A Serial Analysis of Gene Expression and an in vitro Based Approach to Understanding Mouse Definitive Endoderm Development

by

Ali Saleem Hassan

B.Sc. (Honours), McGill University, 2006

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Medical Genetics)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

June 2009

© Ali Saleem Hassan, 2009
ABSTRACT

The definitive endoderm is one of the three germ layers of the embryo that are generated during gastrulation and gives rise to the lungs, liver, pancreas and the gut. Understanding the development of the definitive endoderm is crucial for future in vitro based approaches to regenerative medicine for diabetes or liver regeneration.

However, progress in the understanding of definitive endoderm development has been limited by the lack of genetic markers specific to this tissue. To address this, our lab previously performed gene expression profiling of the definitive endoderm using Serial Analysis of Gene Expression. From this study, a number of genes expressed specifically in this tissue, including Nephrocan and Peptide YY, were uncovered. In an attempt to uncover novel genes expressed in the definitive endoderm, an extended study of this gene expression profiling was done, and two ESTs renamed Endy and NAPS, were further identified. Through whole mount in situ hybridization analysis of the early mouse embryo, expression of Endy and NAPS was seen in different cell populations of the developing definitive endoderm.

In addition to the SAGE analysis, an ES cell differentiation system for definitive endoderm was set up, as a tool to study some of the early signaling pathways leading to the development of the definitive endoderm. Differentiation of mES cells using Activin induced expression of the novel definitive endoderm markers, Nepn and Pyy, in a temporal manner. Inhibition of TGFβ signaling during differentiation resulted in a significant down regulation of these genes. Furthermore, differentiation of mES cells mutant for TGFβ signaling factor Foxh1 revealed an expression pattern for Nepn and Pyy that was inconsistent with what is observed in vivo. Lastly, the ES cell differentiation system was also used to test expression of Endy and NAPS as little molecular information existed for these ESTs.
Collectively, the characterization of these markers in vivo, and manipulations of the ES cell differentiation system to definitive endoderm will facilitate the creation of more accurate fate maps of the definitive endoderm, and address some of the questions regarding early lineage decisions during specification and patterning of this tissue.
# TABLE OF CONTENTS

ABSTRACT .......................................................................................................... ii  
TABLE OF CONTENTS ...................................................................................... iv  
LIST OF TABLES ................................................................................................ vi  
LIST OF FIGURES ............................................................................................. vii  
ACKNOWLEDGEMENTS .................................................................................. ix  

CHAPTER 1 INTRODUCTION ............................................................................ 1  
1.1. Gastrulation and the formation of the definitive endoderm ......................... 1  
   1.1.1. Pre-implantation stages of mouse embryonic development .................... 1  
   1.1.2. Post-implantation (gastrulation) stages of mouse embryonic development ... 2  
   1.1.3. The mouse embryonic organizer ............................................................ 4  
       1.1.3.1. Distinguishing the node and the posterior notochord ....................... 5  
   1.1.4. The definitive endoderm of the mouse .................................................. 7  
       1.1.4.1. Regionalization of the definitive endoderm ...................................... 8  
1.2. The TGFβ/Nodal signaling pathway .......................................................... 10  
1.3. A need for markers specific for definitive endoderm .................................... 13  
   1.3.1. Serial analysis of gene expression to identify novel markers for the definitive endoderm .......................................................... 14  
1.4. Embryonic stem cell differentiation to definitive endoderm ......................... 16  
1.5. Thesis objectives .......................................................................................... 17  

CHAPTER 2 MATERIALS AND METHODS ....................................................18  
2.1. SAGE data analysis ................................................................................. 18  
2.2. ES cell lines ............................................................................................... 19  
2.3. Maintenance and differentiation of ES cells .............................................. 19  
2.4. Reverse transcription polymerase chain reaction (RT-PCR) ....................... 20  
2.5. Whole mount in situ hybridization ............................................................ 21  
2.6. Histological analysis ................................................................................... 23  

CHAPTER 3 RESULTS .......................................................................................24  
3.1. Identification of non-annotated genes enriched for definitive endoderm using serial analysis of gene expression ......................................................... 24  
   3.1.1. Creation of a list of novel tags enriched for expression in definitive endoderm .... 25  
   3.1.2. Tags for ESTs E ndy and NAPS were enriched in the definitive endoderm LongSAGE libraries .......................................................... 27  
       3.1.2.1. Endy is predicted to be protein coding, but may be a non-coding RNA gene or a pseudogene .................................................. 29  
       3.1.2.2. NAPS could be a type of non-coding RNA ..................................... 32
3.1.3. Expression of Endy and NAPS during post-implantation stages of mouse definitive endoderm development .......................................................... 36
  3.1.3.1. Distribution of Endy expression in LongSAGE libraries .............. 36
  3.1.3.2. Transition of Endy expression from specific epiblast cells to the definitive endoderm ................................................................. 38
  3.1.3.3. Distribution of NAPS expression in the LongSAGE libraries ....... 40
  3.1.3.4. NAPS is expressed in cells of the node and the PNC .................. 41
  3.1.3.5. NAPS expression remains unchanged in Hippi−/− mouse embryos 44
3.2.1. Markers used to characterize ES cell differentiation system .......... 46
3.2.2. Definitive endoderm markers identified through the SAGE analysis are expressed in ES cells differentiated to definitive endoderm ................................................. 48
3.2.3. Blocking TGFβ signaling inhibited ES cell differentiation to definitive endoderm ...................................................................................... 53
3.2.4. Nepn expression is lost in ES cell differentiation of Foxh1−/− ES cells ...... 57

CHAPTER 4 DISCUSSION ................................................................................. 62

4.1. Endy and NAPS expression mirrors lineage and fate maps of the definitive endoderm and the node, respectively ......................................................... 62
4.2. Endy and NAPS as non-coding RNA genes ........................................ 64
4.3. Expression patterns of Nepn and Pyy in our ES cell system suggest formation of populations of definitive endoderm ......................................................... 66
4.4. Insights and conclusions ........................................................................ 70
  4.4.1. SAGE as tool to study early mouse development ............................. 70
  4.4.2. ES cell differentiation as model to study early mammalian lineage commitment decisions .............................................................. 72

REFERENCES .................................................................................................. 74
APPENDIX ....................................................................................................... 81
LIST OF TABLES

Table 1. Distribution and T/L ratio values for 42 tag types retrieved through SAGE analysis ...28
LIST OF FIGURES

Figure 1. Schematic diagrams of mouse embryos representing pre- and post implantation stages.3
Figure 2. Diagram depicting the mouse gastrula organizer located at the anterior end of the primitive streak (APS)........................................................................................................................... 6
Figure 3. Formation and development of the mouse definitive endoderm........................................... 9
Figure 4. Simplified diagram of the Nodal signaling pathway......................................................... 11
Figure 5. Overview of how list of non-annotated tags in the definitive endoderm SAGE libraries was selected. ........................................................................................................................................ 26
Figure 6. UCSC genome browser window for Endy (AK014119)...................................................... 31
Figure 7. UCSC Genome Browser window for NAPS (AK084355).................................................. 33
Figure 8. Prediction of potential pre-miRNA stem loops in the NAPS transcript. ......................... 35
Figure 9. Distribution of Endy transcripts in SAGE libraries......................................................... 37
Figure 10. In situ hybridization analysis of Endy during early post-implantation stages of the mouse embryo.......................................................................................................................... 39
Figure 11. Distribution of NAPS transcripts in the SAGE libraries.................................................. 41
Figure 12. In situ analysis of NAPS and Nodal in the early post-implantation stages of the mouse embryo........................................................................................................................................ 43
Figure 13. Expression of NAPS in wildtype versus Hippi-/- embryos ............................................. 45
Figure 14. RT-PCR analysis of CCE and LP9 mouse ES cells differentiated as EBs for 9 days in the presence or absence of Activin......................................................................................................................... 50
Figure 15. RT-PCR analysis of CCE ES cells differentiated as EBs for 9 days in the presence of TGFβ signaling inhibitor SB-431542............................................................................................................. 56
Figure 16. RT-PCR analysis of FKO2 (Foxh1+/+) and FKO3 (Foxh1-/-) ES cells differentiated through EBs for 13 days in the presence and absence of Activin.................................................. 61
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>APS</td>
<td>Anterior primitive streak</td>
</tr>
<tr>
<td>A-P</td>
<td>Anterior-posterior</td>
</tr>
<tr>
<td>ES cell</td>
<td>Embryonic stem cell</td>
</tr>
<tr>
<td>EGF-CFC</td>
<td>Epidermal growth factor – Crypto/FRL1/Cryptic</td>
</tr>
<tr>
<td>ICM</td>
<td>Inner cell mass</td>
</tr>
<tr>
<td>MGC</td>
<td>Mammalian gene collection</td>
</tr>
<tr>
<td>ORF</td>
<td>Open reading frame</td>
</tr>
<tr>
<td>PNC</td>
<td>Posterior notochord</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>SAGE</td>
<td>Serial analysis of gene expression</td>
</tr>
<tr>
<td>SINE</td>
<td>Short interspersed nuclear elements</td>
</tr>
<tr>
<td>TGFβ</td>
<td>Transforming growth factor β</td>
</tr>
<tr>
<td>TS</td>
<td>Theiler stage</td>
</tr>
<tr>
<td>WISH</td>
<td>Whole mount in situ hybridization</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I would like to thank my family, my parents Saleem and Nadiya Hassan, and my sister Aafreen Hassan for their continuous and unconditional support throughout the years. Thank you for always being there during my peaks and lows. A special thank you goes to Jeannine Simon, who has been an epitome of support, strength and encouragement. Thank you for being a friend and my second family.

I would like to heartily thank my research supervisor, Dr. Pamela Hoodless, for giving me the opportunity to work with her. Thank you for your infinite wisdom, encouragement, knowledge and patience. Thank you for having faith in me during times when I felt I had lost faith in myself. To my thesis committee, Dr. Connie Eaves and Dr. Keith Humphries, thank you for your invaluable help during the course of my research.

An infinite amount of thanks goes to the members of the Hoodless Lab, working with whom has been an absolute pleasure. I take my hat off to Rebecca Cullum and Rachel Montpetit, two amazing research assistants, for being the structural core of the lab, and for having kept the lab together. I would like to thank my fellow students Sam Lee and Pavle Vrljicak for all the reflections, philosophical and psychological discussions, and the what-ifs and the what-if-nots of the sinusoidal curve that has been graduate school. For their immense knowledge, experience, suggestions and help I would like to thank Drs Juan Hou, Kristen McKnight and Elizabeth Wederell. The experience of working with all of you has been irreplaceable. Thank you for welcoming my idiosyncrasies, and in turn sharing some of yours. The fun filled attitude and atmosphere that you all created rang deep in the heart and gut of my motivation.

I would like to acknowledge Anita Charters from the Genome Science Centre, British Columbia, for generating the list of SAGE tags that was used in the analysis of this thesis. I
would also like acknowledge Florian Kuchenbauer for pointing me in the right direction for the miRNA component of my thesis, and I would like to thank Danny Chui of the Genetic Modeling Core at the Terry Fox Lab, for the LP9 ES cell line that he derived.

I would also like to add a heartfelt special thank you to Dr Juan Hou. You have been an invaluable and patient mentor. Discussions with you inspired me and motivated me in my research endeavors. You always had time to answer my unrelenting barrage of questions, and always seemed to have a solution to my problems. You also provided patient guidance for the majority of the techniques used in this thesis. For all your help and wisdom I am eternally grateful.
Chapter 1

INTRODUCTION

1.1. Gastrulation and the formation of the definitive endoderm

1.1.1. Pre-implantation stages of mouse embryonic development

Embryonic development in the mouse can be divided into stages of pre-implantation, post-implantation (gastrulation), organogenesis and growth of the fetus. The pre-implantation stages of the mouse embryo comprise the formation of the 8-cell morula from a single fertilized egg cell. The morula subsequently undergoes compaction and cavitation to form the blastocyst at embryonic day (E) 2.5, underlying the first lineage decision to form an embryo with two distinct cell types: an inner cell mass (ICM) and an outer trophectoderm (reviewed in Beddington and Robertson, 1999). Trophoderm overlying the ICM and the blastocyst cavity are termed the polar trophectoderm and mural trophectoderm, respectively (Figure 1). The trophectoderm goes on to constitute extra-embryonic ectoderm and the placenta whereas the inner cell mass will give rise to the definitive embryo, and extraembryonic mesoderm and endoderm. At E 4.5 certain cells of the inner cell mass are fated to become the primitive endoderm (hypoblast) and line the surface of the ICM. The hypoblast will go on to contribute to the extraembryonic parietal and visceral endoderm, which constitute the yolk sac of the developing embryo. The remainder of the ICM, now known as the epiblast, will give rise to the entire fetus, including the germ line and all the extraembryonic mesoderm. The embryo also implants into the uterine wall at E 4.5.
1.1.2. Post-implantation (gastrulation) stages of mouse embryonic development

Prior to gastrulation the mouse embryo develops as a bilaminar cup consisting of the inner epiblast and outer visceral endoderm. Gastrulation commences at E 6.5, whereupon the totipotent cells of the epiblast divide, differentiate and rearrange, converting the embryo into a trilaminar cup consisting of an inner layer of ectoderm, a middle layer of mesoderm, and an outer layer of definitive endoderm (Grapin-Botton and Melton, 2000; Tam and Behringer, 1997; Zernicka-Goetz, 2002). The ectoderm forms the central nervous system and skin, the mesoderm forms the skeleton, connective tissues, muscles, blood and kidneys, and the definitive endoderm will form the epithelial lining of the lungs, stomach, esophagus, intestines, and a major part of many glands including pancreas, thyroid, thymus and the liver. By E 9.0, the embryo rotates, turns, and switches to the traditional fetal shape, bringing the ectoderm to the outside and the endoderm to the inside of the body.

Gastrulation is initiated by recruitment of epiblast cells to a transient structure called the primitive streak (Robb and Tam, 2004; Tam et al., 2006; Tam and Behringer, 1997). The anterior-posterior axis immediately becomes morphologically apparent with the appearance of the primitive streak on the posterior side of the embryo. Cells at the primitive streak undergo epithelial to mesenchymal transition and those that ingress in-between the epiblast and visceral endoderm layers become mesoderm while those that intercalate within the visceral endoderm layer become definitive endoderm. Epiblast cells not recruited through the streak go on to form the ectoderm.
Figure 1. Schematic diagrams of mouse embryos representing pre- and post implantation stages


Figure reprinted with permission from Kristen McKnight’s Doctoral Thesis, 2008
1.1.3. The mouse embryonic organizer

Gastrulation also results in the establishment of the embryonic body plan namely the anterior-posterior (A-P), dorsal-ventral (D-V) and left-right (L-R) axes. Germ layer patterning and axis specification activity has been narrowed down to a population of cells referred to as the organizer. Through some landmark transplantation experiments in the 1920’s, Spemann and Mangold showed that the dorsal lip of the blastopore of salamander embryos had the ability to induce and organize a secondary body axis when transplanted to an ectopic location in a host embryo of another specimen from the same species (reviewed in Sander and Faessler, 2001). Similar structures have been described in various other model organisms including the Hensen’s node in rabbit and chicken and the anterior primitive streak (APS) in the mouse (Boettger et al., 2001; Knoetgen et al., 2000; Robb and Tam, 2004). Cells of the organizer in all studied organisms give rise to the axial mesoderm tissues such as the notochord, which are important in patterning of the embryonic axes. The notochord is a rod-like structure that is laid down along the A-P axis of the embryo. It provides the structural core of the embryo and is an important source of signals that pattern the surrounding tissues including the definitive endoderm. One of its critical functions is to pattern the overlying neural tube and create a D-V axis within the central nervous system.

The existence of the mouse embryonic organizer was first shown, and subsequently corroborated, when cells at the anterior end of the primitive streak induced a secondary axis and differentiation of host tissue upon ectopic transplantation (Beddington, 1994). Lineage tracing and cell ablation studies have shown that cells of the APS give rise to tissues of the midline, specifically the prechordal plate, prechordal mesoderm, anterior definitive endoderm and the notochord (review in Robb and Tam, 2004). The APS also expresses organizer specific
transcription factors such as Foxa2 and Goosecoid giving further evidence for organizer specific activity (Ang and Rossant, 1994; Blum et al., 1992).

Transplantation of the cell populations constituting the APS at different stages of the gastrulating embryo have been shown to have different levels of axis inducing activity (reviewed Robb and Tam, 2004). Thus, based on the different stages of the gastrulating embryo the organizer has been given different names: the early gastrula organizer (EGO) identified in the early streak stage of the embryo, the mid gastrula organizer (MGO) identified at the mid-streak stage, and the node identified at the late streak to early bud stage of the embryo (Figure 2). The MGO is capable of inducing a complete secondary axis with anterior and posterior structures whereas the EGO and the node are only able to induce partial axis structures when transplanted (reviewed Robb and Tam, 2004). Furthermore, these cell populations also differ in gene expression with EGO expressing Foxa2 and Goosecoid, and the MGO and node expressing Brachyury, Shh, Chordin and Noggin in addition to Foxa2 and Goosecoid.

1.1.3.1. Distinguishing the node and the posterior notochord

It is important to distinguish the node, the organizer population of the late-streak embryo from the small indentation that first appears around E7.5 in the distal half of the embryo, anterior to the node, and persists for about 24 hrs (Figure 2). Given the proximity of this indent to the node/organizer it has always been referred to as the “node” and thus has garnered the reputation of being the organizer. This indentation has received additional attention in the past decade as the site where vectorial fluid flow termed “nodal flow” generated by motile monocilia gives rise to a
Cells of the APS at different stages of gastrulation (orange) have been shown to have organizer activity. These cells also give rise to midline tissues of the embryos namely the notochord (pink), the anterior definitive endoderm and prechordal plate (pink) and the floor plate of the neural tube (not shown). The node, the late organizer, at 7.5 is a morphologically distinguishable knot of cells contiguous with the posterior notochord (PNC) depicted by an indentation. The PNC is the site of fluid flow driven left-right determination.

Figure reprinted with permission from Kristen McKnight’s Doctoral Thesis, 2008
proposed morphogen gradient resulting in laterality generation in the mouse (Hirokawa et al., 2006). Recent electron micrographical and immunohistological analyses using Goosecoid have identified the node/organizer population as a slight thickening of cells where the germ layers are indistinguishable at the anterior end of the primitive streak at E7.5 (Blum et al., 2007). The node/organizer cells contain no cilia, but are continuous with the indent that harbors the monociliated cells.

The traditional morphological criterion for the organizer is a center where cells are recruited and exist in a disordered array before they are induced and positioned as they migrate out of the organizer (reviewed in Sander and Faessler, 2001). Electron micrographs of the ciliated indent illustrate a very ordered array of cells distinct from the node/organizer just posterior to it. Furthermore Goosecoid expression was excluded from the cells of the indent, thus suggesting that these cells are not part of the organizer. Lineage tracing analysis has also shown the node to be a source for some parts of the dorsal hindgut and the notochord (Lee and Anderson, 2008). Given that the node lays down cells for the notochord, the indent is now considered to be the very posterior end of the emerging notochord (Blum et al., 2007). Thus, it is postulated that the posterior notochord (PNC) is the site of left-right determination in the mouse, but does not harbor organizer activity. A definitive answer will come when the homologous graft experiments, done at E7.5 to elucidate organizer activity of the node region, are repeated with the physical separation of the APS cells and cells of the supposed PNC.

1.1.4. The definitive endoderm of the mouse

Cells fated to become definitive endoderm are recruited through the primitive streak. As these cells emerge from the streak they intercalate into the visceral endoderm layer expanding to cover the entire embryonic region. In this process the sheet of visceral endoderm cells is
displaced anteriorly and proximally towards the extraembryonic regions of the embryo (Figure 3) (Grapin-Botton and Melton, 2000; Tam et al., 2003; Wells and Melton, 1999). The sheet of definitive endoderm formed then undergoes morphogenesis to form the foregut and hindgut pockets and eventually the entire gut tube. The foregut, at the anterior end of the gut tube, forms first at E8.0 and contributes to a number of organs including the lungs, liver, thyroid, thymus, esophagus, trachea and ventral pancreas. The hindgut derived from the posterior most region of the embryo, forms shortly after and contributes to the small and large intestine, bladder and the urogenital tract.

Interestingly, both the definitive endoderm and visceral endoderm form adsorptive epithelial structures and share genetic signatures during development (Lewis and Tam, 2006). For instance, one of the earliest known markers of endoderm, a Sry related HMG box gene, Sox17, is expressed both in the visceral and definitive endoderm (Kanai-Azuma et al., 2002). Recent fate mapping studies have also highlighted that not all visceral endoderm cells are displaced and that several populations contribute to the formation of the embryonic gut as well (Kwon et al., 2008).

1.1.4.1. Regionalization of the definitive endoderm

The final destination, along the A-P axis, of definitive endoderm cells migrating out of the primitive streak is influenced by the timing of their recruitment into the primitive streak (Franklin et al., 2008; Tam et al., 2003). Cells recruited earlier will form the foregut with subsequent recruitments contributing to progressively posterior regions. Once allocated along the A-P axis interactions with surrounding germ layers are critical for the continued differentiation of the definitive endoderm and for the emergence of organ primordia along the gut axis.
Figure 3. Formation and development of the mouse definitive endoderm

For detailed description see section 1.1.4. The definitive endoderm is formed during gastrulation when newly formed definitive endoderm cells (yellow) emerge from the primitive streak. As more definitive endoderm cells are recruited they intercalate between the visceral endoderm cells (green) displacing them to the proximal region of the embryo. Following gastrulation the endoderm undergoes morphogenesis to form gut tube starting with the foregut in the anterior and the hindgut in the posterior. Organs derived from each segment of the gut tube are shown.

Figure reprinted with permission from Kristen McKnight’s Doctoral Thesis 2008
In contrast to the plethora of information on the developmental fate and movement of the definitive endoderm progenitors, little is known of the genetic underpinnings that regulate the specification and differentiation of this tissue. Signaling of the transforming growth factor β (TGFβ) superfamily of secretory ligands, most notably Nodal, are thought to play a crucial role during the earliest stages of definitive endoderm development (Camus et al., 2006; Schier, 2003; Tam et al., 2006).

### 1.2. The TGFβ/Nodal signaling pathway

The TGFβ signaling pathway is involved in many processes throughout development including cell growth and differentiation, apoptosis and other cellular functions (Massagué, 1998). Gene expression is regulated through this pathway when ligand binds to a heteromeric receptor complex to activate intracellular effectors which subsequently enter the nucleus and interact with DNA binding proteins. The TGFβ pathway ligands and intracellular receptors are divided into two families: The TGFβ/Nodal/Activin family of ligands and their intracellular effectors called receptor Smads (R-Smads) Smad2 and Smad3, and the bone morphogenetic protein (BMP) family of ligands with R-Smads Smad1, Smad5 and Smad8.

The Nodal signaling pathway is required to set up the embryonic axes, to induce mesoderm and endoderm, and also determines left-right asymmetry in the embryo (Hirokawa et al., 2006; Weng and Stemple, 2003; Whitman, 2001). Components of the Nodal pathway are illustrated in Figure 4. The Nodal ligand, like all other ligands of the TGFβ superfamily, binds as a dimer to a receptor complex consisting of type I and type II serine/threonine kinase transmembrane receptors. Nodal acts via the type 1 receptor activin receptor-like kinase 4 (ALK4/ActRIB) and the type II activin receptors ActRII or ActRIIB. Unlike other members of
Figure 4. Simplified diagram of the Nodal signaling pathway

See section 1.2 for detailed description. Adapted from Schier and Schen, 2000
In the superfamily, Nodal also requires an EGF-CFC co-receptors Cripto and Cryptic to activate signal transduction events.

Intracellular signals are activated by TGFβ ligands by assembling two type I and type II receptors and an EGF-CFC co-receptor in the case of Nodal. Ligand first binds the type II receptor which causes the type II receptor to phosphorylate and activate the type I receptor (reviewed in Schier, 2003). The activated type I receptors phosphorylate the intracellular R-Smads increasing their affinity for the common mediator Smad4. The R-Smads then form a complex with Smad4 and translocate into the nucleus whereupon the complex is able to bind tissue specific DNA binding co-factors. The cell type specificity of these co-factors is what elegantly confers the ability of TGFβ signaling to control a wide variety of cellular processes.

A well known Smad binding co-factor is the forkhead transcription factor Foxh1, which was first identified in *Xenopus* and shown to mediate activin and TGFβ dependent activation of the Mix.2 gene (Chen et al., 1996). It was then shown to regulate TGFβ-dependent activation of Goosecoid in the mouse (Labbé et al., 1998). Foxh1 on its own is unable to activate transcription of target genes, but requires binding of the R-Smad/Samd4 complex. Foxh1 expression is first detected throughout the epiblast prior to and during gastrulation with low levels detected in the extraembryonic region as well (Ang et al., 1993). At early somite stages Foxh1 is expressed bilaterally in the lateral plate mesoderm, and subsequently restricted to the heart. By E 10.5 Foxh1 is no longer detected in the embryo. Loss of Foxh1 results in embryonic lethality at E 9.5 and the embryo shows a range of defects resulting from a failure to specify the APS due to a deficiency in Nodal signaling (Hoodless et al., 2001; Yamamoto et al., 2001). The defects range from a loss of anterior structures and midline structures, such as the notochord to a failure in orienting the A-P axis correctly.

Nodal expression is first detected in the inner cell mass of the mouse blastocyst during implantation (E 4.5) and persists until the late primitive streak stage (E7.5) (Schier, 2003). In the
mouse, Nodal has been shown to be essential for the formation of the primitive streak (Conlon et al., 1994). In other model organisms, such as zebrafish, chick, and frog, loss of function experiments have also demonstrated an essential role for Nodal in the induction of mesodermal and endodermal cell fates (Schier, 2003). This suggests the role of Nodal in the induction of mesendoderm is evolutionarily conserved. Genetic manipulations of Nodal levels in the mouse embryo have also illustrated that strength of Nodal signaling determines cell fate decisions in the primitive streak. Notably, highest levels of Nodal are required for the specification of prechordal plate and anterior definitive endoderm progenitors, intermediate levels specify axial mesoderm fates, while low levels are required for paraxial and lateral mesoderm. Further studies of mouse mutants have demonstrated that graded Nodal signaling also governs endoderm cell fate decisions with higher levels of Nodal signaling specifying more anterior fates (Vincent et al., 2003). Embryos with hypomorphic alleles of Nodal or Smad2 exhibit defects in the anterior (foregut) endoderm whereas posterior (hindgut) endoderm remains mostly unaffected. This suggests anterior definitive endoderm fates are sensitive to reduced Nodal activity (Shen, 2007). The role of Nodal activity in midgut and hindgut formation are as yet unclear, however, it is clear that Nodal signaling is required to regulate certain endoderm cell fate decisions and graded levels of Nodal may also govern the spatial contributions of definitive endoderm progenitors to the gut.

1.3. A need for markers specific for definitive endoderm

Understanding of definitive endoderm formation has lagged behind that of ectoderm and mesoderm germ layers. This is in part due to the paucity of markers specific for the definitive endoderm. Many of the genes known to be expressed in the definitive endoderm such as Foxa2,
Hex, and Sox17 are also expressed in the extraembryonic visceral endoderm (Ang et al., 1993; Kanai-Azuma et al., 2002; Keng et al., 1998). Furthermore, these markers are not expressed in the entire definitive endoderm at any point during gastrulation. Thus, there is a strong need for the identification of novel definitive endoderm markers to further our understanding of definitive endoderm formation and development. In addition, the identification of genes exclusively expressed in the definitive endoderm will be useful in characterizing and isolating cell populations produced from embryonic stem cell differentiation protocols to develop potential therapeutic agents.

1.3.1. Serial analysis of gene expression to identify novel markers for the definitive endoderm

Previous studies in the Hoodless laboratory described a systematic screen for genes expressed in the definitive endoderm using Serial Analysis of Gene Expression (SAGE) libraries (Hou et al., 2007). The availability of whole genome sequences of humans and major model organisms such as the mouse, and collection of human and mouse cDNAs and expressed sequence tag (EST) sequences have revolutionized biomedical research. However, there is a continued enterprise to annotate genes to the genome and to create large-scale gene expression screens to eventually simulate biological processes, in conjunction with functional data. SAGE, initially developed for gene expression profiling, has become a powerful tool in transcript detection (Anisimov, 2008; Velculescu et al., 1997). It is a high throughput sequencing based technique that provides a quantitative picture of the transcript distribution within a particular tissue sample at a given developmental stage.
The original SAGE protocol made use of a short sequence of 14 bases that could be assigned to a single gene as a SAGE tag following the 3’ most NlaIII recognition site (5’-CATG-3’) in a transcript sequence. The previous study made use of the improved version of SAGE, termed LongSAGE, which generates tags of 21 bases, thus improving the number of tags that can be assigned to a single genomic position, and enhancing the accuracy of tag-to-gene mapping. In contrast to microarray hybridization techniques, SAGE is a system that does not require prior knowledge of sequences to measure gene transcript levels. In this regard, SAGE not only has the ability to identify genes that have been annotated in various databases, but also has the power to identify novel and uncharacterized genes that may be expressed in the sample of interest.

Previous comparative analysis of three mouse definitive endoderm LongSAGE libraries, against more than 150 mouse LongSAGE libraries from various embryonic stages and tissues, successfully identified several annotated genes as new markers for the definitive endoderm, including Nephrocan (Nepn), Peptide YY (Pyy) (Hou et al., 2007). Nepn is a secreted small leucine rich protein that is specifically expressed in the midgut definitive endoderm, whereas Pyy, a short peptide hormone, is expressed highly in a population of cells in the foregut, which appears to correspond to the liver and ventral pancreas progenitors. This study, however, limited itself to screening for tag sequences that mapped to transcript databases, and omitted a list of non-annotated and potentially novel tag sequences, possibly representing novel genes, enriched for expression in the definitive endoderm. A further extended study of this list is warranted for continued acquisition of specific markers for definitive endoderm to facilitate the molecular understanding of its development.
1.4. Embryonic stem cell differentiation to definitive endoderm

Much of our current knowledge of definitive endoderm induction and lineage decisions is based on findings from studies using model systems such as *Xenopus* and zebrafish. These organisms provide easy access to early embryonic stages of development at a time when these important decisions are taking place. The mouse embryo is much less accessible and limited in tissue availability. The *in vitro* differentiation of embryonic stem (ES) cells provides an attractive alternative to address questions relating to early lineage commitment. Under appropriate conditions in culture, ES cells will differentiate and form aggregates of cells called embryoid bodies (EBs) that have been shown to consist of cells from lineages of all three germ layers (Desbaillets et al., 2000). The majority of *in vitro* studies so far have focused on the development of the mesoderm and ectoderm derivatives, but recently several groups have established culture conditions for the efficient induction of endoderm (Kubo et al., 2004; Tada et al., 2005; Yasunaga et al., 2005). Activin is another member of the TGFβ superfamily of ligands and binds the same receptors as Nodal with the exception of the co-receptor Cripto. It has played an important role in these ES cell differentiation studies due to its ability to form endoderm and mesoderm from ES cells *in vitro*, under different concentrations. Exogenous Nodal has been shown to induce some endoderm and mesoderm *in vitro*, but with considerable less potency (Tada et al., 2005). Activin was initially identified as a potential regulator of early developmental decisions based on its ability to induce mesoderm and endoderm in *Xenopus* animal caps *in vitro* (Ninomiya et al., 1999; Smith et al., 1990). However, its activity in experimental situations does not reflect a role during early development as targeting studies in mice suggest that endogenous Activin does not seem to be required for mesoderm and endoderm induction (Vassalli et al.,
1994). Nonetheless, its propensity to bind the same receptors as Nodal has made it an ideal candidate for \textit{in vitro} based induction of definitive endoderm.

Several studies describing definitive endoderm formation from ES cells have come under scrutiny due to their inability to distinguish the origin of the cell lineage. As previously mentioned, the lack of markers distinguishing visceral and definitive endoderm make it difficult to identify the lineage through which the ES cells are differentiating. Given our success in identifying novel markers for definitive endoderm, establishing an ES cell system to definitive endoderm using our new markers aims to resolve this issue. In addition, the system can be further manipulated to address questions relating to signaling events in the early embryo that initiate the induction and specification of definitive endoderm development.

1.5. Thesis objectives

- Given the paucity of markers specific for the definitive endoderm, the first objective is to identify novel genes enriched in expression in the definitive endoderm during early post implantation mouse development.

- Since ES cell differentiation systems are able to provide access to normally inaccessible early lineages of mouse development, the second objective is to establish and manipulate an ES cell system to definitive endoderm as a step in understanding the molecular mechanisms governing its formation and specification.
Chapter 2

MATERIALS AND METHODS

2.1. SAGE data analysis

Bioinformatic analysis was carried out on definitive endoderm SAGE libraries generated through the Mouse Atlas of Gene Expression project (Siddiqui et al., 2005). SAGE data was analyzed using DiscoverySpace software (Robertson et al., 2007). 21 bp tags were filtered to a 95% sequence quality cut-off and were mapped to the RefSeq transcripts, Ensembl transcripts and Mouse Gene Collection (MGC) databases using DiscoverySpace. Tags that didn’t map to these databases were subsequently mapped to the genome using the UCSC genome browser. Tag ‘position’ was determined by sequentially numbering NlaIII restriction enzyme sites from the 3’-most end of the transcript (position 1) onward (i.e. next 5’ tag would be position 2, and so on). A tag was considered unambiguous if it mapped to a single gene in a sense position and ambiguous if it mapped to multiple genes in a sense position. Tags that were only represented once in the pooled definitive endoderm libraries (single tags) were omitted from the analysis. Online tools and databases used to characterize novel genes are as follows: NCBI BLAST, UCSC Genome Browser, InterPro (Mulder et al., 2005), NCBI ORF Finder and Web Map Preferences (http://pga.mgh.harvard.edu/web_apps/web_map/start).
2.2. ES cell lines

CCE ES cells derived from the mouse strain 129/Sv (Robertson et al., 1986), LP9 ES cells derived from C57BL/6 mice (derived in house by Danny Chui) and FKO2 and FKO3 ES cells derived from Foxh1+/+ and Foxh1−/− 129 mouse strain respectively (Hoodless et al., 2001), were used for in vitro differentiation experiments.

2.3. Maintenance and differentiation of ES cells

Undifferentiated ES cells were maintained on irradiated (100 Grays) primary mouse embryonic fibroblasts in media containing Dulbecco’s Modified Eagle’s Medium (Stem Cell Technologies Inc) supplemented with 15% Fetal Bovine Serum (Stem Cell Technologies Inc), 0.1 mM non-essential amino acids (Stem Cell Technologies Inc), 1 mM sodium pyruvate (Stem Cell Technologies Inc), 2 mM L-glutamine (Stem Cell Technologies Inc), 100 U/ml penicillin, 100 µg/ml Streptomycin (Stem Cell Technologies Inc), 100 µM monothioglycerol and 10 ng/ml murine leukemia inhibitory factor (Invitrogen). Cells were incubated in a 5% CO2-air mixture at 37 °C.

Prior to initiation of differentiation, ES cells were passaged twice on gelatin-coated tissue culture dishes to deplete the population of feeder cells. ES cells were then dissociated using trypsin and seeded at a density of 12,000 cells/ml in 35 mm low adherence tissue culture grade Petri dishes (Stem Cell Technologies Inc) to allow EB formation. The differentiation media consisted of Iscoves’s Modified Dulbecco’s medium supplemented with 10% Fetal Bovine Serum, 0.1 mM non-essential amino acids (Stem Cell Technologies Inc), 1 mM sodium pyruvate.
(Stem Cell Technologies Inc), 2 mM L-glutamine (Stem Cell Technologies Inc), 100U/ml penicillin, 100 µg/ml Streptomycin and 150 µM monothioglycerol. For some experiments recombinant human Activin A (R&D systems) or SB-431542 (Sigma-Aldrich), small molecule inhibitor of TGF β superfamily type 1 receptor, was added to the EB cultures at day 3 of differentiation (Activin A at 20 ng/ml and SB-431542 at 10 µM). Cells were collected at selected days into TRIZOL reagent (Invitrogen) for reverse transcription PCR (RT-PCR) analysis.

2.4. Reverse transcription polymerase chain reaction (RT-PCR)

Cultured cells were placed into TRIZOL reagent (Invitrogen) on selected days of EB differentiation and stored at -80°C. RNA was extracted using MaxTract phase-lock gels (Qiagen) following manufacturer’s instructions. 1 g total RNA was treated with RNase-free DNase I (Invitrogen) prior to reverse transcription to cDNA. For reverse transcription 0.2 g RNA was mixed into cDNA with 0.3 g random hexamer primers (Invitrogen), 0.5 mM dNTPs (Invitrogen) and DEPC-treated water, and incubated for 5 minutes at 65°C then placed on ice. 1X 1st strand synthesis buffer (Invitrogen), 0.02 mM DTT and 50 Units of RNaseOUT (Invitrogen) were then added and samples were incubated at 42°C for 2 minutes. The samples were divided into two equal volumes and 200 Units of Superscript III reverse transcriptase (Invitrogen) was added to one set of tubes while no enzyme was added to the other set of tubes (-RT) to serve as a negative control. Samples were incubated at 42°C for 50 minutes followed by incubation at 70°C for 15 minutes. Reactions were diluted to 50 1 with autoclaved MilliQ water and 2 1 was used per PCR reaction. PCR was performed with Hot Star Taq polymerase (Invitrogen) in PCR buffer and dNTPs. Cycling conditions were as follows: 1 cycle of 95°C for 15 minutes followed by 35-40 cycles of amplification (95°C denaturation for 5 minutes, 95°C
annealing for 30 seconds, 72°C elongation for 1 minute 15 seconds) with final incubation at 72°C for 10 minutes. GAPDH was used as the invariant control. The sequences of primers used are as follows (forward and reverse): Sox17; 5’-AAGATTGAGAAAACACGCATGAC-3’ and 5’-TTTGTGTATAAGCCCGAGATGG-3’, Foxa2; 5’-AGCACCATTACGCCTCAACA-3’ and 5’-CTCCACTCAGCCTCTCATTCC-3’, Nepn; 5’-GCATTCCCTCCACCACCAAGAG-3’ and 5’-TGACAGGTAAAGATGGGACAGGTTTC-3’, Pyy; 5’-TCCTGCTCATCTTGCTTCCG-3’ and 5’-TGAACACACACAGCCCTCCAG-3’, Foxf1; 5’-CACCACAACAGTCACAACCGG-3’ and 5’-GGCATTTGAAAGACAACTCC-3’ Foxh1; 5’-CTGTCCCAGCATGGCCTCAGGGCTG-3’ and 5’-TAATCTGAGCCAGTTTCAGGC-3’, Pax6; 5’-CAGTGAATGGGCGGAGTTAT-3’ and 5’-ACTTGGACCGGAACGTGACAC-3’, Rex1; 5’-TGAAAGTGAGATTAGCCCCGAG-3’ and 5’-GTCCCATCCAATAGCAC-3’, Endy; 5’-GGACAGATACTGCGATTACAGACG-3’ and 5’-CCATCCTGAGGTGTTTTTCCG-3’, NAPS; 5’CCTATCAATCAGCATACACCACCG-3’ and 5’-GGAATCGCCGTGTTGTTTCAGGAAC-3’, Pdx1; 5’-AGCAAGATTGTCGGGTGACC-3’ and 5’-AGTTTGGAGCCCAGGTTGTC-3’

2.5. Whole mount in situ hybridization

The murine cDNA for Endy and NAPS was used as a template for generating Digoxigenin – UTP (DIG) RNA probes using an RNA labeling kit (Roche). Briefly, 10 μg of template DNA was linearized with 10 Units of an appropriate enzyme and purified. RNA probes were in vitro transcribed using 1 μg of linearized template and 20 Units of SP6, T3 or T7 RNA polymerase (Roche). RNA probes were purified using G-50 sephadex columns (Amersham) and
stored at -80°C. Whole mount in situ hybridization (WISH) was performed on whole mouse embryos ranging from E6 to E9. Embryos were dissected in 1X PBS and fixed overnight in 4% paraformaldehyde (PFA) in PBS at 4°C. Fixation was stopped by washing embryos twice in PBS with 0.1% Tween-20 (PBT) for 5 minutes. Subsequently, embryos were washed serially with 50%, 75% methanol/PBT solution and then twice with 100% methanol before being stored at -80°C for no more than a month. Embryos were rehydrated through 75%, 50% methanol/PBT solutions, incubated in 6% H₂O₂ in PBT at room temperature for 1 hour and rinsed thrice with PBT. Embryos were treated with 10 µg/ml proteinase K (Invitrogen) in PBT at room temperature for 12 mins for E6.5-7.0 embryos, 25 mins for E7.5-8.0 embryos and 42 minutes for E8.5-9.0 embryos. The reaction was stopped by carefully rinsing the embryos twice in PBT. Embryos were re-fixed for 20 minutes in 4% PFA with 0.2% glutaraldehyde (Invitrogen) and rinsed and washed twice with PBT. Pre-hybridization and hybridization steps were performed using the following solution: 50 % formamide, 1.3X SSC (pH 5), 5mM EDTA (pH8), 50 µg/ml Yeast tRNA (Invitrogen), 0.2% Tween-20, 0.5% CHAPS, 100 µg/ml Heparin. Embryos were rinsed with 0.5 ml hybridization solution and incubated horizontally for more than 1 hour in 0.5 ml at 65°C. Samples were incubated overnight at 65°C with 1 µl of the DIG-labeled RNA probes. Post-hybridization washes were performed as follows: 3 rinses followed by 2 washes at 65°C for 30 minutes with hybridization solution, one 10-minute wash with 1:1 mix of hybridization solution and MABT (100 mM maleic acid, 150 mM NaCl, pH 7.5, 0.1% Tween-20), followed by 2 rinses and 1 wash with MABT. Embryos were then blocked for 1 hour at room temperature with 2% Blocking Reagent (Roche) dissolved in MABT and for 2 hours at room temperature with 2% Blocking Reagent and 20% heat inactivated goat serum (Invitrogen) in MABT. Embryos were incubated with anti-Digoxigenin antibody (1:2000) (Roche) in 2% Blocking Reagent and 20% heat inactivated goat serum in MABT overnight at 4°C. They were rinsed and washed 3 times for 1 hour with MABT followed by 2 washes in NTMT (5M NaCl, 2M
TrisHCL (pH9.5), 2M MgCl₂, 10% Tween-20) for 10 minutes. Embryos were incubated in BM purple alkaline phosphate substrate (Roche) at 4°C or room temperature. Staining reaction was stopped by washes in PBS with 0.1% Tween-20.

2.6. Histological analysis

Following WISH, embryos were processed through a graded PBT:Glycerol series to 50% glycerol. Images were taken using a Leica MZFLIII microscope and OpenLab software. After imaging, the embryos were treated with a 1:1 mixture of 60% sucrose and Tissue-Tek OCT embedding medium (Sakura Finetek) sucrose overnight at 4°C. Embryos were embedded in 100% OCT in 10x10 mm plastic molds and frozen at -80°C overnight. The molds were then cryosectioned at 10 µm with a cryostat and placed on Superfrost Plus glass slides (VWR) and allowed to dry overnight.Slides were washed twice in PBS for 5 minutes per wash and mounted using 60% glycerol and glass coverslips (VWR). Sections were imaged using Zeiss Axioplan microscope and OpenLab software.
Chapter 3

RESULTS

3.1. Identification of non-annotated genes enriched for definitive endoderm using serial analysis of gene expression

Nepn and Pyy were two of several genes previously identified through a detailed analysis of SAGE data, with an aim to identify known genes enriched for expression in definitive endoderm during early development, to use as markers for this tissue (Hou et al., 2007). The list of tags analyzed came from three definitive endoderm LongSAGE libraries sequenced up to a depth of 100,000 tags: 0-6 somites stage whole definitive endoderm, Somite 8-12 foregut, and Somite 8-12 hindgut. In individual cells, the abundance of specific transcripts can vary, with higher abundance transcripts tending to be from a limited number of genes with housekeeping functions, and lower abundance transcripts tending to be from most genes with specialized function (Anisimov, 2008). In order to capture as much of the low-abundant transcripts and potentially novel genes, the existing whole definitive endoderm and hindgut library was further sequenced to a depth of 300,000 tags. Thus, for the analysis described in this thesis, the foregut library sequenced to a depth of 100,000 was used, while the definitive endoderm and hindgut libraries sequenced to a depth of 100,000 were added to the 300,000 tag libraries for these tissues to generate a total of 400,000 tags for the analysis.
3.1.1. Creation of a list of novel tags enriched for expression in definitive endoderm

A summary of the steps taken to create the list of novel tags enriched for expression in definitive endoderm is shown in Figure 5. To create this list of tags, the three definitive endoderm libraries were pooled. The three libraries consisted of a total of 770,221 tags, which comprised a total 164,319 tag types. Tag types represented just once in the pooled libraries were excluded leaving 42,917 tag types. These were then mapped to the genome using DiscoverySpace software. 22,553 tag types were found to map to either of the Refseq transcripts, Ensembl transcripts, MGC and UCSC genome databases. All the tags mapped to the genome, but approximately 50% of the 22,553 tag types mapped to any one of the RefSeq transcripts, Ensembl transcripts and MGC databases. 13% of the remaining tags mapped to multiple places on the genome leaving 37% of the tags that mapped once to the genome through the UCSC genome browser, but not the RefSeq transcripts, Ensembl transcripts or MGC databases. The 37% of tags that mapped to the genome only were selected as the list of non-annotated tags for further investigation.
Figure 5. Overview of how list of non-annotated tags in the definitive endoderm SAGE libraries was selected.

Flow chart representing the various steps taken to create a list of tags that mapped unambiguously to the UCSC genome database (highlighted by red circle), but not to the other transcriptome databases. The LongSAGE libraries sequenced to a depth of 100,000 tags analysed were: 8-12 somites (Theiler stage (TS) 13) foregut. LongSage libraries sequenced to a depth of 400,000 tags were as follows: 0-6 somites (TS 12) whole definitive endoderm and 8-12 somite (TS 13) hindgut. For full description see section 3.1.
3.1.2. Tags for ESTs Endy and NAPS were enriched in the definitive endoderm LongSAGE libraries

The previous study to identify SAGE tags that are selectively expressed in the definitive endoderm made use of the T/L ratio method (Hou et al., 2007), where “L” is the total number of SAGE libraries in which a particular tag type is present, and “T” is the total number of times that same tag type is found in the libraries examined. It was rationalized that a higher T and a lower L value would typify a higher degree of enrichment for the definitive endoderm. Using this T/L ratio method, a list of tags with $T > 4$ and $L < 58$ was compiled giving a total of 1,114 tags types. Tags with a T/L ratio of less than 0.25 were further omitted leaving a total of 253 tag types enriched in definitive endoderm. These tags were then systematically mapped to the genome using the BLAT feature of the UCSC genome browser. 211 of the 253 tags were found to map to opposite strands of annotated genes, suggestive of anti-sense transcripts produced. While such complementary transcripts are increasingly recognized as possible players that exert various regulatory functions, these tags were not followed on in the current study (Quere et al., 2004; Shendure and Church, 2002). Out of the remaining 42 tag sequences, the top 4 tags, with respect to their T/L ratios, mapped to the genome with EST information in the region (Table 1). 3 of these tags represented alternate tag positions (Refer to Material and Methods for description of position number) of the EST AK014119. The remaining 1 represented tag position 1 of the EST AK084355. For simplicity’s sake, I have named ESTs AK014119 and AK084355 Endynamodermin or Endy and Node And PNC Specific or NAPS, respectively. Given that tags for these ESTs had the highest T/L ratios, they were selected as candidates for further analysis.
Table 1. Distribution and T/L ratio values for 42 tag types retrieved through SAGE analysis

For description of how Endy and NAPS were chosen for analysis see section 3.1.2.
3.1.2.1 Endy is predicted to be protein coding, but may be a non-coding RNA gene or a pseudogene

The position 1 tag sequence representing Endy with a T/L ratio 3.51 mapped to chromosome 8 of the mouse genome, and the EST was cloned from an E13 mouse embryo head (Carninci and Hayashizaki, 1999). The UCSC genome browser map showed several interesting features for this gene (Figure 6). The transcript consisted of 4 exons with the 4th exon having a predicted open reading frame (ORF), implying protein coding potential. The predicted protein sequence consisted of 81 amino acids. Mapping of the protein sequence to various protein motif databases using InterproScan indicated that the sequence may harbor a trans-membrane motif (Quevillon et al., 2005). However, the predicted trans-membrane domain did not match any homologous trans-membrane motifs in the rest of the database.

The 5' half of Endy, including the immediate promoter region, harbored moderate levels of conservation across several organisms including humans (Figure 5). However the 3' region that carries the predicted protein coding exon was the least conserved part of the gene with homology only to the rat genome. This dramatic difference in conservation can be interpreted in two ways.

1. The predicted ORF does not give rise to a protein and Endy is in fact a novel non-coding RNA gene. If Endy is non-coding, then the region of conservation in the 5' half of the gene could represent sequences that will give rise to some form of non-coding RNA such as a miRNA after processing of the transcript. It has also been shown that non-coding RNA genes have many of the upstream promoter features common to protein coding genes, which sometimes show higher levels of conservation than those seen with protein coding genes (Bompfunewerer et al., 2005). Thus, if Endy is a non-coding RNA gene the high conservation seen in the region 5' of the
promoter is not surprising. (2) Endy could conceivably be a pseudogene. Pseudogenes are typically defunct relatives of known genes that have through some mechanism lost their protein-coding ability or are no longer expressed in a cell (Mighell et al., 2000). Different mechanisms can lead to a pseudogene such as mutations, insertions, deletions or transposition of repeat elements. A look at the repeat elements track in the UCSC browser map for Endy revealed several repetitive elements such as SINEs in the 3’ half of the gene corresponding to the regions of least conservation, with almost no repetitive elements found in the 5’ half of the gene. Endy transcripts are produced in the cell since they were detected in the SAGE library, but they may not have protein coding ability due to insertion of these elements in a region that may have had an exon, rendering Endy non-functional. A BLAST search of the entire gene sequence revealed no homology to any known gene, which may suggest that Endy may not have come from functional relative, but may be a gene that was disabled before any duplication event. There are a number of cases of genes rendered non-functional before any duplication event such as the caspase-12 gene in humans (Xue et al., 2006).
Endy maps to chromosome 8 of the genome. It codes for 4 exons separated by 3 introns. The 4th exon at the 3’ end of the gene is predicted to have an open reading frame (Black circle), and thus may be protein coding. In addition Endy has several gene trapped ES cell lines available as seen below the International Gene Trap Consortium Sequence Tag Alignments track. A region of high conservation across multiple organisms including humans is found in the 5’ half of the gene which includes the first 3 exons, whereas the 3’ half of the gene, which includes the predicted coding exon is only highly conserved in the rat. Red square highlights some SINE repetitive elements in the 3’ half of gene, which may explain the loss of conservation in the region.
3.1.2.2. NAPS could be a type of non-coding RNA

The tag sequence representing EST NAPS with a T/L ratio 2.20 maps to chromosome 12 of the mouse genome, and the EST was cloned from an E12 mouse eyeball. The NAPS transcript consisted of 4 exons and 3 introns and contained 4 distinct peaks of high conservation across the transcript (Figure 7). Inserting the NAPS sequence in two ORF prediction programs (NCBI ORF finder and Web Map Preferences) indicated no definitive ORF with the longest possible ORF reading 67 amino acids. This protein sequence was analyzed using InterproScan (Quevillon et al., 2005), revealing no motif or homology to any other protein families. These two results are highly suggestive of non-coding type RNA.

3.1.2.2.1. NAPS may be a potential miRNA precursor

MicroRNAs (miRNA) are single stranded non-coding RNA molecules of ~21-23 nucleotides in length, and have the ability to regulate gene expression. MiRNAs are known to arise from various regions of transcription units (Singh et al., 2008). They have been known to arise from introns and exons of non-coding RNA and in many instances from introns of protein coding genes. Normally both coding and non-coding transcripts fold to form secondary structures such as the stem loops once they are transcribed. Coding transcripts leave the nucleus for translation but non-coding primary transcripts (known as pri-miRNA if a mature miRNA is produced) that will give rise to a mature mRNA are first processed in the nucleus into short ~70-80 nucleotide stem-loop structures called precursor miRNAs (pre-miRNA) before they are processed into the mature miRNA molecules in the cytoplasm.
Figure 7. UCSC genome browser window for NAPS (AK084355)

NAPS maps to chromosome 12 of the mouse genome. It consists of 4 exons separated by 3 introns. There is no predicted ORF for NAPS. Red ovals highlight peaks of high conservation across the transcript.
The UCSC genome browser map for NAPS showed distinct peaks of conservation spanning the introns rather than the exons, and together with the lack of a definitive ORF suggests that the NAPS transcript may represent a form of non-coding RNA. The existence of potential miRNA precursors in its transcript was explored bioinformatically by inserting the full length mRNA sequence including introns into an online RNA folding program called mFold (Zuker, 2003) (Figure 8). A given sequence of RNA is folded into the most thermodynamically stable conformation that could hypothetically exist within the cell. To identify potential pre-miRNA stem loops in NAPS, every sequence of the several stem loops formed through the mFold program was inserted into an online prediction algorithm, MiPred, which classifies such sequences as ‘real’ or ‘pseudo’ precursors of miRNA (Jiang et al., 2007). The intron regions were included in this analysis because many miRNA species are produced from stem loops formed within the introns of nascent mRNA transcript before they are spliced out (Brown et al., 2008; Kim and Kim, 2007). MiPred predicted 12 hairpin sequences out of a total of 144 possible as being potentially “real” miRNA hairpins across the NAPS transcript. Two of these sequences spanned the first exon whereas the remaining mapped to the intronic regions. However, these intronic miRNA precursor sequences did not correspond to the regions of conservation. A BLAST search of each of these sequences against the SANGER miRNA database showed no homology to any of the known mouse miRNA species. These data suggest that NAPS may be a novel miRNA species or some other form of small non-coding RNA.
The NAPS sequence, including the introns, was inserted into a program called mFOLD, which folds a sequence into the most thermodynamically stable structure that could exist in the cell. Sequences of the secondary structures known as stem-loops (black circles in the mFOLD output) were then plugged into a miRNA precursor prediction program, MiPred, which predicted twelve 50-85 bp potentially “real” miRNA precursors in NAPS (red lines in the genome browser map). The first two predicted sequences (overlapping with the 5’ end of NAPS) are the only sequences that overlap with the regions of conservation.
3.1.3. Expression of Endy and NAPS during post-implantation stages of mouse definitive endoderm development

The study of the definitive endoderm has lagged behind that of the mesoderm and ectoderm. One reason for this has been the lack of markers specific for the definitive endoderm. Gene expression profiling studies such as SAGE allow for the discovery of markers enriched for the definitive endoderm, such as Endy and NAPS. To develop the full potential of Endy and NAPS as novel markers of this tissue, the next critical step was to characterize their spatial and temporal expression patterns during early definitive endoderm development. Determining what populations of definitive endoderm cells express these markers will enable the creation of more accurate fate maps for the formation of this tissue, and serve as indicators for in vitro derived definitive endoderm.

3.1.3.1. Distribution of Endy expression in LongSAGE libraries

To determine at what stages Endy expression begins, we analyzed LongSAGE libraries using DiscoverySpace. Endy expression was detected in only 39 out of the 153 embryonic LongSAGE libraries, suggesting a specific and restricted expression (Figure 9). The highest level of transcripts were observed between Thieler stage (TS) 12 (E7.5-8) and TS13 (E8-9) corresponding to early post-implantation stages of embryonic development. Spatially, Endy was specifically observed in the embryonic regions of the conceptus and the early definitive endoderm. At much lower levels, transcripts of Endy were also seen in kidneys, spleen, intestine and the heart. No tags were observed in libraries corresponding to the pre-implantation stages development except in ES cells.
Figure 9. Distribution of Endy transcripts in SAGE libraries

The data is represented as tag count for Endy normalized to the library size and expressed as tags per 100,000 total tags. TS: Theiler stage.
3.1.3.2. Transition of Endy expression from specific epiblast cells to the definitive endoderm

Since Endy showed highest levels of gene expression between E6.5 –E8.5 in the SAGE analysis, I focused my expression analysis on these embryonic time periods using whole mount in situ hybridization. Endy was first observed in a small population of epiblast cells in the distal region of the embryo at E7.0 (Figure 10: panels A, B and C). The expression was specific to the epiblast and was excluded from the embryonic visceral endoderm layer that surrounds the epiblast cells.

Towards the end of gastrulation at E7.25 the expression in the distal most epiblast cells declined (Figure 10: panel D). Upon completion of gastrulation, cells in the anterior, lateral and posterior embryonic regions of the early head fold stage embryo started to express Endy (Figure 10: panels E, F and G). Expression remained excluded from the extraembryonic region, and a section through the early head fold embryo illustrated that expression of Endy was restricted to the cells of the outermost definitive endoderm layer of the embryo (Figure 10: panels H and I). An anterior view of the E8.0 embryo highlighted a distinct crescent shaped expression domain in the anterior most region of the embryo below the headfolds, while no expression was seen in axial and paraxial regions, and distal-wards from the apex of the crescent (Figure 10: panels J, K and L). Endy expression was also seen as a short streak in midline of the embryo marking some ectodermal tissue (Figure 10: Panel K), and disappeared from that region by E8.5. The crescent-shaped expression was contiguous with the expression in the lateral wings of the embryo, and continued towards the posterior most part of the embryo. Towards the distal and posterior side of the embryo two distinctly symmetrical regions of expression were seen extending from an area
Figure 10. In situ hybridization analysis of Endy during early post-implantation stages of the mouse embryo

A-D: Expression of Endy in late streak E7.0-7.5 embryos. Lateral (A) view and frontal (anterior) (B) view and transverse (C) section of E7.0 embryo. Endy was expressed in the distal most cells of the epiblast. Lateral (D) view of E7.25 embryo. Endy expression reduced. E-I: Expression of Endy in E7.5 embryo. Lateral (E), frontal (F), distal (G), cross section (J-I) views of E7.5 embryo. Endy expression was seen in the definitive endoderm and excluded from the extra-embryonic region. A rim-like expression was seen around the embryo with two symmetrical lines of expression flanking the node. Cross sections (J-I) revealed Endy expression (black arrowheads) in the outer definitive endoderm layer. J-N: Endy expression became stronger in the E8.0 embryo. Lateral (J), frontal (K), distal (L) and cross section (M-N) of E8.0 embryo. Endy showed a crescent like pattern of expression in the anterior of the embryo while slight expression was seen in the midline. The distal view revealed Endy expression in the PNC (posterior notochord) and symmetrically flanking the node. Cross sections confirmed expression in the definitive endoderm (black arrowheads). O-S: Lateral (O), frontal (P), ventral (Q) and cross section (R-S) of E8.5 embryo. Endy expression was observed refined to the ventral part of the foregut invagination and dorsal part of the hindgut invagination. Expression pattern in embryo for each stage is representative of 5 embryos.
of strong expression in the indent referred to as the posterior notochord (PNC) (Figure 10: panel L).

The fact that Endy was not expressed in the entire definitive endoderm suggests molecular heterogeneity of definitive endoderm progenitors. As the embryo develops at E8.5, this molecular heterogeneity is translated to the developing gut tube. Consistently expression of Endy became restricted to the foregut and hindgut pocket of the developing gut tube (Figure 10: panels O, P and Q). A cross section through the foregut and hindguts pockets illustrated that Endy was further restricted to the ventral-lateral region of the foregut pocket and dorsal region of the hindgut pocket (Figure 10: panels R and S). This pattern of expression of Endy during gastrulation and definitive endoderm development was reminiscent of Trh (thyrotropin releasing hormone), another early definitive endoderm marker identified through SAGE (McKnight et al., 2007). Like Endy, Trh was first seen in the epiblast prior to gastrulation, excluded from the extraembryonic region of the conceptus, and during gastrulation was exclusively expressed in the definitive endoderm. The expression of Trh then shifted to the foregut and hindgut regions of the developing gut tube. These observations suggest that Trh and Endy may be regulated by similar mechanisms.

3.1.3.3. Distribution of NAPS expression in the LongSAGE libraries.

The NAPS transcript was detected in only 6 of the 153 LongSAGE libraries (Figure 11). The highest level of expression was observed in the hindgut of the definitive endoderm with almost negligible levels in the posterior region of the neural tube. These transcripts in the neural tube library might have been due to slight hindgut tissue contamination during the initial collection of the neural tube tissue for the particular library.
Figure 11. Distribution of NAPS transcripts in the SAGE libraries

The data is represented as tag count for NAPS normalized to the library size and expressed as tags per 100,000 total tags. TS: Theiler stage.

3.1.3.4. NAPS is expressed in cells of the node and the PNC

The expression pattern for NAPS was much more restricted than Endy, and was first detected at E7.0 in the distal most tip of the embryo (Figure 12 panels A, B, C and D). As discussed earlier, the node and PNC first become evident in this region at around E7.5 of the mouse embryo. The node has been shown to consist of cells that can produce a partial secondary axis, and is thought to harbor progenitors for the axial mesoderm, including the notochord, which is required for the formation of the mouse body plan (Beddington, 1994). In contrast, the small indent represented by the PNC is a well known site of left right axis determination in the mouse (Hirokawa et al., 2006). Several genes have been shown to mark this region of the embryo, with Nodal and Goosecoid being of note (Ang and Rossant, 1994; Houde et al., 2006; Schier, 2003;
Wang and Ware, 2009) (Figure 12). Nodal has been shown to be required for the initial breaking of symmetry in the PNC, as well as for transferring the initial asymmetric signal to the left side of the embryo aided by the monociliated cells of the PNC (Hirokawa et al., 2006).

A sagittal section of a gastrulated embryo highlighted that NAPS marked cells of the node and the PNC (Figure 12: panel F). The node at this stage represents the population of anterior primitive streak cells that harbor organizer activity, while the PNC just distal and anterior to it represents the indent where left-right axis determination is known to take place. The node was seen as a disordered mass of cells where the germ layers were indistinguishable, and was continuous with the bilayered structure of the PNC. As the embryo developed, NAPS expression gave the appearance of migrating posteriorly (Figure 12: panel E). This was indicative of the regression of the node and primitive streak to the posterior most end of the embryo, concurrent with the elongation of the notochord as it is laid down by the node to define the axis of the embryo. At E9.0, NAPS expression was observed in the posterior most dorsal hindgut seen through a cross section of the region (Figure 12: panel G, H, I and J). Interestingly, some NAPS expression was also seen in a few cells of the posterior most end of notochord, and reflected perhaps, the PNC population of cells. This temporal expression pattern was consistent with lineage analysis of the node/PNC region where cells of this region end up in the posterior end of the notochord and gut endoderm (Kinder et al., 2001).

NAPS expression overlapped with the temporal expression pattern of Nòdal (Figure 12). At E7.0 Nòdal expression was seen in the node/PNC region and comparing Figure 12 panels D and L showed that the two expression patterns seemed to complement each other, with Nodal expression found on either side of the PNC and NAPS expression domain. This raised the possibility that NAPS and Nòdal might share some regulatory mechanisms or may even regulate the expression domain of each other. Goosecoid has also been shown to be expressed in the node/PNC region between E7-7.5 (Blum et al., 1992). As the embryo develops, prechordal cells
arising from the node begin to express Goosecoid as they migrate to the anterior of the embryo. NAPS expression, in contrast, moves towards the posterior along with the node. Again, there is possibility that NAPS and Goosecoid share regulatory mechanisms.

Figure 12. In situ analysis of NAPS and Nodal in the early post-implantation stages of the mouse embryo.

A, C: NAPS expression in E7.0 embryo. Lateral (A) and frontal (C) views. NAPS expression was seen in the node/PNC region in the distal tip of the embryo. B, D, and F: NAPS expression in E 7.5 embryo. Lateral (B), frontal (D), and sagittal section (F) views. NAPS continued to be expressed in the node/PNC region. The sagittal section revealed that NAPS expression is seen in cells transitioning from the node to the PNC. E, G-J: NAPS expression in E8.5-9.0 embryos. Lateral (E, G), ventral (H) and cross section (I-J) views. NAPS expression domain ended up in the posterior most region of the embryo. Cross sections revealed expression of NAPS in hindgut and some mesoderm derived notochord cells. J-M: Nodal expression in E7.0-8.0 embryos. Lateral (J-K, M) and frontal (L) views. Expression of Nodal followed that of NAPS implying a possibility of regulation. Expression pattern at each stage representative of 5 embryos.
3.1.3.5. **NAPS expression remains unchanged in Hippi \(^{-/-}\) mouse embryos**

Hippi \((\text{Huntington interacting protein-1 protein interactor})\) has been shown to contribute to apoptotic death of striatal neurons in Huntington’s disease (Gervais et al., 2002). It has also recently been reported to be essential for cilia formation in the node as Hippi mutants display a loss of cilia in the node (Houde et al., 2006). The determination of left-right axis in the embryo is reliant on the cilia in the node. The exact mechanism for this determination is still a matter of debate, but an asymmetric flow of morphogen generated by the beating of the monocilia of the node has been implicated (Nonaka et al., 2005).

Mutations of various other genes, such as \(\text{Kif3A}, \text{Kif3B} \text{ and } \text{iv}\), required for proper cilia development have lead to left-right patterning defects (Eggenschwiler and Anderson, 2007). In accordance with these results Hippi deficient mice exhibit morphological defects in left-right patterning such as abnormal heart looping (Houde et al., 2006). Furthermore genes expressed in asymmetric fashion, such as \(\text{Nodal}\), are expressed on both sides of the embryo in Hippi mutants. Several of the phenotypic abnormalities observed in Hippi\(^{-/-}\) embryos overlap with those seen in embryos lacking members of the Sonic hedgehog (Shh) signaling pathway, thus Hippi is thought to function in the Shh signaling pathway (Houde et al., 2006). To test whether the lack of Hippi disrupts NAPS expression, whole mount in situ hybridization was done on Hippi \(^{-/-}\) embryos. The expression of NAPS remained unchanged in Hippi \(^{-/-}\) embryos (5/5 embryos; Figure 13). This would suggest that Shh signaling is not required for the expression of NAPS, and that NAPS is not required for ciliagenesis.
Figure 13. Expression of NAPS in wildtype versus Hippi⁻/⁻ embryos

A-C: Lateral views in wildtype E7.5-8.5 embryos expressing NAPS. NAPS expression was restricted to the node/PNC cell population throughout. D-F: Lateral views of Hippi⁻/⁻ embryos expressing NAPS. Expression of NAPS remained unchanged.
3.2. Establishment and manipulation of an ES cell differentiation system to definitive endoderm

3.2.1. Markers used to characterize ES cell differentiation system

The mouse is much less accessible as a model organism to study the important lineage commitment decisions taking place during early embryonic development. This has led to a recent surge in the development of ES cell differentiation protocols to definitive endoderm in order to address questions relating to early lineage commitment (Kubo et al., 2004; Tada et al., 2005; Yasunaga et al., 2005). To investigate some of the early genetic events that play a role in specification and patterning of the definitive endoderm, I used an ES cell differentiation system. Fetal bovine serum used in culturing ES cells has been shown to be sufficient to induce development of endoderm and mesoderm from ES cells owing to growth factors present in serum (Kubo et al., 2004). However, it has been reported that addition of specific levels of Activin promotes differentiation of ES cells to definitive endoderm fates. I evaluated the ability of exogenous Activin to specifically promote definitive endoderm development in EBs during their differentiation in the presence of serum. The development of endoderm was tracked by monitoring the expression of Foxa2 and Sox17, traditionally used markers for general endoderm, and Nepn and Pyy, the markers identified as for definitive endoderm. I also monitored the expression of Endy and NAPS to determine if these genes exhibited any temporal expression patterns. These markers were compared with those of Rex1, Foxf1 and Pax6, genes indicative of pluripotency, mesoderm and neurectoderm cells, respectively.
The \textit{Rex1} gene is an acidic zinc finger gene and is a well recognized marker for the pluripotent state of both ES cells and embryonic carcinoma (EC) cells (Ben-Shushan et al., 1998). Down regulation of \textit{Rex1} either through differentiation of ES and EC cells with the treatment of retinoic acid or through knockout mutations has been shown to differentiate the cells to parietal endoderm-like cells (Thompson and Gudas, 2002). Thus, \textit{Rex1} has a role in regulating the pluripotent state.

\textit{Foxf1} encodes a forkhead transcription factor that is expressed in mesodermal tissues. Expression of \textit{Foxf1} is first detected in the primitive streak stage of the conceptus in extraembryonic yolk sac mesoderm, and in the mesoderm of the posterior primitive streak and the lateral plate mesoderm of the embryo proper (Mahlapuu et al., 2001). During organogenesis it is mainly expressed in the mesenchyme adjacent to the epithelium lining the gut and also detected in splanchnic mesoderm that gives rise to certain regions of the heart. \textit{Foxf1} null embryos develop normally up until the early somite stage, but die thereafter due to multiple defects in extraembryonic mesoderm derived structures such as the yolk sac. \textit{Foxf1} null embryos also exhibit a reduced amount of mesoderm owing to improper development of the mesoderm in the posterior part of the primitive streak.

The neurectodermal marker \textit{Pax6} is a member of the \textit{Pax} gene class and is a homeobox transcription factor. \textit{Pax6} expression is first observed when the somites are first formed and the neural folds begin to close in the cervical region at ~ E8.0 (Callaerts et al., 1997). Expression is specifically seen in the presumptive prosencephalon (forebrain) and rhombencephalon (hindbrain), in the developing spinal cord, and in the broad region of the head surface ectoderm covering the prosencephalon. \textit{Pax6} null embryos show a wide range of defects from defects in eye development to severe craniofacial and forebrain defects.
3.2.2. Definitive endoderm markers identified through the SAGE analysis are expressed in ES cells differentiated to definitive endoderm

To examine ES cell differentiation to definitive endoderm, I used an ES cell line derived from a 129 mouse strain, CCE ES cells (Robertson et al., 1986), and an ES cell line derived from the C57Bl/6 strain, LP9 ES cells (derived in house by Danny Chui). ES cells were allowed to grow with serum on irradiated and senesced primary mouse embryonic fibroblasts, which help maintain the undifferentiated state of the cells. At about 60% confluency, the ES cells were dissociated and replated onto gelatin-coated plates. This allowed selective removal of the fibroblasts to obtain a highly enriched population of ES cells. The passaging of the ES cells on gelatin was repeated once more to ensure complete removal of the fibroblasts. When the ES cells had again reached ~ 60% confluency, they were dissociated and plated at appropriate densities (see Materials section) into low adherence dishes to stimulate EB formation. The EBs were allowed to differentiate for 9 days with 20 ng/ml Activin added on day 3 of differentiation as described (Kubo et al., 2004; Yasunaga et al., 2005). Cells were then collected for RT-PCR analysis at selected time points (Figure 14).

The results showed that Rex1 was strongly expressed prior to the onset of differentiation at day 0. As differentiation proceeded its expression was reduced, but transcripts continued to be present in both cell lines (Figure 14). This suggested that not all ES cells in the culture were losing their pluripotency as the EBs developed under conditions used. The most prominent feature of Rex1 expression was seen in its difference in expression pattern between the CCE and the LP9 ES cells. The addition of Activin in LP9 ES cells reduced Rex1 expression in comparison to the no Activin control. In contrast, CCE ES cells displayed no difference in Rex1 expression pattern whether or not Activin was added to the EB cultures. Given that the serum
levels were the same for both cultures; this might suggest that LP9 cells are more responsive to Activin.

In general, markers for endoderm were all induced and upregulated as differentiation proceeded. The expression of Sox17 increased upon differentiation, and was maintained at high levels after day 5 in the CCE ES cells. In contrast, Sox17 levels in LP9 cells get stronger as the days proceed. Foxa2 maintains an increasing level of expression as differentiation proceeds in both CCE and LP9 cells. However the presence of Activin in LP9 cells induced Foxa2 more strongly, implying once again, that the LP9 cells are more responsive to exogenously added Activin. This temporal pattern of Sox17 and Foxa2 is consistent with previous ES cell differentiation to definitive endoderm RT-PCR analyses that showed an induction of these two traditional endoderm markers upon differentiation of ES cells in the presence of Activin or serum (Kubo et al., 2004; Tada et al., 2005). Both Sox17 and Foxa2 are required for the specification of early endoderm. Thus induction and upregulation of these markers in the developing EBs generated here is an indication that endoderm development, be it visceral or definitive, was being initiated and maintained. Given the presence of these markers, this system enables the opportunity to isolate cell populations expressing these markers for further study of endoderm development.

The development of definitive endoderm derived cells in the developing EBs was confirmed with the induction of Nepn and Pyy in both CCE and LP9 cells. Some Pyy was detected at day 0 in CCE cells suggesting the presence of some differentiated cells in the undifferentiated population and was corroborated by finding slight levels of Sox17 and Foxa2 transcripts at day 0. Pyy expression was induced quite strongly by day 7 of differentiation in both cell lines; however Nepn induction was slightly delayed in the LP9 cells. As was the case with Rex1 and Foxa2 expression, the responsiveness of LP9 cells to Activin was evident with stronger
Figure 14. RT-PCR analysis of CCE and LP9 mouse ES cells differentiated as EBs for 9 days in the presence or absence of Activin

RT-PCR of Gapdh (housekeeping gene), Sox17, Foxa2 (visceral and definitive endoderm markers), Nepn (midgut endoderm marker), Pyy (foregut invagination marker), Foxf1 (extraembryonic mesoderm, embryonic lateral plate mesoderm and posterior primitive streak mesoderm marker), Pax6 (neurectoderm marker), Rex1 (pluripotency marker), Endy and NAPS. CCE and LP9 ES cells were differentiated for 9 days in the presence or absence of 20 ng/ml Activin. Activin was added on day 3 of differentiation. Cells were collected for RT-PCR analysis on days 0,2,5,7 and 9 of differentiation. Briefly, markers for endoderm were induced and upregulated upon differentiation of both cell lines. Markers for mesoderm and neurectoderm were not induced as strongly, while levels of the pluripotency marker Rex1 were reduced as differentiation proceeds in both cell lines. LP9 ES cells also seemed to be more responsive to Activin with markers for endoderm showing stronger induction and markers of mesoderm, neurectoderm and pluripotency showing stronger reduction in the presence of Activin. CCE ES cell data represents 3 biological replicates while LP9 ES cell data represents 2 biological replicates. For detailed description of expression trends see section 3.2.3.
levels of Nepn and Pyy induced upon differentiation. The expression of Nepn and Pyy between day 5 and 7 of differentiation probably represents the time during development when the definitive endoderm is being regionalized into its different derivatives. Furthermore, the initial induction of Sox17 and Foxa2, followed by the induction of Nepn and Pyy later during EB differentiation mimics the timeline of in vivo development and specification of definitive endoderm.

Levels of Foxf1 transcripts were faint in CCE ES cells regardless of Activin addition, which suggested that the production of at least some mesoderm derivatives of the epiblast (such as the extraembryonic mesoderm and lateral plate mesoderm) were not being favored. Slightly higher levels of Foxf1 expression were seen in the control group of the LP9 cells, but the presence of Activin reduced these levels dramatically, implying that, in the presence of added Activin, LP9 cells favor differentiation to endodermal fates. In CCE cells, Pax6 was induced in contrast to expression of Foxf1. This suggested that a number of cells in the EBs were differentiating down the neurectodermal pathway. However, Pax6 was maintained at low levels in LP9 cells, giving further evidence for differences between cell lines in the control of their differentiation responses.

Expression of Endy and NAPS was also investigated to determine whether they followed temporal patterns similar to that seen in their in vivo expression analysis. Interestingly, Endy was expressed at high levels throughout differentiation even prior to the onset of differentiation suggesting it is active in undifferentiated ES cells. This result corroborates the SAGE data that showed Endy transcripts in the ES cell libraries (Figure 9). Although Endy was found to be expressed in the definitive endoderm in the embryo, the fact that Endy is so highly expressed in ES cells makes it a difficult marker to use on its own to identify definitive endoderm cells as distinct from ES cells, differentiation studies. However, it may be a useful marker in conjunction with other specific markers of the definitive endoderm.
In contrast to Endy, NAPS was induced at day 2 of differentiation in both cell lines. This response may be correlated with the expression of NAPS seen in the node/PNC region during gastrulation. The in situ expression analysis (Figure 12) showed that cells in the node/PNC region continually express NAPS during gut morphogenesis and axis elongation. Expression then ends up in the posterior most ends of the hindgut and notochord. This continued expression was represented in the CCE ES cell EBs where NAPS expression was maintained once it was induced. Whether this expression in CCE ES cells was notochord- or hindgut-specific is difficult to determine. NAPS was induced more strongly in LP9 cells at day 2 and then seemed to drop off before returning to similar levels in day 9 of the no Activin condition. However, NAPS expression remained at strong levels when Activin was added until day 9 when it was reduced. The NAPS expression pattern observed in the absence of Activin addition to LP9 EBs could be a result of cell-line specific differences and their potential to differentiate to certain cell types. Perhaps at day 2, several of the node/PNC-specific cells are being produced, and upon further differentiation NAPS expression is downregulated as certain cell populations are fated to become other lineages. The subsequent return of NAPS expression from day 5 onwards could represent the populations of cells now having the appropriate signals to express NAPS once more. Alternatively, in vivo, during migration of NAPS expression to the posterior of the embryo, certain cell populations may be shutting off NAPS expression and switching it on again upon reaching their final destination. This expression pattern may be reflected by the above expression pattern seen in LP9 ES cells in the absence of Activin addition.

The expression patterns of these various markers highlights the differences that exist in differentiation potential between cell lines and their responsiveness to external factors, likely a result of differences in their genetic backgrounds. The lack of response of CCE cells to exogenous Activin may suggest that endogenous levels of Activin-inducible TGFβ signaling are already saturated. This could be due to the production of saturating concentrations of Activin
produced by the ES cells, themselves, or to a limitation of TGFβ signaling components, such as
the receptors of various Smads. In the case of LP9 cells, Activin was able to induce a response
that was favorable to production of definitive endoderm cell fates upon ES cell differentiation. In
addition, since Nepn and Pyy are excluded from the visceral endoderm in vivo, they serve as
excellent indicators of true definitive endoderm development in these ES cell differentiation
protocols to definitive endoderm. The expression of Nepn and Pyy in the EBs indicated that
definitive endoderm derivatives were being generated, and together with the expression patterns
of Sox17 and Foxa2 were following a temporal molecular pattern expected from the in vivo
specification of definitive endoderm.

3.2.3. Blocking TGFβ signaling inhibited ES cell differentiation to definitive
endoderm

To investigate early lineage decisions that take place during the formation and
specification of the definitive endoderm, I used the ES cell system to test whether blocking
TGFβ signaling during EB differentiation of ES cells would hinder the differentiation to
definitive endoderm. As discussed before, members of the TGFβ signaling pathway, most
notably Nodal, are thought to play an important role in the formation and specification of
definitive endoderm development. Small molecule inhibitors have proven extremely useful for
investigating signal transduction pathways and have the potential for development into
therapeutics for inhibiting signal transduction pathways whose activities cause diseases. I
therefore also used SB-431542, a potent small molecule inhibitor of the TGF β signaling
pathway (Inman et al., 2002), to examine the role of TGFβ signaling pathways in definitive
endoderm specification. This inhibitor selectively inhibits type I ALK receptors, ALK4, ALK5
and ALK7, to which Nodal/Activin and other TGFβ super family of ligands are known to bind. Nodal/Activin and related ligands are known to signal through ALK4 whereas other TGFβ ligands are known to signal through ALK5 (Massagué, 1998). Recently it has been demonstrated that Nodal also signals through ALK7; however, this receptor was shown not to be an essential mediator of Nodal signaling during embryogenesis or left-right patterning in the mouse (Jornvall et al., 2004).

I chose to block TGFβ signaling in CCE ES cells since they showed high levels of induction of definitive endoderm markers regardless of Activin addition suggesting a saturated level of endogenous TGFβ signaling (Figure 14). Cells were once again grown on irradiated primary mouse embryonic fibroblasts and passaged twice on gelatin to remove the fibroblasts. The ES cells were then dissociated and allowed to form EBs in low adherence dishes for 9 days. On day 3 of differentiation 10 μM of SB-431542 was added. Cells were again collected on selected days for RT-PCR analysis (Figure 15).

The addition of the inhibitor dramatically reduced induction of Nepn and Pyy (Figure 15). The continued expression of Sox17 and Foxa2, even in the presence of the inhibitor suggests that perhaps only visceral endoderm formation was being initiated and maintained, as Sox17 and Foxa2 are also expressed in visceral endoderm cells. Taken together, the continued presence of Sox17 and Foxa2 transcripts, and the negligible expression of Nepn and Pyy, suggests that early progenitors of the definitive endoderm were being produced, but are not being specified into the regional fates of the definitive endoderm. Alternatively, it is possible that only visceral endoderm cells were being generated, thus explaining the expression of Sox17 and Foxa2, but not of Nepn and Pyy.

Foxf1 transcript levels remained unchanged, indicating that induction of some mesoderm-derived tissues may not be as sensitive to TGFβ signaling as the formation of definitive endoderm. Pax6 levels, interestingly, increased as differentiation proceeded in the presence of
the inhibitor, implying more cells are differentiating down the neurectodermal pathway. Several studies have suggested that neural fates are the default state of cells during embryological development and of ES cells, suggesting that the formation of the neural lineage is under inhibitory control with a tendency towards the neural fate being mitigated through various intercellular signals including TGFβ signals (Munoz-Sanjuan and Brivanlou, 2002). This concept is supported by the demonstration that blocking TGFβ signals during differentiation of mouse and human ES cells augments evidence of neural differentiation (Munoz-Sanjuan and Brivanlou, 2002; Tropepe et al., 2001). The increase in \( \text{Pax6} \) expression seen in the presence of the inhibitor would also be consistent with this.

Expression of the pluripotency marker \( \text{Rex1} \) was upregulated in the presence of the inhibitor, perhaps indicating a tendency of the cells to proliferate in a pluripotent state as signals to differentiate were being blocked. The inhibitor had no effect on \( \text{Endy} \) expression consistent with its expected expression in ES cells as well as in definitive endoderm cells. \( \text{NAPS} \) expression was upregulated on day 5, but reduced on day 7 in the presence of the inhibitor as compared to the control. This suggests that \( \text{NAPS} \) expression is not under positive regulation by TGFβ signals. The increased levels of \( \text{NAPS} \) transcripts in the presence of the inhibitor could be attributed to an accumulation of those cells that mark the node/PNC region of the embryo which have not been fully specified. Furthermore, the downregulation seen from day 7 onwards could indicate a change of cells either to a neural fate or back to a pluripotent state.

The dramatic downregulation of \( \text{Nepn} \) and \( \text{Pyy} \) in the presence of the inhibitor suggests that the generation of definitive endoderm derived cells expressing these markers is under the control of TGFβ signaling. Together with the upregulation of \( \text{Pax6} \) and \( \text{Rex1} \), these results corroborate the \( \text{in vivo} \) studies that implicate TGFβ signaling in the inhibition of neural fates and the specification of the definitive endoderm (Camus et al., 2006; Munoz-Sanjuan and Brivanlou, 2002; Tropepe et al., 2001).
Figure 15. RT-PCR analysis of CCE ES cells differentiated as EBs for 9 days in the presence of TGFβ signaling inhibitor SB-431542.

RT-PCR of Gapdh (housekeeping gene), Sox17, Foxa2 (visceral and definitive endoderm markers), Nepn (midgut endoderm marker), Pyy (foregut invagination marker), Foxf1 (extraembryonic mesoderm, embryonic lateral plate mesoderm and posterior primitive streak mesoderm marker), Pax6 (neurectoderm marker), Rex1 (pluripotency marker), Endy and NAPS. Differentiation was induced in CCE ES cells. The inhibitor was added on day 3 of differentiation. Briefly, Sox17 and Foxa2 expression was induced and maintained in the presence of inhibitor whereas the definitive endoderm markers Nepn and Pyy were dramatically reduced. Pax6 and Rex1 levels were upregulated in the presence of the inhibitor whereas as Foxf1 and Endy levels remained unchanged. NAPS was expressed at day 5, but was gone upon further differentiation in the presence of the inhibitor. Data representative of two biological replicates.
3.2.4. Nepn expression is lost in ES cell differentiation of Foxh1⁻/⁻ ES cells

Foxh1 is an important cofactor involved in the TGFβ signaling cascade during definitive endoderm development, and mediates an important positive feedback loop to Nodal signaling during early embryo development (Norris et al., 2002). Knocking out Foxh1 in mouse embryos results in a failure to form the APS, as shown with a loss of expression of the APS markers, encoded by Foxa2 and Goosecoid, as well as loss of APS derived structures such as the notochord (Hoodless et al., 2001). Chimera studies have shown that Foxh1-null ES cells failed to contribute to definitive endoderm layer, suggesting a requirement of Foxh1 in the formation of the definitive endoderm (Hoodless et al., 2001). However, a small number Foxh1-null ES cells were found to colonize the hindgut implying that formation of these populations of definitive endoderm may not require Foxh1. In addition, recent analysis of Foxh1 mutant embryos has shown that midgut and hindgut formation is not affected by the loss of Foxh1 (Hou, McKnight and Hoodless unpublished). Nepn, our novel midgut marker, was shown to be strongly expressed in Foxh1-null embryos and Sox17, a marker restricted to the hindgut endoderm by E8.5 remained unaffected in Foxh1 mutant embryos. However, the foregut endoderm was severely affected with significantly reduced levels of Pyy and Trh (markers that are normally highly expressed in the foregut invaginations) in Foxh1-null embryos (Hou, McKnight and Hoodless unpublished). These observations suggest an alternate signaling pathway aiding the formation of the midgut and hindgut definitive endoderm, which is either dependent on Foxh1-independent TGFβ signaling or other signaling pathways.

In order to challenge our established in vitro differentiation system to definitive endoderm further, I induced Foxh1-mutant ES cells to differentiate and compare the pattern of gene expression changes with those seen in vivo Foxh1. I monitored the expression Nepn and
Pyy during differentiation of a Foxh1 knockout ES cell line, FK03, and a wildtype line derived from the same mouse strain, FK02 (Hoodless et al., 2001). Differentiation was again induced in the presence or absence of 20 ng/ml Activin. Activin was used to determine whether it would rescue any effects caused by a reduction in Nodal signaling, as Activin binds the same receptors as Nodal. Differentiation was allowed to proceed over 13 days to cover a wider time course because Foxh1-null embryos are often delayed in their embryological development compared to their wildtype counterparts and a similar effect might have been anticipated in the differentiation of Foxh1 mutant ES cells (Hoodless et al., 2001). EBs were collected on selected days for RT-PCR analysis (Figure 16).

Upon differentiation of wildtype and Foxh1 mutant ES cells, levels of Sox17 transcripts were the same in the two cell lines suggesting that induction of Sox17 expression is not sensitive to loss of Foxh1. Nepn transcripts, however, were absent upon differentiation of Foxh1 mutant ES cells; although in the presence of Activin, Nepn expression was minimally induced on day 13 of differentiation. Surprisingly, Pyy expression remained the same in the two cell lines and the addition of Activin induced Pyy expression slightly in the mutant ES cells. Foxf1 expression was not induced upon differentiation of the mutant ES cells suggesting that induction of some mesoderm derived tissues is reliant on Foxh1 signaling. This is consistent with the fact the Foxh1-dependent signaling is required for development of certain types of mesoderm, including splanchnic mesoderm, which was shown to contribute extensively to the formation of the heart and to express Foxf1 (Mahlapuu et al., 2001; von Both et al., 2004). Thus, the loss Foxh1 prevents the formation of cardiac splanchnic mesoderm like cells, which in turn accounts for the loss of Foxf1 expression.

Pax6 and Rex1 expression was induced more strongly in Foxh1-mutant ES cells suggesting once again that cells not receiving appropriate differentiation signals enter a neur ectodermal pathway or remain as proliferating pluripotent cells. Interestingly, the mutant
cells express Pancreatic and duodenal homeobox 1 (Pdx1), an early pancreatic marker. At E9.0, Pdx1 is known to be expressed in the dorsal pancreas bud (derived from the midgut) and the ventral pancreas bud (derived from the ventral foregut). In the in vivo Foxh1 mutant analysis, Pdx1 expression was ectopically detected in the entire gut midline (Juan, Kris and Hoodless unpublished). Expression of Pdx1 in the mutant cells is consistent with this observation. This, together with the expression of Pyy, implies some definitive endoderm was being formed. However, the expression patterns of Pyy and Nepn in Foxh1 null ES cells were inconsistent with the in vivo observations, suggesting that in vitro induction of definitive endoderm from ES cells does not completely reflect its in vivo development.

Endy expression was observed throughout differentiation, but seemed to be downregulated at day 7 with subsequent upregulation beyond day 7 of differentiation in Foxh1+/+ ES cells, regardless Activin addition. Perhaps these ES cells were undergoing differentiation more readily as Rex1 levels were already much lower than at Day 0. As such, between day 2 and day 7 differentiation, Endy expression may have been reduced due to the differentiation of the majority of ES cells to epiblast like cells, and then upregulated when those cells became specified into definitive endoderm. This Endy transcript expression pattern was not observed in Foxh1-null cells where it was seen to remain at strong levels in the absence of Activin. Interestingly, Endy expression in the presence of Activin was decreased between days 7 and 13 of differentiation. This would suggest that Activin-induced signalling promotes ES cells differentiation resulting in a decrease in Endy expression. In contrast to Endy, almost no NAPS expression was detected upon differentiation of wild-type ES cells, whereas Activin did seem to induce NAPS expression in the Foxh1-null cells. This could, once again, indicate the potential of different cell lines to differentiate into certain cell types more easily than others.

The above results highlight that Foxh1-null ES cells are able to form some definitive endoderm cells that express Pyy and Nepn; however the complete absence of Nepn is
inconsistent with in vivo findings for Foxh1-mutant embryos (Hou, Mcknight and Hoodless unpublished). One of the limitations of in vitro differentiation studies is the inability to replicate the 3D embryo like structure that allows for interspecific interactions between germ layers. The continued presence of Nepn expression in the Foxh1 mutant embryos may be a result of signals from the other germ layers, which may not occur during the differentiation of Foxh1-mutant ES cells in vitro.
Figure 16. RT-PCR analysis of FKO2 (Foxh1+/+) and FKO3 (Foxh1−/−) ES cells differentiated through EBs for 13 days in the presence and absence of Activin.

RT-PCR of Gapdh (housekeeping gene), Sox17, Foxa2 (visceral and definitive endoderm markers), Nepn (midgut endoderm marker), Pyy (foregut invagination marker), Pdx1 (early pancreatic matter) Foxf1 (extraembryonic mesoderm, embryonic lateral plate mesoderm and posterior primitive streak mesoderm marker), Pax6 (neurectoderm marker), Rex1 (pluripotency marker), Endy and NAPS. Differentiation was carried out in FKO2 (Foxh1+/+) and FKO3 (Foxh1−/−) ES cells for 13 days and 20 ng/ml Activin was added on day 3 of differentiation. For detailed description of expression patterns see Section 3.2.4. Data representative of two biological replicates.
4.1. Endy and NAPS expression mirrors lineage and fate maps of the definitive endoderm and the node, respectively

Recent fate-mapping studies have helped identify developmental fates of definitive endoderm cells during gastrulation (Franklin et al., 2008; Tam et al., 2007; Tam et al., 2007). Prior to gastrulation, in the pre-streak embryo, cells in the endoderm layer have been shown to contribute almost exclusively to the extraembryonic visceral endoderm. Once gastrulation commences, epiblast cells fated to become the gut endoderm of the embryo are recruited in the vicinity of the newly formed primitive streak. Lineage tracing of the gastrulating embryo has demonstrated that precursors of the posterior most endoderm and the dorsal endoderm of the anterior most foregut are recruited first at the early streak stage. Subsequently, additional precursors of the posterior-most endoderm and the ventral portion of the anterior-most foregut are recruited. Precursors for the rest of the fore-, mid- and hindgut are thought to originate during the late streak to late-bud stage. Part of the mid- and hindgut may also be generated shortly after gastrulation by the expansion of the anterior and posterior-most precursors of the definitive endoderm (Lewis and Tam, 2006).

I showed here that Endy first marked a small population of epiblast cells in the very distal region of the embryo (Figure 10: panel A). Lineage tracing will be required to predict if these cells are allocated to the definitive endoderm, as they may not be recruited through the primitive
streak, and may thus form part of the ectoderm. At the early head-fold stage, shortly after the completion of gastrulation, all progenitors of the embryonic gut are presumed to be in place. The anterior endoderm which lies beneath the head folds is fated to form the dorsal and ventral-lateral parts of the foregut. At this stage Endy marked the anterior most region of the anterior endoderm and regions immediately lateral to it, visualized as a crescent in the in situ analysis (Figure 10: panel J). Consistent with the fate mapping analyses, expression of Endy converged to the ventrallateral regions of the foregut in the E8.5 embryo (Figure 10: panel O). The expression of Endy in the posterior endoderm followed suit whereby its expression in the posterior and lateral regions of the developing definitive endoderm were regionalized to the dorsal and ventral regions of the developing hindgut. Interestingly, expression of Endy in the midline ectoderm of the embryo at the early head fold stage (Figure 10: panel K) could represent progeny of epiblast cells in the distal most region of embryo that expressed Endy (Figure 10: panels A). These data suggest that Endy may not be a definitive endoderm lineage marker as it is seen in a small population of ectoderm cells as well. However, Endy could be a positional marker for the different progenitors based on their location in the embryo. Regardless, the transition of definitive endoderm cells expressing Endy from the early head fold stage to E9.0 reflects and corroborates the regionalization of these progenitors observed through others’ fate mapping studies.

In contrast, the expression of NAPS was less dynamic and much more restricted than that of Endy. At the late streak stage NAPS was expressed in cells of the node and the PNC (Figure 12: panel F). This expression then migrated towards the posterior of the developing embryo, ending up in the posterior-most hindgut and notochord. Live imaging studies have illustrated that a stationary and transient population of cells constitute the node (Yamanaka et al., 2007). The transient population of node-derived cells migrates out of the node at the early head-fold stage towards the posterior of the embryo to make up the tail notochord. The stationary node-derived
population gives rise to cells of the PNC and the trunk notochord. As the trunk notochord elongates, the node/PNC is passively pushed towards the posterior most end of the embryo. The early migration of the transient node-derived cells occurs before the push of the node/PNC to the posterior (Yamanaka et al., 2007). How these two populations are reconciled at the posterior end to form the tail notochord and posterior-most endoderm is not known. Expression of NAPS migrates to the posterior, but importantly overlaps with the movement of the node itself (Figure 12: panels G, H, I and J). This suggests NAPS may be expressed by the stationary cells of the node/PNC. In addition, it has been suggested that the node/PNC region carries axial progenitor stem cells that give rise to the trunk and tail notochord (Cambray and Wilson, 2007). It may be that NAPS is a marker for this axial progenitor population. This population is pushed towards posterior while laying down differentiated notochord cells, and ends up in the tail of the embryo where it is further involved in the formation of the tail-bud. Lineage tracing analysis for this population expressing NAPS could yield important answers as to their developmental fate.

4.2. Endy and NAPS as non-coding RNA genes

The role of non-coding RNA in the regulation of gene expression has come to the forefront in recent years, especially with the discovery of miRNA species in eukaryotes. Genes encoding miRNA are very abundant, comprising an estimated 1-3% of the animal genomes, with over 400 miRNA species identified to date in the mouse (Bompfunewerer et al., 2005). A limited amount of complementarity between miRNA and target mRNA is required for efficient miRNA function, and because of this, an even greater number of potential miRNA targets have been predicted (Brennecke et al., 2005). Despite these predictions, only a small number of in vivo targets have been identified (Lin et al., 2006). The study of the involvement of miRNA during
early embryo development is still in its infancy. However, recent findings of miRNA regulation of the Nodal pathway have become of significant interest to the developmental biology field (Martello et al., 2007). For example, specific miRNA species have been shown to regulate different components of Nodal signaling during *Xenopus* and zebrafish embryo development. The involvement of miRNA regulation of Nodal signaling has not yet been explored in the mouse embryo, but it is quite conceivable that such a mechanism would be evolutionarily conserved.

The lack of a definitive ORF gives *NAPS* transcripts potential non-coding RNA status. Furthermore, *NAPS* transcripts may contain precursors that form mature miRNA implying a role in the regulation of gene expression. The temporal and spatial overlap of *NAPS* expression with that of Nodal also suggests a potential for cross talk between these two transcript species. Furthermore, if *NAPS* proves to encode a functional miRNA species, then the question of miRNA mediated control of Nodal comes into play. However, proving the existence of a novel miRNA species is an arduous task as strict criteria have to be met for a gene to be annotated as a miRNA precursor (Ambros et al., 2003). The lack of homology of any of the potentially ‘real’ stem-loops of *NAPS* to other known stem-loops provided by the SANGER database do not entirely disprove the idea of *NAPS* being a miRNA gene, but make it difficult to ascribe the label. A better way to identify a miRNA potential of *NAPS* would be to discover a 21-23nt mature miRNA in a tissue sample expressing *NAPS*, and to map the sequence of that miRNA back to the *NAPS* transcript. Detection of a mature *NAPS* miRNA using techniques such as sequencing of size-fractionated RNA from tissues expressing *NAPS* would be a big step towards proving its status as a miRNA-encoding gene (Berezikov et al., 2006). If *NAPS* turns out not to encode a miRNA, its highly restricted expression pattern suggests that it may be involved in some form of regulation, and may be another form of the multitude of small non-coding RNAs that are thought to exist (Mattick and Makunin, 2006).
The Endy gene contains a hypothetical ORF; however the part of the Endy transcript that carries the predicted protein coding information is poorly conserved among different organisms. This brings into question the functional relevance of this gene during development. As discussed earlier, a number of repetitive element insertions in the 3’ half of the gene may have rendered Endy nonfunctional, and thus a transcribed pseudogene. However, it is also possible that Endy does encode a protein, which is only a requirement to mouse and rat development, since the rat genome showed homology to the predicted protein coding region. Regardless of whether or not it is a functional gene, Endy was expressed in the definitive endoderm and, more importantly in specific populations of the definitive endoderm of the mouse. As such, it could be a useful marker to use in characterization of other functional definitive endoderm factors. In addition, the availability of several gene trap ES cell lines gives an opportunity to create a knockout mouse to elucidate the functional relevance of Endy during definitive endoderm development.

4.3. Expression patterns of Nepn and Pyy in our ES cell system suggest formation of populations of definitive endoderm

ES cells are not only the beginnings for a number of potential regenerative therapies for diseases of the endoderm, but can also be useful tools for answering basic molecular questions towards the development of this tissue. It has become increasingly important to distinguish whether ES cells are differentiating through the visceral or definitive endoderm lineage. Identifying novel markers for definitive endoderm prompted us to establish an ES cell system to analyze the expression of these markers. Through a systematic study of definitive endoderm SAGE libraries, Nepn and Pyy were previously identified as markers exclusive to the definitive endoderm (Hou et al., 2007). Using these markers I was able to demonstrate that ES cells
induced to differentiate in the presence of serum and Activin indeed produced definitive endoderm. Recent studies have removed the signaling factor from serum by using serum-free media in combination with Activin to try and generate purer populations of endoderm (Gouon-Evans et al., 2006). Unfortunately in our hands, the use of serum-free media caused dramatically reduced colony formation, increased cell death, and almost eliminated EB formation (data not shown). The success of other studies in using serum-free media could be attributed to the fact that the majority of these studies used genetically modified cell lines that had been selected to either grow or differentiate to endoderm under serum-free conditions. Our goal was to use genetically unmanipulated ES cells, thus the presence of serum was found to be necessary.

In the ES cell differentiation system I used, I found that exogenous Activin increased the expression of the novel definitive endoderm markers, Nepn and Pyy, in LP9 ES cells as compared to control cells not exposed to Activin. Furthermore, the presence of Activin reduced expression of genes marking the other two germ layers and pluripotent cells. Given that the levels of serum were the same in the cultures to which Activin was added, or not, these expression patterns suggest that Activin can promote differentiation towards the definitive endoderm. The lack of such expression patterns in CCE ES cells is not surprising as differences in the levels of signaling, due to limiting factors such as the number of receptors or intracellular factors, are known to exist between different ES cell lines.

The use of ES cells also made it possible to test the expression of Endy and NAPS during the process of definitive endoderm formation. Consistent with the SAGE library analysis, Endy expression was observed in undifferentiated and differentiating ES cells. However, the reduced but continued expression of Rex1 during the differentiation of the ES cells suggested that undifferentiated ES cells persisted in these cultures. Thus, the continued expression of Endy during differentiation could be attributed to these cells. However, through both SAGE and in situ analysis, I showed that Endy is also expressed in specific populations of the definitive endoderm.
Hence, the expression of Endy observed upon induction of ES cell differentiation is likely attributable to both persisting ES cells and appearing definitive endoderm cells. Since Endy is valuable as a marker for specific endoderm cells, a longer time course of differentiation to ensure maximum differentiation of ES cells, aided by the use of several pluripotency markers, would allow the use of Endy to isolate those cell populations for further study. In contrast to Endy, NAPS expression was induced upon differentiation of the ES cells. From the in situ analysis, I found that NAPS and Endy expression overlapped in the very posterior region of the hindgut at E9.0. Thus, in conjunction, Endy and NAPS should be valuable markers to isolate those specific hindgut populations in future studies.

The developmental potential of cell populations cannot be determined through RT-PCR analysis on its own. Unequivocal identification of definitive endoderm induction in vitro would require isolation of cells expressing these markers, growth of these cells and further transplantation to see whether they are able to generate endoderm-derived tissue. Isolating definitive endoderm cells has been difficult due to the lack of cell surface markers for identification. Recently many groups have utilized CXCR4, a chemokine receptor, to isolate endoderm cells from ES cells differentiated towards definitive endoderm (D'Amour et al., 2005; Yasunaga et al., 2005). However, CXCR4 is not an ideal marker for definitive endoderm as it is expressed in a wide variety of tissues during gastrulation (McGrath et al., 1999). The generation of transgenic tagged cell lines using our novel markers might thus offer a superior strategy to allow definitive endoderm precursors to be isolated for further differentiation and analysis studies.

I also tested the effects of blocking the TGFβ signaling pathway on definitive endoderm specification. Inhibition of the TGFβ signaling pathway using the small molecule inhibitor SB-431542 caused a reduction in the expression of Nepn and Pyy during ES cell differentiation. This was consistent with previous studies that demonstrated the importance of TGFβ signaling in the
formation of the primitive streak and the subsequent recruitment of definitive endoderm progenitors as discussed earlier (Robertson et al., 2003; Weng and Stemple, 2003). Interestingly, NAPS was still expressed in the presence of the inhibitor suggesting that the NAPS may not be positively regulated by TGFβ signaling. The negligible expression of Nepn and Pyy coupled with an increased expression of Rex1 and Pax6 in the presence of the TGFβ inhibitor indicates that cells either remained in a pluripotent state or were encouraged to differentiate down the neurectodermal pathway. This result corroborates various studies in mouse and human ES cells whereby the absence of specific intercellular signals promotes ES cells to differentiate to neurectoderm lineages (Munoz-Sanjuan and Brivanlou, 2002). Together, these results show that the ES cell system used here is amenable to manipulation and useful for investigating the mechanisms that govern definitive endoderm specification. Inhibition of other signaling pathways, such as the Notch, Wnt, and FGF, important during development, will further aid in understanding this process.

Much of the role of TGFβ signaling factors during endoderm development has been inferred from chimeric analysis performed with mutant ES cell lines. For instance, ES cells mutant for Foxh1 fail to colonize the definitive endoderm suggesting a role for Foxh1 in the formation of the definitive endoderm. However, the effects of loss of Foxh1 on gut development is not that clear as Nepn is still strongly expressed in Foxh1 mutant embryos, whereas the expression of Pyy is reduced to just a few cells (Hou, McKnight and Hoodless unpublished). These data suggest the Foxh1 is required for the production of certain populations of the definitive endoderm. To see how our ES cell differentiation system compared to what occurs in the Foxh1 mutant embryo I analyzed differentiated Foxh1- mutant ES cells. Contrary to in vivo observations, Nepn expression was not induced, while Pyy expression persisted in the absence of Foxh1. It could be inferred that upon differentiation of Foxh1-mutant ES cells, the Pyy expressing cells are representative of the small population of in vivo Pyy expressing cells in the
Foxh1 mutant embryos. The in vitro conditions may represent the level of Nodal signaling required to give rise to the cells that express Pyy in vivo because knocking out Foxh1 affects the positive feedback loop of Nodal signaling. In contrast, the lack of Nepn expression may be a result of the limited structure of the EBs. Positional signals from the other germ layers may be an important requirement for the induction Nepn in the Foxh1 mutant embryos, and these signals may be absent during the differentiation of Foxh1-mutant ES cells.

Despite the discrepancy in the expression patterns of Nepn and Pyy between the in vitro and in vivo approaches, the presence of Pyy and of the early pancreatic marker Pdx1 upon differentiation of Foxh1-mutant ES cells, and the strong presence of Nepn in the Foxh1 mutant embryos implies that there is some Foxh1-independent TGFβ signaling or TGFβ-independent signaling giving rise to populations of the definitive endoderm. Together these results highlight that in vitro analysis does not always correlate with in vivo observations, and thus cannot be exclusively used to study the definitive endoderm. In synergy with in vivo analyses, however, in vitro approaches can be powerful tools to further our understanding of the development and specification of the definitive endoderm.

4.4. Insights and conclusions

4.4.1. SAGE as tool to study early mouse development

Large scale systematic gene expression analyses of early mouse embryo development provide useful information to identify genes differentially expressed in a given cell or tissue type at a particular time. The focus has slowly shifted from trying to sequence every possible gene to elucidating the biological relevance of those genes. Large-scale gene expression studies allow
one the opportunity to study various cellular processes and signaling pathways that are occurring to ultimately use in functional studies. Large-scale sequence-based techniques such as SAGE have recently stood out over fluorescence-based arrays because SAGE not only provides a large scale and quantitative analysis of the transcriptome, but also provides an opportunity to identify novel transcripts (Anisimov, 2008). Furthermore, SAGE is already being overtaken by next generation sequencing technologies that can now be extended to single cell gene expression profiling (Gupta, 2008; Kurimoto et al., 2006). Since the developing embryo consists of a multitude of different cell types translating to a molecularly heterogenous development, single cell level monitoring of genome-wide gene expression will be a powerful tool to identify cell-specific genetic markers. Such technologies will be powerful tools to gain an in depth understanding of molecular mechanisms that regulate early development, as well as mechanisms that become aberrant in disease models. Understanding both enables the design of more specific disease therapeutics

There is a continuous need for tissue and cell-specific genetic markers in order to understand embryonic development. These markers serve to characterize defects in knockout embryos, to analyze gene function in specific cell lineages, to monitor cell populations induced during differentiation of ES cells and for lineage tracing studies used to create fate maps of embryonic progenitors. Understanding the formation of mouse definitive endoderm has lagged behind that of mesoderm and ectoderm due to a lack of genetic markers specific to this tissue. As discussed earlier, several of the traditionally used definitive endoderm markers such as \( \text{Sox17} \) and \( \text{Foxa2} \), are also expressed in the extraembryonic visceral endoderm (Ang et al., 1993; Kanai-Azuma et al., 2002). Our lab has recently used SAGE to identify novel markers, such as \( \text{Nepn} \) and \( \text{Pyy} \), of the definitive endoderm (Hou et al., 2007). The extended study of SAGE described here has uncovered additional markers, such as \( \text{Endy} \) and \( \text{NAPS} \), which mark different progenitors of the definitive endoderm at different stages of development. \( \text{Endy} \) and \( \text{NAPS} \)
should be useful in the characterization of mutant embryos deficient for various signaling pathways that affect the development of the definitive endoderm.

The availability of these markers specific to the definitive endoderm also allow excellent means to create tools, such as the creation of GFP-tagged constructs that will allow the isolation of specific in vivo derived populations of definitive endoderm through the use of Fluorescence Activated Cell Sorting (FACS). Since definitive endoderm formation is truly a heterogeneous process in molecular terms, such tools will allow isolation and in vitro manipulations of specific definitive endoderm populations leading to a better understanding of its formation and development. Furthermore, such molecular tools will facilitate the creation of more accurate fate maps of the formation of the definitive endoderm.

4.4.2. ES cell differentiation as model to study early mammalian lineage commitment decisions

The establishment of ES cell lines has opened many new experimental avenues towards the understanding of mammalian developmental biology. ES cells are totipotent cells derived from the inner cell mass of the blastocyst. ES cells are able to spontaneously differentiate and generate various lineages associated with embryological development under appropriate culture conditions. Differentiation of ES cells in vitro provides an attractive and powerful model for addressing some of the question regarding early lineage commitments.

Using an ES model over an in vivo approach in study of mouse development has multiple advantages. For instance, generating lineages from ES cells in vitro provides access to early precursors that are almost impossible to access in vivo. In addition, ES cells carrying targeted mutations in genes required for development can be analyzed for their developmental and
differentiation potential. This approach is much more favorable when the analysis of such mutations in vivo is complicated due to early embryonic lethality of such genes.

Using novel markers identified through SAGE, \( \text{Nepn} \) and \( \text{Pyy} \) we successfully established and characterized an ES cell system to definitive endoderm. Furthermore, we were able to manipulate the system using a small molecule inhibitor against the TGFβ signaling pathway, thought to be a requirement for the development of definitive endoderm, and showed that blocking the signaling downregulated the expression of \( \text{Nepn} \) and \( \text{Pyy} \). Thus, ES cell differentiation systems are valuable models to study signaling pathways during early embryonic development.

Together, the further in vivo characterization of our novel definitive endoderm markers, and that of the in vitro derived endoderm will enhance our understanding of the mechanisms that regulate the specification and development of the definitive endoderm.
REFERENCES


APPENDIX

The following are the animal care certificates required during this thesis.

ANIMAL CARE CERTIFICATE A05-1852

THE UNIVERSITY OF BRITISH COLUMBIA

ANIMAL CARE CERTIFICATE

Application Number: A05-1852

Investigator or Course Director: Pamela Hoodless

Department: Medical Genetics

Animals:

- Mice B6.Cg-Tg(Alb-Cre)21Mgn/J 100
- Mice Foxa2tm1Dnl 400
- Mice CD1 500
- Mice R26R-YFP 100
- Mice B6.129S7-Twist1tm1Bhr/J 400
- Mice Foxh1tm1Jlw 400
- Mice C57BL6/J 500

Start Date: January 20, 2006

Approval Date: February 17, 2009

Funding Sources:

Funding Agency: Genome Canada
<table>
<thead>
<tr>
<th>Funding Title</th>
<th>Dissecting Gene Expression Networks in Mammalian Organogenesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Funding Agency</td>
<td>Canadian Institutes of Health Research (CIHR)</td>
</tr>
<tr>
<td>Funding Title</td>
<td>Characterization of regulatory regions, modules and elements in mammalian genomes</td>
</tr>
<tr>
<td>Funding Agency</td>
<td>Genome British Columbia</td>
</tr>
<tr>
<td>Funding Title</td>
<td>Dissecting Gene Expression Networks in Mammalian Organogenesis</td>
</tr>
</tbody>
</table>

| Unfunded title | N/A |

The Animal Care Committee has examined and approved the use of animals for the above experimental project.

This certificate is valid for one year from the above start or approval date (whichever is later) provided there is no change in the experimental procedures. Annual review is required by the CCAC and some granting agencies.

A copy of this certificate must be displayed in your animal facility.

Office of Research Services and Administration
102, 6190 Agronomy Road, Vancouver, BC V6T 1Z3
Phone: 604-827-5111 Fax: 604-822-5093
ANIMAL CARE CERTIFICATE

Application Number: A05-1854

Investigator or Course Director: Pamela Hoodless

Department: Medical Genetics

Animals:

- Mice Foxh1tm1Jlw 400
- Mice C57BL6/J 150
- Mice CD-1 (ICR) 150
- Mice Hippi 300
- Mice Foxa2tm1Dnl 400
- Mice TGIF 400

Start Date: July 1, 2002

Approval Date: February 17, 2009

Funding Sources:

- Funding Agency: National Cancer Institute of Canada
- Funding Title: TGFB signal transduction pathways in development

Unfunded title: N/A
The Animal Care Committee has examined and approved the use of animals for the above experimental project.

This certificate is valid for one year from the above start or approval date (whichever is later) provided there is no change in the experimental procedures. Annual review is required by the CCAC and some granting agencies.

A copy of this certificate must be displayed in your animal facility.

Office of Research Services and Administration
102, 6190 Agronomy Road, Vancouver, BC V6T 1Z3
Phone: 604-827-5111 Fax: 604-822-5093
ANIMAL CARE CERTIFICATE

Application Number: A08-0660

Investigator or Course Director: Pamela Hoodless

Department: Medical Genetics

Animals:

<table>
<thead>
<tr>
<th>Mice ICR 150</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mice R26R-YFP 100</td>
</tr>
<tr>
<td>Mice Hippi 50</td>
</tr>
<tr>
<td>Mice Foxh1tm1Jlw 150</td>
</tr>
<tr>
<td>Mice Nepn-Cre-ER 100</td>
</tr>
<tr>
<td>Mice Foxa2tm1Dnl 50</td>
</tr>
<tr>
<td>Mice R26R-lacZ 100</td>
</tr>
<tr>
<td>Mice C57BL6/J 150</td>
</tr>
<tr>
<td>Mice Nepn-cre 100</td>
</tr>
</tbody>
</table>

Start Date: October 1, 2008  Approval Date: October 27, 2008

Funding Sources:

Funding Agency: Canadian Institutes of Health Research (CIHR)
Funding Title: Formation and patterning of the definitive endoderm
The Animal Care Committee has examined and approved the use of animals for the above experimental project.

This certificate is valid for one year from the above start or approval date (whichever is later) provided there is no change in the experimental procedures. Annual review is required by the CCAC and some granting agencies.

A copy of this certificate must be displayed in your animal facility.

Office of Research Services and Administration
102, 6190 Agronomy Road, Vancouver, BC V6T 1Z3
Phone: 604-827-5111 Fax: 604-822-5093
ANIMAL CARE CERTIFICATE A08-0703

THE UNIVERSITY OF BRITISH COLUMBIA

ANIMAL CARE CERTIFICATE
BREEDING PROGRAMS

Application Number: A08-0703

Investigator or Course Director: Pamela Hoodless

Department: Medical Genetics

Animals:

- Mice Nepn-Cre-ER 150
- Mice B6.129-Twist1tm1Bhr/J 50
- Mice ICR 250
- Mice Foxh1tm1Jlw 400
- Mice TGIF 50
- Mice Foxa2tm1Dnl 100
- Mice R26R-LacZ 150
- Mice Hippi 100
- Mice C57BL6/J 250
- Mice Nepn-cre 150
- Mice R26R-YFP 150

Approval Date: October 27, 2008

Funding Sources:
Funding Agency: Canadian Institutes of Health Research (CIHR)
Funding Title: Formation and patterning of the definitive endoderm

Unfunded title: N/A

The Animal Care Committee has examined and approved the use of animals for the above breeding program.

This certificate is valid for one year from the above approval date provided there is no change in the experimental procedures. Annual review is required by the CCAC and some granting agencies.

A copy of this certificate must be displayed in your animal facility.

Office of Research Services and Administration
102, 6190 Agronomy Road, Vancouver, BC V6T 1Z3
Phone: 604-827-5111 Fax: 604-822-5093