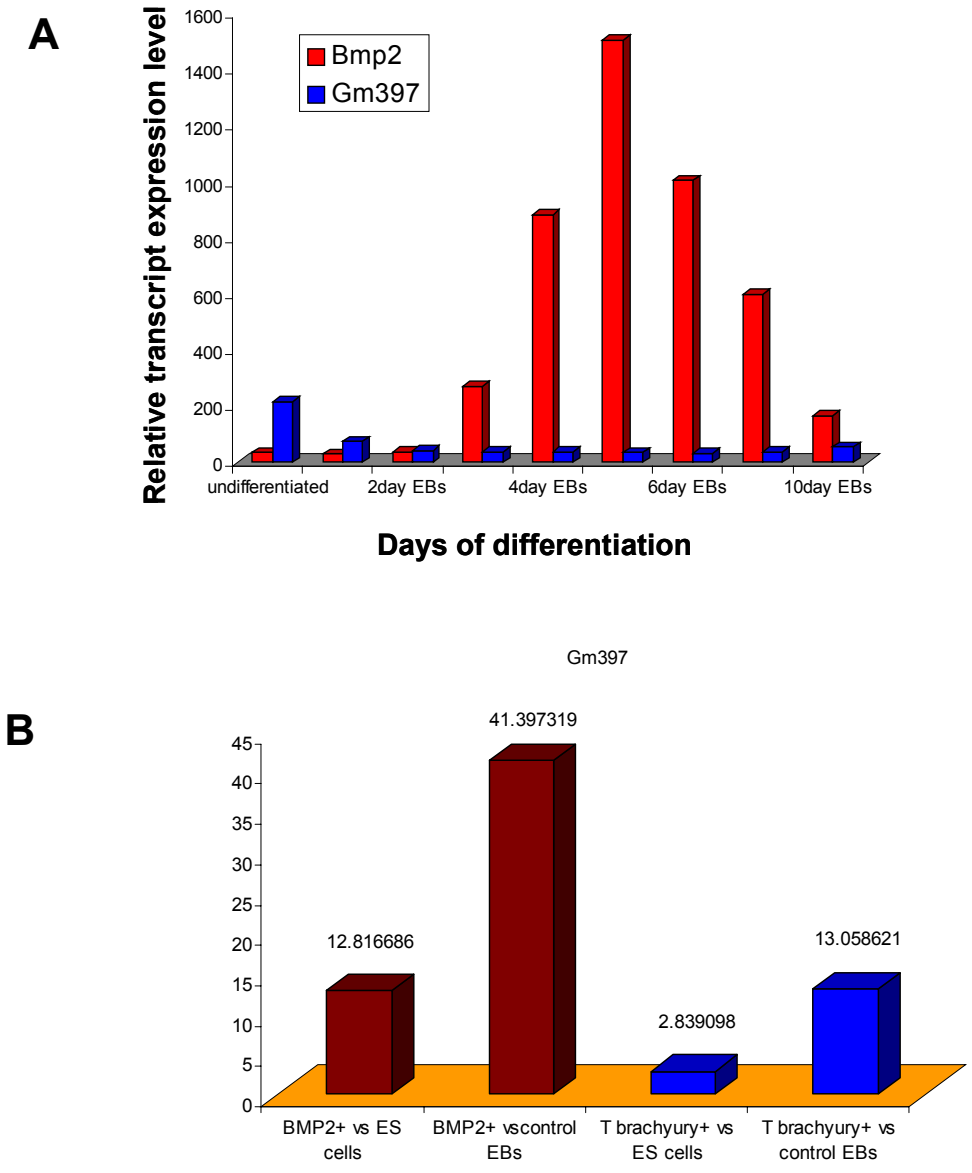


Figure.78 A., Expression of *Gm397* transcript during the course of differentiation. **B.,** Fold change of the *Gm397* transcript in pure T Brachyury and BMP2 mesodermal cells in comparison to ES cells and the control same day EBs as obtained from transcriptomic profiling.

siRNA knockdown:

Knockdown by siRNA targeted against *Gm397* did not cause any significant effect as shown in the figure 79. There were no morphological differences observed.

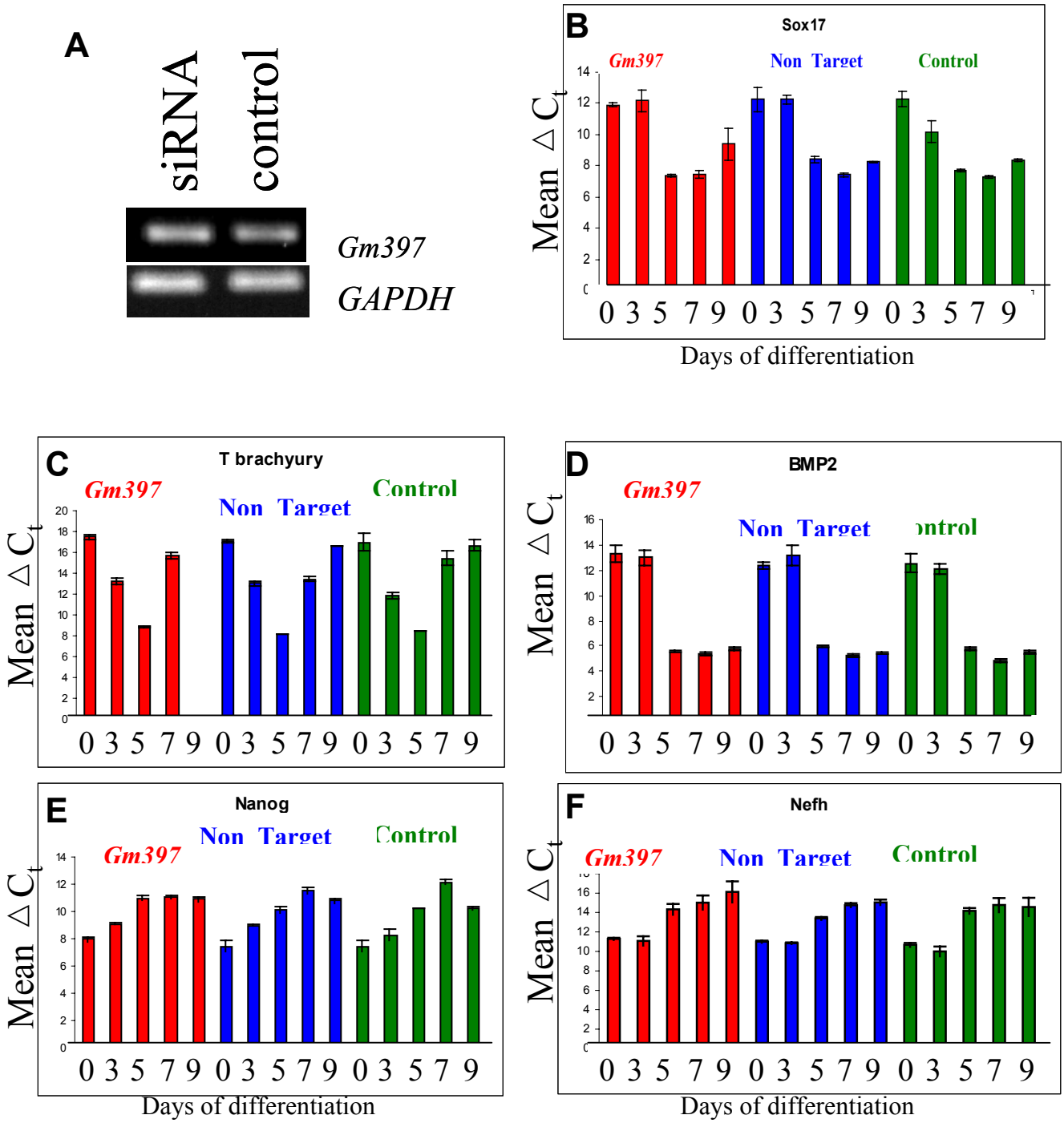


Figure 79. A., RT-PCR analysis of siRNA mediated knockdown of *Gm397* in ES cells 24 hours post siRNA transfection. (B-F) Quantitative PCR analysis of the key developmental markers in the siRNA transfected samples targeted against *Gm397* and the non-target genes and the control samples which were not treated with siRNA or Lipofectamine 2000.

Overexpression

Full Length cDNA Clone IRAKp961F07111Q encoding the Gm397 was purchased from RZPD, German and the Gm397 cDNA was subcloned into the pPyCAGIP over expression plasmid. The resulting plasmid was transfected into T Brachyury ES cell clone and generated the overexpression clone. Hanging drop EBs made from the Gm397 overexpressing ES cell clones were analyzed in terms of the lineage specific transcript levels including BMP-2. There was no significant effect was observed as shown in Figure 80.

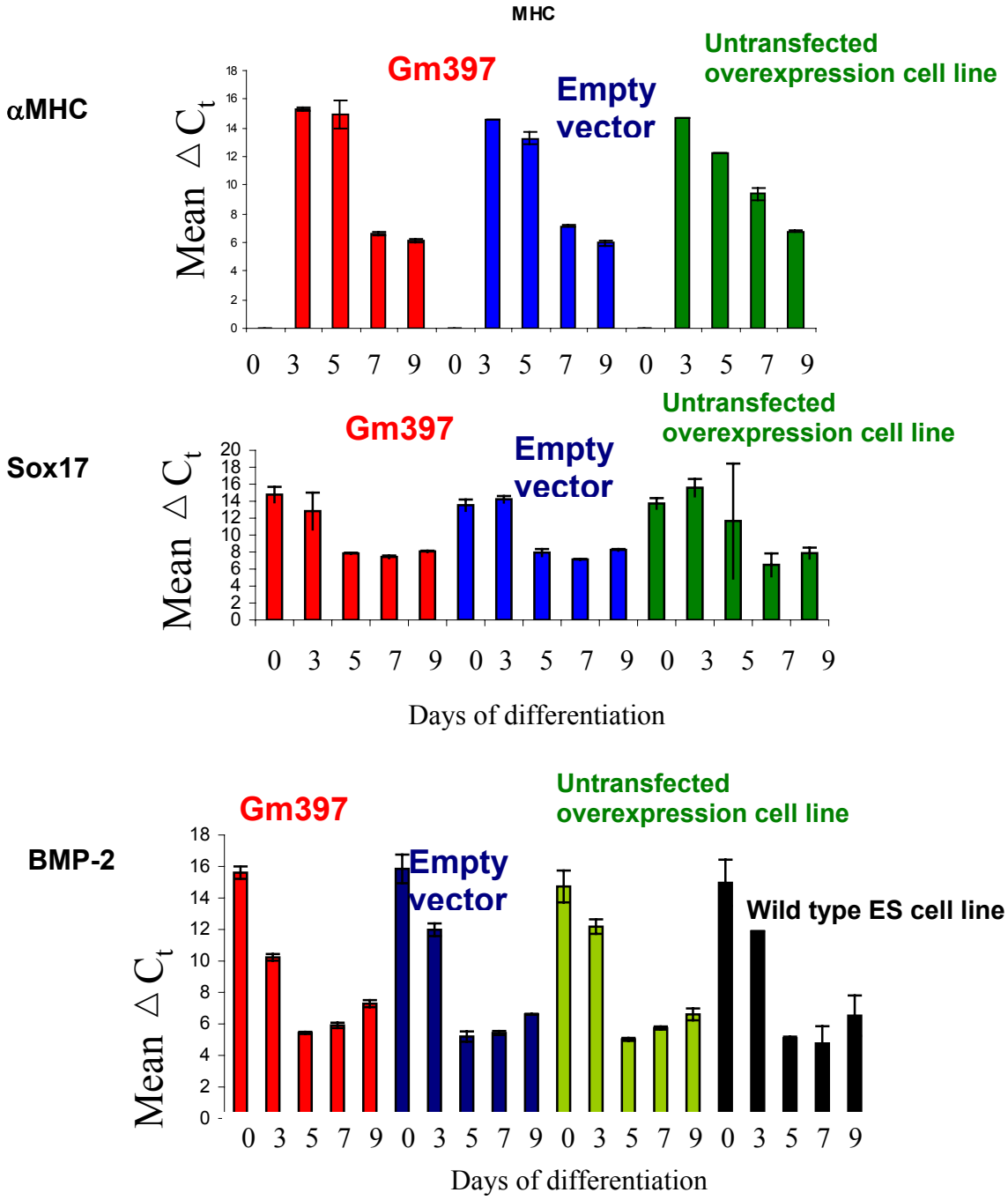


Figure 80. Quantitative PCR analysis of the key developmental markers in the episomally Gm397 overexpressing differentiating EBs with controls.

5.Conclusion and Outlook

Conclusion and Outlook

The transcriptomic analyses of the mesodermal cells and cardiomyocytes have defined unique cellular and genetics identities of the respective lineages towards understanding the inducible fate of mesodermal patterning and predicting to some extent how the prospective mesodermal cells integrate the various signals they receive and how they resolve this information to regulate their morphogenetic behaviors and cell-fate decisions. They in turn significantly contribute in building lineage specific transcriptomic atlas, which will be of immense application for the developmental biologists and also for promoting the protocols for the derivation of selective lineages for the future cell replacement therapy.

The isolated T Brachyury^{+} and BMP-2^{+} cells are able to prime themselves to become cardiomyocytes. While the ES cell-derived T Brachyury lineage also gives rise to endodermal lineages, the BMP-2^{+} lineages are completely devoid of any endodermal derivatives. In addition, early BMP-2^{+} lineages exhibit neural crest stem cell properties. Thus, while T Brachyury cells represent lineage non-committed early stage mesodermal cells, BMP-2 cells represent the late stage mesodermal population that can autonomously give rise to cardiomyocytes and are already lineage committed. Comparison of both the lineages points out that they exhibit certain plasticity at the early stage. Even the T Brachyury^{+} cells can be maintained in culture for longer time with optimal amount of Leukemia inhibitor factors (LIF) and simultaneous puromycin selection. The EBs made from these cultures gave rise to contracting cardiomyocytes, even though the process of cardiomyogenesis was delayed. This represents a model to study patterning of mesodermal cells into cardiomyocytes in a selective approach without starting from ES cells in which case, the interplay of the ectodermal derivatives in mesodermal patterning would make the investigation very complex. In addition, the transcriptomic analysis of T Brachyury cells characterizes the cellular and genetics identity of the mesodermal cells, which have not yet been reported till date. This would significantly complement the better understanding of the mesodermal patterning. In vitro ES derived BMP-2^{+} cell population contains neural crest stem cells and their derivatives-smooth muscle cells, melanocytes, cardiomyocytes and vascular and hematopoietic progenitors. Enrichment of the neural crest cells and their lineages by BMP-2 promoter driven selection marker paves the way for ES cell-derived neural crest stem cell biology. These cells are the attractive candidates for the future cell replacement therapy due their plasticity into any tissue depending upon the local environment existing within the injected tissue. This study enlists the cells expressing BMP-2 and gives an initial clue how these BMP-2 secreting cells play an important role during the development. The characterization of ES cell-derived cardiomyocytes proves that they

behave more or less similar to native *in vivo* cardiomyocytes in terms of the transcripts involved in physiological processes and other development processes.

The future approach will be to further analyze the transcripts to elucidate the underlying mechanisms using several approaches like finding the preferred transcription binding sites by most of the transcription factors to get a clue about the molecular details about the mesodermal patterning. The novel transcripts whose functions have not yet been annotated but enriched in these lineages have to be functionally decoded to fill up the missing links in the developmental process by loss of function and gain of function studies using ES cells and also *in vivo* zebra fish models. In addition, the T Brachyury^{+} cells can be used as starting culture for directing into cardiomyocytes to decipher the molecular mechanisms of cardiomyocytes more closely in an efficient way.

6. Bibliography

Bibliography

1. Slack,J.M. Inducing factors in Xenopus early embryos. *Curr. Biol.* **4**, 116-126 (1994).

2. Zernicka-Goetz,M. Developmental cell biology: cleavage pattern and emerging asymmetry of the mouse embryo. *Nat Rev Mol Cell Biol* **6**, 919-928 (2005).

3. Loebel,D.A., Watson,C.M., De Young,R.A. & Tam,P.P. Lineage choice and differentiation in mouse embryos and embryonic stem cells. *Dev. Biol.* **264**, 1-14 (2003).

4. Tam,P.P., Williams,E.A. & Chan,W.Y. Gastrulation in the mouse embryo: ultrastructural and molecular aspects of germ layer morphogenesis. *Microsc. Res. Tech.* **26**, 301-328 (1993).

5. Haub,O. & Goldfarb,M. Expression of the fibroblast growth factor-5 gene in the mouse embryo. *Development* **112**, 397-406 (1991).

6. Hebert,J.M., Boyle,M. & Martin,G.R. mRNA localization studies suggest that murine FGF-5 plays a role in gastrulation. *Development* **112**, 407-415 (1991).

7. Pelton,T.A., Sharma,S., Schulz,T.C., Rathjen,J. & Rathjen,P.D. Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. *J. Cell Sci.* **115**, 329-339 (2002).

8. Rogers,M.B., Hosler,B.A. & Gudas,L.J. Specific expression of a retinoic acid-regulated, zinc-finger gene, Rex-1, in preimplantation embryos, trophoblast and spermatocytes. *Development* **113**, 815-824 (1991).

9. Hogan,B.L. & Tilly,R. Cell interactions and endoderm differentiation in cultured mouse embryos. *J. Embryol. Exp. Morphol.* **62**, 379-394 (1981).

10. Tam,P.P. & Behringer,R.R. Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev.* **68**, 3-25 (1997).

11. Tam,P.P., Goldman,D., Camus,A. & Schoenwolf,G.C. Early events of somitogenesis in higher vertebrates: allocation of precursor cells during gastrulation and the organization of a meristic pattern in the paraxial mesoderm. *Curr. Top. Dev. Biol.* **47**, 1-32 (2000).

12. Gadue,P., Huber,T.L., Nostro,M.C., Kattman,S. & Keller,G.M. Germ layer induction from embryonic stem cells. *Exp Hematol.* **33**, 955-964 (2005).

13. Rodda,S.J., Kavanagh,S.J., Rathjen,J. & Rathjen,P.D. Embryonic stem cell differentiation and the analysis of mammalian development. *Int. J. Dev. Biol.* **46**, 449-458 (2002).

14. Snow,M.H. & Bennett,D. Gastrulation in the mouse: assessment of cell populations in the epiblast of tw18/tw18 embryos. *J. Embryol. Exp. Morphol.* **47**, 39-52 (1978).

15. Rodda,S.J., Kavanagh,S.J., Rathjen,J. & Rathjen,P.D. Embryonic stem cell differentiation and the analysis of mammalian development. *International Journal of Developmental Biology* **46**, 449-458 (2002).

16. Palmieri,S.L., Peter,W., Hess,H. & Scholer,H.R. Oct-4 transcription factor is differentially expressed in the mouse embryo during establishment of the first two extraembryonic cell lineages involved in implantation. *Dev. Biol.* **166**, 259-267 (1994).

17. Nichols,J. *et al.* Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379-391 (1998).

18. Niwa,H., Miyazaki,J. & Smith,A.G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.* **24**, 372-376 (2000).

19. Voss,A.K. *et al.* Taube nuss is a novel gene essential for the survival of pluripotent cells of early mouse embryos. *Development* **127**, 5449-5461 (2000).

20. Russ,A.P. *et al.* Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* **404**, 95-99 (2000).

21. Rossant,J. & McKerlie,C. Mouse-based phenogenomics for modelling human disease. *Trends Mol Med* **7**, 502-507 (2001).

22. Nagy,A., Rossant,J., Nagy,R., bramow-Newerly,W. & Roder,J.C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. U. S. A* **90**, 8424-8428 (1993).

23. Tanaka,M. *et al.* Vertebrate homologs of tinman and bagpipe: roles of the homeobox genes in cardiovascular development. *Dev. Genet.* **22**, 239-249 (1998).

24. Yan,J. *et al.* Retinoic acid promotes differentiation of trophoblast stem cells to a giant cell fate. *Dev Biol* **235**, 422-432 (2001).

25. Watson,C.M. & Tam,P.P. Cell lineage determination in the mouse. *Cell Struct. Funct.* **26**, 123-129 (2001).

26. Herrmann,B.G. Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. *Development* **113**, 913-917 (1991).

27. Kispert,A. & Herrmann,B.G. Immunohistochemical analysis of the Brachyury protein in wild-type and mutant mouse embryos. *Dev. Biol.* **161**, 179-193 (1994).

28. Kinder,S.J. *et al.* The orderly allocation of mesodermal cells to the extraembryonic structures and the anteroposterior axis during gastrulation of the mouse embryo. *Development* **126**, 4691-4701 (1999).

29. Fehling,H.J. *et al.* Tracking mesoderm induction and its specification to the hemangioblast during embryonic stem cell differentiation. *Development* **130**, 4217-4227 (2003).

30. Yoshikawa,Y., Fujimori,T., McMahon,A.P. & Takada,S. Evidence that absence of Wnt-3a signaling promotes neuralization instead of paraxial mesoderm development in the mouse. *Dev. Biol.* **183**, 234-242 (1997).

31. Galceran,J., Farinas,I., Depew,M.J., Clevers,H. & Grosschedl,R. Wnt3a-/--like phenotype and limb deficiency in Lef1(-/-)Tcf1(-/-) mice. *Genes Dev.* **13**, 709-717 (1999).

32. Yamaguchi,T.P., Bradley,A., McMahon,A.P. & Jones,S. A Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* **126**, 1211-1223 (1999).

33. Galceran,J., Hsu,S.C. & Grosschedl,R. Rescue of a Wnt mutation by an activated form of LEF-1: regulation of maintenance but not initiation of Brachyury expression. *Proc. Natl. Acad. Sci U. S A* **98**, 8668-8673 (2001).

34. Chapman,D.L. & Papaioannou,V.E. Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature* **391**, 695-697 (1998).

35. Ciruna,B. & Rossant,J. FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. *Dev. Cell* **1**, 37-49 (2001).

36. Ciruna,B.G., Schwartz,L., Harpal,K., Yamaguchi,T.P. & Rossant,J. Chimeric analysis of fibroblast growth factor receptor-1 (Fgfr1) function: a role for FGFR1 in morphogenetic movement through the primitive streak. *Development* **124**, 2829-2841 (1997).

37. Sun,X., Meyers,E.N., Lewandoski,M. & Martin,G.R. Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev.* **13**, 1834-1846 (1999).

38. Yamaguchi,T.P., Harpal,K., Henkemeyer,M. & Rossant,J. fgfr-1 is required for embryonic growth and mesodermal patterning during mouse gastrulation. *Genes Dev.* **8**, 3032-3044 (1994).

39. Winnier,G., Blessing,M., Labosky,P.A. & Hogan,B.L. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev.* **9**, 2105-2116 (1995).

40. Wilson,V., Rashbass,P. & Beddington,R.S. Chimeric analysis of T (Brachyury) gene function. *Development* **117**, 1321-1331 (1993).

41. Mishina,Y., Suzuki,A., Ueno,N. & Behringer,R.R. Bmpr encodes a type I bone morphogenetic protein receptor that is essential for gastrulation during mouse embryogenesis. *Genes Dev.* **9**, 3027-3037 (1995).

42. Miyazawa,K., Shinozaki,M., Hara,T., Furuya,T. & Miyazono,K. Two major Smad pathways in TGF-beta superfamily signalling. *Genes Cells* **7**, 1191-1204 (2002).

43. Yang,Y., Guillot,P., Boyd,Y., Lyon,M.F. & McMahon,A.P. Evidence that preaxial polydactyly in the Doublefoot mutant is due to ectopic Indian Hedgehog signaling. *Development* **125**, 3123-3132 (1998).

44. Sirard,C. *et al.* The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. *Genes Dev.* **12**, 107-119 (1998).

45. Bachiller,D. *et al.* The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature* **403**, 658-661 (2000).

46. Ding,J. *et al.* Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo. *Nature* **395**, 702-707 (1998).

47. Meno,C. *et al.* Mouse Lefty2 and zebrafish antivin are feedback inhibitors of nodal signaling during vertebrate gastrulation. *Mol. Cell* **4**, 287-298 (1999).

48. Fishman,M.C. & Chien,K.R. Fashioning the vertebrate heart: earliest embryonic decisions. *Development* **124**, 2099-2117 (1997).

49. Fishman,M.C. & Olson,E.N. Parsing the heart: genetic modules for organ assembly. *Cell* **91**, 153-156 (1997).

50. Bruneau,B.G. Transcriptional regulation of vertebrate cardiac morphogenesis. *Circ. Res* **90**, 509-519 (2002).

51. Zaffran,S. & Frasch,M. Early signals in cardiac development. *Circ. Res.* **91**, 457-469 (2002).

52. Tam,P.P., Parameswaran,M., Kinder,S.J. & Weinberger,R.P. The allocation of epiblast cells to the embryonic heart and other mesodermal lineages: the role of ingression and tissue movement during gastrulation. *Development* **124**, 1631-1642 (1997).

53. Cohen-Gould,L. & Mikawa,T. The fate diversity of mesodermal cells within the heart field during chicken early embryogenesis. *Dev. Biol.* **177**, 265-273 (1996).

54. DeHaan,R.L. Migration patterns of the precardiac mesoderm in the early chick embrvo. *Exp. Cell Res.* **29**, 544-560 (1963).

55. Kelly,R.G. & Buckingham,M.E. The anterior heart-forming field: voyage to the arterial pole of the heart. *Trends Genet.* **18**, 210-216 (2002).

56. Schultheiss,T.M., Burch,J.B. & Lassar,A.B. A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev.* **11**, 451-462 (1997).

57. Yamada,M., Revelli,J.P., Eichele,G., Barron,M. & Schwartz,R.J. Expression of chick Tbx-2, Tbx-3, and Tbx-5 genes during early heart development: evidence for BMP2 induction of Tbx2. *Dev. Biol.* **228**, 95-105 (2000).

58. Sparrow,D.B., Kotecha,S., Towers,N. & Mohun,T.J. Xenopus eHAND: a marker for the developing cardiovascular system of the embryo that is regulated by bone morphogenetic proteins. *Mech. Dev.* **71**, 151-163 (1998).

59. Reifers,F. *et al.* Fgf8 is mutated in zebrafish acerebellar (ace) mutants and is required for maintenance of midbrain-hindbrain boundary development and somitogenesis. *Development* **125**, 2381-2395 (1998).

60. Nakagawa,O. *et al.* Members of the HRT family of basic helix-loop-helix proteins act as transcriptional repressors downstream of Notch signaling. *Proc. Natl. Acad. Sci U. S A* **97**, 13655-13660 (2000).

61. Piccoli,D.A. & Spinner,N.B. Alagille syndrome and the Jagged1 gene. *Semin Liver Dis* **21**, 525-534 (2001).

62. Doss,M.X., Koehler,C.I., Gissel,C., Hescheler,J. & Sachinidis,A. Embryonic stem cells: a promising tool for cell replacement therapy. *J. Cell Mol. Med.* **8**, 465-473 (2004).

63. Irizarry,R.A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* **31**, e15 (2003).

64. Eisen,M.B., Spellman,P.T., Brown,P.O. & Botstein,D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci U. S. A* **95**, 14863-14868 (1998).

65. Herrmann,B.G., Labeit,S., Poustka,A., King,T.R. & Lehrach,H. Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617-622 (1990).

66. Kubo,A. *et al.* Development of definitive endoderm from embryonic stem cells in culture. *Development* **131**, 1651-1662 (2004).

67. Zhang,H. & Bradley,A. Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. *Development* **122**, 2977-2986 (1996).

68. Weinhold,B. *et al.* Srf(-/-) ES cells display non-cell-autonomous impairment in mesodermal differentiation. *EMBO J.* **19**, 5835-5844 (2000).

69. Ghatpande,S., Brand,T., Zile,M. & Evans,T. Bmp2 and Gata4 function additively to rescue heart tube development in the absence of retinoids. *Dev Dyn* (2006).

70. Schultheiss,T.M. & Lassar,A.B. Induction of chick cardiac myogenesis by bone morphogenetic proteins. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 413-419 (1997).

71. Kimura,N., Matsuo,R., Shibuya,H., Nakashima,K. & Taga,T. BMP2-induced apoptosis is mediated by activation of the TAK1-p38 kinase pathway that is negatively regulated by Smad6. *J. Biol. Chem.* **275**, 17647-17652 (2000).

72. Iwasaki,S., Hattori,A., Sato,M., Tsujimoto,M. & Kohno,M. Characterization of the bone morphogenetic protein-2 as a neurotrophic factor. Induction of neuronal differentiation of PC12 cells in the absence of mitogen-activated protein kinase activation. *J Biol Chem* **271**, 17360-17365 (1996).

73. Hattori,A. *et al.* Bone morphogenetic protein-2 promotes survival and differentiation of striatal GABAergic neurons in the absence of glial cell proliferation. *J. Neurochem.* **72**, 2264-2271 (1999).

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

75. Tian,T. & Meng,A.M. Nodal signals pattern vertebrate embryos. *Cell Mol. Life Sci.* (2006).

76. Pfendler,K.C., Catuar,C.S., Meneses,J.J. & Pedersen,R.A. Overexpression of Nodal promotes differentiation of mouse embryonic stem cells into mesoderm and endoderm at the expense of neuroectoderm formation. *Stem Cells Dev.* **14**, 162-172 (2005).

77. Ariizumi,T. & Asashima,M. Control of the embryonic body plan by activin during amphibian development. *Zoolog. Sci.* **12**, 509-521 (1995).

78. Minchiotti,G., Parisi,S., Liguori,G.L., D'Andrea,D. & Persico,M.G. Role of the EGF-CFC gene cripto in cell differentiation and embryo development. *Gene* **287**, 33-37 (2002).

79. Takahashi,Y., Kitajima,S., Inoue,T., Kanno,J. & Saga,Y. Differential contributions of Mesp1 and Mesp2 to the epithelialization and rostro-caudal patterning of somites. *Development* **132**, 787-796 (2005).

80. Kitajima,S., Miyagawa-Tomita,S., Inoue,T., Kanno,J. & Saga,Y. Mesp1-nonexpressing cells contribute to the ventricular cardiac conduction system. *Dev Dyn* **235**, 395-402 (2006).

81. Heussler,H.S. & Suri,M. Sonic hedgehog. *Mol Pathol.* **56**, 129-131 (2003).

82. Joseph,N.M. *et al.* Neural crest stem cells undergo multilineage differentiation in developing peripheral nerves to generate endoneurial fibroblasts in addition to Schwann cells. *Development* **131**, 5599-5612 (2004).

83. Abzhanov,A., Tzahor,E., Lassar,A.B. & Tabin,C.J. Dissimilar regulation of cell differentiation in mesencephalic (cranial) and sacral (trunk) neural crest cells in vitro. *Development* **130**, 4567-4579 (2003).

84. Tomita,Y. *et al.* Cardiac neural crest cells contribute to the dormant multipotent stem cell in the mammalian heart. *J Cell Biol* **170**, 1135-1146 (2005).

85. Shah,N.M. & Anderson,D.J. Integration of multiple instructive cues by neural crest stem cells reveals cell-intrinsic biases in relative growth factor responsiveness. *Proc. Natl. Acad. Sci. U. S. A* **94**, 11369-11374 (1997).

86. Stemple,D.L. & Anderson,D.J. Isolation of a stem cell for neurons and glia from the mammalian neural crest. *Cell* **71**, 973-985 (1992).

87. Le Douarin,N.M., Creuzet,S., Couly,G. & Dupin,E. Neural crest cell plasticity and its limits. *Development* **131**, 4637-4650 (2004).

88. Le Douarin,N.M. & Dupin,E. Multipotentiality of the neural crest. *Curr. Opin. Genet. Dev.* **13**, 529-536 (2003).

89. Dorsky,R.I., Moon,R.T. & Raible,D.W. Control of neural crest cell fate by the Wnt signalling pathway. *Nature* **396**, 370-373 (1998).

90. Mackenzie,M.A., Jordan,S.A., Budd,P.S. & Jackson,I.J. Activation of the receptor tyrosine kinase Kit is required for the proliferation of melanoblasts in the mouse embryo. *Dev. Biol.* **192**, 99-107 (1997).

91. Orr-Urtreger,A. *et al.* Developmental expression of c-kit, a proto-oncogene encoded by the W locus. *Development* **109**, 911-923 (1990).

92. Kleber,M. *et al.* Neural crest stem cell maintenance by combinatorial Wnt and BMP signaling. *J. Cell Biol.* **169**, 309-320 (2005).

93. Willette,R.N. *et al.* BMP-2 gene expression and effects on human vascular smooth muscle cells. *J. Vasc. Res.* **36**, 120-125 (1999).

94. Gissel,C. *et al.* An optimized embryonic stem cell model for consistent gene expression and developmental studies: a fundamental study. *Thromb. Haemost.* **94**, 719-727 (2005).

95. Yuasa,S. *et al.* Transient inhibition of BMP signaling by Noggin induces cardiomyocyte differentiation of mouse embryonic stem cells. *Nat. Biotechnol.* **23**, 607-611 (2005).

96. Behfar,A. *et al.* Stem cell differentiation requires a paracrine pathway in the heart. *FASEB J* **16**, 1558-1566 (2002).

97. Zhao,G.Q. Consequences of knocking out BMP signaling in the mouse. *Genesis.* **35**, 43-56 (2003).

98. Sawano,A. *et al.* Flt-1, vascular endothelial growth factor receptor 1, is a novel cell surface marker for the lineage of monocyte-macrophages in humans. *Blood* **97**, 785-791 (2001).

99. Mishra,L., Derynck,R. & Mishra,B. Transforming growth factor-beta signaling in stem cells and cancer. *Science* **310**, 68-71 (2005).

100. Guilak,F. *et al.* The pericellular matrix as a transducer of biomechanical and biochemical signals in articular cartilage. *Ann. N. Y. Acad. Sci* **1068**, 498-512 (2006).

101. Lefebvre,V., Huang,W., Harley,V.R., Goodfellow,P.N. & de Crombrugghe,B. SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. *Mol Cell Biol* **17**, 2336-2346 (1997).

102. Zammit,P.S. *et al.* Muscle satellite cells adopt divergent fates: a mechanism for self-renewal? *J. Cell Biol.* **166**, 347-357 (2004).

103. Yablonka-Reuveni,Z. & Rivera,A.J. Temporal expression of regulatory and structural muscle proteins during myogenesis of satellite cells on isolated adult rat fibers. *Dev. Biol.* **164**, 588-603 (1994).

104. Seale,P. & Rudnicki,M.A. A new look at the origin, function, and "stem-cell" status of muscle satellite cells. *Dev. Biol.* **218**, 115-124 (2000).

105. Shefer,G., Wleklinski-Lee,M. & Yablonka-Reuveni,Z. Skeletal muscle satellite cells can spontaneously enter an alternative mesenchymal pathway. *J. Cell Sci.* **117**, 5393-5404 (2004).

106. Okada,H., Danoff,T.M., Kalluri,R. & Neilson,E.G. Early role of Fsp1 in epithelial-mesenchymal transformation. *Am. J. Physiol* **273**, F563-F574 (1997).

107. Sasmono,R.T. *et al.* A macrophage colony-stimulating factor receptor-green fluorescent protein transgene is expressed throughout the mononuclear phagocyte system of the mouse. *Blood* **101**, 1155-1163 (2003).

108. Rosen,E.D. & Spiegelman,B.M. PPARgamma : a nuclear regulator of metabolism, differentiation, and cell growth. *J. Biol. Chem.* **276**, 37731-37734 (2001).

109. Yokomori,N., Tawata,M. & Onaya,T. A transcriptional repressor regulates mouse GLUT4 gene expression during the differentiation of 3T3-L1 cells. *Diabetes* **48**, 2471-2474 (1999).

110. Stanley,E.R., Berg,K.L., Einstein,D.B., Lee,P.S. & Yeung,Y.G. The biology and action of colony stimulating factor-1. *Stem Cells* **12 Suppl 1**, 15-24 (1994).

111. Doi,H. *et al.* A new preadipocyte cell line, AP-18, established from adult mouse adipose tissue. *Tohoku J. Exp. Med.* **207**, 209-216 (2005).

112. Kolossov,E. *et al.* Identification and characterization of embryonic stem cell-derived pacemaker and atrial cardiomyocytes. *FASEB J* **19**, 577-579 (2005).

113. Banach,K., Halbach,M.D., Hu,P., Hescheler,J. & Egert,U. Development of electrical activity in cardiac myocyte aggregates derived from mouse embryonic stem cells. *Am. J Physiol Heart Circ. Physiol* **284**, H2114-H2123 (2003).

114. Liu,W. *et al.* beta-adrenergic modulation of L-type Ca2+-channel currents in early-stage embryonic mouse heart. *Am. J Physiol* **276**, H608-H613 (1999).

115. Maltsev,V.A., Rohwedel,J., Hescheler,J. & Wobus,A.M. Embryonic stem cells differentiate in vitro into cardiomyocytes representing sinusnodal, atrial and ventricular cell types. *Mech Dev* **44**, 41-50 (1993).

116. Maltsev,V.A., Wobus,A.M., Rohwedel,J., Bader,M. & Hescheler,J. Cardiomyocytes differentiated in vitro from embryonic stem cells developmentally express cardiac-specific genes and ionic currents. *Circ. Res* **75**, 233-244 (1994).

117. Kolossov,E. *et al.* Identification and characterization of embryonic stem cell-derived pacemaker and atrial cardiomyocytes. *FASEB J* **19**, 577-579 (2005).

118. Maltsev,V.A., Rohwedel,J., Hescheler,J. & Wobus,A.M. Embryonic stem cells differentiate in vitro into cardiomyocytes representing sinusnodal, atrial and ventricular cell types. *Mech Dev* **44**, 41-50 (1993).

119. Maltsev,V.A., Wobus,A.M., Rohwedel,J., Bader,M. & Hescheler,J. Cardiomyocytes differentiated in vitro from embryonic stem cells developmentally express cardiac-specific genes and ionic currents. *Circ. Res* **75**, 233-244 (1994).

120. Kolossov,E. *et al.* Identification and characterization of embryonic stem cell-derived pacemaker and atrial cardiomyocytes. *FASEB J* **19**, 577-579 (2005).

121. Metzger,J.M., Lin,W.I., Johnston,R.A., Westfall,M.V. & Samuelson,L.C. Myosin heavy chain expression in contracting myocytes isolated during embryonic stem cell cardiogenesis. *Circ. Res.* **76**, 710-719 (1995).

122. Lyons,G.E., Schiaffino,S., Sassoon,D., Barton,P. & Buckingham,M. Developmental regulation of myosin gene expression in mouse cardiac muscle. *J Cell Biol.* **111**, 2427-2436 (1990).

123. Kolossov,E. *et al.* Identification and characterization of embryonic stem cell-derived pacemaker and atrial cardiomyocytes. *FASEB J* **19**, 577-579 (2005).

7. Annexures

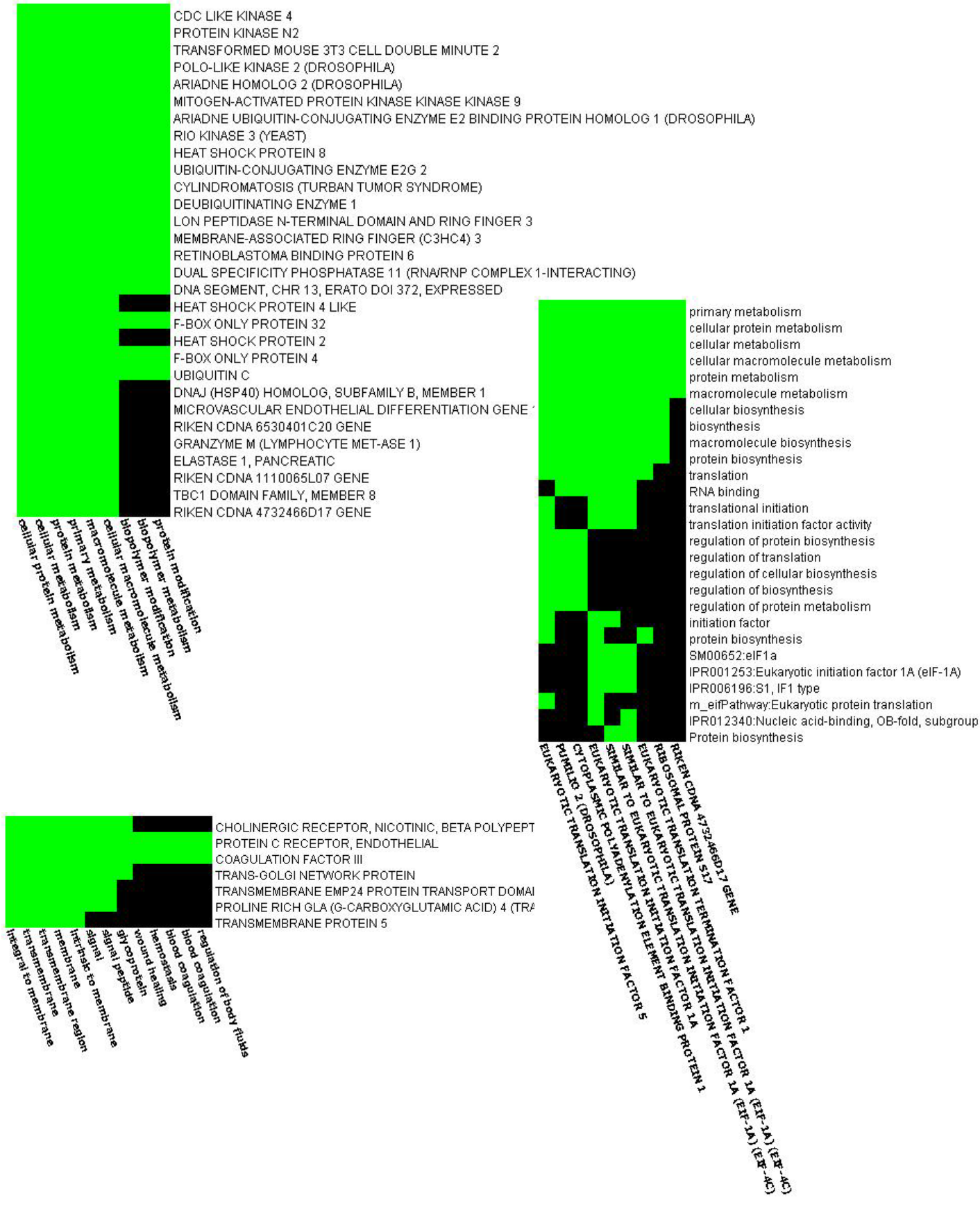
Annexure-1

Pathways differentially regulated in T Brachyury^{+} cells in comparison to T Brachyury ES cells and T Brachyury control EBs.

Category	Term	Count	%	PValue
KEGG_PATHWAY	MMU04010: MAPK SIGNALING PATHWAY	95	1.69%	0.03802829
KEGG_PATHWAY	MMU04510: FOCAL ADHESION	82	1.46%	0.00410464
KEGG_PATHWAY	MMU04360: AXON GUIDANCE	56	1.00%	0.00211058
KEGG_PATHWAY	MMU04310: WNT SIGNALING PATHWAY	54	0.96%	0.05802459
KEGG_PATHWAY	MMU04110: CELL CYCLE	45	0.80%	0.00725305
KEGG_PATHWAY	MMU04350: TGF-BETA SIGNALING PATHWAY	41	0.73%	8.38E-04
KEGG_PATHWAY	MMU04512: ECM-RECEPTOR INTERACTION	40	0.71%	0.00293432
KEGG_PATHWAY	MMU04540: GAP JUNCTION	37	0.66%	0.03179355
KEGG_PATHWAY	MMU00310: LYSINE DEGRADATION	27	0.48%	0.03009477
KEGG_PATHWAY	MMU00051: FRUCTOSE AND MANNOSE METABOLISM	24	0.43%	0.02629444
KEGG_PATHWAY	MMU00500: STARCH AND SUCROSE METABOLISM	24	0.43%	0.07417547
KEGG_PATHWAY	MMU00280: VALINE, LEUCINE AND ISOLEUCINE DEGRADATION	22	0.39%	0.07152007
KEGG_PATHWAY	MMU00632: BENZOATE DEGRADATION VIA COA LIGATION	21	0.37%	0.03611389
KEGG_PATHWAY	MMU00620: PYRUVATE METABOLISM	20	0.36%	0.01976104
KEGG_PATHWAY	MMU00052: GALACTOSE METABOLISM	18	0.32%	0.01122606
KEGG_PATHWAY	MMU00640: PROPANOATE METABOLISM	17	0.30%	0.04698314
KEGG_PATHWAY	MMU00251: GLUTAMATE METABOLISM	17	0.30%	0.01374867
KEGG_PATHWAY	MMU00040: PENTOSE AND GLUCURONATE INTERCONVERSIONS	13	0.23%	0.00141937
KEGG_PATHWAY	MMU00220: UREA CYCLE AND METABOLISM OF AMINO GROUPS	13	0.23%	0.08003552
KEGG_PATHWAY	MMU00910: NITROGEN METABOLISM	13	0.23%	0.01400915
KEGG_PATHWAY	MMU00903: LIMONENE AND PINENE DEGRADATION	13	0.23%	0.08003552
KEGG_PATHWAY	MMU00100: BIOSYNTHESIS OF STEROIDS	11	0.20%	0.00712206
KEGG_PATHWAY	MMU05060: PRION DISEASE	8	0.14%	0.02168277
BIOCARTA	m_arfPathway:Tumor Suppressor Arf Inhibits Ribosomal Biogenesis	9	0.16%	0.005623
BIOCARTA	m_badPathway:Regulation of BAD phosphorylation	9	0.16%	0.093991
BIOCARTA	m_telPathway:Telomeres	11	0.20%	0.050115
BIOCARTA	m_plcePathway:Phospholipase C-epsilon pathway	5	0.09%	0.077789
BIOCARTA	m_p53Pathway:p53 Signaling Pathway	11	0.20%	0.009843

Annexure-2

Heat map of the transcripts upregulated specifically in T Brachyury^{+} cell population that are not included in GOs of development , transcription and signaling.



Annexure-3

Development related transcripts that are upregulated in T Brachyury$^+$ cells than ES cells and 6d control EBs but show differential expression among ES cells and 6 day EBs. The fold change is expressed relative to the lowest value of expression that is normalized to 1.

NCBI ID	Gene Name	ES cells	6day Control EBs	T Brachyury$^+$ cells
NM_013562	interferon-related developmental regulator 1	2.60	1.00	7.28
NM_009964	crystallin, alpha B	1.00	5.74	13.16
NM_009255	serine (or cysteine) peptidase inhibitor, clade E, member 2	1.00	5.90	14.33
NM_013749	tumor necrosis factor receptor superfamily, member 12a	1.00	2.07	11.37
NM_011674	UDP galactosyltransferase 8A	4.22	1.00	9.39
NM_011580	thrombospondin 1	1.00	14.95	30.90
NM_008343	insulin-like growth factor binding protein 3	1.00	3.36	9.74
NM_009573	zinc finger protein of the cerebellum 1	1.00	4.78	11.86
NM_011526	transgelin	1.00	6.55	14.00
NM_010029	DEAD (Asp-Glu-Ala-Asp) box polypeptide 4	18.81	1.00	92.42
NM_010118	early growth response 2	1.00	2.80	7.38
NM_013657	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	1.00	2.03	24.21
NM_176930	RIKEN cDNA C130076O07 gene	1.00	4.47	10.74
NM_139298	wingless-type MMTV integration site 9A	1.00	2.96	15.99
NM_010516	cysteine rich protein 61	1.00	2.48	5.31
NM_008655	growth arrest and DNA-damage-inducible 45 beta	1.00	2.44	11.38
NM_011817	growth arrest and DNA-damage-inducible 45 gamma	1.00	7.53	26.97
NM_011536	T-box 4	1.00	2.03	4.67

Annexure-4

Transcripts that are downregulated specifically in T-brachyury population and play a role in development. The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

			6d Control	
NCBI ID	Gene Name	ES cells	EBs	T Brachyury$^+$
NM_009468	dihydropyrimidinase-like 3	2.91	4.17	1.00
NM_013685	transcription factor 4	3.65	3.43	1.00
NM_008195	glycogen synthase 3, brain	2.90	4.71	1.00
NM_011299	ribosomal protein S6 kinase, polypeptide 2	2.71	3.57	1.00
NM_013865	N-myc downstream regulated gene 3	2.26	2.58	1.00
NM_009384	T-cell lymphoma invasion and metastasis 1	3.57	2.07	1.00
NM_007548	PR domain containing 1, with ZNF domain	11.85	7.18	1.00
NM_007792	cysteine and glycine-rich protein 2	3.42	5.90	1.00
NM_007936	Eph receptor A4	2.24	2.04	1.00
NM_008713	nitric oxide synthase 3, endothelial cell	2.90	2.85	1.00
NM_007554	bone morphogenetic protein 4	4.86	3.89	1.00
NM_033524	sprouty protein with EVH-1 domain 1, related sequence	2.39	2.10	1.00
NM_010875	neural cell adhesion molecule 1	4.13	5.60	1.00
NM_009929	procollagen, type XVIII, alpha 1	2.67	3.37	1.00
NM_133719	meteorin, glial cell differentiation regulator	2.33	2.11	1.00
NM_019413	roundabout homolog 1 (Drosophila)	3.43	3.40	1.00
NM_011652	titin	2.33	2.33	1.00
NM_008957	patched homolog 1	4.36	8.67	1.00
NM_016809	RNA binding motif protein 3	2.12	2.20	1.00
NM_008914	protein phosphatase 3, catalytic subunit, beta isoform	2.29	2.68	1.00
NM_198127	abl-interactor 2	2.14	3.56	1.00
NM_011858	odd Oz/ten-m homolog 4 (Drosophila)	2.89	2.59	1.00
---	---	2.51	2.81	1.00
NM_033509	loop tail associated protein	2.05	2.78	1.00
NM_010233	fibronectin 1	4.06	5.68	1.00
NM_001033299	zinc finger protein 217	2.53	2.36	1.00
NM_146194	Phosphatidylinositol binding clathrin assembly protein	2.01	2.64	1.00
NM_013864	N-myc downstream regulated gene 2	5.84	5.54	1.00
NM_019588	phospholipase C, epsilon 1	6.53	7.01	1.00
NM_178688	actin-binding LIM protein 1	2.25	3.48	1.00
NM_008342	insulin-like growth factor binding protein 2	2.60	3.09	1.00
NM_030732	transducin (beta)-like 1X-linked receptor 1	3.94	2.80	1.00
NM_020259	Hedgehog-interacting protein	4.21	3.30	1.00
NM_009706	Rho GTPase activating protein 5 (Arhgap5), mRNA	2.29	2.33	1.00
NM_010164	eyes absent 1 homolog (Drosophila)	9.12	5.64	1.00
NM_010162	Exostoses (multiple) 1, mRNA	2.46	2.02	1.00
XM_486096	deleted in liver cancer 1	3.16	2.09	1.00

Annexure-5

Transcription factors that are downregulated specifically in T-brachyury population. The fold change is expressed relative to the lowest value of expression that is normalized to 1.

NCBI ID	Gene Name	ES cells	6d Control Ebs	T Brachyury^{+}
NM_009122	special AT-rich sequence binding protein 1	2.90	2.47	1.00
NM_013685	transcription factor 4	3.65	3.43	1.00
NM_007548	PR domain containing 1, with ZNF domain	11.85	7.18	1.00
NM_133210	SERTA domain containing 3	2.11	2.22	1.00
NM_011249	retinoblastoma-like 1 (p107)	5.20	4.98	1.00
NM_011605	thymopoietin	2.24	2.68	1.00
NM_021559	zinc finger protein 191	2.23	2.28	1.00
NM_019574	zinc finger protein 278	2.22	2.23	1.00
NM_027642	PHD finger protein 6	2.67	3.00	1.00
NM_023144	non-POU-domain-containing, octamer binding protein	2.30	2.03	1.00
NM_001015099	RIKEN cDNA 6030408C04 gene	5.38	5.88	1.00
NM_146066	G1 to S phase transition 1	3.33	2.32	1.00
NM_053202	Forkhead box P1 (Foxp1), mRNA	2.98	2.01	1.00
NM_028399	cyclin T2	5.83	3.11	1.00
XM_125901	cullin associated and neddylation disassociated 1	3.21	2.02	1.00
NM_011308	nuclear receptor co-repressor 1	2.32	2.29	1.00
NM_177358	RIKEN cDNA A630033E08 gene	2.72	3.07	1.00
NM_001024910	septin 10	2.56	3.84	1.00
NM_001013391	cleavage and polyadenylation specific factor 6	2.36	2.83	1.00
NM_177359	RIKEN cDNA 6030490I01 gene	3.19	3.66	1.00
NM_153198	High mobility group box transcription factor 1 (Hbp1)	2.14	2.33	1.00
NM_172586	Zinc finger protein 322a, mRNA	2.20	2.88	1.00
NM_138755	PHD finger protein 21A (Phf21a), mRNA	3.80	3.69	1.00
NM_138679	Ash1(absent, small,or homeotic)-like (Drosophila)(Ash1l)	2.16	2.15	1.00
NM_013843	zinc finger protein 53	2.49	3.03	1.00
NM_011546	Zinc finger homeobox 1a (Zfhx1a), mRNA	2.22	2.72	1.00
NM_172153	Mlr1 mRNA for transcription factor MLR1	2.51	2.56	1.00
NM_009366	TSC22-related inducible leucine zipper 1b (Tilz1b)	2.12	2.00	1.00
NM_023868	ryanodine receptor 2, cardiac	2.03	2.79	1.00
NM_030887	Jun dimerization protein 2	3.15	2.50	1.00
NM_030732	transducin (beta)-like 1X-linked receptor 1	3.94	2.80	1.00
NM_139141	Zinc finger protein 192 (Zfp192), mRNA	2.68	5.27	1.00
NM_010164	eyes absent 1 homolog (Drosophila)	9.12	5.64	1.00
NM_152895	jumonji, AT rich interactive domain 1B (Rbp2 like)	2.64	2.56	1.00

Annexure-6

A. GTPase regulators that are downregulated in T-brachyury population and also differentially expressed among ES cells and 6day control EBs. The fold change is expressed relative to the lowest value of expression that is normalized to 1.

NCBI ID	Gene Name	ES cells	6d Control EBs	T Brachyury$^+$
NM_023884	Ral GEF with PH domain and SH3 binding motif 2	2.05	2.14	1.00
NM_009384	T-cell lymphoma invasion and metastasis 1	3.57	2.07	1.00
XM_129913	dedicator of cytokinesis 10	2.34	4.07	1.00
NM_001004364	development and differentiation enhancing factor 2	2.17	2.37	1.00
XM_147847	spermatogenesis associated 13	2.52	4.49	1.00
NM_172579	signal-induced proliferation-associated 1 like 1	4.36	3.84	1.00
NM_019588	phospholipase C, epsilon 1	6.53	7.01	1.00
XM_130797	TRAF2 and NCK interacting kinase	2.13	3.27	1.00
NM_009706	Rho GTPase activating protein 5 (Arhgap5), mRNA	2.29	2.33	1.00
XM_486096	deleted in liver cancer 1	3.16	2.09	1.00

B. Development related transcripts that are downregulated in T-brachyury population and also differentially expressed in ES cells and 6dcontrol EBs. The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

NCBI ID	Gene Name	ES cells	6d Control Ebs	T Brachyury$^+$
NM_008480	laminin, alpha 1	11.45	5.18	1
NM_013690	endothelial-specific receptor tyrosine kinase	4.18	25.16	1
NM_026820	interferon induced transmembrane protein 1	7.59	2.98	1
NM_177546	phosphate cytidylyltransferase 1, choline, beta isoform	9.82	2.27	1
---	PREDICTED: Mus musculus similar to hypothetical protein FLJ39502	10.54	2.13	1
NM_013834	secreted frizzled-related sequence protein 1	6.72	2.95	1
NM_019922	cartilage associated protein	4.29	2.07	1
NM_016697	glypican 3	3.54	46.18	1
XM_136212	GLI-Kruppel family member GLI2	4.93	2.18	1

Annexure-7

Development related transcripts that are upregulated in T Brachyury$^+$ than ES cells but downregulated than EBs . The fold change is expressed relative to the lowest value of expression that is normalized to 1.

NCBI ID	Gene Name	ES cells	6d Control EBs	T Brachyury$^+$
NM_153319	angiomotin	1.00	279.17	25.10
XM_129060	serologically defined colon cancer antigen 33	1.00	48.90	10.54
NM_007553	bone morphogenetic protein 2	1.00	85.99	7.98
NM_008259	forkhead box A1	1.00	43.79	6.39
NM_011514	suppressor of variegation 3-9 homolog 1 (Drosophila)	1.00	18.57	6.05
NM_001013833	protein kinase, cGMP-dependent, type I	1.00	12.63	5.80
NM_008239	Forkhead box Q1 (Foxq1), mRNA	1.00	13.89	5.80
NM_011098	paired-like homeodomain transcription factor 2	1.00	57.25	5.25
NM_010514	insulin-like growth factor 2	1.00	234.66	4.21
NM_010882	necdin	1.00	27.67	3.89
NM_007441	aristaless 3	1.00	7.80	3.77
NM_018826	Iroquois related homeobox 5 (Drosophila)	1.00	8.55	3.73
NM_010449	homeo box A1	1.00	7.41	3.51
NM_009851	CD44 antigen	1.00	43.29	3.51
NM_010109	ephrin A5	1.00	8.94	3.50
NM_009367	transforming growth factor, beta 2	1.00	59.11	3.42
NM_178804	slit homolog 2 (Drosophila)	1.00	32.94	3.41
NM_007471	amyloid beta (A4) precursor protein	1.00	13.74	3.40
NM_007585	annexin A2	1.00	8.93	3.32
NM_013904	hairy/enhancer-of-split related with YRPW motif 2	1.00	14.82	3.26
NM_013598	kit ligand	1.00	30.30	3.15
NM_010258	GATA binding protein 6	1.00	26.13	3.14
NM_008495	lectin, galactose binding, soluble 1	1.00	7.15	2.96

Annexure-7 (continued)

Development related transcripts that are upregulated in T Brachyury$^+$ than ES cells but downregulated than EBs (continued). The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

NCBI ID	Gene Name	ES cells	6d Control EBs	T Brachyury$^+$
NM_010423	hairy/enhancer-of-split related with YRPW motif 1	1.00	17.37	2.91
NM_008737	neuropilin 1	1.00	264.22	2.65
NM_020510	frizzled homolog 2 (Drosophila)	1.00	18.61	2.54
NM_008321	inhibitor of DNA binding 3	1.00	11.50	2.54
NM_008413	Janus kinase 2	1.00	5.06	2.46
NM_021457	frizzled homolog 1 (Drosophila)	1.00	11.20	2.44
NM_029770	unc-5 homolog B (C. elegans)	1.00	12.28	2.36
NM_007664	cadherin 2	1.00	17.86	2.35
NM_008592	forkhead box C1	1.00	5.05	2.28
NM_011594	tissue inhibitor of metalloproteinase 2	1.00	10.88	2.25
NM_008937	prospero-related homeobox 1	1.00	4.48	2.21
NM_015753	zinc finger homeobox 1b	1.00	7.32	2.20
NM_016769	MAD homolog 3 (Drosophila)	1.00	4.75	2.14
NM_009021	retinoic acid induced 1	1.00	5.58	2.12
NM_011864	3'-phosphoadenosine 5'-phosphosulfate synthase 2	1.00	5.03	2.09
NM_010612	kinase insert domain protein receptor	1.00	36.35	2.09
NM_011415	snail homolog 2 (Drosophila)	1.00	18.78	2.09
NM_008091	GATA binding protein 3	1.00	5.11	2.04
NM_145506	erythrocyte protein band 4.1-like 5	1.00	5.44	2.03
NM_017464	neural precursor cell expressed, developmentally down-regulated gene 9	1.00	16.42	2.02
NM_009876	cyclin-dependent kinase inhibitor 1C (P57)	1.00	15.54	2.01

Annexure-8

Development related transcripts that are downregulated in T-brachyury population than the ES cells but upregulated than 6day control EBs. The fold change is expressed relative to the lowest value of expression that is normalized to 1.

NCBI ID	Gene Name	ES cells	6d Control EBs	T Brachyury^{+}
NM_007430	nuclear receptor subfamily 0, group B, member 1	240.16	1.00	90.02
XM_132755	Nanog homeobox	112.17	1.00	49.75
NM_019448	DNA (cytosine-5-)-methyltransferase 3-like	142.34	1.00	37.64
NM_013633	POU domain, class 5, transcription factor 1	87.33	1.00	29.06
NM_010202	fibroblast growth factor 4	175.48	1.00	25.68
NM_133197	mcf.2 transforming sequence	62.18	1.00	12.30
NM_011562	teratocarcinoma-derived growth factor	54.26	1.00	10.71
NM_172496	cordon-bleu	21.12	1.00	9.83
NM_009575	zinc finger protein of the cerebellum 3	15.38	1.00	6.64
NM_172303	PHD finger protein 17	13.35	1.00	6.25
NM_175641	latent transforming growth factor beta binding protein 4	14.59	1.00	5.67
NM_019926	X-linked myotubular myopathy gene 1	10.74	1.00	5.21
NM_138944	POU domain, class 4, transcription factor 2	64.31	1.00	4.47
NM_008126	gap junction membrane channel protein beta 3	16.88	1.00	4.47
NM_023794	ets variant gene 5	8.95	1.00	4.34
NM_010128	epithelial membrane protein 1	10.77	1.00	4.33
NM_033077	DNA segment, Chr 1, Pasteur Institute 1	11.53	1.00	4.29
NM_019984	transglutaminase 1, K polypeptide	26.74	1.00	3.65
NM_008381	inhibin beta-B	32.21	1.00	3.18
NM_010262	gastrulation brain homeobox 2	12.93	1.00	2.99
NM_011593	tissue inhibitor of metalloproteinase 1	8.66	1.00	2.68
NM_009328	transcription factor 15	8.43	1.00	2.58
NM_010425	forkhead box D3	7.42	1.00	2.54
NM_020581	angiopoietin-like 4	6.25	1.00	2.53
NM_028199	plexin domain containing 1	5.90	1.00	2.38
NM_144547	anti-Mullerian hormone type 2 receptor	5.45	1.00	2.36
NM_011518	spleen tyrosine kinase	4.73	1.00	2.31
NM_007905	polyhomeotic-like 1 (Drosophila)	4.76	1.00	2.29
NM_030886	Ankyrin repeat domain 17, mRNA	6.47	1.00	2.23
NM_133195	bruno-like 4, RNA binding protein (Drosophila)	6.56	1.00	2.22
NM_008234	helicase, lymphoid specific	5.97	1.00	2.19
NM_011898	sprouty homolog 4 (Drosophila)	14.23	1.00	2.10
NM_008601	microphthalmia-associated transcription factor	19.18	1.00	2.04
NM_007868	dystrophin, muscular dystrophy	6.16	1.00	2.01

Annexure-9

Transcription factors that are downregulated in T-Brachyury$^+$ population than the ES cells but upregulated than 6day control EBs. The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

NCBI ID	Gene Name	ES cells	6d Control Ebs	T-brachyury$^+$
NM_001017525	BTB (POZ) domain containing 11	6.61	1.00	3.21
NM_009482	undifferentiated embryonic cell transcription factor 1	12.05	1.00	5.92
NM_023258	PYD and CARD domain containing	12.32	1.00	3.94
NM_007430	nuclear receptor subfamily 0, group B, member 1	240.16	1.00	90.02
NM_013633	POU domain, class 5, transcription factor 1	87.33	1.00	29.06
NM_023755	transcription factor CP2-like 1	6.25	1.00	2.44
NM_010827	musculin	24.23	1.00	4.07
NM_010262	gastrulation brain homeobox 2	12.93	1.00	2.99
NM_030676	nuclear receptor subfamily 5, group A, member 2	19.41	1.00	6.72
NM_008578	myocyte enhancer factor 2B	4.96	1.00	2.27
NM_010425	forkhead box D3	7.42	1.00	2.54
NM_011542	transcription elongation factor A (SII), 3	13.36	1.00	2.71
NM_019448	DNA (cytosine-5-)-methyltransferase 3-like	142.34	1.00	37.64
NM_023794	ets variant gene 5	8.95	1.00	4.34
NM_178935	CXORF15	14.19	1.00	6.77
NM_011934	estrogen related receptor, beta	68.86	1.00	24.75
NM_138944	POU domain, class 4, transcription factor 2	64.31	1.00	4.47
XM_139740	gene model 313, (NCBI)	6.59	1.00	3.07
NM_023184	Kruppel-like factor 15	5.71	1.00	2.35
NM_011487	signal transducer and activator of transcription 4	12.64	1.00	2.42
NM_008452	Kruppel-like factor 2 (lung)	40.84	1.00	11.63
NM_009328	transcription factor 15	8.43	1.00	2.58
NM_008601	microphthalmia-associated transcription factor	19.18	1.00	2.04

Annexure-10

The biological processed represented by the transcripts downregulated in BMP-2^{+} cells in comparison to both ES cells and 7 day control EBs.

Category	Term	Count	%	PValue
GOTERM_BP_ALL	double-strand break repair via homologous recombination	2	1.05%	0.042403
GOTERM_BP_ALL	double-strand break repair	2	1.05%	0.090936
GOTERM_BP_ALL	negative regulation of neuron differentiation	2	1.05%	0.066983
GOTERM_BP_ALL	DNA replication	9	4.71%	1.86E-05
GOTERM_BP_ALL	DNA repair	11	5.76%	1.62E-06
GOTERM_BP_ALL	DNA recombination	4	2.09%	0.008177
GOTERM_BP_ALL	chromosome segregation	3	1.57%	0.040413
GOTERM_BP_ALL	cell division	18	9.42%	2.42E-13
GOTERM_BP_ALL	recombinational repair	2	1.05%	0.042403
GOTERM_BP_ALL	pyridine nucleotide biosynthesis	2	1.05%	0.090936
GOTERM_BP_ALL	DNA metabolism	24	12.57%	1.68E-10
GOTERM_BP_ALL	cell cycle	24	12.57%	3.11E-09
GOTERM_BP_ALL	response to endogenous stimulus	11	5.76%	1.57E-05
GOTERM_BP_ALL	base-excision repair	3	1.57%	0.012677
GOTERM_BP_ALL	cellular physiological process	99	51.83%	0.007573
GOTERM_BP_ALL	nucleobase, nucleoside, nucleotide and nucleic acid metabolism	36	18.85%	0.045319
GOTERM_BP_ALL	sterol metabolism	3	1.57%	0.091987
GOTERM_BP_ALL	cellular metabolism	75	39.27%	0.004404
GOTERM_BP_ALL	cholesterol metabolism	3	1.57%	0.081602
GOTERM_BP_ALL	protein modification	19	9.95%	0.088599
GOTERM_BP_ALL	steroid biosynthesis	4	2.09%	0.015016
GOTERM_BP_ALL	cellular catabolism	8	4.19%	0.091393
GOTERM_BP_ALL	regulation of progression through cell cycle	9	4.71%	0.010783
GOTERM_BP_ALL	M phase of mitotic cell cycle	11	5.76%	1.43E-07
GOTERM_BP_ALL	cytokinesis	4	2.09%	6.42E-04
GOTERM_BP_ALL	M phase	13	6.81%	5.59E-08
GOTERM_BP_ALL	biopolymer modification	20	10.47%	0.07165
GOTERM_BP_ALL	macromolecule metabolism	52	27.23%	7.08E-04
GOTERM_BP_ALL	biopolymer methylation	3	1.57%	0.069186
GOTERM_BP_ALL	biopolymer metabolism	44	23.04%	2.51E-06
GOTERM_BP_ALL	metabolism	78	40.84%	0.008402
GOTERM_BP_ALL	mitotic cell cycle	12	6.28%	6.46E-07
GOTERM_BP_ALL	response to DNA damage stimulus	11	5.76%	8.67E-06
GOTERM_BP_ALL	regulation of cell cycle	9	4.71%	0.010958
GOTERM_BP_ALL	mitosis	11	5.76%	1.33E-07
GOTERM_BP_ALL	primary metabolism	69	36.13%	0.020493
GOTERM_BP_ALL	DNA-dependent DNA replication	4	2.09%	0.007714

Annexure-11

The development related transcripts downregulated in BMP-2^{+} cells comparison to both ES cells and 7day control EBs. The fold change is expressed relative to the lowest value of expression that is normalized to 1.

Unigene ID	Gene Name	ES cells	7d Control Ebs	BMP-2+
Mm.90003	gap junction membrane channel protein beta 3	5.99	2.60	1.00
Mm.4269	transcription factor 4	4.56	2.01	1.00
Mm.3488	aurora kinase B	2.14	4.88	1.00
Mm.175661	interferon induced transmembrane protein 1	13.78	5.07	1.00
Mm.4974	neural cell adhesion molecule 1	5.16	2.38	1.00
Mm.228798	patched homolog 1	2.79	6.63	1.00
Mm.250185	eyes absent 1 homolog (Drosophila)	3.00	6.93	1.00
Mm.281691	secreted frizzled-related sequence protein 1	11.14	4.68	1.00

Annexure-12

A) Development related transcripts that are upregulated in BMP-2 cells more than the ES cells but downregulated than 7 day control EBs . The fold change is expressed relative to the lowest value of expression that is normalized to 1.

Unigene ID	Gene Name	ES cells	7d Control EBs	BMP-2+
Mm.110	inhibitor of DNA binding 3	1.00	34.11	10.14
Mm.181959	early growth response 1	1.00	9.47	2.37
Mm.257437	cadherin 2	1.00	11.93	3.35
Mm.197	homeo box A1	1.00	4.39	2.19
Mm.246513	FBJ osteosarcoma oncogene	1.00	16.95	5.49
Mm.233903	neuronatin	1.00	62.79	7.14
Mm.246804	paired-like homeodomain transcription factor 2	1.00	17.55	7.78
Mm.193925	guanine nucleotide binding protein, alpha 13	1.00	5.03	2.26
Mm.38378	Fras1 related extracellular matrix protein 2	1.00	10.08	2.27

B) **Development related transcripts that are upregulated more than the 7 day control EBs in BMP-2$^+$ cells but downregulated than the ES cells.** The fold change is expressed relative to the lowest value of expression that is normalized to 1.

Unigene ID	Gene Name	ES cells	7d Control Ebs	BMP-2+
Mm.15105	inositol polyphosphate-5-phosphatase D	17.65	1.00	3.51
Mm.286536	filamin binding LIM protein 1	10.82	1.00	2.15
Mm.280641	deleted in azoospermia-like	20.08	1.00	9.34
Mm.2581	Eph receptor A2	5.73	1.00	2.79
Mm.13433	DNA (cytosine-5-)-methyltransferase 3-like	59.10	1.00	14.01
Mm.3092	inhibin beta-B	39.70	1.00	2.34
Mm.6047	Nanog homeobox	27.15	1.00	2.68
Mm.272251	latent transforming growth factor beta binding protein 4	9.73	1.00	3.40
---	inactive X specific transcripts	16.58	1.00	3.00
Mm.108054	DNA segment, Chr 1, Pasteur Institute 1	13.35	1.00	2.42
Mm.288474	secreted phosphoprotein 1	25.58	1.00	5.59
Mm.3881	transcription factor 15	5.26	1.00	2.56
Mm.5090	teratocarcinoma-derived growth factor	60.58	1.00	9.00

Annexure-13

Biological processes overrepresented by differentially regulated transcripts in αMHC$^+$ cells in comparison to αMHC ES cells and 15day old αMHC control EBs.

Category	Term	Count	%	PValue
KEGG_PATHWAY	MMU04010:MAPK SIGNALING PATHWAY	128	1.66%	0.007806
KEGG_PATHWAY	MMU04510:FOCAL ADHESION	109	1.42%	4.53E-04
KEGG_PATHWAY	MMU00230:PURINE METABOLISM	76	0.99%	0.004314
KEGG_PATHWAY	MMU04110:CELL CYCLE	70	0.91%	1.33E-07
KEGG_PATHWAY	MMU04360:AXON GUIDANCE	68	0.88%	0.006319
KEGG_PATHWAY	MMU04310:WNT SIGNALING PATHWAY	68	0.88%	0.089166
KEGG_PATHWAY	MMU00190:OXIDATIVE PHOSPHORYLATION	65	0.84%	0.001586
KEGG_PATHWAY	MMU04670:LEUKOCYTE TRANSENDOTHELIAL MIGRATION	60	0.78%	0.045594
KEGG_PATHWAY	MMU04530:TIGHT JUNCTION	57	0.74%	0.097827
KEGG_PATHWAY	MMU04350:TGF-BETA SIGNALING PATHWAY	52	0.68%	2.27E-04
KEGG_PATHWAY	MMU00240:PYRIMIDINE METABOLISM	51	0.66%	0.002039
KEGG_PATHWAY	MMU04512:ECM-RECEPTOR INTERACTION	48	0.62%	0.007577
KEGG_PATHWAY	MMU04540:GAP JUNCTION	46	0.60%	0.043825
KEGG_PATHWAY	MMU04070:PHOSPHATIDYLINOSITOL SIGNALING SYSTEM	44	0.57%	0.048484
KEGG_PATHWAY	MMU04520:ADHERENS JUNCTION	41	0.53%	0.017819
KEGG_PATHWAY	MMU04730:LONG-TERM DEPRESSION	40	0.52%	0.038833
KEGG_PATHWAY	MMU04662:B CELL RECEPTOR SIGNALING PATHWAY	35	0.45%	0.062506
KEGG_PATHWAY	MMU00500:STARCH AND SUCROSE METABOLISM	33	0.43%	0.01421
KEGG_PATHWAY	MMU00760:NICOTINATE AND NICOTINAMIDE METABOLISM	33	0.43%	0.007227
KEGG_PATHWAY	MMU00632:BENZOATE DEGRADATION VIA COA LIGATION	28	0.36%	0.008187
KEGG_PATHWAY	MMU00260:GLYCINE, SERINE AND THREONINE METABOLISM	27	0.35%	0.059186
KEGG_PATHWAY	MMU00620:PYRUVATE METABOLISM	24	0.31%	0.027854
KEGG_PATHWAY	MMU04320:DORSO-VENTRAL AXIS FORMATION	19	0.25%	0.032942
KEGG_PATHWAY	MMU00220:UREA CYCLE AND METABOLISM OF AMINO GROUPS	18	0.23%	0.013778
KEGG_PATHWAY	MMU00970:AMINOACYL-TRNA SYNTHETASES	18	0.23%	0.022143
KEGG_PATHWAY	MMU00040:PENTOSE AND GLUCURONATE INTERCONVERSIONS	14	0.18%	0.005913
KEGG_PATHWAY	MMU05010:ALZHEIMER'S DISEASE	14	0.18%	0.056557
KEGG_PATHWAY	MMU00670:ONE CARBON POOL BY FOLATE	12	0.16%	0.035696
KEGG_PATHWAY	MMU00100:BIOSYNTHESIS OF STEROIDS	12	0.16%	0.019641
KEGG_PATHWAY	MMU00900:TERPENOID BIOSYNTHESIS	5	0.06%	0.088275

Annexure-14

Development related transctirpts upregulated in αMHC$^+$ cardiomyocytes compared to both ES cells and 15day old control EBs and which in turn show diffrential expression among ES cells and control EBs. The fold change is expressed relative to the lowest value of expression that is normalized to be 1.

Refseq	Genes	ES cells	15d control EBs	α MHC$^+$
NM_009468	dihydropyrimidinase-like 3	1.00	2.60	5.68
NM_009964	crystallin, alpha B	1.00	12.43	58.96
NM_019564	HtrA serine peptidase 1	1.00	11.80	44.56
NM_024223	cysteine rich protein 2	1.00	3.23	16.86
NM_015814	dickkopf homolog 3 (Xenopus laevis)	1.00	9.99	45.07
NM_010856	myosin, heavy polypeptide 6, cardiac muscle, alpha	1.00	2.15	81.41
NM_026131	PDZ and LIM domain 7	1.00	2.67	7.09
NM_009365	transforming growth factor beta 1 induced transcript 1	1.00	3.67	9.34
NM_020493	serum response factor	1.00	2.18	5.91
NM_010801	myeloid leukemia factor 1	1.00	8.68	61.38
NM_007664	cadherin 2	1.00	98.44	329.84
NM_007561	bone morphogenic protein receptor, type II	1.00	4.51	9.24
NM_025282	myocyte enhancer factor 2C	1.00	31.69	75.13
NM_007555	bone morphogenetic protein 5	1.00	2.04	5.65
NM_009814	calsequestrin 2	4.78	1.00	13.09
NM_009406	troponin I, cardiac	1.00	12.88	110.42
NM_009324	T-box 2	1.00	7.69	22.95
NM_010255	guanidinoacetate methyltransferase	1.00	2.02	4.81
NM_007911	ephrin B3	1.00	2.99	6.30
NM_010197	fibroblast growth factor 1	1.00	2.12	7.11
NM_009367	transforming growth factor, beta 2	1.00	21.25	48.11
NM_019972	sortilin 1	1.00	2.42	8.30
NM_008441	kinesin family member 1B	1.00	2.25	5.68
NM_011537	T-box 5	1.00	2.31	17.52
NM_145136	myocardin	1.00	12.15	116.41
NM_008481	laminin, alpha 2	1.00	5.24	29.44
NM_011486	signal transducer and activator of transcription 3	1.00	2.02	4.87
NM_010043	desmin	1.00	2.59	9.79
XM_110852	amyotrophic lateral sclerosis 2 (juvenile) chromosome	1.00	2.09	5.18
NM_028004	titin	1.00	20.06	148.28
NM_010118	early growth response 2	1.00	15.50	43.66
NM_175606	homeobox only domain	1.00	6.87	45.81
NM_026163	plakophilin 2	1.00	2.01	7.85
NM_029273	SLIT and NTRK-like family, member 5	1.00	3.05	23.01
NM_009116	paired related homeobox 2	1.00	3.05	9.54
NM_024124	histone deacetylase 9	1.00	5.79	31.55
NM_009861	cell division cycle 42 homolog (S. cerevisiae)	1.00	2.05	6.37
XM_484197	PREDICTED: integrin beta 8	1.00	10.60	73.78

Annexure-14 (continued)

Development related transctirpts upregulated in αMHC$^+$ cardiomyocytes compared to both ES cells and 15day old control EBs and which in turn show differential expression among ES cells and control EBs. The fold change is expressed relative to the lowest value of expression that is normalized to 1.

Refseq	Genes	ES cells	15d control	α MHC$^+$
NM_018781	early growth response 3	1.00	6.49	23.64
NM_175013	phosphoglucomutase 5	1.00	5.14	16.78
NM_008937	prospero-related homeobox 1	1.00	2.07	7.54
NM_008664	myomesin 2	1.00	2.13	40.73
NM_010788	methyl CpG binding protein 2	1.00	3.69	7.49
NM_021320	Netrin 4 (Ntn4), mRNA	1.00	2.25	24.35
NM_010789	Myeloid ecotropic viral integration site 1, mRNA	1.00	6.66	15.86
NM_145519	FERM, RhoGEF and pleckstrin domain protein	1.00	2.35	5.01
NM_009762	SET and MYND domain containing 1	1.00	3.65	63.17
NM_010141	Eph receptor A7, mRNA (cDNA clone	1.00	2.67	7.34
NM_008854	Protein kinase, cAMP dependent, catalytic	1.00	2.11	5.33
NM_019684	serine/threonine kinase 23	1.00	7.98	89.16
NM_013613	nuclear receptor subfamily 4, group A, member	1.00	2.61	12.09
NM_133201	mitofusin 2	1.00	2.17	8.06
NM_013864	N-myc downstream regulated gene 2	1.00	2.72	7.08
NM_080728	myosin, heavy polypeptide 6, cardiac muscle	1.00	4.20	88.11
NM_016798	PDZ and LIM domain 3	1.00	3.44	7.55
NM_008700	NK2 transcription factor related, locus 5	1.00	2.05	15.95
NM_020332	progressive ankylosis	1.00	3.35	7.42
NM_013593	myoglobin	1.00	7.49	164.09
NM_145584	spondin 1, (f-spondin) extracellular matrix	1.00	22.59	100.57
NM_010518	insulin-like growth factor binding protein 5	1.00	113.32	283.56
XM_485171	myosin, light polypeptide 9, regulatory	1.00	32.50	292.93
NM_020496	T-box 20	1.00	26.76	99.84
NM_011817	growth arrest and DNA-damage-inducible 45	1.00	18.14	50.32
NM_008601	microphthalmia-associated transcription factor	2.24	1.00	6.13
NM_011935	estrogen-related receptor gamma	1.00	2.90	8.39
NM_021566	junctophilin 2	1.00	2.59	14.01
NM_008904	peroxisome proliferative activated receptor	1.00	3.04	33.29
NM_172862	Fras1 related extracellular matrix protein 2	1.00	11.24	36.72
NM_008737	neuropilin 1	1.00	2.26	13.03
NM_013808	cysteine and glycine-rich protein 3	1.00	9.40	42.77

Annexure-15

The transcripts involved in MAPK signalling that are upregulated in αMHC$^+$ cardiomyocytes compared to both ES cells and 15day old control EBs and which in turn show diffrential expression among ES cella and control EBs. The fold change is expressed relative to the lowest value of expression that is normalized to 1.

Refseq	Genes	ES cells	15d control EBs	α MHC$^+$
NM_009393	troponin C, cardiac/slow skeletal	1	179.1639	561.1686
NM_023129	phospholamban	1	11.68295	147.3904
NM_011406	solute carrier family 8 (sodium/calcium exchanger), member 1	1	5.729582	83.51378
NM_023868	ryanodine receptor 2, cardiac	1	4.845286	71.77675
NM_175441	RIKEN cDNA D830007F02 gene	1	4.465861	53.47837
NM_009722	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2	1	7.55059	21.29664
NM_021415	calcium channel, voltage-dependent, T type, alpha 1H subunit	1	5.802638	17.26068
NM_009781	calcium channel, voltage-dependent, L type, alpha 1C subunit	1	2.16628	16.40157
NM_028981	calcium channel, voltage-dependent, L type, alpha 1D subunit	1	2.8751	8.586728
NM_008854	Protein kinase, cAMP dependent, catalytic, alpha, mRNA (cDNA clone MGC:6169 IMAGE:3497908)	1	2.109265	5.325959

Annexure-16

Development related transctirpts downregulated in αMHC^{+} cardiomyocytes compared to both ES cells and 15day old control EBs and which in turn show diffrential expression among ES cella and control EBs. The fold change is expressed relative to the lowest value of expression that is normalized to 1.

Refseq	Genes	ES cells	15d control	α MHC^{+}
NM_019768	mortality factor 4 like 2	2.3827	2.958439	1
NM_019641	stathmin 1	3.665087	2.471618	1
NM_011896	sprouty homolog 1	4.695359	7.37134	1
NM_011418	SWI/SNF related, matrix	3.678515	2.537717	1
NM_011379	signal-induced proliferation	2.567472	3.708008	1
NM_011370	cytoplasmic FMR1	3.910966	3.667405	1
NM_010798	macrophage migration	2.028321	2.090147	1
NM_020006	CDC42 effector protein (Rho	2.308123	3.62208	1
NM_007971	enhancer of zeste homolog 2	5.537979	3.649182	1
NM_019676	phospholipase C, delta 1	2.000191	2.975177	1
NM_011801	craniofacial development	2.291356	2.132353	1
NM_020581	angiopoietin-like 4	3.244748	2.817706	1
NM_008709	v-myc myelocytomatosis	7.085321	5.742079	1
NM_007892	E2F transcription factor 5	2.075681	2.942454	1
NM_011524	transforming, acidic coiled-	13.6565	8.250881	1
NM_027324	sideroflexin 1	5.945818	4.819637	1
NM_133762	leucine zipper protein 5	12.29549	8.990721	1
NM_009929	procollagen, type XVIII	4.134619	5.437193	1
NM_009616	a disintegrin and	2.317457	3.697396	1
NM_031877	WASP family 1	3.031342	2.529039	1
NM_008714	Notch gene homolog 1	2.668646	3.994116	1
NM_013690	endothelial-specific receptor	3.957614	7.05991	1
NM_019971	platelet-derived growth	3.171103	3.837322	1
NM_009506	vascular endothelial growth	3.247194	3.516279	1
NM_010067	DNA methyltransferase 2	2.662462	2.787407	1
NM_181327	myosin, heavy polypeptide 9	2.020944	2.373787	1
NM_033608	immunoglobulin superfamily	3.08558	4.373959	1
NM_011035	p21 (CDKN1A)-activated	3.385995	2.963145	1
NM_013471	annexin A4	3.348172	5.559627	1
NM_010590	ajuba	7.858155	4.474931	1
NM_011898	sprouty homolog 4	2.969037	2.137937	1
NM_008713	nitric oxide synthase 3	3.88712	2.938937	1
NM_011917	5'-3' exoribonuclease 2	4.326286	2.268112	1
NM_007554	bone morphogenetic protein	6.715097	3.609592	1
NM_010066	DNA methyltransferase	7.919341	4.175482	1
NM_033524	sprouty protein with EVH-1	2.974918	3.244007	1
NM_021409	par-6 (partitioning defective	2.090105	3.959838	1
NM_009361	transcription factor Dp 1	2.333865	2.6165	1
NM_008245	hematopoietically expressed	3.668262	4.426747	1
NM_008031	fragile X mental retardation	2.730245	2.358289	1

Annexure-16 (continued)

Refseq	Genes	ES cells	15d control EBs	α MHC$^+$
NM_023699	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 4	2.022953	2.995325	1
NM_013685	transcription factor 4	2.250863	2.778532	1
NM_011496	aurora kinase B	12.10002	10.32018	1
NM_009689	baculoviral IAP repeat-containing 5	15.20271	8.3278	1
NM_010734	leukocyte specific transcript 1	4.257557	2.664464	1
NM_144841	orthodenticle homolog 2 (Drosophila)	22.24234	11.34875	1
NM_010136	eomesodermin homolog (Xenopus laevis)	2.686802	2.165604	1
NM_009863	cell division cycle 7 (S. cerevisiae)	10.40622	8.770731	1
NM_009261	spermatid perinuclear RNA binding protein	3.879285	2.596944	1
NM_178598	transgelin 2	4.323942	5.032029	1
NM_010233	fibronectin 1	3.40462	3.832651	1
XM_127565	filamin, beta	2.455687	2.397417	1
NM_008995	peroxisome biogenesis factor 5	2.709981	2.447557	1
NM_011952	mitogen activated protein kinase 3	2.292793	2.6013	1
XM_129658	centromere autoantigen F	7.776469	5.802609	1
NM_008255	3-hydroxy-3-methylglutaryl-Coenzyme A reductase	3.750631	3.472402	1
NM_019413	roundabout homolog 1 (Drosophila)	5.078589	3.458883	1
NM_033320	glucuronyl C5-epimerase	4.0164	2.771649	1
NM_172256	dynein 2 light intermediate chain	2.605158	2.456556	1
NM_008957	patched homolog 1	7.428647	6.097137	1
NM_025881	Luc7 homolog (S. cerevisiae)-like	2.012057	2.410539	1
NM_028006	epsilon-tubulin 1	8.215935	6.593246	1
NM_008722	nucleophosmin 1	3.533406	2.709068	1
NM_009696	apolipoprotein E	14.30636	12.07333	1
NM_013511	erythrocyte protein band 4.1-like 2	3.407721	4.066809	1
NM_008914	protein phosphatase 3, catalytic subunit, beta isoform	2.505594	2.6818	1
NM_008662	myosin VI	3.800492	5.806802	1
NM_021284	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	3.434407	2.719002	1
NM_007488	aryl hydrocarbon receptor nuclear translocator 2	2.013441	3.209756	1
NM_011858	odd Oz/ten-m homolog 4 (Drosophila)	2.212676	2.341704	1
NM_010153	v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)	4.986544	2.673098	1
NM_010939	neuropilin 2	2.331395	3.52589	1
NM_178444	EGF-like domain 7	2.144922	3.349859	1

Annexure-16 (continued)

Refseq	Genes	ES cells	15d control EBs	α MHC$^+$
NM_010422	hexosaminidase B	8.088736	4.208275	1
NM_010774	Methyl-CpG binding domain protein 4, mRNA (cDNA clone MGC:36060 IMAGE:5356016)	3.687663	2.012679	1
NM_011916	5'-3' exoribonuclease 1 (Xrn1), mRNA	3.289551	2.281186	1
NM_010907	nuclear factor of kappa light chain gene enhancer in B-cells inhibitor, alpha	2.194143	2.21978	1
NM_019424	Hermansky-Pudlak syndrome 1 homolog (human)	2.851202	2.343379	1
NM_009376	tetratricopeptide repeat domain 10	3.291793	2.176994	1
NM_008021	forkhead box M1	2.956777	2.062402	1
NM_007795	cardiotrophin 1	2.075084	2.550527	1
NM_010929	Notch gene homolog 4 (Drosophila)	2.593436	2.678144	1
NM_011976	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4G	2.023384	2.20944	1
NM_008057	frizzled homolog 7 (Drosophila)	7.683741	5.192648	1
NM_030732	transducin (beta)-like 1X-linked receptor 1	2.045006	2.036298	1
NM_021383	rcd1 (required for cell differentiation) homolog 1 (S. pombe)	5.690296	3.416921	1
NM_025564	RIKEN cDNA 2010012C16 gene	3.312168	2.341818	1
NM_017405	liver-specific bHLH-Zip transcription factor	2.49742	3.485917	1
NM_020331	general transcription factor II I repeat domain-containing 1	2.347795	2.465841	1
NM_010472	HIV-1 Rev binding protein	5.916857	3.363449	1
NM_030147	bromodomain containing 8	2.167344	2.16721	1
NM_021099	kit oncogene	11.94732	10.62623	1
NM_145990	CDK5 regulatory subunit associated protein 2	2.899983	2.186952	1
NM_019830	heterogeneous nuclear ribonucleoproteins methyltransferase-like 2 (S. cerevisiae)	3.635778	2.852337	1
NM_008021; NM_018858; XM_132869	forkhead box M1; phosphatidylethanolamine binding protein; RIKEN cDNA 4933413G19 gene	7.520145	3.884855	1
NM_008342	insulin-like growth factor binding protein 2	23.85178	14.84272	1
NM_011623	topoisomerase (DNA) II alpha	19.12018	11.34437	1
NM_011759	zinc finger protein 41	3.585899	3.539256	1
NM_175930	Rap guanine nucleotide exchange factor (GEF) 5	2.501675	3.914852	1
NM_008449	kinesin family member 5C	4.011012	3.035083	1
NM_020259	Hedgehog-interacting protein	3.825948	3.437845	1
XM_109868	tensin 3	7.237513	3.782385	1
NM_022987	Opr	10.61627	6.16661	1
NM_010164	eyes absent 1 homolog (Drosophila)	3.689654	4.226483	1
NM_153138	Wiskott-Aldrich syndrome protein interacting protein	3.001951	5.039865	1

Annexure-17

Developmental related transcripts that are upregulated in α MHC $^{+}$ cells more than the ES cells but downregulated than 15day EBs. The fold change is expressed relative to the lowest value of expression that is normalized to 1.

Refseq	Genes	ES cells	15d control EBs	α MHC^{+}
NM_013598	kit ligand	1.00	21.97	7.97
NM_008973	pleiotrophin	1.00	15.14	2.63
NM_010825	myeloid ecotropic viral integration site-related gene 1	1.00	16.59	5.53
NM_019521	growth arrest specific 6	1.00	12.18	3.24
NM_001012477	chemokine (C-X-C motif) ligand 12	1.00	35.60	6.75
NM_009876	cyclin-dependent kinase inhibitor 1C (P57)	1.00	28.29	11.56
NM_008520	latent transforming growth factor beta binding protein 3	1.00	7.45	3.51
NM_010151	nuclear receptor subfamily 2, group F, member 1	1.00	37.98	9.42
NM_008259	forkhead box A1	1.00	26.40	2.33
NM_011658	twist gene homolog 1 (Drosophila)	1.00	12.75	3.81
NM_026162	plexin domain containing 2	1.00	6.72	2.85
NM_007585	annexin A2	1.00	8.89	2.15
NM_008813	ectonucleotide pyrophosphatase/phosphodiesterase 1	1.00	4.97	2.27
NM_008495	lectin, galactose binding, soluble 1	1.00	9.82	2.45
NM_010111	ephrin B2	1.00	9.84	3.41
NM_001008702; NM_023118	disabled homolog 2 (Drosophila)	1.00	45.13	5.28
NM_011580	thrombospondin 1	1.00	112.16	3.71
NM_010496	inhibitor of DNA binding 2	1.00	7.86	3.69
NM_015753	zinc finger homeobox 1b	1.00	43.18	17.61
NM_009821	runt related transcription factor 1	1.00	10.32	2.32
NM_130448	protocadherin 18	1.00	26.56	3.87
NM_009573	zinc finger protein of the cerebellum 1	1.00	41.27	2.83
NM_015784	periostin, osteoblast specific factor	1.00	130.64	31.78
NM_010517	insulin-like growth factor binding protein 4	1.00	60.34	2.63

Annexure-17(continued)

Refseq	Genes	ES cells	15d control EBs	α MHC$^+$
NM_009851	CD44 antigen	1.00	22.93	5.77
NM_008482	laminin B1 subunit 1	1.00	5.87	2.70
NM_178804	slit homolog 2 (Drosophila)	1.00	21.73	2.37
NM_011098	paired-like homeodomain transcription factor 2	1.00	13.71	4.77
NM_144783	Wilms tumor homolog	1.00	7.34	2.04
NM_009371	transforming growth factor, beta receptor II	1.00	19.97	2.15
NM_011544	transcription factor 12	1.00	4.54	2.06
NM_013657	sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3C	1.00	26.93	11.16
---	Transcribed locus	1.00	10.90	5.00
NM_008687	nuclear factor I/B	1.00	4.47	2.15
NM_009524	wingless-related MMTV integration site 5A	1.00	19.80	5.26
NM_017464	neural precursor cell expressed, developmentally down-regulated gene 9	1.00	15.73	2.40
NM_021457	frizzled homolog 1 (Drosophila)	1.00	29.16	6.83
NM_007904	endothelin receptor type B	1.00	40.31	2.48
NM_010512	Insulin-like growth factor 1	1.00	36.10	2.24
NM_009154	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A	1.00	13.29	5.21
NM_007963	ecotropic viral integration site 1	1.00	14.66	2.09
NM_011415	snail homolog 2 (Drosophila)	1.00	15.10	2.86
NM_010514	insulin-like growth factor 2	1.00	264.94	113.88
NM_010417;	hephaestin	1.00	14.34	2.52
NM_019932	chemokine (C-X-C motif) ligand 4	1.00	33.96	8.95
NM_007483	ras homolog gene family, member B	1.00	15.30	4.90
NM_010612	kinase insert domain protein receptor	1.00	14.87	2.05
NM_009472	unc-5 homolog C (C. elegans)	1.00	12.73	2.40
NM_021459	ISL1 transcription factor, LIM/homeodomain (islet 1)	1.00	7.34	2.33
NM_016697	glypican 3	1.00	11.59	2.46
NM_011756	zinc finger protein 36	1.00	13.73	5.44
NM_175260	myosin, heavy polypeptide 10, non-muscle	1.00	4.19	2.03
NM_153319	angiomotin	1.00	64.44	12.68
NM_011243	retinoic acid receptor, beta	1.00	29.81	4.00
NM_016707	B-cell CLL/lymphoma 11A (zinc finger protein)	1.00	4.03	2.00
NM_007441	aristaless 3	1.00	7.28	3.50

Annexure-18

Developmental related transcripts that are upregulated in α MHC $^{+}$ more than the 15d control EBs but downregulated than the ES cells. The fold change is expressed relative to the lowest value of expression that is normalized to 1.

Refseq	Genes	ES cells	15d control EBs	α MHC^{+}
NM_009447	tubulin, alpha 4	6.15	1.00	2.98
NM_019812	sirtuin 1 ((silent mating type information regulation 2, homolog) 1 (S. cerevisiae)	6.24	1.00	2.27
NM_031386	testis expressed gene 14	14.05	1.00	2.05
XM_485238	hypothetical LOC433593	6.36	1.00	2.82
NM_023844	junction adhesion molecule 2	12.96	1.00	2.13
XM_488684	RIKEN cDNA D630045M09 gene	12.83	1.00	3.73
NR_001463	inactive X specific transcripts	8.07	1.00	2.18
NM_025768	GH regulated TBC protein 1, mRNA (cDNA clone MGC:27905 IMAGE:3500563)	7.73	1.00	3.44
---	DNA segment, Chr 14, ERATO Doi 725, expressed	4.64	1.00	2.12
---	hypothetical protein 9630027E11	8.94	1.00	2.15
XM_130232	PREDICTED: nebulin [Mus musculus], mRNA sequence	6.58	1.00	2.36
NM_016754	myosin light chain, phosphorylatable, fast skeletal muscle	15.54	1.00	2.36
NM_009477	uridine phosphorylase 1	5.45	1.00	2.41
NM_030697	ankyrin repeat domain 47	5.01	1.00	2.41
NM_146100	internexin neuronal intermediate filament protein, alpha	5.06	1.00	2.33
NM_022409	zinc finger protein 296	23.45	1.00	2.44
---	RIKEN cDNA 2310043M15 gene	7.82	1.00	3.33
NM_001003918	Ubiquitin specific peptidase 7 (Usp7), mRNA	4.19	1.00	2.00
---	RIKEN cDNA A230083H22 gene	6.85	1.00	2.36
---	16 days embryo head cDNA, RIKEN full-length enriched library, clone:C130032P17 product:unclassifiable, full insert sequence	10.15	1.00	2.34
---	DNA segment, Chr 13, ERATO Doi 787, expressed	7.19	1.00	2.37

Annexure-19

NCBI Sequence Viewer v2.0

NCBI Entrez Nucleotide

My NCBI [Sign In] [Register]

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Nucleotide for Go Clear

Limits Preview/Index History Clipboard Details

Display GenBank Show 5 Send to Hide: sequence all but gene, CDS and mRNA features

Range: from begin to end Reverse complemented strand Features: + Refresh

1: NM_001037926. Reports Mus musculus simi...[gi:85702313] Links

Comment Features Sequence

```
LOCUS       NM_001037926            2825 bp    mRNA    linear   ROD 18-NOV-2006
DEFINITION  Mus musculus similar to zinc finger protein 665 (MGC117846), mRNA.
ACCESSION   NM_001037926 XM_485208 XM_894676 XM_901790 XM_901798 XM_901800
            XM_901804 XM_901807 XM_901809 XM_901813 XM_901815 XM_901817
            XM_901820
VERSION     NM_001037926.1  GI:85702313
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2825)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  CONSRTM   Mammalian Gene Collection Program Team
  TITLE     Generation and initial analysis of more than 15,000 full-length
```

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=85702313 (1 of 3)20/02/2007 15:04:19

Annexure-19 (continued)

NCBI Sequence Viewer v2.0

human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED 12477932
COMMENT PREDICTED REFSEQ: The mRNA record is supported by experimental evidence; however, the coding sequence is predicted. The reference sequence was derived from BC108336.1.
On or before Jan 26, 2006 this sequence version replaced gi:82895782, gi:82895784, gi:82895786, gi:82895788, gi:82895790, gi:82895792, gi:82895794, gi:82895796, gi:82895798, gi:82895800, gi:82895802, gi:82895804.

FEATURES Location/Qualifiers
source 1..2825
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/chromosome="4"
gene 1..2825
/gene="MGC117846"
/note="similar to zinc finger protein 665"
/db_xref="GeneID:626316"
CDS 188..2089
/gene="MGC117846"
/codon_start=1
/product="hypothetical protein LOC626316"
/protein_id="NP_001033015.1"
/db_xref="GI:85702314"
/db_xref="GeneID:626316"
/translation="MSVCLVNTPQGLLTFKDVALDFSLEEWECLSFAQRSLYMDVMLE
NYNNLLFVENHCIRGKNGKVLDQDSQYIVHEHMNIQEKSSKWDKLSNVILESLQCTPY
KTNHSSDALQFSNQKRLKPRNTKEVCKYNDSVNSLSLFSTISLNQGINMQKKKHNRNA
ELDKIFVSKHKVMVKKDNIAGNPYKCSEFHKYLTQRENLQSQQRIYHGKKPYKSSKSD
KCFTHQIHLSIHQEIHTEEKIYKCSECDKCFKNKFNLTMHQRIHTGEKPYKCSECDKC
FTKQSNLSIHQRIHTGEKPYKCSECDKCFTQESYLSIHQRIHTGEKPYKCSECGKCFT
EKSSLRIHQRIHTGEKPYKCSECDKCFTQQSHLSIHQRIHTGEKPYKCSECDKCFTQQ
SNLSIHQRIHTGEKPYKCSECGKCFTEKSSLRIHQRIHTGEKPYKCSECGKCFTEKSS
LRIHQRIHTGEKPYKCNECGKCFTEKSSLRIHQRIHTGEKPYKCSECGKCFTVKSSLR
IHQKIHTGEKPYKCSECDKCFTKPSHLSIHQRIHTGEKPYKCSECEKCFNEKNILKIH
QRIHTGERPYKCSECDKCFSRKFHLGIHQRIHTGKKPYKCSECDKCFTQKSHLSIHQK
IHTGEKPYK"

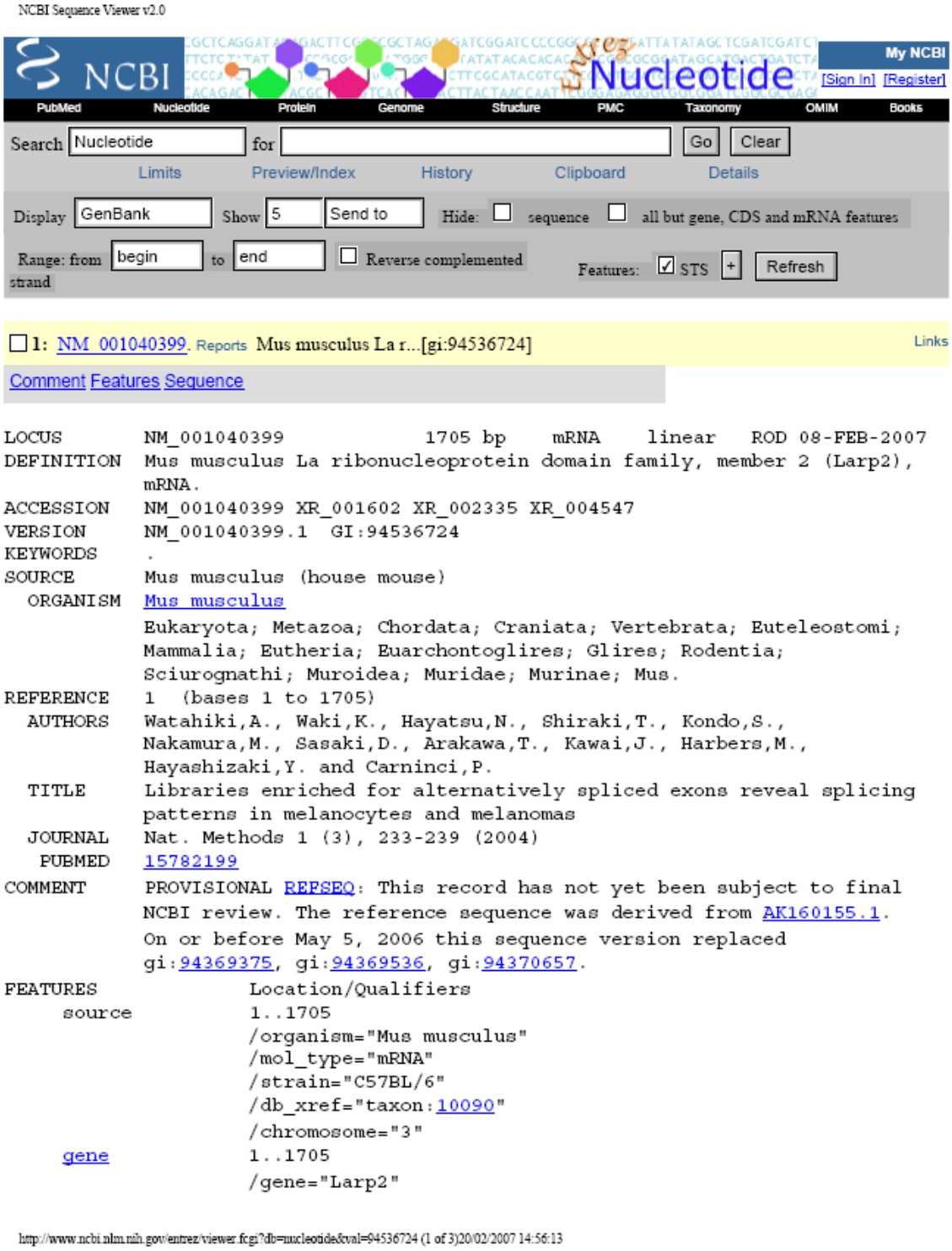
ORIGIN
1 agaggtaatc cgagactagt gcctagtgtt tgcagaacct gaaaacttcc ttaaactaga
61 gttgctgtgc tggcctccta ggatattcag cagctggtcc tgtcaccttc cagaccttga
121 tgtgaaattt caagaagaat tcaaactttg ttcttcactg gaccctggga gaaggacaac
181 aacaagaatg agtgtgtgtc tggtaaatac accacagggt ctgttaactt tcaaggacgt
241 ggccttggac ttctcactgg aggaatggga atgctcagt tttgctcaga ggtcattgta
301 catggatgtg atgttggaga attacaacaa tctgttgttt gtggaaaatc attgcatacg
361 tggaaagaat gggaaggtat tggatcaaga cagccagtac attgtccatg aacacatgaa

Annexure-19 (continued)

NCBI Sequence Viewer v2.0

```
     421 tattcaagag aagtcttcta aatgggacaa gcttagcaat gtcattcttg aatccctcca
     481 gtgtacacct tacaaaacta atcacagctc tgatgctctt caattctcga accaaaaaag
     541 acttaagccc aggaacacta aagaagtttg caaatacaat gactctgtaa acagtttaag
     601 tttgtttttct accattagtc tcaatcaagg aattaacatg cagaagaaaa aacacaaccg
     661 aaatgcagaa ttagacaaga tttttgtttc taaacataaa gtcatggtga aaaaagacaa
     721 tattgcaggg aacccttaca aatgtagtga atttcacaaa tatcttaccc agagagaaaa
     781 cctacaaagt cagcagagaa tttatcatgg aaagaaacct tacaaaagta gtaaaagtga
     841 caaatgcttt acccatcaga tccatcttag tattcatcag gaaattcata cagaagagaa
     901 aatttacaaa tgtagtgaat gtgacaaatg ttttaaaaac aaattcaatc ttactatgca
     961 tcagagaatt catacaggag agaaacctta caaatgcagt gaatgtgaca aatgctttac
    1021 caaacaatcc aatcttagta ttcatcagag aattcataca ggagagaaac cttacaaatg
    1081 cagtgaatgt gacaaatgct ttacccaaga atcctatctt agtattcatc agagaattca
    1141 tacaggagag aaaccttaca aatgcagtga atgtggcaaa tgctttactg aaaaaagcag
    1201 tctgagaatt catcagagaa ttcatacagg agagaaacct tacaaatgca gtgaatgtga
    1261 caaatgcttt acccaacaat cccatcttag tattcatcag agaattcata caggagagaa
    1321 accttacaaa tgcagtgaat gtgacaaatg ctttacccaa caatccaatc ttagtattca
    1381 tcagagaatt catacaggag agaaacctta caaatgcagt gaatgtggca aatgctttac
    1441 tgaaaaaagc agtctgagaa ttcatcagag aattcataca ggagagaaac cttacaaatg
    1501 cagtgaatgt ggcaaatgct ttactgaaaa aagcagtctg agaattcatc agagaattca
    1561 tacaggagag aaaccttaca aatgcaatga atgtggcaaa tgctttactg aaaaaagcag
    1621 tctgagaatt catcagagaa ttcatacagg agagaaacct tacaaatgca gtgaatgtgg
    1681 caaatgcttt actgtaaaaa gcagtctgag aattcatcag aaaattcata caggagagaa
    1741 accttacaaa tgcagtgaat gtgacaaatg ctttaccaaa ccatcccatc ttagtattca
    1801 tcagagaatt catacaggag agaaacctta caaatgcagt gaatgtgaaa aatgctttaa
    1861 tgaaaaaaac attctgaaaa ttcatcagag aattcataca ggagagagac cttacaaatg
    1921 tagtgaatgt gacaaatgct tttcccgaaa attccatctt ggtattcatc agagaattca
    1981 tacaggaaag aaaccttaca aatgtagtga atgtgacaaa tgctttaccc agaaatccca
    2041 tcttagtatt catcagaaaa ttcatacagg agagaaacct tacaaataga atgtatgtga
    2101 caaatacaat actcacaatg tcagtcttag aaatcatcag ttatttcata cagtagagaa
    2161 tctttataaa tgaagtgaac gagacaatcc tttaccaaga tatatcatct caaaattcat
    2221 tagagaatac ttagagaaga caaacttttc aaatatattg taggtatgaa tcctttcctt
    2281 ttggcctatt tcctagactg taacagtgaa tacttacagg ctataaactt aactcatgca
    2341 ttgtatgaag aaatatcttt gacttcattt caggtcagag agaacatcta agaacttttt
    2401 agggtaaagt atttctgttg gggaaaatgt tgtattcttg acaacacttg aggctccaca
    2461 ctagagaatt agtatgatta tgaaaatctt gtgaaaatat tcatgcaatg ttcaaatctt
    2521 agataatgtc aactatttat gttgggagaa actctaaaaa tgtgaaaatg tagaaagctt
    2581 ttcatgacac ctttttactt gttcaccctc aaacgttata agatgaagat aagttaggaa
    2641 gtcagaagat agtattgtat aacactttca ttgagacaga aataatccat ttcaaaagga
    2701 aactatttaa ggtctttaat acaaccctca aaataggtaa taaaactggt agcattttgt
    2761 atatttttat tgtcctgtaa ctaactccaa taaaactgaa agattcaaca aaaaaaaaaa
    2821 aaaaa
//
```

Annexure-20



NCBI Sequence Viewer v2.0

NCBI Nucleotide [Sign In] [Register] My NCBI

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books

Search Nucleotide for Go Clear

Limits Preview/Index History Clipboard Details

Display GenBank Show 5 Send to Hide: sequence all but gene, CDS and mRNA features

Range: from begin to end Reverse complemented strand Features: STS + Refresh

1: NM_001040399. Reports Mus musculus La r...[gi:94536724] Links

Comment Features Sequence

```
LOCUS       NM_001040399            1705 bp    mRNA    linear   ROD 08-FEB-2007
DEFINITION  Mus musculus La ribonucleoprotein domain family, member 2 (Larp2),
            mRNA.
ACCESSION   NM_001040399 XR_001602 XR_002335 XR_004547
VERSION     NM_001040399.1  GI:94536724
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1705)
  AUTHORS   Watahiki,A., Waki,K., Hayatsu,N., Shiraki,T., Kondo,S.,
            Nakamura,M., Sasaki,D., Arakawa,T., Kawai,J., Harbers,M.,
            Hayashizaki,Y. and Carninci,P.
  TITLE     Libraries enriched for alternatively spliced exons reveal splicing
            patterns in melanocytes and melanomas
  JOURNAL   Nat. Methods 1 (3), 233-239 (2004)
   PUBMED   15782199
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from AK160155.1.
            On or before May 5, 2006 this sequence version replaced
            gi:94369375, gi:94369536, gi:94370657.
FEATURES             Location/Qualifiers
     source          1..1705
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /chromosome="3"
     gene            1..1705
                     /gene="Larp2"
```

http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=nucleotide&val=94536724 (1 of 3)20/02/2007 14:56:13

Annexure-20 (continued)

NCBI Sequence Viewer v2.0

```
                     /note="La ribonucleoprotein domain family, member 2;
                     synonyms: AA690246, FLJ10378, MGC38253, MGC117760,
                     1700108L22Rik, 4933421B21Rik"
                     /db_xref="GeneID:214048"
                     /db_xref="MGI:1914604"
     CDS             132..1139
                     /gene="Larp2"
                     /codon_start=1
                     /product="La ribonucleoprotein domain family, member 2"
                     /protein_id="NP_001035489.1"
                     /db_xref="GI:94536725"
                     /db_xref="GeneID:214048"
                     /db_xref="MGI:1914604"
                     /translation="MANWPTPGELVNTGSQSVINQGNKKPQIRKEKEEKIEKRSNSES
                     KENREAKLDGPTENISEDEAQSSSQRKRANKHKWVPLHLDDVRPDSQERPGSRNSSRC
                     QPEANKPTHSNRKSDTRSWRREREKRDDQDEVSSVRSEGGTIRGSTRGRGRGRGRGRG
                     RGRGNPRLNPDYSYGYREPGERTEQPFSTELNTSMMYYYDDGTGVRVYPVEETLLKEY
                     IKRQIEYYFSTENLERDFFLRRKMDEQGFLPISLIAGFHRVQALTTNLNLILEALKDS
                     TEVEIVDEKMRKKIEPEKWPIPGPPPRNVPQTDFSQLIDCPEFIPGQAFGSHTVRVMI
                     Y"
     STS             443..649
                     /gene="Larp2"
                     /db_xref="UniSTS:236420"
     STS             486..580
                     /gene="Larp2"
                     /standard_name="AA690246"
                     /db_xref="UniSTS:159106"
ORIGIN
        1 tgtccaccgt gggcaccggc gtgccgtcgc ccgctcagga ctccccggag gccgagctca
       61 gttcccccaa aattataaaa gcgggaaaac tcaagacaaa gaaatacaac aaggctagtg
      121 acttcagtga tatggcaaac tggccaacac caggtgaatt ggttaacacg gggtctcaga
      181 gtgttatcaa ccaaggaaat aaaaagccac aaattagaaa agaaaaagaa gagaagattg
      241 aaaaaagaag caacagtgag agcaaagaga accgggaagc aaaactagat ggtcctactg
      301 agaacatcag cgaggatgag gctcagtcaa gcagccagag gaagagagct aataagcaca
      361 aatgggtacc actccactta gatgatgtga gaccagacag ccaggaaaga cctgggtccc
      421 ggaacagctc tagatgtcaa cctgaagcaa ataaaccaac gcacagtaat aggaaaagtg
      481 atacacgaag ttggagacgg gagagagaaa agagagatga tcaagatgaa gtgtccagtg
      541 tgagaagtga gggtggtacc atccgaggtt ccactagagg tcggggacga ggccgggccc
      601 gaggaagagg ccgaggccga gggaaccctc gattgaactt tgattattca tatggttatc
      661 gagagcctgg ggaaaggact gagcagccat tttctacaga gcttaatacc agcatgatgt
      721 attactatga tgatggcaca ggtgtgcgtg tgtatcctgt tgaagaaaca ctgcttaaag
      781 agtacattaa gcgccaaatt gagtattact tcagcacaga gaacttggaa cgggacttct
      841 ttcttcgaag aaagatggac gagcaggggt tcctgcccat ttccctgatt gctggttttc
      901 accgtgtgca ggctctcacc acaaacctta atctcatctt ggaggcactt aaggatagca
      961 cagaagtaga aattgtggat gagaaaatga ggaaaaagat agaaccagaa aaatggccaa
     1021 ttccaggccc tcctccacga aatgtgccac aaacagactt ctctcagttg attgattgtc
     1081 cggagttcat acctggccag gcctttggct cccatacagt cagggtgatg atctactgat
     1141 gctcttggat gtcatatgat cggaatatac agtggagtag gtgactccct gaaataccat
```

Annexure-20 (continued)

NCBI Sequence Viewer v2.0

```
     1201 gtaaatatga gagcaaatcg tgctgcagaa gtgagagtga ccatactgtc ttctgtgcaa
     1261 gaatgtctac gcccaccccc ttgtggagaa ggttgggcca gcctgcagtg gagaaaacca
     1321 cctttacgat aatgatgtga gacagaccag aggactgcgg ttactgtagg agcctctcct
     1381 ttctatgtca ttagcttagg ctgcacgggc tacttcattc atggtgggaa agttaattgt
     1441 tatcattggg gaaatactgc aaattttatc ttttgtttcc aggcatttat actttgccca
     1501 gatcaataaa gcagaaaaag agaggattga gtagcatttc ccaaattgtt aggtcctttg
     1561 ctatctgact tgctctttaa tatagtatac tttgcataac tcacatttgt catttttcac
     1621 atgtactata agatgtatta atgtgaatac ttaaggggta agaaatttaa gttaaataaa
     1681 taagggcatt tttcctttat gcttc
//
```

Annexure-21

NCBI Sequence Viewer v2.0

1: NM_001013765. Reports Mus musculus gene...[gi:118131157] Links

Comment Features Sequence

```
LOCUS       NM_001013765            2276 bp    mRNA    linear   ROD 18-NOV-2006
DEFINITION  Mus musculus gene model 397, (NCBI) (Gm397), mRNA.
ACCESSION   NM_001013765 XM_142517
VERSION     NM_001013765.2  GI:118131157
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
            Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2276)
  AUTHORS   Katayama,S., Tomaru,Y., Kasukawa,T., Waki,K., Nakanishi,M.,
            Nakamura,M., Nishida,H., Yap,C.C., Suzuki,M., Kawai,J., Suzuki,H.,
            Carninci,P., Hayashizaki,Y., Wells,C., Frith,M., Ravasi,T.,
            Pang,K.C., Hallinan,J., Mattick,J., Hume,D.A., Lipovich,L.,
            Batalov,S., Engstrom,P.G., Mizuno,Y., Faghihi,M.A., Sandelin,A.,
            Chalk,A.M., Mottagui-Tabar,S., Liang,Z., Lenhard,B. and
            Wahlestedt,C.
  CONSRTM   RIKEN Genome Exploration Research Group; Genome Science Group
            (Genome Network Project Core Group); FANTOM Consortium
  TITLE     Antisense transcription in the mammalian transcriptome
  JOURNAL   Science 309 (5740), 1564-1566 (2005)
   PUBMED   16141073
COMMENT     PREDICTED REFSEQ: The mRNA record is supported by experimental
            evidence; however, the coding sequence is predicted. The reference
            sequence was derived from AC166368.3.
            On Nov 17, 2006 this sequence version replaced gi:62000641.

            Sequence Note: The RefSeq transcript and protein were derived from
            genomic sequence to make the sequence consistent with the reference
            genome assembly. The genomic coordinates used for the transcript
            record were based on transcript alignments.
```

Annexure-21 (continued)

NCBI Sequence Viewer v2.0

```
FEATURES             Location/Qualifiers
     source          1..2276
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /chromosome="7"
     gene            1..2276
                     /gene="Gm397"
                     /note="gene model 397, (NCBI); synonym: XM_142517"
                     /db_xref="GeneID:245109"
                     /db_xref="MGI:2685243"
     CDS             205..1725
                     /gene="Gm397"
                     /go_component="intracellular; nucleus"
                     /go_function="metal ion binding; nucleic acid binding;
                     transcription factor activity; zinc ion binding"
                     /go_process="regulation of transcription, DNA-dependent"
                     /codon_start=1
                     /product="hypothetical protein LOC245109"
                     /protein_id="NP_001013787.1"
                     /db_xref="GI:62000642"
                     /db_xref="GeneID:245109"
                     /db_xref="MGI:2685243"
                     /translation="MASQQAPAKDLQTNNLEFTPTDSSGVQWAEDISNSPSAQLNFSP
                     SNNGCWATQELQSLWKMFNSWLQPEKQTKEQMISQLVLEQFLLTGHCKDKYALTEKWK
                     ASGSDMRRFMESLTDECLKPPVMVHVSMQGQEALFSENMPLKEVIKLLKQQQSATRPT
                     PDNEQMPVDTTQDRLLATGQENSENECNNSCNATEANVGESCSGNEMDSLLIIQKEQH
                     PEHEEGNVVCQFPHGARRASQGTPSHHVDFPSAPTTADVPMEEQPKDLSRENISEDKN
                     NCYNTSRNAATQVYSGDNIPRNKSDSLFINKRIYHPEPEVGDIPYGVPQDSTRASQGT
                     STCLQESLGECFSENDPREVPGLQSRQEQPISDPVLLGKNHEANLPCESHQKRFCRDA
                     KLYKCEECSRMFKHARSLSSHQRTHLNKKSELLCVTCQKMFKRVSDRRTHEIIHMPEK
                     PFKCSTCEKSFSHKTNLKSHEMIHTGEMPYVCSLCSRRFRQSSTYHRHLRNYHRSD"
ORIGIN
        1 gacacacagt gcctccctgg gcttcttggc atcacccttg aagttcaccg gagaaagcag
       61 tgaggtggag gaataggtaa actttccttc ctagtggtct tgaatgtctt ttacagtaca
      121 tccatcaact gttagcattt tcgtaaagtc acaaaacaga tattaaacta ctatagttga
      181 atctttcaca ccattgtcac cacaatggct tcacagcagg caccagcaaa agaccttcag
      241 accaacaatt tagagtttac tccaactgat agttctggtg tgcagtgggc agaagacatc
      301 tctaactcac caagtgctca gctaaacttt tccccaagta acaatggctg ctgggcaact
      361 caggagctgc aaagtctctg gaagatgttc aactcctggt tgcagccaga aaagcagact
      421 aaggagcaga tgatttctca actggtcttg gagcagtttc tcctcactgg gcactgcaag
      481 gacaagtatg ctttgacaga gaagtggaaa gccagtggta gcgatatgag gagattcatg
      541 gagagtctga ctgatgagtg cttgaagcct cctgtcatgg tccatgtttc aatgcaagga
      601 caagaagccc tcttttctga aaacatgcca ttaaaagaag tcatcaagct tttgaaacaa
      661 cagcaatctg caacaaggcc aacaccagat aatgagcaga tgccagtaga caccacacaa
      721 gatagattat tggccacagg acaagaaaac agtgaaaatg aatgcaacaa ctcttgtaat
      781 gctactgaag caaatgttgg tgaaagctgt agtggaaatg aaatggactc ccttcttatt
```

Annexure-21(continued)

NCBI Sequence Viewer v2.0

```
     841 atccagaaag aacagcaccc tgagcatgaa gaggggaatg ttgtttgtca attccctcat
     901 ggtgccagaa gagcaagtca aggcaccccc agtcatcatg tagacttccc gagtgctccg
     961 actactgccg atgtccccat ggaggaacaa ccaaaggatt tatccagaga aaacatctct
    1021 gaggacaaga acaattgcta taacacttcc agaaatgcag ctactcaagt atatagtggt
    1081 gataatattc ccaggaacaa gtcagactcc cttttcatta acaagagaat atatcatcct
    1141 gagcctgagg tgggagatat tccttatgga gttcctcagg attctacaag agcaagtcaa
    1201 ggaacatcta catgcctgca agagtcactt ggggaatgtt tttctgaaaa cgacccaagg
    1261 gaggtaccag ggttgcagtc taggcaagag cagcctatct ctgatcctgt ccttcttggt
    1321 aagaatcatg aggcaaactt accatgtgaa agtcatcaaa agagattctg tagagatgcc
    1381 aaactataca agtgtgaaga atgttctagg atgttcaaac atgccaggag cctttcatcc
    1441 caccagagaa ctcacctgaa taagaagagt gaattgcttt gtgtcacctg tcagaaaatg
    1501 ttcaaacgag tctctgaccg ccgcacccat gagatcatac acatgccaga aaagcctttc
    1561 aagtgcagca catgtgaaaa gtccttcagc cacaagacca acctgaagtc tcatgagatg
    1621 attcacacag gagaaatgcc ttatgtctgt tccctatgta gccgtcgctt tcgccaatca
    1681 tccacttacc atcgtcacct gaggaattac cacagatctg actgaactat ctaacatcct
    1741 cagcagagac tggtagggct tcagcctcag tatgtcatct tcaaagagag aagaatgttg
    1801 caagtaaatt gtactgtccc aataatgata taacatgctt gtggattgcc acttttatgt
    1861 tttgttttgt tttttatttt gtgtgtgtgt gtatgtaatt ttttgtctgt atttccatat
    1921 ttccacagca taagttatta gaatactttg ctgttaattc ttgagttgct tcttgctttt
    1981 agacagtgtc tttctggttg gcagctttat aaacctgtct ttctggcact agagtttcca
    2041 aacattttct gatctccact tttattctct acagtgttct tgacagaagc ctggcattcc
    2101 ctctgacatt ttctacatgt tggggttttc atcccaagtc ttagggttgc aagttaaatg
    2161 cattgcctct tcagacatct catgccatgt ctactgctta cagttcaaga atatttctct
    2221 acattactag aacgacgttc aaagtggaat aataaataaa taaataatca acaatt
//
```

Teilpublicatione

Publications:

1. **Doss MX**, Chen S, Odenthal M, Wickenhauser C, Balaraman S, Hippler-Altenburg R, Huebner N, Schulz H, Hescheler J, Sachinidis A. **"Transcriptomic and phenotypic analysis of embryonic stem cell-derived BMP-2^{+} lineage cells: an insight into mesodermal patterning"**

   (Communicated to Genome Biology)

2. **Doss MX**, Winkler J, Altenburg R, Huebner N, Schulz H, Pfannkuche K,Hescheler J, Sachinidis A. **"Global transcriptome analysis of murine embryonic stem cells-derived cardiomyocytes."**

   (*Communicated to Genome Biology)*

Curriculum Vitae

Name : Michael Xavier Doss.J

Address : Simmerer Strasse 27,
50935 Cologne.
Germany.

Permanent address : 3/603 Bagath Singh Road,
Paramakudi-623707.
Tamil Nadu.
India.
Email: xavi7677@yahoo.com

Date of Birth : 7th June 1978.

Nationality : Indian

Schooling

1983-1988	Primary School Studies	Paramakudi, India
1988-1993	Secondary School Studies	Paramakudi, India
1993-1995	Higher Secondary School Studies	Paramakudi, India

College/University Education:

1995-1998	Bachelor of Science in Biochemistry	Madurai Kamaraj University, India.
1998-2000	Master of Science in Biotechnology	University of Calicut, India.
2000-2001	Pre-Ph.D course work	Jawaharlal Nehru University, India.

Research

2001-2002 Junior Research Fellow (sponsored by Council for Scientific and Industrial Research, Govt., of India) at National Institute of Virology (NIV), Pune, India.
2002-2003 Ph.D student at Medizinische Hochschule Hannover, Hannover.Germany.
2003-2007 Ph.D at Institute of Neurophysiology, University of Cologne. Germany.

Merits:

1. University 2nd rank in Bachelor of Science in Biochemistry, 1998.
2. Qualified "All India combined entrance exam" for admission into Master of Science in Biotechnology with fellowship from Department of Biotechnology, Govt. of India. 1998.
3. Qualified GATE (Graduate Aptitude Test in Engineering) conducted by Indian Institute of Technology, Govt.of India in 2000.
4. Awarded combined Junior Research Fellowship-cum-Lectureship by Council for Scientific and industrial Research (CSIR), Govt. of India and University Grants Commission (UGC), in 2000.
5. Awarded Junior Research Fellowship by Indian Council of Medical Research, India in 2001.
6. Cleared TOEFL (CAT) (score:253/300) 2001.

Cologne. 01.02.2007 (Michael Xavier Doss.J)

LEBENSLAUF

Name : Michael Xavier Doss.J

Addresse : Simmerer Strasse 27,
50935 Köln.
Deutschland.

Heimataddresse : 3/603 Bagath Singh Road,
Paramakudi-623707.
Tamil Nadu.Indien.
Email: xavi7677@yahoo.com

Geburtsdatum : 7. Juni 1978.

Staatsangehörigkeit : Indisch

Schulausbildung

1983-1988	Grundschule	Paramakudi, Indien
1988-1995	Gymnasium und Abitur	Paramakudi, Indien

Universität Studium

1995-1998	Bachelor of Science in Biochemie	Madurai Kamaraj Universität, Indien
1998-2000	Diplomarbeit in Biotechnologie	Calicut Universität, Indien.
2000-2001	Vorbereitungskurs für Doktoranden	Jawaharlal Nehru Universität, Indien.

Forschung

2001-2002 Forschungsstipendiat (gefördert von dem staatlichen Rat der wissenschaftlichen und industrieellen Forschung (CSIR) am Nationalen Institut für Virologie (NIV) in Pune, Indien.
2002-2003 Doktorand an der Medizinischen Hochschule in Hannover, Deutschland.
2003-2007 Doktorand am Institut für Neurophysiologie, Universität Köln, Deutschland.

Leistungen, Qualifikationen

1. 2. Platz beim Abschluß als Bachelor of Science in Biochemie an der Madurai Kamaraj Universität in Indien, 1998.
2. Qualifizierter staatlicher Aufnahmetest für das Studium als Stipendiat an ausgewählten Universitäten für Biotechnolgie in Indien, 1998
3. Qualifizierter Abschluß des staatlichen, akademischen Eignungstest (GATE) für das Ingenieurwesen, Indien 2000.
4. Staatliche Auszeichnung als Forschungsstipendiat mit Lehrerlaubnis durch den Rat der wissenschaftlichen und industriellen Forschung (CSIR) und der Universitäts Antrags Kommission (UGC), Indien 2000.
5. Staatliche Auszeichnung als Forschungsstipendiat durch den indischen Rat der medizinischen Forschung (ICMR), Indien 2001.
6. TOEFL Test (Ergebnis 253 von 300 Punkten), 2001.

Köln, 01.02.2007. Unterschrift