# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Abbreviations</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables and Figures</td>
<td>iv</td>
</tr>
<tr>
<td>Abstract</td>
<td>vi</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>II. GENECHIP ANALYSIS OF HUMAN EMBRYONIC STEM CELL DIFFERENTIATION INTO HEMANGIOBLASTS: AN IN SILICO DISSECTION OF MIXED PHENOTYPES</td>
<td>29</td>
</tr>
<tr>
<td><a href="#">Genome Biology 2007;8(11):R240.</a></td>
<td></td>
</tr>
<tr>
<td>III. MICROARRAY ANALYSIS OF BLADDER SMOOTH MUSCLE FROM PATIENTS WITH MYELOMENINGOCOELE</td>
<td>84</td>
</tr>
<tr>
<td><a href="#">British Journal of Urology International. 2008 Sep;102(6):741-6.</a></td>
<td></td>
</tr>
<tr>
<td>IV. THE EFFECTS OF ETHANOL ON AMNIOTIC FLUID-DERIVED STEM CELLS</td>
<td>110</td>
</tr>
<tr>
<td>[Submitted to Alcoholism (3/2009)]</td>
<td></td>
</tr>
<tr>
<td>V. TRANSCRIPTIONAL PROFILING OF AMNIOTIC FLUID-DERIVED STEM CELL IDENTIFIES MULTIPLE LINEAGE PROGRAMS</td>
<td>146</td>
</tr>
<tr>
<td><a href="#">In preparation</a></td>
<td></td>
</tr>
<tr>
<td>VI. Discussion</td>
<td>179</td>
</tr>
<tr>
<td>Scholastic Vita</td>
<td>189</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

ESC, Embryonic Stem Cell
IPS cell, Induced Pluripotent Stem cell
EBs, Embryoid Bodies
BCs, hemangioblast cells
AFS cell, Amniotic Fluid-derived Stem cell
NSC: Neural Stem Cell
HSC, Hematopoietic Stem Cell
hUVEC, human Umbilical Vein Endothelial Cell
MSC, mesenchymal stem cell
DAVID, Database for Annotation, Visualization, and Integrated Discovery
U133A, Affymetrix Human 133A GeneChip
MAS5, Microarray Suite 5.0
RMA, Robust Multichip Average
LIMMA, Linear Model of MicroArray data
EASE, Expression Analysis Systematic Explorer
GenMAPP, Gene Map Annotator and Pathway Profiler
MM, Myelomeningocele
SMC, Smooth Muscle Cell
EtOH, Ethanol
GO, Gene Ontology
OPN, Osteopontin
LIST OF TABLES AND FIGURES

Chapter I:
Table 1: Summary of methods for generating pluripotent stem cells ........23

Chapter II:
Table 1: Characterization of hES-BCs by Immunocytochemistry (IC) and Affymetrix arrays .................................................................36
Figure 1: Validation of differentially expressed genes by RT-PCR in human ESCs, EBs, and BCs .................................................................37
Table 2: Gene ontologies for up-regulated processes in BCs vs ESCs......41
Table 3: Gene ontologies of up-regulated processes in BCs versus epithelial cells .........................................................................................45
Figure 2: GenMAPP of complement and coagulation ...........................47
Figure 3: GenMAPP of heme biosynthesis ............................................48
Table 4: Gene ontologies for up-regulated processes in BCs versus leukocytes .........................................................................................51
Table 5: Gene ontologies for up-regulated processes in BCs versus endothelial cells ..................................................................................53
Table 6: Gene ontologies for up-regulated processes in BCs versus stromal cells .........................................................................................56
Figure 4: Ingenuity pathway analysis of network of genes expressed in EB and BC data sets from level II analysis ........................................61
Figure 5: Ingenuity pathway analysis of genes expressed in EB and BC data sets associated with VEGF ...........................................................62
Figure 6: Ingenuity pathway analysis of genes expressed in EB and BC data sets associated with GATA2 .........................................................63
Table 7: Primer sequences .....................................................................74

Chapter III:
Table 1: GO analysis of gene over-expressed in MMSMCs ....................92
Figure 1: Schematic depiction of the algorithm used for illustrating prominent biological themes and pathways ...........................................93
Figure 2: GeneMapp analysis of muscle development and contraction with genes over- and under-expressed in MMSMCs ..........................96
Figure 3: GeneMapp pathways of transmission of nerve impulse with genes that were over- or under-expressed in MMSMCs .........................98
Figure 4: GenMAPP analysis of gene over- and under-expressed in MM SMCs.................................................................................................................... 99

Figure 5: Ingenuity analysis was used to categorize the top 10 biological systems, and the development and function of up-regulated genes in MM SMCs .................................................................................................................... 101

Chapter IV:
Figure 1: Effect of ethanol on cell proliferation and viability ......................... 122
Table 1: Enrichment of biological themes of ethanol-responsive genes .... 126
Figure 2: Effect of ethanol on OPN expression in AFSCs.............................. 130
Figure 3: Effect of ethanol on alkaline phosphatase activity in AFSCs ...... 133
Figure 4: Alkaline phosphatase activity when AFSCs were exposed to ethanol at the midpoint of differentiation........................................ 135
Figure 5: The effect of ethanol on calcium deposition................................. 137

Chapter V:
Figure 1: Experimental design of microarray experiment............................ 155
Table 1: Gene ontologies of genes that were up- and down-regulated during osteogenic differentiation of AFSCs by direct comparisons....... 158
Table 2: Enrichment analysis of multiple reference-define global gene expression profiles of undifferentiated AFSCs............................. 162
Figure 2: Enrichment analysis of reference-define global gene expression profiles during osteogenic differentiation................................. 167
ABSTRACT

Jennifer A. Hipp

APPLICATIONS OF MICROARRAYS AND STEM CELLS FOR REGENERATIVE MEDICINE

Dissertation under the direction of
Shay Soker, Ph.D., Associate Professor of Regenerative Medicine and Surgical Sciences

Stem cells are viewed as a potential resource for cell-based therapy and regenerative medicine. Understanding the genes that govern the special properties of stem cells has implications for both stem cell biology and regenerative medicine. This information will ensure the quality of the cells, the specificity of differentiation, and the assessment of mixed phenotypes. Microarray analysis measures the global expression of genes and can provide insight into the genetic programs expressed in stem cells. A novel in silico dissection method based on comparisons of microarray data between stem cells and multiple cell types was developed to better understand the properties of human embryonic stem cells, tissue progenitor cells, and amniotic fluid-derived stem cells. By comparing microarray data of tissue-specific reference files to microarray data of stem cells, we were able to identify similarities in particular phenotypes while revealing other novel signatures. This method can be used to better understand stem cell differentiation. By comparing stem cells and their osteogenic derivatives to multiple reference cell types, we found that stem cells express multiple lineage-specific programs. In silico dissection can also be used
to evaluate the therapeutic potential of stem cells derived from a diseased organs and individuals for tissue engineering. We demonstrate the use of this method to study stem cells in a model of fetal alcohol spectrum disorders. In conclusion, we propose that stem cell differentiation requires both up-regulation of a lineage-specific program and the elimination of other lineages.
CHAPTER I

INTRODUCTION

SOURCES OF STEM CELLS FOR REGENERATIVE MEDICINE

Jennifer A. Hipp and Anthony Atala

The following manuscript was published in *Stem Cell Reviews* 2008 Spring;4(1):3-11 and is reprinted with permission. Stylistic variations are due to the requirements of the journal.
Abstract:

The shortage of organ donors for regenerative medicine has stimulated research on stem cells as a potential resource for cell-based therapy. The development of innovative methods to generate stem cells from different sources suggests that there may be new alternatives for cell-based therapies. Here, we provide an overview of human embryonic stem cells (ESCs) and the methods for obtaining these cells and other broadly multipotent or pluripotent cell types. These methods include somatic cell nuclear transfer, single cell embryo biopsy, arrested embryos, altered nuclear transfer, and reprogramming somatic cells. We also discuss the use of amniotic-fluid derived stem cells (AFSCs) for potential therapies.
Introduction

Patients with diseased organs can be treated by organ transplantation. However, the number of patients in need of new organs far exceeds the organ supply, and this organ shortage is expected to worsen as the aging population increases. The goal of regenerative medicine is to replace or restore normal function of cells, tissues, and organs that are damaged by disease. Regenerative medicine strategies usually fall into three categories: (1) cell based therapy, (2) the use of biomaterials (scaffolds) alone, or (3) the use of scaffolds seeded with cells.

Cells used for tissue engineering are obtained from a small biopsy of tissue which is dissociated in culture. The resulting cell population is expanded, seeded onto a matrix, and implanted back into the host. The source of donor tissue can be allogenic (donor derived) or autologous (the host’s cells), but autologous cells are preferred because they are not rejected by the immune system, and the use of immunosuppressant drugs is avoided. However, inherent difficulty of ex vivo expansion is a major limiting factor for use of some autologous cells. Even though some organs have a very high regeneration rate in vivo (for example, the liver), cells from these organs can be difficult to expand in vitro, which hinders their clinical potential. By studying the conditions that regulate and guide in vivo regeneration after injury and exploring conditions that promote cell differentiation, one can overcome this limitation and achieve extensive expansion in vitro. Studies have now shown that urothelial cells, which previously had limited cell expansion in vitro, can be expanded to cover 4,202 m²
(roughly the size of a football field). This is achieved by isolating and expanding the progenitor cells contained in the cell culture and maintaining their undifferentiated state(1). Due to these advances, it is now possible to isolate cells from a patient’s bladder, to expand them in culture, and to implant them back into the patient in sufficient quantities for regenerative purposes (2,3).

While autologous cells are recognized as the ideal transplantation resource, some patients with end-stage organ disease do not produce enough cells for transplantation. In this case, allogenic cells may be advantageous. Furthermore, some primary cells, whether autologous or allogenic, cannot be expanded from particular organs, such as the pancreas. In these situations, pluripotent stem cells are envisioned as an alternative source of cells from which the desired tissue can be derived. Pluripotent stem cells represent an endless source of versatile cells that could lead to novel sources of replacement organs.

Human embryonic stem cells (hES) are pluripotent (the ability to differentiate into most tissues of the embryo) and retain the ability to self-renew. However, the derivation of ESCs requires the destruction of embryos, and as a result, they are ineligible for federal funding in the United States at this time. Although President Bush rejected a bill that would have expanded research funding and allowed creation of new ESC lines, he issued an executive order which gives support to research on alternative sources of pluripotent stem cells that can be derived without creating or destroying embryos (3).
Adult Stem Cells and Tissue Progenitor Cells

Adult stem cells are, especially in the area of hematopoietic stem cells, better understood than any other aspect of stem cell biology(4). In fact, cell based therapy dates back to the first bone marrow transplant in 1956(5), though the first evidence that a specific, “master” cell type might be responsible for reconstituting the bone marrow was gained from experience with persons exposed to lethal doses of radiation in 1945. Till and McCulloch analyzed the bone marrow in the early 1960s to determine which cells were responsible for marrow reconstitution, and discovered cells with the ability to renew themselves as well as the ability to differentiate into various cell types(6, 7). These two characteristics are still used to define stem cells today. Later, gastrointestinal stem cells (crypt cells) were discovered as scientists studied the regeneration of the intestinal mucosa after sublethal radiation doses, and this discovery ignited the search for stem cells in other tissues.

Today, that search is still an intense area of study, as the potential for progenitor cell therapy is applicable to a myriad of degenerative conditions. Within the last decade, it has been found that progenitor cell populations are present in many adult tissues other than the bone marrow and the gastrointestinal tract, including the brain, skin, and muscle. The discovery of such tissue specific progenitors has opened up new avenues for research.

Adult stem cells tend to be tissue specific, self-renewing populations of cells which can differentiate into cell types associated with the organ system in which they reside (8, 9). They are quite rare, on the order of 1 in 10,000 cells
within the tissue of interest (10, 11). Currently, it is known that niches of stem cells exist in many tissues, such as bone marrow, brain, liver, skin, skeletal muscle, the gastrointestinal tract, the pancreas, the eye, blood, and dental pulp (8, 9, 12, 13). Of these, the most studied are CD34+ hematopoietic stem cells isolated from bone marrow. These cells are capable of producing cells of the lymphoid and myeloid lineages in blood. CD34+ cells are the only currently available therapeutic application of stem cells and are used for a variety of purposes, most often the reestablishment of the immune system after a disease or toxic therapy has damaged it.

A notable exception to the tissue specificity of adult stem cells is the mesenchymal stem cell, or what is more recently called the multipotent adult progenitor cell. This cell type is derived from bone marrow stroma (9, 11, 14, 15). Such a cell has been shown to differentiate in vitro into numerous tissue types and to also differentiate developmentally if injected into a blastocyst. Multipotent adult progenitor cells will develop into multiple tissues including neuronal, adipose, muscle, liver, lungs, spleen, and gut, but notably, not bone marrow or gonads (15).

However, research on adult stem cells has been slow, largely because great difficulty has been encountered in maintaining adult non-mesenchymal stem cells in culture. There are challenges involved in maintaining and expanding long term cultures of adult stem cells in large numbers. Isolation has also proven to be quite problematic as these cells are present in extremely low numbers in the adult tissue. Such cells are often selected utilizing Fluorescent
Activated Cell Sorting (FACS) or Magnetic Activated Cell Sorting (MACS) against surface markers specific to the stem cell of interest (16), but sometimes there is no known marker specific for a type of stem cell, so these methods cannot be used. However, new markers are being described at a rapid pace.

While current use of adult stem cells is quite limited, there is great potential in future utilization of such cells for the use of tissue specific regenerative therapies. The advantage of adult stem cells is that they can be used in autologous therapies, thus avoiding any immune rejection complications.

**Embryonic Stem Cells**

According to data from the Centers for Disease Control, as many as 1 million Americans will die every year from disease that, in the future, may be treatable with tissues derived from stem cells (17). Diseases that might benefit from embryonic stem cell-based therapies included diabetes, heart disease, cerebrovascular disease, liver and renal failure, spinal cord injuries and Parkinson’s disease.

In 1981, pluripotent cells were found in the inner cell mass of the mouse embryo, and the term “embryonic stem cell” was coined (18). These cells are able to differentiate into all cells, excluding placental cells (only cells from the morula are totipotent; that is, able to develop into all cell types).

The political controversy surrounding stem cells began in 1998 with the creation of ESCs derived from discarded, non-transferred human embryos. ESCs were isolated from the inner cell mass of a blastocyst (an embryo 5 days post-fertilization) using an immunosurgical technique. Using this technique, the
blastocyst was incubated with antibodies specific to the trophectoderm. Complement proteins then resulted in lysis of the trophectoderm so that the only surviving cells were the inner cell mass (19). Given that some cells cannot be expanded ex vivo, ES cells could be the ideal resource for tissue engineering because of their fundamental properties: the ability to self-renew indefinitely and the ability to differentiate into cells from all three embryonic germ layers. However, their clinical application is limited because they represent an allogenic resource and thus have the potential to evoke an immune response. New stem cell technologies (such as somatic cell nuclear transfer and reprogramming) promise to overcome this limitation.

**Somatic Cell Nuclear Transfer**

Somatic cell nuclear transfer (SCNT) entails the removal of an oocyte nucleus in culture, followed by its replacement with a nucleus derived from a somatic cell obtained from a patient. Activation with chemicals or electricity stimulates cell division up to the blastocyst stage, at which time the inner cell mass is isolated and cultured, resulting in ES cells that are genetically identical to the patient. It has been shown that nuclear transferred ES cells derived from fibroblasts, lymphocytes, and olfactory neurons are pluripotent and generate live pups after tetraploid blastocyst complementation, showing the same developmental potential as fertilized blastocysts (20-23). The resulting ES cells are perfectly matched to the patient’s immune system and no immunosuppressants would be required to prevent rejection.
Although ES cells derived from SCNT contain the nuclear genome of the donor cells, mitochondrial DNA inherited by the oocyte could lead to immunogenicity after transplantation. To assess the histocompatibility of nuclear transfer-generated tissue, the nucleus of a bovine skin fibroblast was microinjected into an enucleated oocyte (24). Although the blastocyst was implanted, the purpose was to generate renal, cardiac and skeletal muscle cells, which were then harvested, expanded in vitro, and seeded onto biodegradable scaffolds. These scaffolds were then implanted into the donor from whom the cells were cloned to determine if cells were histocompatible. Analysis revealed that cloned renal cells showed no evidence of T-cell response, suggesting that rejection will not necessarily occur in the presence of oocyte-derived mtDNA.

Although promising, SCNT has certain limitations that require further improvement before its clinical application, in addition to ethical considerations regarding the potential of the resulting embryos to develop into cloned embryos if implanted into a uterus. Many animal studies have shown that blastocysts generated from SCNT can give rise to a liveborn infant that is a clone of the donor when implanted into a uterus. In 1997, for example, a sheep named Dolly was derived from an adult somatic cell using nuclear transfer (25). This is known as reproductive cloning, which is banned in most countries for human applications. In contrast, therapeutic cloning is used to generate only ES cell lines whose genetic material is identical to that of their source. In this case blastocysts are allowed to grow until a 100 cell-stage where ES cells can be obtained, and thus, the blastocysts are never implanted into a uterus. In
addition, this technique has not been shown to work in humans. The initial failures and fraudulent reports of nuclear transfer in humans reduces enthusiasm for human applications (26-28). However, non-human primate ES cell lines have been generated by SCNT of nuclei from adult skin fibroblasts (29, 30). This group used a modified SCNT approach that avoids the use of Hoechst 33342 and ultraviolet light, which is detrimental to the quality of the cytoplasts, to completely remove spindles containing nuclear DNA (30). They used an Oosight spindle imaging system to efficiently enucleate primate oocytes. A total of 304 oocytes yielded 35 blastocysts, from which two ES cell lines were derived. Both lines demonstrated typical ES cell morphology. They also demonstrate self-renewal and express OCT4, SSEA4, LEFTYA, TDGF, TRA1-60 and TRA1-80. To test their differentiation potential, the cells were exposed to cardiomyocyte differentiation conditions, and this produced contracting aggregates that expressed markers of cardiac muscle tissue. Neural differentiation resulted in the expression of microtubule-associated protein 2 (MAP2), B-tubulin, and tyrosine hydroxylase. When injected into SCID mice, SCNT-derived ES cells induced teratomas which contained differentiated cell types representing all three embryonic germ layers. More importantly, microsatellite single nucleotide polymorphism and mtDNA analysis confirmed that nuclear DNA was identical to the donor somatic cells and that the mtDNA originated from the oocyte.

Before SCNT-derived ES cells can be used as clinical therapy, careful assessment of quality of the lines must be determined. Unfortunately, one of the lines generated by SCNT revealed a translocation consisting of an
isochoromosome comprised of two copies of the long arm of the Y chromosome. It is not known whether chromosomal abnormalities in SCNT-derived ES cells originate from aneuploid embryos or occurred during ES cell isolation and culture.

The low efficiency of SNCT (0.7%) and the inadequate supply of human oocytes further hinder the therapeutic potential of this technique. Although promising, SCNT has limitations that require further improvement before clinical application, including ethical considerations regarding the potential of the generated blastocysts to develop into cloned embryos if implanted into a uterus. Furthermore, the destruction of embryos is not an ethically acceptable means to generate pluripotent stem cells. On the other hand, this study renews the hope that ES cell lines could be generated by SCNT in humans to generate patient-specific stem cells with the potential to cure/treat many human diseases that are currently untreatable.

**Single Cell Embryo Biopsy**

Since the major objection to human embryonic stem cell research is the destruction of embryos, it would be advantageous to develop a method of isolating these cells from and embryo without destroying it. In 2006, Chung et al. were the first to report the generation of mouse embryonic stem cell lines without destroying the embryo (31). This alternative method of generating ES cell lines is based on a technique used to obtain a single cell embryo biopsy for preimplantation genetic diagnosis (PGD) of genetic defects (32). Blastomere-derived ES cells differentiated into derivatives of all three germ layers *in vitro* and
as well as into teratomas in vivo. In addition, the biopsied mouse embryos developed to term without a reduction in their developmental potential. Experiments are being carried out to determine whether human ES cells can be derived from single blastomeres (33). Additional studies will be needed to determine whether blastomere-derived ESC lines are different from conventional ESC lines in their ability to form functional differentiated cell types. It has been recently reported that both blastomere-derived and conventional ESCs can form functional hemangioblasts, which are capable of forming both hematopoietic and endothelial cell types, suggesting that blastomere-derived ESCs can be used as a cell therapy (34).

Although this technique does not involve the destruction of human embryos, concerns have been raised as to whether individual eight-cell-stage blastomeres could potentially generate a human being. Eight to sixteen cell-stage blastomeres have not been shown to have the intrinsic capacity to generate a complete organism in most mammalian species (35). Another technical concern is that ESCs derived from blastomeres require co-culture with a previously established ESC line. With further studies, it is likely that this technical hurdle will be resolved.

**Arrested embryos**

With the current restrictions surrounding ESC work, researchers are investigating methods to derive ESCs lines without destroying human embryos. It has been shown that human ESCs can be derived from arrested embryos (36). During in vitro fertilization, only a small number of all zygotes produced will
develop successfully to the morula and blastocyst stages, and well over half of the embryos stop dividing (37, 38) and are therefore considered dead embryos (39). Such embryos have unequal or fragmented cells and blastomeres and are usually discarded. However, not all the cells within these arrested embryos are abnormal (36, 40). In one study, one hundred and sixty one embryos were donated from IVF clinics. Out of these, 119 embryos arrested early at day 3-5 (4-10 cells stage) and 13 embryos arrested late at day 6-7 (16-24 cell stage). One stable and fully characterized ESC line was derived from the 13 late arrested embryos. This ESC line expressed OCT4, NANOG, REX1, TRA-1-61, TRA-1-81, which are associated with stem cells, and showed normal karyotype. These cells were injected into immunodeficient mice and formed teratomas, just as typical ESCs would. This data demonstrates that arrested embryos represent a novel source of pluripotent stem cells.

One of the major concerns with this technique is the ambiguity over what constitutes a dead embryo. Arrested embryos were used in this study when no cell or blastomere from the embryo had undergone any cleavage during the last 24 to 48 hours. However, the identification of molecules associated with the loss of embryonic viability will be helpful in defining dead or terminally arrested embryos (41). Another concern at this time is the quality of the cell lines derived from arrested embryos. More studies are needed to characterize the full proliferation and differentiation potential of ESCs derived from arrested embryos.
Altered Nuclear Transfer

Altered nuclear transfer is a variation of SCNT in which a genetically modified nucleus from a somatic cell is transferred into a human oocyte. This embryo, which contains a deliberate genetic defect, is capable of developing into a blastocyst but the induced defect prevents the blastocyst from implanting in the uterus. This process has the potential to generate customized ESCs from the blastocyst stage (42). It is hypothesized that human embryos with this genetic defect would lack the capacity to develop into human beings, due to their inability to implant, thus providing an alternative source of stem cells. This concept was proven in mice by Meissener et al., in 2006 (43). Here, scientists inactivated Cdx2 gene, which is crucial for trophectoderm development. The trophectoderm is required for the fetal-maternal interface within the placenta (44). The experiment used a lentiviral vector to insert a floxed Cdx2 short hairpin RNA sequence into the cells. Embryos resulting from nuclear transfers containing the Cdx2 deficiency showed no delay in developing into the early blastocyst stage, but the blastocysts were morphologically abnormal and could not implant into the uteri of pseudopregnant females. In culture, however, even though the Cdx2 knockdown blastocysts did not produce a trophectoderm, they did produce an inner cell mass structure from which ESCs could be obtained. These ESC lines were injected into diploid blastocysts and formed chimeras with extensive contributions to most tissues, except for the intestines, consistent with reports that Cdx2 is required for the normal development of the gastrointestinal tract (45). Interestingly, restoration of Cdx2 expression by Cre-mediated deletion of
the shRNA vector created ESC lines that could generate all somatic tissues including normal intestinal cells.

Despite the fact that these altered cells and embryos have little to no potential to form a functional organism, the ethics of the situation is still being debated. In addition, it is not clear whether human CDX2-deficient embryos die at the same stage as those from mice, or whether this mutation restricts their developmental potential into certain lineages. While much research must be done before such a therapy could ever enter the clinic, at this time ESCs derived from altered nuclear transfer can provide opportunities to study pluripotentiality in ESCs.

**Reprogramming**

Reprogramming is a technique that involves de-differentiation of adult somatic cells to produce patient-specific pluripotent stem cells without the use of embryos. Cells generated by reprogramming would be genetically identical to the somatic cells and would not be rejected by the donor. This method also avoids the technical limitations of nuclear transfer into oocytes. Yamanaka was the first to discover that mouse embryonic fibroblasts (MEFs) and adult mouse fibroblasts can be reprogrammed into an induced pluripotent stem cells (IPS) (46). They examined 24 genes that were thought to be important for embryonic stem cells and identified 4 key genes that were required to bestow embryonic stem cell-like properties on fibroblasts. Mouse embryonic fibroblasts and adult fibroblasts were co-transduced with retroviral vectors, each carrying Oct3/4, Sox2, c-Myc, and Klf4. Reprogrammed cells were selected via drug resistance.
In this case, a downstream gene of Oct4, Fbx15, was replaced with a drug resistance gene via homologous recombination. The resultant IPS cells possessed the immortal growth characteristics of self-renewing ESCs, expressed genes specific for ESCs, and generated embryoid bodies in vitro and teratomas in vivo. When the IPS cells were injected into mouse blastocysts they contributed to a variety of diverse cell types, demonstrating their developmental potential. Although IPS cells selected by Fbx15 were pluripotent they were not identical to ESCs. Unlike ESCs, chimeras of IPS cells did not result in full-term pregnancies. Gene expression profiles of the IPS cells showed that they possessed a distinct gene expression signature compared to ESCs. The epigenetic state of the IPS cells was somewhere between their somatic origins and fully reprogrammed ESCs, suggesting that the reprogramming was incomplete.

These results were improved significantly by Wernig and Jaenisch in July 2007 (47). Fibroblasts were infected with retroviral vectors and selected for the activation of endogenous Oct4 or Nanog genes. Results from this study showed that DNA methylation, gene expression profiles, and chromatic state of the reprogrammed cells were similar to those of ESCs. Teratomas induced by these cells contained differentiated cell types representing all three embryonic germ layers. Most importantly, the reprogrammed cells from this experiment were able to form viable chimeras and contribute to the germ line like ESCs, suggesting that these IPS cells were completely reprogrammed. This may be due to the fact that Wernig et al. observed that the number of reprogrammed colonies increased.
when drug selection was initiated later (day 20). This suggests that reprogramming is a slow and gradual process and may explain why the process of using Fbx15 activation on day 3 post-transfection may result in incomplete reprogramming.

It has recently been shown that reprogramming by transduction of four defined factors can be done with human cells (48, 49) Yamanaka’s group began by optimizing the transduction efficiencies of human dermal fibroblasts (HDFs) and determined that the introduction of a mouse receptor for retroviruses into HDFs using a lentivirus improved the transduction efficiency from 20% to 60%. Yamanaka then showed that retrovirus-mediated transfection of OCT3/4, SOX2, KLF4, and c-MYC generates human IPS cells that are similar to ESCs in terms of morphology, proliferation, gene expression, surface markers, and teratoma formation. In contrast, Thompson’s group showed that retroviral transduction of OCT4, SOX2, NANOG, and LIN28 could generate pluripotent stem cells without introducing any oncogenes (c-MYC). Both studies showed that human IPS cells were similar but not identical to ESCs. Another concern is that these IPS cells contain three to six retroviral integrations (one for each factor) which may increase the risk of tumorigenesis.

These studies used retroviral transduction to induce reprogramming of somatic cells into a pluripotent state. Yamanaka et al. studied the tumor formation in chimeric mice generated from Nanog-IPS cells and found 20% of the offspring developed tumors due to the retroviral expression of c-myc (50). An alternative approach would be to use a transient expression method, such as
adenovirus-mediated system, since both Jaenisch and Yamanaka showed strong silencing of the viral-controlled transcripts in IPS cells (50, 47). This indicates that they are only required for the induction, not the maintenance, of pluripotency. Another concern is the use of transgenic donor cells for reprogrammed cells in the mouse studies. In both mouse studies, iPS cells were isolated by selecting for the activation of a drug-resistant gene inserted into endogenous Fbx15, Oct3/4, or Nanog. The use of genetically modified donors hinders its clinical applicability for humans.

To assess whether iPS cells can be derived from genetically unmodified donor cells, MEF and adult skin cells were retrovirally transduced with Oct3/4, Sox2, c-Myc, and Klf4 and ES-like colonies were isolated by morphology, without the use of drug selection for Oct4 or Nanog (51). IPS cells from unmodified donor cells formed teratomas and generated live chimeras. This study suggests that genetically modified donor cells are not necessary to generate IPS cells.

Although this is an exciting phenomenon, it is unclear why reprogramming adult fibroblasts and mesenchymal stromal cells have similar efficiencies (46). It would seem that cells that are already multipotent could be reprogrammed with greater efficiency, since the more undifferentiated donor nucleus the better SCNT performs (52). This further emphasizes our limited understanding of the mechanism of reprogramming, yet the potential for this area of study is exciting.

**Amniotic Fluid-derived Stem Cells**

The amniotic fluid is known to contain a heterogeneous population of cell types derived from the developing fetus (53, 54). Cells from this heterogeneous
population contain adipocytes, osteocytes, neurogenic and endothelial cells (55-58). The isolation of human and mouse amniotic fluid-derived stem cells (AFSCs) that are capable of extensive self-renewal and give rise to adipogenic, osteogenic, myogenic, endothelial, neurogenic and hepatogenic lineages was reported (60). AFSCs represent approximately 1% of the cells found in the amniotic fluid. These cells are immuno-selected for cells that express the surface antigen C-KIT (CD117), the receptor for stem cell factor (61). AFSCs express embryonic markers such as OCT4 and SSEA4. They also express mesenchymal and/or neuronal markers such as CD29, CD44, CD73, CD90, and CD105. More importantly, clonal analyses using human lines verified by retroviral marking confirmed that differentiated cells of various types can descend from a single cell (60).

In addition to demonstrating that differentiated cells express lineage-specific markers, we have shown that such cells give rise to specialized functions. Cells differentiated down a neural pathway secreted glutamine or expressed G-protein-gated inwardly rectifying potassium channels. Cells of the hepatic lineage secreted urea and α-fetoprotein, while osteogenic cells produced mineralized calcium.

AFSCs were cultured in neuronal differentiation medium for a time and then grafted into the lateral cerebral ventricles of control mice and the ventricles of the twitcher mouse model, in which a progressive loss of oligodendrocytes leads to massive demyelination and neuronal loss. AFSCs integrated into the brains of both strains, and survived efficiently for at least 2 months. Interestingly,
more of the AFSCs integrated into the injured twitcher brains (70%) than into the normal brains (30%), hinting at the potential for CNS therapies. From a tissue engineering perspective, osteogenic AFSCs were embedded in an alginate/collagen scaffold and implanted subcutaneously into immunodeficient mice. By 18 weeks after implantation, highly mineralized tissues and blocks of bone-like material were observed in the recipient mice using micro CT. These blocks displayed a density somewhat greater than that of mouse femoral bone.

The recent discovery of a stem cell population in the amniotic fluid offers a promising alternative source of stem cells for cellular therapy. The full range of adult somatic cells that AFSCs can produce remains to be determined, but their ability to differentiate into cells of all three embryonic germ layers and their high proliferation rate are two advantages. AFSCs represent a new class of stem cells with properties somewhere between embryonic and adult stem cell types. However, unlike ES cells, AFSCs do not form teratomas. AFSCs could be use for both autologous and allogenic therapy through matching of histocompatible donor cells with recipients.

**The Future of Stem Cells for Regenerative Medicine**

Before stem cells can be used as any type of clinical therapy, strict guidelines must be established to ensure the quality of the cells, the specificity of differentiation, and the assessment of mixed phenotypes. While lineage-specific gene expression and cell surface markers are commonly used to describe a differentiated phenotype, it is difficult to determine whether cells are bona fide
neurons or merely neuronal-like cells. To address this question, we are developing high throughput methodologies using microarrays to evaluate new stem cell derivatives. In these experiments, ESCs were differentiated into retinal pigmented epithelial (RPE) cells, the site of the major lesions in macular degeneration. Microarrays were used to demonstrate similarities between ESC-derived RPE and freshly isolated RPE (62).

Another concern about the clinical potential of stem cells is their tendency to form mixed phenotypes in most differentiation protocols. We have recently reported a novel approach to assess heterogeneous populations, also using microarray technologies (63). The therapeutic potential of stem cells largely relies on efficient and controlled differentiation towards a specific cell type and the generation of homogeneous cell populations. Many differentiation protocols utilize the formation of progenitors through a stepwise approach. Thus, characterizing and understanding mixed populations of progenitor stages will be of increasing importance in stem cell research. We developed a method to assess and identify tissue-specific genetic signatures within a heterogeneous population with microarrays through biologically relevant in silico comparisons of data sets. Even once these issues are resolved, stem cell therapy still has many hurdles to overcome before it will become a viable and widely used clinical option.

**Conclusion**

At this time, it is unclear which type of stem cell will provide the best approach for cell therapy. Depending on the clinical scenario, one cell may be
better than another. For example, tissue progenitor cells are excellent
candidates for tissue engineering organs such as the bladder. Even though
tissue progenitor cells have restricted growth and differentiation potential, this is
advantageous because they have a lower tendency to form tumors and mixed
phenotypes. On the other hand, ESCs have the ability to grow indefinitely and
differentiate into cells of all three germ layers. However, ethical concerns
regarding the use of embryos, the formation of teratomas, and the potential of
ESCs to evoke an immune response currently dampen enthusiasm in their
clinical potential.

Methods for generating alternative sources of pluripotent stem cells have
their advantages and disadvantages (Table 1). Somatic cells can be
reprogrammed by nuclear transfer into enucleated oocytes and generate
pluripotent and patient-specific stem cells. The disadvantages of this method
include the use of oocytes, which are in limited supply, and the destruction of a
cloned blastocyst. In addition, it is unclear whether SCNT can be successfully
performed in humans. Single cell embryo biopsy is a promising approach
because it is based on a common, well-established technique in IVF and has
been successful in thousands of children.

The ability to reprogram an adult somatic cell into a pluripotent stem cell is
a scientific breakthrough, but the use of retroviral transduction and limited
understanding hinders their clinical potential. On the other hand, AFSCs are
easily accessible and do not require technical manipulations, but they may not be
as nimble as other pluripotent cells. In order to determine the best source of cells for a given application, it will be important to consider all types of stem cells.

Table 1. Summary of methods for generating pluripotent stem cells.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Advantages</th>
<th>Limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somatic cell nuclear transfer</td>
<td>-customized stem cells</td>
<td>-requires oocytes</td>
</tr>
<tr>
<td></td>
<td>-been shown to work in non-human primates</td>
<td>-not been shown to work in humans</td>
</tr>
<tr>
<td>Single cell embryo biopsy</td>
<td>-patient-specific to embryo</td>
<td>-allogenic cell types</td>
</tr>
<tr>
<td></td>
<td>-does not destroy or create embryos</td>
<td>-not known if single cells are totipotent</td>
</tr>
<tr>
<td></td>
<td>-has been done in humans</td>
<td>-requires coculturing with a previous established hES line</td>
</tr>
<tr>
<td>Arrested embryos</td>
<td>-from discarded embryos</td>
<td>-allogenic cell types</td>
</tr>
<tr>
<td></td>
<td>-has been done in humans</td>
<td>-quality of lines maybe questionable</td>
</tr>
<tr>
<td>Altered nuclear transfer</td>
<td>-customized stem cells</td>
<td>-ethical concerns regard embryos with no potential</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-modified genome</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-not been done with human cells</td>
</tr>
<tr>
<td>Reprogrammed/IPS cells</td>
<td>-customized stem cells</td>
<td>-retroviral transduction</td>
</tr>
<tr>
<td></td>
<td>-no embryos or oocytes needed</td>
<td>-oncogenes</td>
</tr>
<tr>
<td></td>
<td>-has been done with human cells</td>
<td></td>
</tr>
<tr>
<td>Amniotic fluid-derived stem</td>
<td>-express embryonic and adult stem cell markers</td>
<td>-full potential not known</td>
</tr>
<tr>
<td>cells</td>
<td>-does not form teratomas</td>
<td>-patient-specific to embryo</td>
</tr>
<tr>
<td></td>
<td>-easily accessible</td>
<td>-allogenic and autologous if banked</td>
</tr>
<tr>
<td></td>
<td>-non-invasive approach</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-has been done in humans</td>
<td></td>
</tr>
</tbody>
</table>
References

3. Bush, G. Executive order 13435


29. Byrne, J., Pedersen, D., Clepper, L., Nelson, M., Sanger, W., Gokhale, S.,
31. Chung, Y., Klimanskaya, I., Becker, S., Marh, J., Lu, S. J., Johnson, J.,
    Nature. 344,768-770.
    Nature. 444,481-485.
34. Lu, S. J., Feng, Q., Caballero, S., Chen, Y., Moore, M. A., Grant, M. B. and
40. Martinez, F., Rienzi, L., Iacobelli, M., Ubaldi, F., Mendoza, C., Greco, E. and
44. Strumpf, D., Mao, C. A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K.,
45. Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J. and
47. Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger,


CHAPTER II

GENECHIP ANALYSIS OF HUMAN EMBRYONIC STEM CELL DIFFERENTIATION INTO HEMANGIOBLASTS: AN IN SILICO DISSECTION OF MIXED PHENOTYPES

Shi-Jiang Lu & Jennifer A. Hipp, Qiang Feng, Jason Hipp, Robert Lanza and Anthony Atala

The following manuscript was published in *Genome Biology* 2007;8(11):R240 and is reprinted with permission. Stylistic variations are due to the requirements of the journal.
Abstract

Background: Microarrays are being used to understand human embryonic stem cell (ESC) differentiation. Most differentiation protocols use a multi-stage approach that induces commitment along a particular lineage; each stage represents a more mature and less heterogeneous phenotype. Therefore, characterizing heterogeneous progenitor populations upon differentiation are of increasing importance. Here we describe a novel method of data analysis using a recently developed differentiation protocol involving the formation of functional hemangioblasts from ESCs.

Results: Large-scale transcriptional analysis was performed at distinct time points of ESC differentiation (undifferentiated ESCs, embryoid bodies (EBs), and blast cells, the later generates both hematopoietic and endothelial progenies). Identifying genes enriched in blast cells relative to ESC revealed a genetic signature indicative of erythroblasts, suggesting that erythroblasts are the predominant cell type in the blast cell population. Because of the heterogeneity of blast cells, numerous comparisons were made to publicly available data sets in silico, some of which blast cells are capable of differentiating into, to assess and characterize the blast cell population. Biologically relevant comparisons masked particular genetic signatures within the heterogeneous population and identified genetic signatures indicating the presence of endothelia, cardiomyocytes, and hematopoietic lineages in the blast cell population.

Conclusion: The significance of this microarray study is in its ability to assess and identify cellular populations within a heterogeneous population through
biologically relevant in silico comparisons of publicly available data sets. In conclusion, multiple in silico comparisons were necessary to characterize tissue-specific genetic signatures within a heterogeneous hemangioblast population.

**Background:**

The establishment of human embryonic stem cells (ESC) raised the possibility of treating/curing a lot of human disease most of which are nowadays untreatable. These therapeutic potentials, however, largely rely on efficient and controlled differentiation of ESCs towards a specific cell type and the generation of homogeneous cell populations. Many differentiation protocols utilize the formation of progenitors through a stepwise approach. Thus, characterizing and understanding mixed populations of progenitor stages will be of increasing importance in stem cell research.

ESCs have been shown to be able to differentiate into a variety of cell types, including hematopoietic precursors and endothelial cells, in vitro under various culture conditions [1-9]. Hemangioblasts are the precursors of both hematopoietic and endothelial cells [10]. The existence of hemangioblasts was first demonstrated using in vitro differentiation system of mouse ESCs. Replating of embryonic bodies (EB) of mouse ESCs resulted in the formation of blast colony forming cells (BL-CFC), which possessed hemangioblastic characteristics: BL-CFC generated both hematopoietic and endothelial cells upon transfer to appropriate conditions [11,12]. Cells with hemangioblastic characteristics have
been reported in both mouse and human adult tissues [13-18]. In human ESC system, Wang et al [3] have found that a fraction of a percent (0.18%) of CD45\(^{\text{neg}}\)FVP cells with hemangioblast-like properties in ESC derived EBs. Zambidis et al [8] demonstrated the formation of multi-potential colonies from human EBs, although it is unclear whether these colonies can be expanded and/or whether they have any functional activity in vivo. Umeda et al[19] also identified the presence of CD34+/KDR+ bipotential cells in non-human (Cynomolgus) ESCs. Kennedy et al [20] recently reported the generation of BL-CFCs from human ESCs. However, the rarity of the cells with hemangioblast properties both from adult tissues and from ESC systems precluded comprehensive analysis of gene expression and comparison with other populations.

We have recently developed a two-step strategy that can efficiently and reproducibly generate blast colonies (BCs), the human counterparts of BL-CFCs, from human ESCs [21]. These BC cells expressed gene signatures characteristic of hemangioblasts, and could be differentiated into multiple hematopoietic cell lineages as well as endothelial cells. When the BC cells were injected into animals with spontaneous type II diabetes rats or ischemia/reperfusion injury of the retina, they homed to the site of injury and showed robust reparative function of the damaged vasculature. The cells also showed a similar regenerative capacity in NOD/SCID \(\beta2/-\) mouse models of both myocardial infarction (50% reduction in mortality rate) and hind limb ischemia,
with restoration of blood flow in the later model to near normal levels, demonstrating their functional properties of hemangioblast in vivo [21]. In contrast to previous studies, these cells could be readily obtained in large scale, which allow us to perform comprehensive gene expression in these cells and compare with other cell populations from whom the BC cells originated.

Microarrays assess the total amount of RNA in a population and can be influenced by a predominating cell type. Variation in the homogeneity of the population can influence the number of genes identified as differentially expressed. Here we will show how comparisons to publicly available tissues in silico can identify differentially expressed genes representative of the various cell types within a heterogeneous population.

In the current study, we analyzed the global gene expression profiles with Robust Multi-chip Average (RMA), to provide a relative value of gene expression between two samples. The first analysis consisted of direct comparisons with ESCs and their derivatives (EBs and BCs). Genes enriched in BCs relative to ESCs revealed a genetic signature indicative of erythroblasts, suggesting that erythroblasts are the predominant cell type in the BC population. The next analysis consisted of multiple but biologically meaningful in silico comparisons to publicly available data set which identified other progenitor cell types within the BC population. The significance of this microarray study is in its ability to assess
and identify heterogeneous cellular populations through biologically relevant in silico comparisons.

Results

Strategy

Microarrays assess the total amount of RNA in a population and can therefore be influenced by a predominating cell type. Variations in the homogeneity of the population can influence the number of genes identified as differentially expressed, especially if both populations are relatively homogeneous. Here we will show how comparisons to publicly available tissues in silico can identify differentially expressed genes representative of the various cell types within the heterogeneous population of BCs.

We will describe our method of assessing heterogeneous samples in three levels of analysis. The first level consists of making direct comparisons within the ESCs and their differentiated derivatives (EBs and BCs). The advantage of this technique is that it provides a kinetic-like relationship of changes in gene expression upon differentiation. The second level of analysis consists of indirect comparisons to a baseline, or reference tissue. Breast epithelium was chosen as a reference tissue because it represents a genetically distinct cell type which BCs are not capable of differentiating into. ESCs, EBs, and BCs were compared to breast epithelial tissue and differentially expressed genes were compared and contrasted to each other. Because genes that are up-regulated in BCs when
compared to breast epithelia could represent those that are under-expressed in breast tissue, we removed those that were also up-regulated in a genotypically similar but different cell type (ESCs), when compared to breast epithelia. The third level of analysis consists of comparing BCs to tissues in which they are capable of differentiating into as a way to mask that cell type’s “genetic signature” and revealing signatures of the more minor cell types. Samples were chosen based on type of GeneChip and their public availability- leukocytes, endothelial, and stromal cells. These biologically relevant comparisons identified tissue specific genetic signatures that would have otherwise been missed in the Level I and Level II analyses.

The reliability of the microarray data generated from our multi-comparison analysis is demonstrated by the consistent identification of a set of genes among multiple comparisons of which a subset of genes were confirmed by immunocytochemistry (Table 1) and RT-PCR (Figure 1). To summarize, comparing BCs to leukocytes identified genes involved in vasculogenesis, to endothelial cells identified genes involved in hematopoiesis, and to stromal cells identified genes involved in heart development.
Table 1. Characterization of ESC-BCs by Immunocytochemistry (IC) and Affymetrix arrays. The reliability of the microarray data generated from our multi-comparison analysis is demonstrated by the consistent identification of a set of genes among multiple comparisons of which a subset of genes were confirmed by immunocytochemistry. For immunocytochemistry: +, moderate to strong staining; -, negative staining; +/−, very weak staining. For Affymetrix arrays, +, detected as up-regulated in BCs, --, not detected as up-regulated in BCs.
Figure 1. Validation of differentially expressed genes by RT-PCR in human ESCs, EBs and BCs. Total RNA from human ESCs, EBs and BCs was used to construct cDNA pools, and the expression of genes was examined by semi-quantitative PCR (A). The number at the top of each lane indicates the amount (microliters) of cDNA used in the 50-µL PCR reaction. M = 100 bp DNA ladder. Direct (B) and indirect (C) analysis of differentially expressed genes matched the expression patterns obtained by RT-PCR. The fold change data is presented on the y-axis using logarithm-base-10.
**Level I Analysis:**

*Genes Down-regulated in ESCs → EBs and BCs*

We began our data analysis by verifying the expression of “stemness” genes that are down-regulated in ESCs upon differentiation into EBs and BCs. We identified 87 genes that were down-regulated upon differentiation into EBs. Genes with the highest fold change include SOX2, LEFTY1, GAL, NODAL, OCT4, and THY1, which play a critical role in maintaining undifferentiated status of ESCs [22-24]. To uncover enriched processes, data sets were analyzed by Database for Annotation, Visualization and Integrated Discovery (DAVID) and Expression Analysis Systematic Explorer (EASE), a web-based tool that identifies over-represented biological themes in a data set based on their gene ontologies (GO). GO was provides consistent descriptions of genes in terms of biological processes and molecular function. When these genes were clustered with DAVID based on their GO, processes involved in development, cell differentiation, and proliferation were identified. The genes identified in the development ontology were DNA methyltransferase 3B, FGF2, THY1, SFRP2, LEFTY1, GREM1, and NODAL.

We also identified 267 genes that were down-regulated upon differentiation from ESCs to BCs. These genes included GAL, TDGF, NANOG, LEFTY1, and OCT4, most of them are stemness genes [22-24]. When genes were clustered with DAVID using their gene ontology, processes include development, cell differentiation, and morphogenesis. This data demonstrates that OCT4, NODAL,
GAL, and THY1 are initially down-regulated in stage 1 (ESCs→EBs) and are further down-regulated in stage 2 (EBs→BCs).

*Genes Up-regulated upon ESCs→EBs*

While the focus of this paper is to evaluate the pathways involved in hemangioblast differentiation, we begin by identifying those genes that were up-regulated in early stage of differentiation into EBs (day 3.5) from which BCs were derived [21]. We identified 128 genes that were up-regulated upon ESC differentiation into EBs. These genes include HAND1, WNT5, HEY1, LMO2, BMP4, TBX3, MYL4. Clustering these genes with EASE identified processes involved in development, transcription, organ development and system development; and some of which are related to hemangioblastic differentiation. These genes include SOX9, HOXB2, HOXB3, Neuregulin 1, LMO2 [25] and GATA2[26]. This data set also included numerous in the transcription factors such as MESP1, HAND1, TBX3, GATA2, SOX7, SOX9, HOXB2, and HOXB3.

*Genes Down-regulated upon EBs→BCs*

When EBs were compared to BCs, 185 genes were identified as down-regulated upon differentiation. This data set contained processes that were similar to those that were down-regulated upon ESC differentiation into EBs, such as tissue and organ development. The most significantly down-regulated genes included NANOG, WNT5, OCT4, GAL, TDGF1, BMP4, endothelin receptor B, and VEGF. This data demonstrates that BMP4, WNT5, and HEY1 are initially up-regulated
upon differentiation into EBs but then down-regulated upon further differentiation into BCs.

**Genes Up-regulated upon EBs → BCs**

In contrast, 82 genes were up-regulated upon the differentiation of EBs into BCs. The genes with the greatest fold change were hemoglobin genes and erythropoietic genes, such as hemoglobins γ, ζ, α, and ε, Alas2, Afp, TUBB1, GYPα, and RHAG (FC 31x-886x). Genes with moderate increases in expression (FC 6.2x-7.2x) were KLF1, TAL1/SCL, GATA1 and CD71. When genes were clustered with DAVID using their gene ontology, process characteristic of erythropoiesis (heme and porphyrin biosynthesis and oxygen transport) were identified.

**Genes Up-regulated upon ESCs → BCs**

107 genes were up-regulated upon the differentiation of ESCs into BCs. Similar to the data set above (EBs → BCs), the genes with the greatest fold change (FC 29x-810x) were involved in hemoglobin synthesis; hemoglobins γ, ε, and α, ALA2, GYPα, TAL1/SCL, similarly to the comparison of EBs to BCs. In addition, this data set contained many key transcription factors involved in hemangioblastic differentiation which were increased, such as GATA2[26], LMO2[25] and TAL1/SCL genes[27,28]. GATA2 and MYL4 were up-regulated upon differentiation into EBs and remained at a constant level upon differentiation into BCs. EASE analysis of up-regulated genes in BC identified
biologically relevant themes such as oxygen and gas transport, and development (TAL1/SCL, KLF1, LMO2, GATA1, Table 2).

**Table 2.** Gene ontologies for up-regulated processes in BCs vs ESCs. EASE analysis of up-regulated genes in BC identified biologically relevant themes such as oxygen and gas transport, and development.

<table>
<thead>
<tr>
<th>Gene Category</th>
<th>List Hits</th>
<th>EASE score</th>
</tr>
</thead>
<tbody>
<tr>
<td>oxygen transport</td>
<td>(erythrocytic) 6</td>
<td>1.69E-09</td>
</tr>
<tr>
<td>gas transport</td>
<td>(erythrocytic) 6</td>
<td>1.69E-09</td>
</tr>
<tr>
<td>transcription from Pol II promoter</td>
<td>7</td>
<td>1.64E-02</td>
</tr>
<tr>
<td>regulation of transcription from Pol II promoter</td>
<td>5</td>
<td>2.38E-02</td>
</tr>
<tr>
<td>development</td>
<td>(developmental) 15</td>
<td>2.56E-02</td>
</tr>
</tbody>
</table>

**Level II Analysis**

*Genes Enriched in ESCs*

Since genes that are up-regulated in ESCs when compared to breast epithelia could represent those genes that are under-expressed in breast epithelium, we filtered out those that were also up-regulated in BC when compared to breast epithelium. This analysis identified 2,108 genes which consisted of gene ontologies involved in cell cycle, DNA and RNA metabolism, and DNA replication as expected. Genes with the highest fold change include TDGF1, GAL, LEFTY1/2, OCT4, NANOG (FC 130.0x, 69.6x, 68.7x, 47.6x, 43.8x, and 31.5x). When this data set was clustered based on their GO, processes involved in development, cell differentiation and nervous system development were identified. This data set was then analyzed with GenMAPP and then used for pathway analysis. Each genetic signature was assigned a color, ESCs (green),
EBs (orange), and BCs (red). The Embryonic Stem Cell pathway confirms that most of the embryonic genes were not removed when we compared them to breast epithelial cells (data not shown).

**Genes Enriched in EBs**

Since genes up-regulated in EBs when compared to breast epithelia could similarly represent those that are under-expressed in breast epithelium, we also filtered out those that were also up-regulated in ESCs when compared to breast epithelium. We identified 939 genes as up-regulated in EBs relative to breast epithelium and filtered out those that are enriched in ESCs relative to breast epithelium. When these genes were clustered with DAVID, processes involved in development, transcription, Wnt and frizzle signaling, cell cycle and blood vessel morphogenesis (KDR, VEGF, Neuropilin-1 and 2, and FLT1) were identified. The EBs also express genes involved in organ development suggesting a heterogeneous mixture of cell types (GATA2/4/5/6, BMP4, NCAM1, NOG, ISL2, Nkx2-5) and mesoderm genes (HAND1, T-brachyury, MESP1). Further examination of the data set identified multiple genes involved in BMP signaling pathway in the differentiation of blood and endothelial cells. These genes include BMPR1A, BMP4, T-brachyury, KDR, GATA2, TAL1/SCL [26,28].

**Genes Enriched in BCs**

We identified 2,735 genes that were up-regulated in BCs relative to breast epithelium after removing genes that were enriched in ESC when compared to
breast epithelium and genes enriched in BCs when compared to breast epithelium. When genes were clustered based on their gene ontology, we identified processes characteristic of lymphocytic cells (response to stimulus, defense response, immune response), erythrocytes (heme and porphyrin biosynthesis), coagulation, neurophysiology, development, and mesoderm and heart development (Table 3).

Genes that were up-regulated in BCs with respect to epithelia were characteristic of hematopoiesis (CD markers 5/6/9/38/41/48/55/71/74/84/244, EPOR, GATA1/2/4/5, Tcr-α, natural cytotoxicity triggering receptor-1, 2 and 3), coagulation (coagulation factor II, V, VII, XII, coagulation factor 2 receptor like 2,3/thrombin receptors, antithrombin 3, cyclooxygenase 1, plasminogen), cardiac muscle (NKX2-5, HAND1/2, GATA4, SOX6, TBX5), smooth/skeletal muscle (NOTCH1, smoothelin, acetylcholinesterase, desmin, SOX6), synaptic markers (cholinergic receptor, muscarinic 2 and 5, adrenergic α-1A-receptor, dopamine receptor D2, serotonin receptor 1B and 4, glutamate receptor Nmda 1 and 2A/B and C, gaba A receptor β1 and 2, purinergic receptor P2X 1 and 2), and hemangioblasts (GATA2[26], RUNX1[29], LMO2[25] and TAL1/SCL [27,28]. Some of the genes identified in the coagulation ontology are not only involved in coagulation but also angiogenesis, such as thrombin [30], plasminogen [31], and possibly coagulation factor 2 receptor like 2/3 [30].
Table 3. Gene ontologies of up-regulated processes in BCs vs epithelial cells. When genes were clustered based on their gene ontology, we identified processes characteristic of lymphocytic cells (response to stimulus, defense response, immune response), erythrocytes (heme and porphyrin biosynthesis), coagulation, neurophysiology, development, and mesoderm and heart development.
Table 3

<table>
<thead>
<tr>
<th>Gene Category</th>
<th>List Hits</th>
<th>EASE score</th>
</tr>
</thead>
<tbody>
<tr>
<td>response to external stimulus</td>
<td>(lymphocytic)</td>
<td>146</td>
</tr>
<tr>
<td>cellular process</td>
<td></td>
<td>490</td>
</tr>
<tr>
<td>response to biotic stimulus</td>
<td>(lymphocytic)</td>
<td>94</td>
</tr>
<tr>
<td>ion transport</td>
<td></td>
<td>68</td>
</tr>
<tr>
<td>defense response</td>
<td>(lymphocytic)</td>
<td>87</td>
</tr>
<tr>
<td>development</td>
<td>(developmental)</td>
<td>165</td>
</tr>
<tr>
<td>hemostasis</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>blood coagulation</td>
<td>(coagulation)</td>
<td>18</td>
</tr>
<tr>
<td>cell-cell signaling</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>response to abiotic stimulus</td>
<td>(lymphocytic)</td>
<td>60</td>
</tr>
<tr>
<td>cyclic-nucleotide-mediated signaling</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>morphogenesis</td>
<td>(developmental)</td>
<td>101</td>
</tr>
<tr>
<td>response to light</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>natural killer cell activation</td>
<td>(lymphocytic)</td>
<td>6</td>
</tr>
<tr>
<td>second-messenger-mediated signaling</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>perception of light</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>synaptic transmission</td>
<td>(neurophysiology)</td>
<td>31</td>
</tr>
<tr>
<td>cell communication</td>
<td></td>
<td>244</td>
</tr>
<tr>
<td>cation transport</td>
<td></td>
<td>46</td>
</tr>
<tr>
<td>transmission of nerve impulse</td>
<td>(neurophysiology)</td>
<td>31</td>
</tr>
<tr>
<td>transport</td>
<td></td>
<td>159</td>
</tr>
<tr>
<td>response to radiation</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>organogenesis</td>
<td>(developmental)</td>
<td>88</td>
</tr>
<tr>
<td>metal ion transport</td>
<td></td>
<td>36</td>
</tr>
<tr>
<td>perception of abiotic stimulus</td>
<td>(lymphocytic)</td>
<td>35</td>
</tr>
<tr>
<td>vision</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>immune response</td>
<td>(lymphocytic)</td>
<td>68</td>
</tr>
<tr>
<td>perception of external stimulus</td>
<td>(lymphocytic)</td>
<td>37</td>
</tr>
<tr>
<td>cell surface receptor linked signal transduction</td>
<td></td>
<td>93</td>
</tr>
<tr>
<td>xenobiotic metabolism</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>sensory perception</td>
<td></td>
<td>34</td>
</tr>
<tr>
<td>heme biosynthesis</td>
<td>(erythrocytic)</td>
<td>5</td>
</tr>
<tr>
<td>response to xenobiotic stimulus</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>skeletal development</td>
<td>(development)</td>
<td>18</td>
</tr>
<tr>
<td>potassium ion transport</td>
<td></td>
<td>19</td>
</tr>
<tr>
<td>G-protein signaling</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>pigment biosynthesis</td>
<td>(erythrocytic)</td>
<td>6</td>
</tr>
<tr>
<td>monovalent inorganic cation transport</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>heme metabolism</td>
<td>(erythrocytic)</td>
<td>5</td>
</tr>
<tr>
<td>negative regulation of natural killer cell activity</td>
<td>(lymphocytic)</td>
<td>3</td>
</tr>
<tr>
<td>regulation of natural killer cell activity</td>
<td>(lymphocytic)</td>
<td>3</td>
</tr>
<tr>
<td>anion transport</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>heart development</td>
<td>(heart development)</td>
<td>6</td>
</tr>
<tr>
<td>histogenesis</td>
<td>(developmental)</td>
<td>16</td>
</tr>
<tr>
<td>porphyrin biosynthesis</td>
<td>(erythrocytic)</td>
<td>5</td>
</tr>
<tr>
<td>pigment metabolism</td>
<td>(erythrocytic)</td>
<td>6</td>
</tr>
<tr>
<td>pigmentation</td>
<td>(erythrocytic)</td>
<td>6</td>
</tr>
<tr>
<td>mesoderm development</td>
<td>(mesoderm development)</td>
<td>7</td>
</tr>
<tr>
<td>cellular defense response</td>
<td>(lymphocytic)</td>
<td>13</td>
</tr>
<tr>
<td>response to pest/pathogen/parasite</td>
<td>(lymphocytic)</td>
<td>43</td>
</tr>
</tbody>
</table>
This was also recapitulated by DAVID analysis which identified the following pathways as statistically over-represented: porphyrin metabolism, acute myocardial infarction, hematopoietic cell lineage pathways and calcium signaling pathways. GenMAPP was then used for pathway analysis. Each genetic signature was assigned a color, ESCs (green), EBs (orange), and BCs (red). GenMAPP of pathways involved in whole blood, bone marrow, coagulation and complement, heme and porphyrin synthesis are indicative of hematopoietic cell types (Figure 2). Genes identified in GenMAPP’s heme biosynthesis pathway are indicative of erythroblasts (Figure 3). We also identified myogenic (cardiac and smooth) pathways which correlate with the GO analysis.
Figure 2. GenMAPP of Complement and Coagulation. Genes that were up-regulated in BCs in red, up-regulated genes in EBs in orange, and up-regulated genes in ESCs in green with breast epithelia as baseline were mapped onto a pre-existing pathways. This pathway contains genes that are mostly up-regulated in BCs relative to breast epithelia.
Figure 3. GenMAPP of Heme Biosynthesis: Genes that were up-regulated in BCs in red, up-regulated genes in EBs in orange, and up-regulated genes in ESCs in green with breast epithelia as baseline were mapped onto a Heme Biosynthesis pathway. This pathway contains genes that are up-regulated in BCs relative to breast epithelia.
Level III Analysis

*Enriched BCs vs Leukocytes*

We identified 2,101 genes that were up-regulated in BCs relative to leukocytes (after removing genes that were enriched in both ESCs and BCs when compared to leukocytes. When these genes were clustered based on their GO, we identified processes involved in development, nervous system development, blood vessel development and angiogenesis, and erythrocytes (Table 4). The presence of development ontology not only indicates a “progenitor” status of BCs, but contains genes involved in hemangioblast development such as LMO2, TAL1/SCL, and RUNX1. This comparison identified genes that are characteristic of endothelia cells (PECAM1, VE-Cadherin, CD34, vWF, EPOR [32], endothelin 1 [33], thrombin receptor). There were also genes that indicate the presents of erythrocytes (GATA1, spectrin and ankyrin), and blood vessel development (neuropilin-1 and 2, stabilin 1 and 2, EGFR, FGF1 and 6, NOTCH4), neurons/neuronal junctions (glutamate receptor 1/6, serotonin receptor 1e/6, nestin, neurogenic differentiation 4, neuroligin 2, myelin basic protein, peripheral myelin protein 22).
Table 4. Gene ontologies for up-regulated processes in BCs vs leukocytes.

When these genes were clustered based on their GO, we identified processes involved in development, nervous system development, blood vessel development and angiogenesis, and erythrocytes.
<table>
<thead>
<tr>
<th>Gene Category</th>
<th>List Hits</th>
<th>EASE score</th>
</tr>
</thead>
<tbody>
<tr>
<td>development</td>
<td>(developmental) 178</td>
<td>6.39E-09</td>
</tr>
<tr>
<td>morphogenesis</td>
<td>(developmental) 116</td>
<td>1.98E-08</td>
</tr>
<tr>
<td>organogenesis</td>
<td>(developmental) 105</td>
<td>5.04E-08</td>
</tr>
<tr>
<td>cellular process</td>
<td>482</td>
<td>6.48E-08</td>
</tr>
<tr>
<td>cell communication</td>
<td>257</td>
<td>4.78E-07</td>
</tr>
<tr>
<td>cell adhesion</td>
<td>68</td>
<td>4.05E-06</td>
</tr>
<tr>
<td>skeletal development</td>
<td>(developmental) 25</td>
<td>4.44E-06</td>
</tr>
<tr>
<td>cell-cell signaling</td>
<td>64</td>
<td>4.86E-06</td>
</tr>
<tr>
<td>ion transport</td>
<td>62</td>
<td>8.10E-05</td>
</tr>
<tr>
<td>circulation</td>
<td>20</td>
<td>9.47E-05</td>
</tr>
<tr>
<td>metal ion transport</td>
<td>38</td>
<td>2.53E-04</td>
</tr>
<tr>
<td>cartilage condensation</td>
<td>6</td>
<td>3.56E-04</td>
</tr>
<tr>
<td>cation transport</td>
<td>46</td>
<td>5.97E-04</td>
</tr>
<tr>
<td>G-protein coupled receptor protein signaling pathway</td>
<td>61</td>
<td>8.03E-04</td>
</tr>
<tr>
<td>regulation of blood pressure</td>
<td>9</td>
<td>9.06E-04</td>
</tr>
<tr>
<td>neurogenesis</td>
<td>(neuronal) 46</td>
<td>1.17E-03</td>
</tr>
<tr>
<td>second-messenger-mediated signaling</td>
<td>19</td>
<td>1.32E-03</td>
</tr>
<tr>
<td>cell surface receptor linked signal transduction</td>
<td>93</td>
<td>1.52E-03</td>
</tr>
<tr>
<td>bone remodeling</td>
<td>9</td>
<td>2.01E-03</td>
</tr>
<tr>
<td>ossification</td>
<td>9</td>
<td>2.01E-03</td>
</tr>
<tr>
<td>cell homeostasis</td>
<td>15</td>
<td>2.07E-03</td>
</tr>
<tr>
<td>cell ion homeostasis</td>
<td>14</td>
<td>2.27E-03</td>
</tr>
<tr>
<td>ion homeostasis</td>
<td>14</td>
<td>2.27E-03</td>
</tr>
<tr>
<td>transition metal ion transport</td>
<td>9</td>
<td>2.40E-03</td>
</tr>
<tr>
<td>transport</td>
<td>151</td>
<td>2.97E-03</td>
</tr>
<tr>
<td>homeostasis</td>
<td>15</td>
<td>3.10E-03</td>
</tr>
<tr>
<td>transition metal ion homeostasis</td>
<td>7</td>
<td>3.37E-03</td>
</tr>
<tr>
<td>receptor mediated endocytosis</td>
<td>7</td>
<td>4.20E-03</td>
</tr>
<tr>
<td>metal ion homeostasis</td>
<td>11</td>
<td>4.48E-03</td>
</tr>
<tr>
<td>homophilic cell adhesion</td>
<td>15</td>
<td>5.91E-03</td>
</tr>
<tr>
<td>di-, tri-valent inorganic cation homeostasis</td>
<td>10</td>
<td>6.86E-03</td>
</tr>
<tr>
<td>gas transport</td>
<td>(erythrocytic) 5</td>
<td>7.51E-03</td>
</tr>
<tr>
<td>oxygen transport</td>
<td>(erythrocytic) 5</td>
<td>7.51E-03</td>
</tr>
<tr>
<td>cation homeostasis</td>
<td>12</td>
<td>8.76E-03</td>
</tr>
<tr>
<td>angiogenesis</td>
<td>(blood vessel development) 8</td>
<td>1.10E-02</td>
</tr>
<tr>
<td>cyclic-nucleotide-mediated signaling</td>
<td>15</td>
<td>1.13E-02</td>
</tr>
<tr>
<td>transmission of nerve impulse</td>
<td>(neurophysiology) 27</td>
<td>1.16E-02</td>
</tr>
<tr>
<td>monovalent inorganic cation transport</td>
<td>28</td>
<td>1.19E-02</td>
</tr>
<tr>
<td>di-, tri-valent inorganic cation transport</td>
<td>14</td>
<td>1.42E-02</td>
</tr>
<tr>
<td>enzyme linked receptor protein signaling pathway</td>
<td>21</td>
<td>1.42E-02</td>
</tr>
<tr>
<td>synaptic transmission</td>
<td>(neurophysiology) 26</td>
<td>1.42E-02</td>
</tr>
<tr>
<td>blood vessel development</td>
<td>(blood vessel development) 8</td>
<td>1.44E-02</td>
</tr>
<tr>
<td>iron ion homeostasis</td>
<td>5</td>
<td>1.64E-02</td>
</tr>
<tr>
<td>blood coagulation</td>
<td>(coagulation) 13</td>
<td>1.80E-02</td>
</tr>
<tr>
<td>tyrosine kinase signaling pathway</td>
<td>15</td>
<td>1.99E-02</td>
</tr>
<tr>
<td>signal transduction</td>
<td>177</td>
<td>2.18E-02</td>
</tr>
<tr>
<td>potassium ion transport</td>
<td>17</td>
<td>2.47E-02</td>
</tr>
<tr>
<td>hemostasis</td>
<td>13</td>
<td>2.77E-02</td>
</tr>
<tr>
<td>pregnancy</td>
<td>8</td>
<td>2.89E-02</td>
</tr>
<tr>
<td>axon guidance</td>
<td>(neurophysiology) 6</td>
<td>4.00E-02</td>
</tr>
</tbody>
</table>
Of particular note is the absence of leukocytic processes such as response to stimulus and defense response identified in the level two analyses. Thus, this comparison allowed for the masking of the “lymphocytic” signature and thus the identification of other endothelial and blood vessels development genes (vWF, bradykinin receptor b1, and thrombin receptor). When this data set was analyzed by GenMAPP, more endothelial genes were mapped to the coagulation cascade pathway (vWF, bradykinin receptor b1, thrombin receptor-pathway (data not shown).

Enriched in BCs vs Endothelial Cells

We identified 904 genes that were up-regulated in BCs relative to prostate-derived endothelium after filtering out those genes that are enriched in ESCs relative to endothelium. Comparing BCs to endothelial cells did not identify as many genes because of their similar origin and thus fewer gene ontologies. However, it did identify more erythrocytic processes (9 in total, Table 5) than the other comparisons. Another predominant theme in this data set was development. By comparing BCs to a more mature yet similar cell type (adult endothelium) we were able to mask the endothelial signature, thus identifying predominantly development genes (indicating stem/progenitor type signature) such as caudal type homeobox transcription factor 2(CDX2), delta-like homolog (DLK1), lamin A/C, secreted frizzled-related protein 5 (SFRP5), patched (PTCH), dishevelled 2 (DVL2), even-skipped homeobox homolog 1 (EVX1).
Table 5

<table>
<thead>
<tr>
<th>Gene Category</th>
<th>List Hits</th>
<th>EASE score</th>
</tr>
</thead>
<tbody>
<tr>
<td>development</td>
<td>(developmental)</td>
<td>84</td>
</tr>
<tr>
<td>oxygen transport</td>
<td>(erythrocytic)</td>
<td>6</td>
</tr>
<tr>
<td>gas transport</td>
<td>(erythrocytic)</td>
<td>6</td>
</tr>
<tr>
<td>heme biosynthesis</td>
<td>(erythrocytic)</td>
<td>5</td>
</tr>
<tr>
<td>heme metabolism</td>
<td>(erythrocytic)</td>
<td>5</td>
</tr>
<tr>
<td>morphogenesis</td>
<td>(developmental)</td>
<td>51</td>
</tr>
<tr>
<td>organogenesis</td>
<td>(developmental)</td>
<td>46</td>
</tr>
<tr>
<td>porphyrin biosynthesis</td>
<td>(erythrocytic)</td>
<td>5</td>
</tr>
<tr>
<td>porphyrin metabolism</td>
<td>(erythrocytic)</td>
<td>5</td>
</tr>
<tr>
<td>transport</td>
<td></td>
<td>77</td>
</tr>
<tr>
<td>pigment biosynthesis</td>
<td>(erythrocytic)</td>
<td>5</td>
</tr>
<tr>
<td>pigment metabolism</td>
<td>(erythrocytic)</td>
<td>5</td>
</tr>
<tr>
<td>pigmentation</td>
<td>(erythrocytic)</td>
<td>5</td>
</tr>
<tr>
<td>hemostasis</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>learning and/or memory</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>skeletal development</td>
<td>(developmental)</td>
<td>10</td>
</tr>
<tr>
<td>response to chemical substance</td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>blood coagulation</td>
<td>(coagulation)</td>
<td>8</td>
</tr>
<tr>
<td>hearing</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>perception of sound</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>synaptic transmission</td>
<td>(neurophysiology)</td>
<td>14</td>
</tr>
<tr>
<td>ion transport</td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>cell-cell signaling</td>
<td></td>
<td>25</td>
</tr>
<tr>
<td>cellular process</td>
<td></td>
<td>204</td>
</tr>
<tr>
<td>transmission of nerve impulse</td>
<td>(neurophysiology)</td>
<td>14</td>
</tr>
<tr>
<td>response to abiotic stimulus</td>
<td>(lymphocytic)</td>
<td>25</td>
</tr>
<tr>
<td>cartilage condensation</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>circulation</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>coenzyme and prosthetic group metabolism</td>
<td>9</td>
<td>4.98E-02</td>
</tr>
</tbody>
</table>

Table 5. Gene ontologies for up-regulated processes in BCs vs endothelial cells. This comparison did not identify as many genes because of their similar origin and thus fewer gene ontologies. However, it did identify more erythrocytic processes than the other comparisons.
**Enriched in BCs vs Stromal Cells**

The BCs were then compared to prostate-derived stromal fibromuscular (CD49a immunoselected) tissue. This comparison identified the most number of genes (3,277 genes), and had the most diverse GO (lymphocytic, developmental, erythrocytic, coagulation, synapses and neurogenesis, and heart development, Table 6).
Table 6. Gene ontologies for up-regulated processes in BCs vs stromal cells. This comparison identified the most number of genes and had the most diverse gene ontologies.
<table>
<thead>
<tr>
<th>Gene Category</th>
<th>List Hits</th>
<th>EASE score</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell communication</td>
<td>328</td>
<td>7.47E-07</td>
</tr>
<tr>
<td>cellular process</td>
<td>617</td>
<td>1.25E-06</td>
</tr>
<tr>
<td>cell-cell signaling</td>
<td>79</td>
<td>4.71E-06</td>
</tr>
<tr>
<td>response to external stimulus</td>
<td>160</td>
<td>1.38E-05</td>
</tr>
<tr>
<td>ion transport</td>
<td>74</td>
<td>4.07E-04</td>
</tr>
<tr>
<td>response to abiotic stimulus</td>
<td>72</td>
<td>4.61E-04</td>
</tr>
<tr>
<td>transport</td>
<td>199</td>
<td>8.97E-04</td>
</tr>
<tr>
<td>cation transport</td>
<td>56</td>
<td>1.08E-03</td>
</tr>
<tr>
<td>defense response</td>
<td></td>
<td>1.09E-03</td>
</tr>
<tr>
<td>metal ion transport</td>
<td>44</td>
<td>1.30E-03</td>
</tr>
<tr>
<td>response to biotic stimulus</td>
<td>101</td>
<td>1.35E-03</td>
</tr>
<tr>
<td>development</td>
<td>192</td>
<td>1.55E-03</td>
</tr>
<tr>
<td>intracellular receptor-mediated signaling pathway</td>
<td>8</td>
<td>1.72E-03</td>
</tr>
<tr>
<td>organogenesis</td>
<td>108</td>
<td>1.98E-03</td>
</tr>
<tr>
<td>cell adhesion</td>
<td>72</td>
<td>2.17E-03</td>
</tr>
<tr>
<td>morphogenesis</td>
<td>118</td>
<td>2.98E-03</td>
</tr>
<tr>
<td>synaptic transmission</td>
<td>35</td>
<td>3.04E-03</td>
</tr>
<tr>
<td>signal transduction</td>
<td>238</td>
<td>4.29E-03</td>
</tr>
<tr>
<td>transmission of nerve impulse</td>
<td>35</td>
<td>5.07E-03</td>
</tr>
<tr>
<td>regulation of blood pressure</td>
<td>9</td>
<td>5.42E-03</td>
</tr>
<tr>
<td>monovalent inorganic cation transport</td>
<td>36</td>
<td>6.29E-03</td>
</tr>
<tr>
<td>steroid hormone receptor signaling pathway</td>
<td>7</td>
<td>6.65E-03</td>
</tr>
<tr>
<td>homeostasis</td>
<td>17</td>
<td>6.77E-03</td>
</tr>
<tr>
<td>cell surface receptor linked signal transduction</td>
<td>113</td>
<td>8.56E-03</td>
</tr>
<tr>
<td>perception of abiotic stimulus</td>
<td>39</td>
<td>1.09E-02</td>
</tr>
<tr>
<td>hemostasis</td>
<td>17</td>
<td>1.10E-02</td>
</tr>
<tr>
<td>second-messenger-mediated signaling</td>
<td>20</td>
<td>1.18E-02</td>
</tr>
<tr>
<td>potassium ion transport</td>
<td>22</td>
<td>1.19E-02</td>
</tr>
<tr>
<td>cyclic-nucleotide-mediated signaling</td>
<td>18</td>
<td>1.24E-02</td>
</tr>
<tr>
<td>blood coagulation</td>
<td>16</td>
<td>1.39E-02</td>
</tr>
<tr>
<td>response to chemical substance</td>
<td>31</td>
<td>1.45E-02</td>
</tr>
<tr>
<td>heme biosynthesis</td>
<td>5</td>
<td>1.48E-02</td>
</tr>
<tr>
<td>gas transport</td>
<td>5</td>
<td>2.00E-02</td>
</tr>
<tr>
<td>oxygen transport</td>
<td>5</td>
<td>2.00E-02</td>
</tr>
<tr>
<td>neurogenesis</td>
<td>51</td>
<td>2.02E-02</td>
</tr>
<tr>
<td>cell homeostasis</td>
<td>15</td>
<td>2.31E-02</td>
</tr>
<tr>
<td>perception of external stimulus</td>
<td>41</td>
<td>2.41E-02</td>
</tr>
<tr>
<td>estrogen receptor signaling pathway</td>
<td>4</td>
<td>2.48E-02</td>
</tr>
<tr>
<td>heme metabolism</td>
<td></td>
<td>2.61E-02</td>
</tr>
<tr>
<td>homophilic cell adhesion</td>
<td>16</td>
<td>2.74E-02</td>
</tr>
<tr>
<td>pigment biosynthesis</td>
<td>6</td>
<td>2.84E-02</td>
</tr>
<tr>
<td>circulation</td>
<td>17</td>
<td>2.94E-02</td>
</tr>
<tr>
<td>olfaction</td>
<td>9</td>
<td>3.02E-02</td>
</tr>
<tr>
<td>response to pest/pathogen/parasite</td>
<td>52</td>
<td>3.19E-02</td>
</tr>
<tr>
<td>acute-phase response</td>
<td>6</td>
<td>3.43E-02</td>
</tr>
<tr>
<td>skeletal development</td>
<td>19</td>
<td>3.72E-02</td>
</tr>
<tr>
<td>sensory perception</td>
<td>37</td>
<td>3.86E-02</td>
</tr>
<tr>
<td>disaccharide metabolism</td>
<td>3</td>
<td>3.86E-02</td>
</tr>
<tr>
<td>enzyme linked receptor protein signaling pathway</td>
<td>24</td>
<td>3.96E-02</td>
</tr>
<tr>
<td>heart development</td>
<td>6</td>
<td>4.09E-02</td>
</tr>
</tbody>
</table>
This data set contained lymphocytic processes according to their GO (response to stimulus, defense response) and numerous lymphocytic markers (CD 6/38/41/43/48/55/61/71/84/244, immunoglobulin genes heavy constant γ1, constant κ, constant λ1, and CD 158A/B/D/F/H (killer cell immunoglobulin-like receptor). This comparison identified genes involved in hemangioblast differentiation (TAL1/SCL, LMO2, RUNX1), endothelial genes (neuropilin 1 and 2), and coagulation (fibrinogen α and β chain, coagulation factor 5, plasminogen) but not KDR, FLT1, CD4, PECAM, VE-Cadherin, vWF). Although EASE analysis for both epithelial and stromal comparisons identified similar heart GOs, the stromal comparison identified different heart development genes such as MEF2C, aortic preferentially expressed (APEG1), POU6F1, TBX1, and ryanodine receptor 2 (cardiac).

When these genes were clustered for pathways analysis with DAVID, we identified Nfat, hypertrophy of the heart and Alk in cardiac myocytes as a statistically over-represented pathway (data not shown). GenMAPP identified a similar pathway involved in myometrial contraction and calcium regulation in the cardiac cell (data not shown). These genetic signatures from this analysis would indicate the presence of progenitors (indicated by the number of developmental genes) of erythrocytes, leukocytes, neurons/neuronal-muscular junctions, and cardiomyocytes. Thus, comparing BCs to stromal cells masked a connective tissue-like signature allowing for the identification of tissue-specific processes.
**Ingenuity Analysis**

To identify signaling pathways involved in hemangioblast differentiation each of the data sets were analyzed by Ingenuity. Ingenuity is a program that converts large data sets into networks containing direct and indirect relationships between genes based on known interactions in the literature. Genetic networks were created using the EB and BC data sets. The EB data set from the level 2 analysis contained a network of genes (VEGF, GATA4, BMP4) that are interconnected and involved in blood vessel development (VEGF), heart development (GATA4), and cellular development (BMP4) (Figure 4a). For example, Bmp-4 has been shown to promote blood vessel development by increasing VEGF production [34] and VEGF induces or binds to KDR FLT1, NRP1, and NRP2 [35-38]. This network suggests that BMP4 inhibits cardiac development by increasing HEY1, a transcriptional repressor of GATA4 and 6 (Figure 5a) [39,40] and inhibits heart development by inducing DKK1 (dickkopf homolog 1) [41], which then inhibits WNT11 mRNA expression, GATA4, and NKX2-5 [42-44]. In conclusion, this network suggests that BMP4 induces blood vessel development through VEGF signaling and inhibits cardiac differentiation through HEY1 and DKK1.

We then looked for these and other signaling pathways in the level II BC data set. Here we identified genes involved in cardiovascular development (SHH [45], RAR-B, TBX5 [46], WNT11 [43]) acting through GATA4 (Figure 4b). However, unlike the EB data set, we did not identify the cardiac repressor HEY1, VEGF
and BMP4 in the BC data set. Instead, the BC network contained cardiac and skeletal genes such as HAND2 and ANKRD1 [47-50] and HIF3A, an inhibitor of VEGF expression [51] (Figure 5b). Thus, these networks demonstrate that when EBs differentiate into BCs, we see a transition from early hemangioblast-like signature to a more mature hemangioblast signature with some cardiac pathways.

Another signaling network we identified as differentially expressed between EBs and BCs is the network containing GATA2. GATA2 has been shown to play a vital role in hemangioblast development [26] by up-regulating BMP4, KDR, and TAL1/SCL expression. In the EB data set, the GATA2 network contained EPOR, TAL1/SCL, TCF3 and PITX2 (Figure 6a). Pitx2 is a homeobox gene involved in regulating the balance between proliferation and differentiation of progenitor cells [52] and is highly expressed in EBs (FC=24x) and is absent in BCs. PITX2 is not only rapidly down-regulated upon HSC differentiation [52], but may also promote hemangioblast differentiation by inducing GATA2 expression [53]. In the BC data set, the GATA2 network contained the hemangioblastic and hematopoietic genes TAL1/SCL and LMO2[25,27,54], FOG/Zfpm1[55], CD41/GPIIB/Igta2B [56] and GATA1 [57] (Figure 6b).

A predominant network we identified in all 3 data sets of the Level III analysis were those involving GATA1. GATA1 is a globin transcription factor and is present in all BC but not EB data sets. For example, we see GATA1 interacting
with other nuclear genes such as TAL1/SCL, LMO2, and KLF1 which induce erythropoietic genes such as the hemoglobin family (HBG1, HBG2, HBE, HBB, and HBZ), heme synthesis (ALAS2), and genes expressed on the cell surface of RBC (Ankyrin 1, Rh blood group, glycophorin A, erythrocyte membrane protein band 4.2) (network not shown).
Figure 4. Ingenuity pathway analysis shows a network of genes expressed in EB and BC data sets from the level II analysis. This network contains nodes (genes/gene products) and edges (relationships between the nodes). The shaded genes, known as focus genes, were identified by microarrays and are the starting point to generate the network. The asterisks indicate duplicate identifies in the dataset. The EB data set contains a network of genes (VEGF, GATA4, BMP4) that are interconnected and involved in blood vessel development (VEGF), heart development (GATA4), and cellular development (BMP4). The BC data set identified a network of genes involved in cardiovascular development (SHH, RARB, TBX5, WNT11) acting through GATA4.
Figure 5. Ingenuity pathway analysis identified a network of genes expressed in EB and BC data sets associated with VEGF. In the EB data set, VEGF is associated with 35 “focus” genes including KDR, FLT1, NRP1, and NRP2. In the BC data set, 13 focus genes were associated with VEGF, even though it is not present in the data set.
Figure 6. Ingenuity pathway analysis shows a network of genes expressed in EB and BC data sets associated with GATA2. In the EB data set, the GATA2 network contained EPOR, TAL1, TCF3 and PITX2. In the BC data set, the GATA2 network contained the hemangioblastic and hematopoietic genes TAL1, LMO2, FOG/ZFPM1, IGTA2B/CD41/GPIIb, and GATA1.
Discussion

To delineate the mechanisms involved in the development of hemangioblasts from murine ES cells, Lugus et al [26] performed a global gene expression profiling of murine BL-CFC cells. However, no such analysis has been reported due to availability of such cells. In the present study, we carried out a large scale transcriptomical analysis to profile undifferentiated human ESCs, early stage EBs, and BCs as an initial effort to understand the differentiation program of ESCs towards hemangioblasts. The availability of a well-annotated genome database for human ESCs, EBs and BCs will provide a foundation that allows one to map and identify genes involved in human hemangioblast development, and to optimize conditions for efficient generation of hemangioblast from ESCs. Our studies in general are consistent with previous reports that a cluster of genes (OCT4, NANOG) expressed at significantly high levels in ESCs were down-regulated in early stage EBs and BCs [23,58-63], whereas some genes restricted to endothelial and hematopoietic cells (neuropilins, LMO2, GATA1 and 2, TAL1/SCL and globins) were up-regulated dramatically in BCs and even in early stage EBs [2,23,26,64].

Our previous study has shown that BCs contain a mixed progenitor population of cells capable of forming hemangioblasts, hematopoietic and endothelial cells [21]. To assess the heterogeneous populations in BCs, comparisons to publicly available data sets were performed in silico. Biologically relevant comparisons allowed us to mask particular genetic signatures within the heterogenous
population allowing us to identify others, such as myogenic, vasculogenic, hematopoietic progenitors within BCs. Our Level I analysis compared ESCs to their differentiated progeny providing a kinetic-like relationship of gene expression. When comparing the ESCs to BCs, many of the down-regulated genes were involved in development, cell differentiation, and morphogenesis. The predominant up-regulated genes identified in this Level I analysis were those involved in the development of hemangioblasts and primitive erythroblasts. These genes include transcription factors TAL1/SCL, LMO2, GATA1 and 2, heme synthesis, erythroblast membrane proteins, and embryonic and fetal globin genes, but very low level of adult globin gene, suggesting the yolk sac primitive status of BCs. This observation is consistent with findings obtained from both mouse and zebrafish models, in which hemangioblasts were first developed from primitive streaks and yolk sacs [65,66]. The results may also represent a fact that BC expansion medium contains several hematopoietic cytokines that pushes the development pathway towards hematopoietic lineage, although BCs harvested at day 6 retain the potential of differentiating into endothelial cells under appropriate conditions [21]. Optimization of BC expansion condition would therefore be valuable to keep these cells bipotential.

When applying this technique of multiple tissue type comparisons, some of the genes that are identified as “up-regulated” in our tissue of interest could be “under expressed” in the reference tissue. We controlled for this by filtering for those “under expressed” genes with a comparison to a genotypically similar but
different tissue type. For example, in our Level II analysis, we identified those genes that are up-regulated in BCs relative to breast epithelial cells (3700 genes) and then removed those genes that were up-regulated in ESCs relative to epithelial cells (2735 genes). This removed 965 genes that could be thought of as being up-regulated in both BCs and ESCs relative to breast epithelium or as genes that are under-expressed in breast epithelium. GO analysis of this data set identified cell cycle genes as the predominant theme. This observation correlates with their biology because the BCs and ESCs are actively dividing cells in vitro, while the breast epithelium is relatively senescent cells freshly isolated from in vivo. Most importantly, we do not identify any tissue-specific processes that we had or would expect to find in BCs.

GO analysis of the genes up-regulated in BCs with respect to epithelial cells after filtering out those that are up-regulated in ESCs relative to breast epithelial cells identified biological themes involved in erythropoiesis as in the Level I analysis (heme and porphyrin biosynthesis), but also angiogenic, coagulation, and synaptic processes. This analysis identified genetic signatures representative of not just erythrocytes as in the above Level I analysis, but also the other cellular components of the BC population such as muscle, cardiac, hematopoietic and hemangioblasts. In silico comparison of BCs to a biologically distinct reference tissue allows one to identify statistically significant genes that might otherwise be missed within a heterogenous population.
In the Level III analysis, biologically relevant comparisons were made in silico to adjust for particular cell types represented in the BC population. Like in the Level II analysis, this analysis also identified genetic signatures of the erythrocytic population in addition to other cellular components of BCs. When BCs were compared to leukocytes in silico, we identified a genetic signature representative of vasculogenesis, endothelial, neurons/synapses, hemangioblasts, and erythrocytes with the relative absence of leukocytic genes. When BCs were compared to endothelia in silico, we identified a signature of erythrocytic and developmental genes with the relative absences of the vasculogenic signature. When BCs were compared to stromal cells in silico, we identified processes involved in hematopoiesis, synapses, angiogenesis/endothelia, development and more genes involved in cardiomyogenesis. Although EASE analysis for both epithelial and stromal comparisons identified similar heart GOs, the stromal comparison identified different heart development genes. We believe that the epithelial comparison revealed more of the mesodermal aspect of BCs while the stromal comparison masked the mesodermal components and revealed more cardiomyocytic genes.

It has been shown that murine BL-CFCs were be able to differentiate into hematopoietic, endothelial and smooth muscle cells, but failed to give rise to cardiomyocytes [67]. Molecular analyses showed that BL-CFC cells expressed genes indicative of the hematopoietic and endothelial lineages, but not cardiomyocyte[26]. Kattman et al [68] recently identified a cardiovascular
progenitor with the same phenotype of BL-CFC progenitor, brachyury+ and Flk-1+ cells from day4.25 EBs, one day later than that of BL-CFC progenitor, of mouse ESCs. These studies strongly suggest that BL-CFC and cardiomyocyte, at least in mouse ESC system, were derived from two different progenitors. While two other studies demonstrate the existence of multipotential progenitors for cardiomyocytes and muscle cells, but with different surface markers [69,70]. In the present study, several genes restricted to cardiomyocyte or its progenitor were detected with relatively high level of expression in BCs. This observation suggests that human BCs may possess the potential of cardiomyocyte development, which will need further investigation. Alternatively, the purified BCs from multiple colonies could contain dissimilar blast clones originated from different differentiation stages, some of them may have the potential of cardiomyocyte development as demonstrated recently by three groups [68-70] in the mouse ESC system. Simultaneous isolation and characterization of functionally distinct colonies and analysis of gene expression in these colonies might serve to determine whether human BCs possess the potentials to differentiate into hematopoietic, endothelial and cardiomyocyte lineages.

Conclusions

The identification and characterization of cell types within a heterogeneous population will be of increasing importance in stem cell research since differentiation protocols require the formation of progenitors through a multi-stage approach. Our previous study has shown that BCs contain a mixed progenitor
population of cells capable of forming hemangioblasts, hematopoietic and endothelial cells [21]. To assess the heterogeneous populations in BCs, comparisons to publicly available data sets were performed in silico. Biologically relevant comparisons allowed us to mask particular genetic signatures within the heterogeneous population allowing us to identify others, such as myogenic, vasculogenic, hematopoietic progenitors within BCs. The significance of this microarray study is in its ability to assess and identify cellular populations within a heterogeneous population with microarrays through biologically relevant in silico comparisons of publicly available data sets. In conclusion, multiple in silico comparisons were necessary to characterize tissue-specific genetic signatures within a heterogeneous hemangioblast population.

Materials and Methods

Human ESC culture and BC growth: Culture of ESCs and growth of blast colonies were as reported previously [21]. In brief, undifferentiated ESCs (H1, H9) were cultured with inactivated mouse embryonic fibroblast (MEF) cells in complete ESC media until they reach 80% confluence. Undifferentiated ESCs were dissociated by 0.05% trypsin-0.53 mM EDTA (Invitrogen) for 2-5 min and collected by centrifugation at 1,000 rpm for 5 minutes. To induce hemangioblast precursor (mesoderm) formation, ESCs (2 to 5 $\times$ 10$^5$ cells/ml) were plated on ultra-low dishes (Corning) in Stemline II media with the addition of BMP-4 and VEGF$_{165}$ (50 ng/ml, R&D Systems) and cultured in 5% CO2. Forty eight hours later, half the media were removed and fresh media were added with the same
final concentrations of BMP-4 and VEGF, plus SCF, Tpo and FLT3 ligand (20 ng/ml, R&D Systems), and PTD-HoxB4(1.5 μg/ml) to expand out BCs and its precursor. After 3.5 days, EBs were collected and dissociated by 0.05% trypsin-0.53 mM EDTA (Invitrogen) for 2-5 min, and single cell suspension was prepared by passing through 22G needle 3-5 times. To expand BCs, single cell suspension derived from differentiation of 2 to 5 X 10^5 ESCs were mixed with 2 ml hemangioblast expansion medium plated on ultra-low dishes and incubated at 37 °C in 5% CO2 for 6 days, and BC cells were then collected and subjected for RNA isolation.

**Affymetrix GeneChip analysis:** Total RNA was isolated from purified BC cells, day3.5-EBs and undifferentiated ESCs (from two human ESC lines, H1 and H9) using the Qiagen RNAeasy kit and amplified as previously described [21]. A total of six microarrays were performed (2 biological replicates per time point using different HESC lines). Fragmented antisense cRNA was used for hybridizing with human U133 Plus 2.0 arrays (Affymetrix, Inc) at the Core Genomic Facility of University of Massachusetts. The validation of differentially expressed genes was confirmed by immunocytochemistry in our previous studies [21] and by semi-quantitative RT-PCR analyses (Figure 1). The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus [71] and is accessible through GEO series accession number GSE8884 and GSE9196, in accordance with MIAME standards. The demographics of the publicly available Genechip data sets are breast-derived epithelial cells (n=7), leukocytes.
(n=6), prostate-derived endothelial cells (n=5), and prostate-derived stromal cells (n=5). These samples were chosen based on their homogeneity (cells were immunoselected or enriched) and the number of replicates. These data sets were downloaded from the NCBI’s Gene Expression Omnibus and are accessible through GEO accession number GSE9086 (breast-derived epithelial cells), GSE9091 (leukocytes), GSE9090 (stromal), and GSE9089 (endothelia).

**Data Analysis:** Raw CEL files were provided by the Core Genome Facility of University of Massachusetts and were then analyzed with a software package AffylmGUI (Affymetrix LIMMA, Linear Models for Microarray Data, Graphical User Interfaces) [72,73]. Within AffylmGUI, gene expression values were summarized with RMA. RMA adjusts for background noise, performs a quantile normalization, transforms the data into log based 2, and then summarizes the multiple probes into one intensity [74-76]. Quantification of relative differences in gene expression among the groups of interest was accomplished using AffylmGUI, the sister package of limmaGUI[72,73]. AffylmGUI reads the raw Affymetrix CEL files directly, summarizes the gene expression values using RMA, and then uses LIMMA to identify statistically significant differences in gene expression[77]. LIMMA fits a linear model for every gene (like ANOVA or multiple regression analysis), and adjusts P-values for multiple testings [77]. Differentially expressed genes were identified with a B statistic > 0. The B statistic, also known as a likelihood of odds (LOD) score is a moderated t-statistic.
with posterior residual standard deviations. Subsequent analyses were performed in Microsoft Excel and Microsoft Access.

**EASE:** The application Expression Analysis Systematic Explorer (EASE) was used to determine biologically relevant themes in a list of differentially expressed genes. EASE identifies over-represented biological themes in terms of their gene ontology (GO)[78]. GO was developed to provide consistent descriptions of genes in terms of biological processes and molecular function. Genes with a B value >0 are incorporated into EASE where each gene is matched to all possible GOs. The results of this analysis is compared to all possible GOs for all genes on the microarray platform and calculates a P-value, based on a conservative variant of the Fischer’s exact probability test. We select those pathways/processes with a P<0.05.

**DAVID:** This program provides tools and statistical methods for uncovering enriched processes and pathways within diverse and disparate gene lists [79]. DAVID also identifies over-represented biological themes in terms of their gene ontology (GO) [78] and provides tools to visualize the distribution of genes on BioCarta and KEGG pathway maps.

**GenMAPP:** This application is designed to visualize gene expression data on maps representing biological pathways and grouping of genes. GenMAPP consists of hundreds of pre-made pathways maps and were used to identify
pathway level changes amongst multiple data sets[80]. In this program, data sets are assigned a color corresponding to cell type and the direction of gene expression changes.

**Ingenuity:** Networks were constructed using Ingenuity Pathways Analysis (Ingenuity ® Systems). A data set containing gene identifiers of genes with a B>0 was uploaded into the applications. These genes, called focus genes, were overlaid onto a global molecular network developed from information contained in the Ingenuity Pathway Knowledge Base. Networks of these focus genes were then algorithmically generated based on their connectivity.

**RNA isolation and gene expression quantification by semi-quantitative PCR:** Total RNA was isolated from human ES cells (ESCs), day3 EBs (EBs) and hemangioblasts (BCs) using an RNAeasy Mini Kit (Qiagen, Valencia, CA) following the procedure recommended by the supplier with the step of DNase I digestion, which will eliminate the contamination of genomic DNA. RNA was subjected to first-strand cDNA synthesis with SMART II and CDS primers (Clontech), using Superscript II reverse transcriptase (Invitrogen), and cDNA pools were constructed using the SMART cDNA synthesis kit (Clontech) as described previously [2,81] Complementary DNA pools generated by the SMART procedure have been shown to preserve the relative abundance relationship of the original mRNA populations [82-84]. The DNA templates in cDNA pools of human ES cells, EBs and hemangioblasts were adjusted to equal amounts based on the relative expression level of the hypoxanthine
phosphoribosyltransferase (HPRT) gene and gene expression quantification by semi-quantitative PCR was performed as described previously [2,81]. The sense and anti-sense primer sequences, and the corresponding cDNA PCR product sizes are shown in Table 7. The conditions for PCR amplification was as described with annealing temperatures and concentrations of MgCl$_2$ for each specific gene as shown below. Ten µl of PCR products were separated on 1.5 to 2.0 % agarose gel and visualized by ethidium bromide staining. The relative expression levels in cDNA from human ES cells, EBs and hemangioblasts were estimated visually.

Table 7  Primer Sequence

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer Sequence</th>
<th>Annealing Tm (C)</th>
<th>MgCl$_2$ Conc (M)</th>
<th>Product Size (bp)</th>
</tr>
</thead>
</table>
| HPRT                  | Sense: 5’-CTTGCAGCCTTGACCACATCTTTTGGA-3’  
                          Antisense: 5’-GGCGTGATGTATAGGTGATGAACC-3’                             | 58               | 1.5              | 467                |
| GATA-2                | Sense: 5’-TATGTGGCGCGGCGCTGCCCACGACTACA-3’  
                          Antisense: 5’-GGCTCTTCTTGCGCGCCGACAGTCTTT-3’                 | 60               | 1.5              | 280                |
| SCL                   | Sense: 5’-GAAGTGCTCCCCCTCTGAAAGTT-3’  
                          Antisense: 5’-GGCTATCTCTCTCTGCACCTCG-3’                         | 58               | 1.5              | 319                |
| OCT4                  | Sense: 5’-GAAGGTATTCCAGCCAAACGAC-3’  
                          Antisense: 5’-GTTACAGAACCACACTCGGA-3’                        | 55               | 2                | 315                |
| Nanog                 | Sense: 5’-TGCAAATGTCTTCTGACGCAGA-3’  
                          Antisense: 5’-GCGTACGCAAATTAAAAGTGCCAGA-3’                 | 55               | 2                | 285                |
| Rex-1                 | Sense: 5’-TGACAGGCAAGAGCTTCCCAG-3’  
                          Antisense: 5’-GCCTGACGCAAATTAAAAGTGAGAA-3’                | 55               | 2                | 350                |
Direct and indirect comparison of microarray data of differentially expressed genes: The direct analysis (level I) consists of making comparison with ESC, EB, and BC. Since there are two possible comparisons for each cell type, for example, ESC relative to EBs or ESC relative to BCs, fold changes were determined by comparing each to a cell type that did not detect that gene. Fold change levels for Oct-4 and Nanog were determined by comparing ES to BC. Fold change levels for Gata-2 was determined by comparing EB to ES and BC to ES. Fold change levels for SCL/Tal1 was determined by comparing BL to ES. Fold change levels for γ/ε-globin were determined by comparing BC to EB. For the indirect (level II) analysis, fold changes were simply determined by comparing each cell type to breast epithelia. If more than one probe-set was identified as differentially express, fold changes were averaged.
References


43. Schulze M, Belema-Bedada F, Technau A, Braun T: Mesenchymal stem cells are recruited to striated muscle by NFAT/IL-4-mediated cell fusion. *Genes Dev* 2005, 19: 1787-1798.


77. Smyth GK: Linear models and empirical bayes methods for assessing
differential expression in microarray experiments. Stat Appl Genet Mol

78. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM et al.:
Gene ontology: tool for the unification of biology. The Gene Ontology

79. Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC et al.:
DAVID: Database for Annotation, Visualization, and Integrated
Discovery. Genome Biol 2003, 4: 3.

80. Dahlquist KD, Salomonis N, Vranizan K, Lawlor SC, Conklin BR:
GenMAPP, a new tool for viewing and analyzing microarray data on

81. Lu S-J, Li F, Vida L, Honig GR: Comparative gene expression in
hematopoietic progenitor cells derived from embryonic stem cells.

82. Zhumabayeva B, Diatchenko L, Chenchik A, Siebert PD: Use of SMART-
generated cDNA for gene expression studies in multiple human

83. Seth D, Gorrell MD, McGuinness PH, Leo MA, Lieber CS, McCaughan
GW et al.: SMART amplification maintains representation of relative
gene expression: quantitative validation by real time PCR and
application to studies of alcoholic liver disease in primates. J

84. Zheng J, Shen W, He DZ, Long KB, Madison LD, Dallos P: Prestin is the
CHAPTER III

MICROARRAY ANALYSIS OF BLADDER SMOOTH MUSCLE FROM PATIENTS WITH MYELOMENINGOCELE

Jennifer A. Hipp, Jason D. Hipp, James J. Yoo, Anthony Atala and Karl-Erik Andersson

The following manuscript was published in British Journal of Urology International 2008 Sep;102(6):741-6 and is reprinted with permission. Stylistic variations are due to the requirements of the journal.
Abstract

Objective

The genetic profiles of smooth muscle cells (SMCs) from healthy bladders and bladders from patients with myelomeningocele (MM) were compared to identify genes that are over- and under-expressed in MM bladder SMCs. We examined whether gene profiles would providing a molecular evaluation of the quality and therapeutic potential in patients with MM.

Material and Methods

Bladder smooth muscle biopsies were obtained from ‘healthy’ subjects undergoing bladder surgeries for vesicoureteral reflux and patients with neurogenic bladder secondary to MM. Bladder SMCs were expanded in vitro and total RNA was isolated and hybridized to Affymetrix U133A GeneChips to evaluate the differential expression levels of 22,283 genes. Differentially expressed genes were identified by two methods. In the first analysis, we made direct comparison of raw data sets of healthy SMCs to those derived from MM patients. In the second analysis, indirect comparisons were made by comparing healthy SMCs and MMSMCs to a reference file to create a genetic signature of genes that are over- and under-expressed in MMSMCs.

Results

The direct analysis identified 240 genes that were over-expressed and 104 genes were under-expressed in MMSMCs. Gene ontology (GO) classifications
were used to identify biological themes and pathways. Genes that were over-expressed in MMSMCs were involved in development: mesenchyme homeobox 2 (fold change (FC) = 9.3); bone morphogenic protein 6 (FC=4.0); fibroblast growth factor 2 (FC = 4.8); inhibin A (FC = 4.2), cartilage oligomeric matrix protein (FC =9.97); collagen 11A (FC =6); collagen 5A2 (FC=3) collagen 1A1 (FC = 2.18). The indirect analysis identified 665 genes that were over-expressed and 1343 genes that were under-expressed in MMSMCs. Pathway-based analysis of these genetic signatures revealed an over-expression of genes involved in muscle development and focal adhesion/extracellular matrix interactions. Genes that were under-expressed in MMSMCs were mapped to muscle contraction, transmission of nerve impulses, and cell-cell adhesion pathways.

**Conclusion**

Our data is consistent with previous studies showing that MM bladders have an excess of extracellular matrix (ECM) deposition, improper contraction, and are developmentally immature relatively to healthy SMCs. The clinical implication of microarray analysis of MMSMCs is that it provides potential targets that could induce muscle differentiation and inhibit ECM production.
Introduction

Myelomeningocele (MM) is the most common form of neural tube defects and a major cause of neurogenic bladder dysfunction. It is a congenital defect caused by the inability of the neural tube to close during early embryo development. This often results in the extrusion of spinal cord and its covering into a sac at the level of the lesion (1). MM patients may present with a spectrum of defects such as lower limb paralysis, bowel and bladder dysfunction (1;2). They often have detrusor-sphincter dyssynergia, and may eventually present with a small, fibrotic, and non-compliant bladder (3). Common complications include urinary incontinence, infections (4) and reflux that often require bladder reconstruction/augmentation (1).

A common feature of neurogenic bladders, including MM, is an increased muscle mass within the bladder wall and a decreased compliance (5). The thickening of the bladder wall is believed to be due to smooth muscle hypertrophy and/ or hyperplasia (6). It has been suggested that the poor elasticity and compliance of neurogenic bladders may be due to an abnormal deposition of extracellular matrix proteins (5;7;8). As a consequence of these changes, alterations in the global gene expression profile may be expected. Understanding the global changes in gene expression is necessary to determine which processes and pathways are involved in MM bladders.
In this study, we compared the genetic profiles of SMCs from healthy bladders and bladders from patients with MM in order to identify genes that are over- and under-expressed in MMSMCs. Pathway-based analysis of genes identified in these comparisons was performed to provide a molecular evaluation of the quality and therapeutic potential of MMSMC tissue. Based on previous studies, we hypothesized that the predominant genetic difference in MMSMCs includes changes in extracellular matrix, growth and developmental processes.

**Material and Methods**

After obtaining parental consent and institutional review board approval, biopsies were obtained from three patients with neurogenic bladder secondary to myelomeningocele. SMCs were isolated and cultured according to existing protocols (9-11). ‘Healthy’ SMCs were obtained from patients undergoing reflux operations whose bladders were otherwise normal. Briefly, bladder tissue specimens obtained from the dome were immediately dissected to isolate the detrusor layer, and then cut into fragments. Muscle tissue fragments of approximately 2 mm in size were placed on a scored cell-culture dish and allowed to briefly air-dry for adherence. Dulbecco’s modified Eagle’s medium (DMEM, Gibco/Life Technologies, Gaithersburg, MD) supplemented with 10% fetal bovine serum (FBS, HyClone Laboratories, Logan, UT), penicillin (100 µ/mL) and streptomycin (100 µg/mL) was added. The cells migrated from the tissue explants were maintained in the same medium under 5% CO₂ at 37°C.
All experiments were performed in triplicate on cells between passages 2 and 5. Total RNA was isolated and hybridization to the Affymetrix U133A GeneChip performed by the Wake Forest University School of Medicine’s Affymetrix core facility using standard protocols. Raw data files were then analyzed with Robust Multichip Array (RMA) (12) and differentially expressed genes were found with LIMMA (Linear Models for Microarray Data) (13) using the graphical user interface provided by affylmGUI, the sister package of limmaGUI (14). Differentially expressed genes were ranked using the B statistics (15,16) and P values were adjusted using the FDR method of Benjamini (17). Genes that had a B value of greater than 0 (Log of Odds score) and a corresponding FDR adjusted P values that were less than 0.0393 were selected as being differentially expressed.

Two comparative methods were used to identify differentially expressed genes as previously described (18,19). In the first analysis, we made direct comparison of raw data sets of healthy SMCs to those derived from MM patients. To identify the genetic “themes” present in the data sets, the probe sets were loaded into Expression Analysis Systematic Explorer (EASE) (20) and Database for Annotation, Visualization and Integrated Discovery (DAVID) (21). DAVID provides statistical methods for uncovering enriched biological themes within data sets. Probe sets are then clustered based on their Gene Ontologies (GO Bio, GO Molecular Function, and GO Cellular Component) and for annotation.
Statistically over-represented pathways were identified by selecting those with Benjamini score of less than 0.05.

The functional analyses were generated through the use of Ingenuity Pathways Analysis (Ingenuity® Systems, www.ingenuity.com) identified biological functions and/or diseases that were most significant to the data set. A data set containing up-regulated genes was uploaded into the applications. Genes from the data set were associated with biological functions and/or diseases in the Ingenuity Pathways Knowledge Base. Fischer’s exact test was used to calculate a p-value determining the probability that each biological function and/or disease assigned to that data set is due to chance alone.

In the second analysis, indirect comparisons were made by comparing healthy SMCs and MMSMCs to a reference file to create a genetic signature of genes that are over- and under-expressed in MMSMCs. The reference file consisted of nine raw fibroblast data sets (22) that were downloaded from NCBI Gene Expression Omnibus from accession number GDS 1503. These data files were analyzed with GenMapp, a program consisting of hundreds of pre-made pathway maps for pathway identification (23).

**Results**

Two comparative methods were used to identify differentially expressed genes in MMSMCs. RMA was used to analyze raw data sets and provided a relative level
of gene expression between two samples. A direct comparison between
‘healthy’ SMCs and those derived from MM patients was performed. Our first
method consists of making direct comparisons with ‘healthy’ SMCs to those
derived from MM patients. This identified 240 genes that were up-regulated and
104 genes that were down-regulated in MMSMCs.

GO analysis of up-regulated genes revealed predominantly genetic themes
involved in development, morphogenesis, cell adhesion and extracellular matrix
(Table 1). Genes identified in the development process included mesenchyme
homeobox 2 (MEOX2; fold change, (FC) = 9.3); bone morphogenic protein
(BMP6; FC=4.0); basic fibroblast growth factor (FGF2; FC = 4.8) and inhibin A
(INHBA FC=4.2). Genes identified in the cell adhesion process included mostly
ECM proteins such as cartilage oligomeric matrix protein (COMP; FC =9.97);
collagen 11A (FC =6); collagen 5A2 (FC=3) collagen 1A1 (FC = 2.18). Although,
down-regulated genes were analyzed by DAVID, no pathways were identified as
statistically significant. This data set comprised about half as many genes as in
the over-expressed data set, which resulted in fewer pathways and genetic
processes. However, genes that were under-expressed in MM were involved in
retinoic acid signaling and included retinoic acid receptor responder 1
(RARRES1; FC = 9.9), aldehyde dehydrogenase A3 (FC = 26.5), retinol binding
protein 1 (RBP1; FC =3.4), retinoic acid receptor responder 2 (RARRES2; FC =
3.2), and’ stimulated by retinoic acid gene 6’ (STRA6; FC = 2.3). Under-
expressed genes also include smooth muscle actin, (FC = 5.2), myosin heavy chain (MYH10; FC = 3.5), troponin C (FC=3.2), and GATA2 (FC = 3.2).

Table 1.

<table>
<thead>
<tr>
<th>Category</th>
<th>Term</th>
<th>Count</th>
<th>Benjamini-Hochberg corrected p-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO Biological Process</td>
<td>development</td>
<td>60</td>
<td>1.70E-04</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>morphogenesis</td>
<td>29</td>
<td>7.70E-04</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>cell adhesion</td>
<td>25</td>
<td>4.10E-02</td>
</tr>
<tr>
<td>GO Biological Process</td>
<td>skeletal development</td>
<td>11</td>
<td>4.9E-02</td>
</tr>
<tr>
<td>GO Cellular Process</td>
<td>extracellular region</td>
<td>36</td>
<td>5.20E-03</td>
</tr>
<tr>
<td>GO Cellular Process</td>
<td>extracellular matrix</td>
<td>16</td>
<td>6.60E-03</td>
</tr>
<tr>
<td>GO Molecular Process</td>
<td>extracellular matrix</td>
<td>11</td>
<td>4.30E-03</td>
</tr>
<tr>
<td></td>
<td>structural constituent</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. GO analysis of gene over-expressed in MMSMCs. Genes were loaded into DAVID and clustered based on their GO Biological Process, and GO Cellular Component, GO Molecular Function. GOs with Benjamini-Hochberg corrected p-values < 0.05 were selected. GO analysis revealed predominantly genetic themes involved in development and extracellular matrix.

The second analysis compared ‘healthy’ SMCs and MMSMCs to fibroblasts, to generate genetic signatures of genes that are over- and under-expressed in MMSMCs. Human fibroblasts were chosen because they lack tissue specific genes which makes it possible to subtract housekeeping genes, such as those involved in cell metabolism and maintenance. This will remove potential genetic
noise. Data sets were compared and contrasted and identified 665 genes that were over-expressed in MMSMCs, 1,343 genes that were under-expressed in MMSMCs, and 409 genes that were expressed in both ‘healthy’ SMCs and MMSMCs (Figure 1). Genes that were identified in both direct and indirect analysis were combined to create a MMSMCs gene expression profile of 905 genes that were up-regulated and 1,447 genes that were down-regulated.

**Figure 1.** Schematic depiction of the algorithm used for illustrating prominent biological themes and pathways
For pathway analysis, genes that were identified as differentially expressed in both direct and indirect analysis were analyzed by GenMapp to visualize the changes of gene expression in biologically relevant pathways and processes. We created a color template by assigning each data set a color; up-regulated genes are labeled red, down-regulated genes are green, and genes that were expressed in both are labeled orange. These genes were then over-laid onto pre-existing pathways. GenMapp analysis of up-regulated genes shows significant involvement of genes involved in muscle development (Figure 2) and focal adhesion/ECM interactions (data not shown). The muscle development pathways also had an equal number of genes that were down-regulated in MMSMCs (14 genes vs 14 genes). However, further examination of the pathway suggests that up-regulated genes in MMSMCs include growth factors (FGF2, IGF1) and genes that regulate smooth muscle contraction (caldesmon 1 and tropomyosin 1 and 2). Down-regulated genes identified are involved in the functional aspects of muscle contraction such as myosin heavy chain (MYH) 3, 6, 7, 11, 13, myosin light chain (MYL) 5, troponin I2 (TNNI2), and troponin T2 (TNNT2).
**Figure 2.** GenMapp analysis of muscle development and contraction with genes over- and under-expressed in MMSMCs (red and green, respectively). Genes over-expressed in MMSMCs mapped to muscle development genes while under-expressed genes were involved in muscle contraction.
**Genes under-expressed in MMSMCs**

To identify genes which are under-expressed in MMSMC, we created a genetic signature by combining the indirect and direct data sets (1,343 and 104 genes). When analyzed by GenMapp, genes that were identified as under-expressed were mapped to muscle contraction (Figure 2), transmission of nerve impulses (Figure 3), and cell-cell adhesion pathways (Figure 4). Genes identified in muscle contraction include MYL5, DESMIN, MYH3, 6, 7, 11, 13, TNNI2, TNNT2, and connexin 43. Under-expressed genes involved in synaptic transmission include procadherin (PCDHB) 3, 4, 5, 6, 10, 12, 14, 16, acetylcholinesterase (AChE), monoamine oxidase A (MAOA), and serotonin receptor 7 (HTR7). Under-expressed genes involved in cell-cell adhesion and cell junction include intracellular adhesion molecule (ICAM) 2, cadherin 15 (CDH15), PCDHB3, 4, 5, 6, 10, 12, 14, 16 and claudin 3 (CLDN3), CLDN11, connexin 43, connexin 46, tight junction protein (TJP) 3, respectively (Figure 4).
Figure 3. GenMapp pathways of transmission of nerve impulse with genes that were over- or under-expressed in MMSMCs (red and green, respectively). More under-expressed genes mapped to this pathway than over-expressed genes.
Figure 4. GenMapp analysis of genes over- and under-expressed in MMSMCs (red and green, respectively). More under-expressed genes mapped to cell-cell adhesion pathway than over-expressed genes.
**Ingenuity**

Differentially expressed genes in MMSMCs were analyzed by Ingenuity to organize genes into processes involved with a particular disease or function. Over twenty processes were identified as statistically significant. Both up- and down-regulated genes were involved in similar categories, such as cardiovascular, connective tissue, skeletal and muscular system, development and function (Figure 5). However, up-regulated genes in MMC had at least twice the number of genes in each category. Interestingly, the ‘tumor morphology’ process identified 23 up-regulated genes in MMSMCs while normal SMC only contained 3 genes.
Figure 5. Ingenuity analysis categorized the top ten biological system, development and function of up-regulated genes in MMSMC. Down-regulated genes were also analyzed by Ingenuity and involved in similar categories, such as cardiovascular, connective tissue, skeletal and muscular development. However, up-regulated genes in MMSMCs had at least twice the number of genes in these categories. Numbers represent a comparison of genes that were identified in a particular process.
Discussion

Patients with myelomeningocele may present with a wide range of defects that translate into a variety of clinical abnormalities. As such, the bladder wall changes usually lead to varying degrees of bladder dysfunction. Recent advances in molecular and gene profiling technologies have provided an ability to identify abnormal gene expressions in pathologic tissues. This study investigated whether analysis of the genetic profiles of smooth muscle cells obtained from MM bladders could provide an opportunity to further understand the global changes in gene expression. Large-scale gene expression analysis of 22,283 genes were used to identify the global changes in gene expression between healthy and MMSMCs. Two methods of analysis allowed us to examine different aspects of gene expression. The direct analysis determined a relative expression of bladder SMCs from healthy individuals and MM patients, while the indirect analysis created a genetic expression profile of healthy SMCs and MMSMCs. Both analyses identified similar characteristics of genes that were up- and down-regulated in MM.

The consistency of our microarray analysis is validated by the identification of genes and processes that have been previously reported. The up-regulation of FGF2 and collagens in MM bladders is consistent with previous reports (6;8). Although it has been shown that neurogenic bladder secondary to MM have an increase expression of collagens (6), our microarray analysis identified numerous genes involved in extracellular matrix production such as collagen 6a1, cartilage
oligogenic matrix protein, fibronectin 1, biglycan, collagen 11A1, laminin gamma 1, collagen 1A1, collagen 5A1, and collagen 5A2. Although MMSMCs showed an up-regulation of ECM proteins, the down-regulation of cell-cell adhesion in MMSMCs may explain other studies that have shown decrease in adhesion capacity in MMSMCs (24). It was shown that MM-derived SMCs showed less adhesive properties than normal SMCs when cell detachment was analyzed from substrates (culture dish) (24).

However, we also identified themes involved in development, suggesting that MMSMCs represent a progenitor stage of smooth muscle cells. Our GenMAPP analysis further demonstrates that MMSMCs are developmental immature relative to healthy SMCs. The over-expression of skeletal and muscle developmental genes in MMSMC provides a molecular explanation of their immature status. This finding is similar to other studies that show the importance of innervation on bladder development and the significance of the stretch-relaxation conditions needed for proper development (25). This is further validated by the under-expression of smooth muscle genes such as connexin 43, desmin, and genes involved in muscle contraction. However, the significance of this should not be over-interpreted because healthy SMC progenitors expanding in vitro, under-express many key SM contraction genes upon in vitro expansion.

It is unclear whether the increase in muscle mass in MM bladder is due to proliferation and/or hypertrophy. It has been shown that MMSMCs and normal
SMCs have similar growth properties up to 5 days after initial seeding (24;26). It was only when cells numbers were examined after 7 days of initial seeding, MMSMCs showed an approximately 2-fold increase in cell number (24). When genes were analyzed by GenMAPP, we identified a slight increase in the number of genes involved in growth and regulation of cell size (28 genes up MMC vs 25 genes that were up in normal and 16 genes vs 13 genes, respectively). This suggests that MMSMCs do not have a significant difference in growth and regulation of cell size as reported by Lin et al (24). In addition, the identification of genes up-regulated in MMSMCs in ‘tumor morphology’ processes, which consists of growth, metastasis, and regulation of cell size, may explain why there are more cells in MM bladders. This is consistent with in vitro experiments showing that MMSMCs have an increase in cell number only when grown at high confluency (24). The lack of contact inhibition is a property that is seen in tumor and stem cells and may contribute to the increase in muscle mass seen in MM bladders (24;27;28).

Patients who need reconstructed bladders have “unhealthy” cells and whether these cells could be expanded and have normal function has been a matter of discussion. Lin et al. reported that cultured MMSMCs have different growth, contraction, and cell-adhesion properties than normal SMCs, suggesting that cells from MM bladders may not be a suitable source for tissue engineering (24). Dozmorov et al. compared the expression profiles between SMCs derived from normal and neuropathic bladders using Altas Human Cancer 1.2, a commercial
nylon microarray (29). This array contains only 1,185 genes and was chosen because of its emphasis on cell cycle, apoptosis, signaling receptors and matrix-related gene. Their analysis only identified 18 differentially expressed genes between cultured normal and neuropathic bladder SMCs. The resulting set of differentially expressed genes was examined by pathway analysis and identified 3 significant pathways such as fibroblast growth factor, PTEN, and integrin signaling. Based on their findings Dozmorov et al. suggested that altered neuropathic bladder SMC phenotypes is maintained in the culture environments and that SMCs derived from diseased bladders may not be appropriate for tissue engineering purpose without modification of pathologically altered genes expression (29).

However, it has been demonstrated that tissue engineered SMCs from normal and diseased bladders have no phenotypic or functional differences (26). Furthermore, positive data (follow up time 22-61 moths) on seven myelomeningocele patients with poor bladder capacity who have been implanted with reconstructed bladders using their own cells were recently reported (30). Thus, SMCs from MM bladders still retain the capacity to expand, seed onto scaffolds, and be implanted into patients to improve bladder function. Although MM-derived SMCs could be tissue engineered to improve bladder capacity, these bladders are still not normal. We identified a difference in the genetic signatures between MMSMCs and normal SMCs. The identification of genetic signatures of diseased tissues could allow the re-engineering of diseased into
healthy tissues at the ex vivo expansion stage. The identification of an up-regulation of genes involved in development and ECM could lead to better protocols and culture conditions to induce myogenic differentiation and inhibit ECM production.

**Conclusion:**

This study provides a molecular evaluation of the quality and therapeutic potential of MMSMCs. Large-scale gene expression profiling identified an up-regulation of genes involved in development, growth, and focal adhesion and a down-regulation of genes involved in muscle contraction and neuronal transmission. Future studies will need to be performed in vitro to evaluate effective methods to differentiate MMSMCs into a more mature and better functioning phenotype.
Reference List


Ref Type: Journal (Full)


CHAPTER IV

THE EFFECTS OF ETHANOL ON AMNIOTIC FLUID-DERIVED STEM CELLS

Jennifer A. Hipp, Jason D. Hipp, Anthony Atala and
Shay Soker
Abstract:

Fetal Alcohol Spectrum Disorder (FASD) is a set of developmental defects caused by prenatal alcohol exposure. Clinical manifestations of FASD are highly variable and include mental retardation and developmental defects of the heart, kidney, muscle, skeleton, and craniofacial structures. Specific effects of ethanol on fetal cells include induction of apoptosis as well as inhibition of proliferation, differentiation, and migration. This complex set of responses suggests that a bioinformatics approach could clarify some of the pathways involved in these responses. In this study, the responses of fetal stem cells derived from the amniotic fluid (AFSCs) to treatment with ethanol have been examined. Large-scale transcriptome analysis of ethanol-treated AFSCs indicates that genes involved in skeletal development and ossification are up-regulated in these cells. Therefore, the effect of ethanol on osteogenic differentiation of AFSCs was studied. Exposure to ethanol during the first 48 hours of an osteogenic differentiation protocol increased in vitro calcium deposition by AFSCs and increased alkaline phosphatase activity. In contrast, ethanol treatment later in the differentiation protocol (day 8) had no significant effect on the activity of alkaline phosphatase. These results suggest that transient exposure of AFSCs to ethanol during early differentiation enhances osteogenic differentiation of the cells.
Introduction

Maternal alcohol (ethanol) consumption during pregnancy may cause abnormal growth and morphogenesis of the conceptus (1). The spectrum of defects caused by maternal alcohol consumption is known as Fetal Alcohol Spectrum Disorder (FASD) and occurs in approximately 0.03-0.15 percent of all live births in the United States (2). At the severe end of the spectrum of alcohol-induced defects is Fetal Alcohol Syndrome (FAS). Affected individuals typically have mild to moderate mental retardation, growth deficiencies, and craniofacial defects (3-5). Prenatal alcohol exposure can also result in defects in multiple organs and tissues, including the heart, eyes, kidneys, and skeleton (6-10).

Considering that, during embryonic and fetal life, various types of stem cells are prevalent and critical for normal development (11), and that interference with their viability or function may represent a little-explored, yet potentially significant teratogenic pathology, the current study has employed human amniotic fluid-derived stem cells (AFSCs) as a model to study the effects of ethanol on stem cell differentiation. AFSCs are multipotent and have extensive self-renewal potential. They are capable of differentiating into bone, muscle, fat, endothelium, liver, and neuron-like cells in vitro (12).

Ethanol has many reported teratogenic mechanisms of actions including induction of apoptosis and inhibition of proliferation, differentiation, migration and
other cellular functions (13-16). Additionally, ethanol exposure affects membrane-associated receptor signaling pathways (17) and cell adhesion (18;19), yields free radical-mediated damage (20), and alters the binding of transcription factors (21). Recognizing this wide range of cellular effects, initial examination of ethanol-mediated insult to AFSCs entailed global gene expression analysis. The results of this work indicate that genes involved in osteogenesis are consistently upregulated in AFSCs exposed to ethanol during early differentiation. An in vitro osteogenic differentiation protocol was subsequently employed to examine the potential effects of ethanol on this process. The data suggest that ethanol exposure during the uncommitted stage lineage restricts AFSCs to osteogenesis.
Materials and Methods

Ethanol Treatment

Human AFSCs were seeded at 3,000 cells/cm² and maintained in culture as described previously (12). They were grown in α-MEM medium (Gibco, Invitrogen, Carlsbad, CA) containing 15% ES-FBS, 1% glutamine, and 1% penicillin/streptomycin (Gibco), supplemented with 18% Chang B and 2% Chang C (Irvine Scientific, Santa Ana, CA). The cells were kept at 37 °C in an atmosphere containing 5% CO₂. For growth and viability studies, the AFSCs were treated with 25 mM, 50 mM, 75 mM, or 100 mM ethanol (Sigma-Aldrich, St. Louis, Missouri) and the plates were sealed with Parafilm. Media was replaced every 24 hours. The ethanol concentrations used in these experiments are equivalent to the blood alcohol concentrations achieved by social drinkers to chronic alcoholics (22;23). They have been utilized in previous in vitro experiments (20). Ethanol concentrations were measured spectrophotometrically using an Ethanol L3K assay (Diagnostic Chemicals Limited, Oxford CT) according to the manufacturer’s instructions.

Cell number and viability

Forty-eight hours after cell seeding, ethanol was added to the culture media and the plates were sealed with Parafilm to prevent evaporation. Untreated cells served as controls. Cell counts were made using a Coulter counter after 48 hours of exposure to the various ethanol concentrations. Cell viability was determined
by propidium iodine (PI) exclusion. Briefly, AFSCs were dissociated using a solution of 0.05% trypsin/EDTA (Gibco), centrifuged at 1,500 RPM for 5 minutes, and resuspended in 1 ml of PBS in 15 ml polypropylene tubes. 50 µl of PI, a nucleotide analogue, was added to each tube followed by incubation for 1 hour at room temperature. Cells were then centrifuged, washed twice with phosphate buffered saline and analyzed by flow cytometry using a FACSCalibur analyzer (BD Biosciences, San Jose, CA) and the FL-1 channel.

RNA Isolation and Real-Time Quantitative PCR
Total RNA was isolated from the AFSCs using the PerfectPure RNA Cultured Cell Kit (5 Prime, Gaithersburg, MD) according to the manufacturer's protocol. A DNA digestion step was included to eliminate contamination by genomic DNA. The quality of total RNA was assessed by measuring the A260/280 ratio using a spectrophotometer. For the reverse-transcriptase reaction, SuperscriptII reverse transcription reagents (Invitrogen, Carlsbad, CA) were used. Briefly, 1 mg of RNA was converted to cDNA, and PCR amplification of this DNA was performed using the TaqMan Universal Master Mix (Applied Biosystems). Reactions were performed in duplicate and consisted of 1 ml of cDNA, 1.25 ml probe, 12 ml Master Mix, and 10 ml DI water. These reactions were performed in a 96-well optical reaction plate (Applied Biosystems, Foster City, CA). Reactions were amplified and quantified using an ABI 7700 sequence detector and the manufacturer's software (Applied Biosystems, Foster City, CA). On demand fluorescent probes were produced for the β-actin and osteopontin genes. The
threshold cycle (Ct) indicates the fractional cycle number at which the amount of amplified target reaches a defined threshold. ΔCt was obtained by subtracting the Ct values of endogenous controls (β-actin) from the Ct values of the target genes.

**Microarray Analysis**

Two microarrays were performed on AFSCs cultured for 48 hours with or without 100 mM of ethanol. Fragmented antisense cRNA was used for hybridizing to human U133 A arrays (Affymetrix, Inc. Santa Clara, CA, USA) at the Core Genomic Facility of Wake Forest University School of Medicine. Raw CEL files were provided by the Microarray Core Facility of the Wake Forest University School of Medicine and were analyzed with AffyImGUI (Affymetrix LIMMA, Linear Models for Microarray Data, Graphical User Interfaces) (24;25). Within AffyImGUI, gene expression values were summarized with RMA. RMA adjusts for background noise, performs a quantile normalization, transforms the data into log base 2, and then summarizes the multiple probes into one intensity (26-28). Quantification of relative differences in gene expression among the groups of interest was accomplished using AffyImGUI (24;25). AffyImGUI reads the raw Affymetrix CEL files directly, summarizes the gene expression values using RMA, and then uses LIMMA to identify statistically significant differences in gene expression (29). LIMMA fits a linear model for every gene (similar to ANOVA or multiple regression analysis) and adjusts P values for multiple testings (29). Differentially expressed genes were identified as those with a fold change > 1.8.
The data discussed in this publication is deposited in NCBI's Gene Expression Omnibus (GEO) and will be accessible through GEO series accession number GSE13569, in accordance with MIAME standards.

**DAVID**

To uncover enriched processes, data sets were analyzed with DAVID (Database for Annotation, Visualization and Integrated Discovery), a web-based tool that provides statistical methods for identifying over-represented biological themes and pathways within diverse and disparate gene lists (30). DAVID also identifies over-represented biological themes in terms of their Gene Ontology (GO) terms and provides tools to visualize the distribution of genes on BioCarta and KEGG pathway maps. (31). GO provides consistent descriptions of genes in terms of biological processes and molecular function. Gene-enrichment analysis computes a modified Fisher exact p-value by comparing the ontological themes identified in our data set to total possible ontological processes present on the U133A chip. Those ontological processes that had a p-value of less than 0.05 were selected.

**Osteogenic Induction**

Human AFSCs were induced to differentiate into osteogenic cell types as described previously (12). As described in De Coppi et al, AFSCs develop into osteoblast-like morphology within 1 week of differentiation. By sixteen days, they form bone-like lamellar structures. Furthermore, AFSCs express mRNA and protein for alkaline phosphatase after one week of osteogenic differentiation.
Functional assays for calcium deposition show strong histological staining by alizarin red. They also show strong histochemical staining for alkaline phosphatase and secrete this enzyme (12). Briefly, AFSCs were cultured in low glucose DMEM containing 10% FBS and supplemented with 100nM dexamethasone (Sigma-Aldrich, St. Louise, MO), 10mM b-glycerophosphate (Sigma-Aldrich, St Louise, MO) and 0.05mM ascorbic acid-2-phosphate (Wako Chemicals, Irving, TX). Cells were grown to confluency and then treated with ethanol for the first 48 hours of osteogenic differentiation to control for the antiproliferative effects of ethanol. Cell number was determined after 24 and 48 hours of ethanol treatment and osteogenic differentiation. Ethanol had no effect on cell number when cells were near confluency (data not shown). Bone differentiation was analyzed by measurement of expression of osteopontin and bone-specific alkaline phosphatase mRNAs, measurement of alkaline phosphatase activity, and with Alizarin Red to quantify extracellular calcium deposition at Day 23 of differentiation.

**Alizarin Red Staining**

The presence of calcium in the cell cultures was determined by alizarin red (Sigma) staining at day 23 of osteogenic differentiation. Cells were fixed with 70% ethanol for 15 min. Fixed cells were incubated with 0.5% alizarin red solution in water (pH adjusted to 4.0) for 1 minute and then washed three times with deionized water and once with 70% ethanol, then allowed to dry. The alizarin red stain was extracted with 100 mM cetylpyridinium chloride (Sigma-
Aldrich, St Louise, MO) at room temperature for three hours. The absorbance of
the extracted alizarin red stain was measured at 540nm. The concentration of
alizarin red staining in the samples was determined by comparing the
absorbance values with those obtained from an alizarin red standard curve.

**Alkaline Phosphatase Activity**

Alkaline phosphatase activity was measured using a p-nitrophenyl phosphate
liquid substrate system (Sigma). Cells grown in 24-well plates were rinsed with
PBS and incubated with 0.15% Triton X-100 for 30 mins. Subsequently, 200 ml
of p-nitrophenyl phosphate solution were added to the Triton-X 100 solution.
Cells were incubated in the dark for 1 hour and absorbance was measured at
405 nm using a spectrophotometer.

**Data Analysis**

Results are expressed as mean ± standard deviation (S.D.) for quantitative data.
Analysis of Variance (ANOVA) was used to identify statistically significant
differences between groups. Alternatively, two-tailed tests of significance were
computed to determine relationships between ethanol-treated and control
groups. Statistical significance was defined as p < 0.05.
Results

Effects of ethanol on number and viability of AFSCs

When AFSCs were exposed to ethanol for 48 hours, a dose-dependent reduction in cell number was observed (Figure 1A). The dose range for these studies was chosen to reflect a physiologically relevant range of blood alcohol concentrations (from a “legal” blood alcohol level to levels found in chronic alcoholics) (22;23). The maximum concentration tested, 100 mM, caused a 33% reduction in the cell count when compared to control, whereas the 25 mM dose (equivalent to blood alcohol level of 0.12%) resulted in a 22% reduction in cell numbers. Although cell numbers continued to increase in the presence of ethanol, this increase was partially attenuated by ethanol exposure.

To determine if the ethanol-induced reduction in cell number was due to cell death, we examined the effect of ethanol exposure on cell viability using propidium iodide exclusion (Figure 1B). Cultures of AFSCs that had not been exposed to ethanol had an average baseline level of 12.4% ± 2.7% non-viable cells. The percentage of non-viable cells in cultures that were exposed to ethanol concentrations ranging from 25mM to 100mM had averages of 9.8% ± 1% to 11.4% ± 4% non-viable cells. These data indicate that ethanol does not have a significant effect on cell viability and suggests that the observed reduction in the number of AFSCs is a result of a reduced proliferation rate.
Figure 1. Effect of ethanol on cell proliferation and viability. AFSCs were cultured with 0, 25, 50, 75, and 100 mM of ethanol. (A) Cell counts were made after 48 hours of ethanol exposure and compared to counts of cultures that were not treated with ethanol. The results are expressed as percentages relative to cells without ethanol (dark grey bars). The values shown are the mean +/- standard deviation (SD, n=5) of three independent experiments (*, p<0.03 by student T-test). (B) The effect of ethanol on the percentage of non-viable cells was determined by 7-AAD and flow cytometry. Light grey bars indicate the percentage of non-viable cells in the presence of various concentrations of ethanol. The data shown represents the mean number of non-viable cells in 10,000 events from two independent experiments. Error bars represent SDs. Data were not significant. (C) Ethanol concentration in the media was measured by spectrophotometry. Results represent the mean +/- SD from three replicates.
Figure 1

A

Cell Number (Percent of Control)

Ethanol Concentration (mM)

B

Number of Apoptotic Cell (Percent of Control)

Ethanol Concentration (mM)

C

Ethanol Concentration (mM)

Time
The effect of ethanol on global gene expression

For large-scale transcriptome analysis, AFSCs were exposed to 100 mM ethanol for 48 hours in growth media. Ethanol concentration present in the culture media did not change over a period of 24 hours (Figure 1C). Standard growth media was used instead of a lineage-specific differentiation medium in order to prevent a bias toward identification of lineage-specific genes. To identify differentially expressed genes, Affymetrix GeneChips were used. The data sets were normalized and subjected to statistical analysis. Using this method, 65 genes that were up-regulated in response to ethanol and 19 genes that were down-regulated in response to ethanol were identified. To uncover enriched processes, data sets were analyzed by DAVID, a web-based tool that identifies over-represented biological themes in a data set based on their Gene Ontology (GO) terms. GO provides consistent descriptions of genes in terms of biological processes and molecular function.

DAVID analysis identified several processes that were up-regulated in response to ethanol, as compared to growth under normal conditions. These included processes involved in skeletal development (5 genes), ossification (5 genes), blood vessel development (4 genes), organ development (12 genes) and developmental processes (17 genes; Table 1). Up-regulated skeletal development genes included osteopontin, osteonectin, ectonucleotide pyrophosphatase, myocyte enhancer factor 2c, and matrix metallopeptidase 14. Osteopontin and osteonectin are secreted phosphoproteins expressed at the
early stage of osteogenic differentiation. They have been shown to mediate cell-
matrix interactions, cell adhesion, and differentiation (32-34). These results
suggest that ethanol exposure may predispose AFSCs towards an osteogenic
lineage.
Table 1. Enrichment of Biological Themes of Ethanol-responsive Genes.

Genes were selected based on the gene ontologies. Gene ontologies with a modified Fisher Exact \( P \)-value < 0.05 were selected. Processes that were identified in genes that were up-regulated in response to ethanol include skeletal development while genes that were down-regulated in response to ethanol included embryonic development.
Table 1

A. Gene ontologies of up-regulated genes in response to ethanol

<table>
<thead>
<tr>
<th>Gene Category</th>
<th>Count</th>
<th>P- value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Molecular function</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>biomineral &amp; ossification</td>
<td>5</td>
<td>2.87E-04</td>
</tr>
<tr>
<td>organ morphogenesis</td>
<td>7</td>
<td>0.002</td>
</tr>
<tr>
<td>organ development</td>
<td>12</td>
<td>0.002</td>
</tr>
<tr>
<td>reproductive developmental process</td>
<td>4</td>
<td>0.0034</td>
</tr>
<tr>
<td>skeletal development</td>
<td>5</td>
<td>0.006</td>
</tr>
<tr>
<td>system development</td>
<td>13</td>
<td>0.009</td>
</tr>
<tr>
<td>blood vessel morphogenesis</td>
<td>4</td>
<td>0.0167</td>
</tr>
<tr>
<td>tyrosine kinase signaling pathway</td>
<td>4</td>
<td>0.0252</td>
</tr>
<tr>
<td>IGF receptor signaling pathway</td>
<td>2</td>
<td>0.0289</td>
</tr>
<tr>
<td>developmental process</td>
<td>17</td>
<td>0.05</td>
</tr>
</tbody>
</table>

B. Gene ontologies of down-regulated genes in response to ethanol

<table>
<thead>
<tr>
<th>Gene Category</th>
<th>Count</th>
<th>P- value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Molecular function</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>negative regulation of cellular process</td>
<td>6</td>
<td>0.0024</td>
</tr>
<tr>
<td>circulatory system process</td>
<td>3</td>
<td>0.0098</td>
</tr>
<tr>
<td>glucose transport</td>
<td>2</td>
<td>0.0288</td>
</tr>
<tr>
<td>embryonic development</td>
<td>3</td>
<td>0.0295</td>
</tr>
<tr>
<td>multi-organism process</td>
<td>3</td>
<td>0.0334</td>
</tr>
<tr>
<td>anatomical structure development</td>
<td>6</td>
<td>0.036</td>
</tr>
<tr>
<td>regulation of biological quality</td>
<td>4</td>
<td>0.043</td>
</tr>
</tbody>
</table>
Genes that were down-regulated in response to ethanol were also organized by their gene ontology. The predominant pathway identified in this data set was embryonic development, which includes the genes encoding dickkopf homolog 1, neuregulin, and endothelin 1. Basic fibroblast growth factor (bFGF/FGF2) was also down-regulated in response to ethanol. bFGF is a potent mitogen and is an important factor in limb development and neurogenesis (35;36). The down-regulation of genes associated with these pathways suggests that ethanol restricts the differentiation potential of AFSCs and may interfere with proper embryonic and fetal development.

The effect of ethanol on osteopontin expression

Osteopontin is a phosphorylated glycoprotein that is secreted in the early stages of osteogenic differentiation. It is abundant in mineralized tissue and may be involved in bone formation (32;34;37). To characterize the effect of ethanol on osteopontin expression, real time-PCR was performed after 24 and 48 hours of exposure to 100 mM ethanol. Although control AFSCs expressed some osteopontin, ethanol-exposed AFSC showed a significant increase in osteopontin mRNA expression (Figure 2A). After 24 hours of ethanol exposure, osteopontin mRNA levels increased by 2.2-fold (p< 0.02). Exposure to ethanol for 48 hours increased the expression of osteopontin mRNA by 2.8-fold (p< 0.036). These results suggest that the ethanol-induced increase in osteopontin expression may restrict AFSCs to an osteogenic lineage.
Next, whether ethanol had similar effects on AFSCs when the cells were cultured in osteogenic differentiation media was examined. To determine whether ethanol interferes with the osteogenic differentiation process, the mRNA expression of osteopontin after a 24 hour exposure to 100 mM ethanol in osteogenic media was analyzed. Ethanol increased the expression of osteopontin by 22 fold in the presence of osteogenic media (Figure 2B). This suggests that ethanol induces osteopontin expression in both uncommitted AFSCs and AFSCs that are committed towards an osteogenic phenotype.
Figure 2. Effect of ethanol on OPN expression in AFSCs. (A) Real-time RT PCR analysis of AFSCs exposed to ethanol for 24 or 48 hours in growth media. CT values were determined from 3 independent experiments from two cell lines and ∆CT values were obtained by subtracting the CT values of β-actin. Mean fold change was determined from four independent experiment-pairs of control and ethanol-exposed AFSCs. Black columns indicate AFSCs without ethanol while grey columns indicate AFSCs exposed to 100 mM ethanol. P-values were determined using a one-tailed paired t-test with significance at 24 hours (P < 0.020, n=3) and 48 hours (P < 0.036, n=5). (B) Real-time PCR analysis of AFSCs exposed to ethanol for 24 hours in osteogenic media. CT values were determined from 3 independent experiments using two cell lines and ∆CT values were obtained by subtracting the CT values of β-actin. Mean fold change was determined from 3 independent experiment-pairs of control and ethanol-exposed AFSCs. P-values were determined using a one-tailed paired t-test with significance at 24 (P < 0.018).
Figure 2

A

![Bar chart showing fold change of OPN in Growth media and Growth media + EtOH after 24 hr and 48 hr.](chart_A)

B

![Bar chart showing fold change of OPN in Osteogenic media and Osteogenic media + EtOH after 24 hr.](chart_B)
The effect of ethanol on alkaline phosphatase activity

The effect of ethanol on alkaline phosphatase activity, which is an established marker of osteoblasts was also studied (38;39). It had previously been shown that AFSCs begin to express alkaline phosphatase activity after 8 days of osteogenic differentiation. To address whether ethanol exposure had an effect on alkaline phosphatase activity in differentiating AFSCs, they were exposed to 100 mM ethanol for 48 hours in osteogenic media, and then the ethanol was removed. Cells were allowed to differentiate further and were assayed for alkaline phosphatase activity at various time points (Figure 3). AFSCs that were not exposed to osteogenic media expressed a basal level of alkaline phosphatase activity. After 8 days of osteogenic differentiation, alkaline phosphatase activity in untreated cells rose above control levels and continued to increase until day 10 of osteogenic differentiation. However, ethanol exposure during the first 48 hours induced a small, but significant, increase in alkaline phosphatase activity at day 9 and 10. These results suggest that ethanol may have a persistent and enhancing effect on osteogenic differentiation that can be seen several days beyond the ethanol exposure period.
Figure 3. Effect of ethanol on alkaline phosphatase activity in AFSCs.

AFSCs were cultured with or without ethanol for the first 48 hours of osteogenic differentiation. Ethanol was removed and AFSCs continued to differentiate until days 7-10. At this time, they were assessed for alkaline phosphatase activity. Alkaline phosphatase activity was determined by spectrophotometric measurement of p-nitrophenol conversion. AFSCs exposed to ethanol showed a modest yet significant increase in alkaline phosphatase activity at day 9 and 10 of osteogenic differentiation (*, p < 0.001; ANOVA and two-tail T-test). The values shown are the mean +/- SD from at least ten replicate cultures and similar patterns were observed in a second cell line (data not shown). Differences between control and treated cultures were evaluated by t-tests. Open squares, osteogenic media+ EtOH (48 hours); Black diamonds, osteogenic media; black circles, growth media.
Figure 3

Alkaline Phosphatase Activity (Percent of Control)

Days in Culture

- Osteogenic media
- Osteogenic media + EtOH
- Growth media

* *
In order to determine whether the effects of ethanol on AFSCs differed depending on the stage of differentiation, AFSCs were cultured in osteogenic media for 8 days without ethanol. Then, beginning on day 8, the cells were treated for 48 hours with 100mM ethanol and the alkaline phosphatase activity was measured. The results showed no statistically significant effect (Figure 4). Collectively, these experiments suggest that ethanol exposure has a significant effect on alkaline phosphatase levels only when it occurs early in the differentiation process.
Figure 4. Alkaline phosphatase activity when AFSCs were exposed to ethanol at the midpoint of differentiation. AFSC were treated for 48 hours with 100 mM ethanol beginning at day 8 of osteogenic differentiation. Black columns indicate AFSCs that were not exposed to ethanol while grey bars indicate AFSCs that were exposed to ethanol. Ethanol exposure during the midpoint of differentiation had no significant effect on alkaline phosphatase activity ($t$ test $p<0.42$). The values shown are the mean +/- SD of twenty cultures.
The effect of ethanol on calcium deposition

Since transient ethanol exposure increased the expression and activation of genes involved in mineralization, the effect of ethanol on calcium deposition was examined. Cells were exposed to osteogenic media with or without 100 mM of ethanol for 48 hours. The ethanol was then removed and the AFSCs were allowed to terminally differentiate. At day 23, calcium deposition in these differentiated cultures was measured by Alizarin Red staining. The effects of ethanol treatment on calcium deposition are shown in Figure 5. AFSCs exposed to ethanol during the first 48 hours of osteogenic differentiation produced 155.06 ± 75.85 ug/mL of calcium, while non-exposed AFSCs produced 77.40 ± 26.85 ug/mL of calcium. Calcium deposition was also detected in non-differentiated AFSCs (52.5 ± 2.223 ug/ml). These data suggest that the effect of transient ethanol exposure on osteogenic genes during early differentiation is directly correlated with the AFSCs' ability to deposit calcium.
Figure 5. The effect of ethanol on calcium deposition. AFSCs were exposed to 100 mM ethanol during the first 48 hours of osteogenic differentiation. At day 23 of differentiation, AFSCs were stained for calcium deposition using alizarin red staining. AFSCs exposed for 48 hours to 100 mM ethanol had significantly increased calcium deposition when compared to controls (155.1 ± 75.8 µg/mL vs 77.4 ± 26.9 µg/mL). Calcium deposition was detected in non-differentiated AFSCs at a basal level of 53.3 ± 2.2 µg/ml. The values shown are the mean +/- SD (n=3-6) and similar results were confirmed in another cell line (data not shown). * One-tailed t-test of significance of p<0.006.
Discussion

This work has employed AFSCs as a unique model system for analysis of ethanol’s cellular effects. In this study, exposure of these cells, at their uncommitted stage, to teratogenic concentrations of ethanol has been shown to result in up-regulation of osteogenic genes. Additionally, in the presence of ethanol, the sequential expression of osteogenic markers as the cells progress from stem cells into those of an osteogenic lineage has been shown.

The effects of ethanol on a limited variety of stem cells have previously been reported, with some studies showing enhanced and some showing reduced differentiation potential. Neural stem (NSCs) and progenitor cells have been the most commonly studied. Ethanol has been shown to alter the differentiation potential of NSCs by enhancing astrocytic and oligodendrocytic differentiation and decreasing neuronal differentiation (40). Adult bone-marrow derived stem cells (BMSCs) have also been employed and, as opposed to the current study, inhibition of osteogenic differentiation has been shown by Gong et al. Another study, which used immortalized human fetal osteoblasts to examine the effect of ethanol on skeletal development by analysis of osteogenic gene expression demonstrated little or no effect of ethanol (41). As for the BMSCs, the response of these immortalized fetal osteoblasts may not reflect that of prenatal stem cells. Ethanol has been shown to enhance cartilage differentiation in embryonic limb mesenchyme cultures (42). While these cells seem to be lineage restricted because they spontaneously differentiate into chondrocytes, naive AFSCs are
not lineage restricted to osteogenesis. Thus, ethanol may act to lineage restrict AFSCs to osteogenesis by elevating the expression of osteogenesis-specific genes.

Clearly, gene array analysis is a powerful tool for exploring the pleotrophic effects of agents, or conditions, on gene expression. For the current study, it has allowed characterization of the global effects of ethanol on gene expression, as well as providing the potential to identify unknown processes and pathways. Although there are many techniques available for large scale evaluation of transcripts, Affymetrix GeneChips were selected because their proven reproducibility and the extensive number of genes included. To identify ethanol exposure-related changes in gene expression, analysis was performed following 48 hours of ethanol exposure. This duration was chosen in order to encompass the entire 36 hr AFSC cell cycle and to allow identification of the late responses of AFSC to ethanol (rather than responses to early oxidative stress). As determined by gene ontology analyses, ethanol exposure was shown to modulate unique pathways that pertain to bone development. This result is consistent with functional analyses that show a predisposition of ethanol-exposed AFSCs to differentiate into an osteogenic lineage.

The conditions necessary for the in vitro induction of osteogenic differentiation of AFSCs have previously been shown to include the presence of ascorbic acid, which is required for collagen synthesis (44); beta-glycerol phosphate, which
provides an organic phosphate for the formation of hydroxyapatite (45;46); and
dexamethasone, a glucocorticoid that induces transcription at the promoters of
osteogenic genes (47). However, these media additives have other functions
that are not limited to osteogenic differentiation. For example, dexamethasone
can also induce adipogenic differentiation in AFSCs (12). With previous reports
showing that, among ethanol’s mechanisms of action is induction of oxidative
stress, the presence in the culture media of ascorbic acid, which has antioxidant
activity, might be expected to negate the effects of ethanol. However, in the
presence of ethanol, a clear effect on induction of osteogenic gene expression in
AFSCs was shown.

Measures of alkaline phosphatase activity, which provided a marker of
osteogenic differentiation, were employed to assess the developmental stage-
dependency of the AFSC’s osteogenic response to ethanol. It was shown that
the stem cells are more susceptible to ethanol prior to and during early
differentiation than at later stages. While the reason for this is unknown, it is
possible that at early stages of cell differentiation, ethanol affects epigenetic
mechanisms, such as DNA methylation and histone methylation and acetylation.
Differentiation requires dynamic epigenetic modification, whereas self-renewal
requires epigenetic factors that preserve constant gene expression during stem
cell division. Although there have been recent studies reporting the effects of
ethanol on epigenetic mechanisms (43), the field is relatively new and much work
remains to be done. AFSCs may provide a useful model system in this regard.
Conclusion

Identification, in this study, of ethanol-induced osteogenic differentiation in AFSCs provides new information that may be of significance for understanding mechanisms involved in the genesis of FASD. These findings are consistent with the growing view that ethanol exposure alters the differentiation potential of stem cells. Although further studies are needed in order to determine whether ethanol enhances differentiation into other lineages, the decrease in the number of AFSCs and their increased sensitivity to differentiation cues suggest that ethanol acts as a differentiation agent. Importantly, AFSCs are uncommitted fetal-derived stem cells and can be obtained by non-invasive means early in pregnancy. Thus they offer a unique model system with significant promise for pertinent future studies. Acquiring a complete picture of the response of AFSCs to ethanol exposure also offers promise for identification of a biomarker of prenatal insult.
Reference List


CHAPTER V

TRANSCRIPTIONAL PROFILING OF AMNIOTIC FLUID-DERIVED STEM CELLS IDENTIFY MULTILINEAGE PROGRAMS

Jennifer A Hipp, Jason D Hipp, Anthony Atala and Shay Soker
Abstract

Understanding the genes that govern the special properties of stem cells has implications for both stem cell biology and regenerative medicine. Efforts to characterize genome-wide expression profiles of stem cells are biased towards the identification of genes that display the most pronounced differential expression. Here we describe a novel method of *in silico* analysis to detect low levels of differential gene expression in amniotic fluid-derived stem cells (AFSCs) and at distinct time points of osteogenic differentiation. The experimental design consisted of direct (conventional) and indirect comparisons. For the indirect comparison, AFSCs and their derivatives were compared to tissue-specific reference cell types to mask particular signatures while revealing others. We show that undifferentiated AFSCs express programs involved in skeletal development, nervous system development, muscle development and blood vessel development. This data suggests that AFSCs express multiple lineage-specific programs. To understand how these multiple lineage-specific programs change during differentiation, microarray analysis was performed at day 20 and day 30 of osteogenic differentiation. As AFSCs differentiate into an osteogenic lineage, the osteogenic program is maintained however, processes associated with nervous system development, muscle development and blood vessel development were reduced to lower levels. The expression of multiple lineage programs prior to differentiation suggests that AFSCs are primed to differentiate into these lineages. This data suggests that rather than being viewed as a
“blank” state, stem cells express multiple lineage-specific genes. We propose that stem cell differentiation requires both up-regulation of a lineage-specific program and the elimination of other lineages.
Introduction

The recent discovery of a stem cell population in the amniotic fluid offers a promising source of stem cells for cell-based therapy (De, Bartsch, Jr. et al., 2007a). AFSCs exhibit extensive self-renewal and multilineage differentiation potential. AFSCs represent approximately 1% of the cells found in the amniotic fluid. These cells are immuno-selected for the expression of surface antigen C-KIT (CD117), express embryonic markers such as OCT4 and SSEA4, but do not form teratomas when injected into SCID mice. They also express mesenchymal and/or neuronal markers such as CD29, CD44, CD73, CD90, CD105 and give rise to adipogenic, osteogenic, myogenic, endothelial, neurogenic and hepatogenic lineages in vitro. Studies showed that AFSCs incorporate into kidney and lung epithelium in vivo and further demonstrates their broad multipotentiality (Carraro, Perin et al., 2008; Perin, Giuliani et al., 2007).

However, the genetic signature of AFSCs is not well known and the mechanism regulating their differentiation is not completely understood. Before AFSCs can be used as any type of clinical therapy, large-scale transcriptional analysis to profile undifferentiated AFSCs and their derivatives are essential.

Microarray analysis allows one to measure the expression levels of thousands of known and unknown genes. The design of a microarray experiment depends on the biological question. Studies that characterize the progression of stem cells into specific lineages can be divided into two types of analysis; direct and indirect
comparisons. The direct comparison, for example, compares sample A to sample B, and identifies genes that are differentially expressed between two samples (Yang and Speed, 2002). This approach is biased towards the identification of genes that display the most pronounced differential expression. Genes that are expressed but not drastically changed are removed from further analysis. The reference design, or indirect comparison, is used because gene expression values can be meaningful in a relative sense (Kerr and Churchill, 2001). The indirect comparison compares microarray data of AFSCs to those of a reference cell type. Since genes that are common in both samples are mutually cancelled, we hypothesize that “normalizing” AFSC to a reference cell type can identify specific components of AFSCs. Robust multi-chip average (RMA) algorithm was used to generate reference-defined global gene expression profiles of AFSCs by comparing microarray data of AFSCs to those of specific reference cell types. By generating reference-defined global gene expression profiles of AFSCs, we found that undifferentiated AFSCs express multiple lineage-specific programs. Reference-defined global gene expression profiles of AFSCs and their osteogenic derivatives showed the skeletal development program as highly expressed throughout differentiation while other tissue-specific programs are down-regulated. The significance of this microarray study is that it challenges the “stem cell is a blank cell” concept. This study also demonstrates that multiple in silico comparisons are needed to define a stem cell phenotype.
Methods and Materials:

AFSC culture

Human AFSCs were seeded at 3,000 cells/cm² and maintained in culture as described previously (De, Bartsch, Jr. et al., 2007b). They were grown in α-MEM medium (Gibco, Invitrogen, Carlsbad, CA) containing 15% ES-FBS, 1% glutamine, and 1% penicillin/streptomycin (Gibco), supplemented with 18% Chang B and 2% Chang C (Irvine Scientific, Santa Ana, CA). The cells were kept at 37 °C in an atmosphere containing 5% CO₂.

Osteogenic Induction

Human AFSCs were induced to differentiate into osteogenic cell types as described previously (De, Bartsch, Jr. et al., 2007b). As described in De Coppi et al, AFSCs develop into osteoblast-like morphology within 1 week of differentiation. By sixteen days, they form bone-like lamellar structures. Furthermore, AFSCs express mRNA and protein for alkaline phosphatase after one week of osteogenic differentiation. Functional assays for calcium deposition show strong histological staining by alizarin red. They also show strong histochemical staining for alkaline phosphatase and secrete this enzyme (De, Bartsch, Jr. et al., 2007b). Briefly, AFSCs were cultured in low glucose DMEM containing 10% FBS and supplemented with 100nM dexamethasone (Sigma-Aldrich, St. Louise, MO), 10mM b-glycerophosphate (Sigma-Aldrich, St Louise, MO) and 0.05mM ascorbic acid-2-phosphate (Wako Chemicals, Irving, TX).
**Affymetrix GeneChip analysis:**

Total RNA was isolated from AFSCs, AFSCs after 20 days of osteogenic differentiation, and AFSCs after 30 days of osteogenic differentiation. 9 chips RNA from three independent lines of AFSCs, each with 3 biological replicates were hybridized to 9 chips. Biological replicates of AFSC after 20 and 30 days of differentiation were hybridized to 2 chips for each timepoint. Fragmented antisense cRNA was used for hybridizing to human U133 A arrays (Affymetrix, Inc. Santa Clara, CA, USA) at the Core Genomic Facility of Wake Forest University School of Medicine. Raw CEL files were provided by the Microarray Core Facility of the Wake Forest University School of Medicine and were analyzed with AffylmGUI (Affymetrix LIMMA, Linear Models for Microarray Data, Graphical User Interfaces) (Wettenhull, Simpson et al., 2006; Wettenhall and Smyth, 2004). The demographics of the publicly available Genechip data sets are human fibroblast (n=8, GSE3860), human ESC (n=3, (Sato, Sanjuan et al., 2003)), human HSC (n=3, GSE3823), and human UVECs (n=4, GSE2639) and were downloaded from the NCBI’s Gene expression Omnibus.

**Pre-processing Analysis:**

Within AffylmGUI, gene expression values were summarized with RMA. RMA adjusts for background noise, performs quantile normalization, transforms the data into log based 2, and then summarizes probes from multiple arrays into one intensity (Bolstad, Irizarry et al., 2003;Irizarry, Bolstad et al., 2003;Irizarry, Hobbs et al., 2003). Quantification of relative differences in gene expression among
the groups of interest was accomplished using AffylmGUI, the sister package of limmaGUI (Wettenhall, Simpson et al., 2006; Wettenhall and Smyth, 2004). AffylmGUI reads the raw Affymetrix CEL files directly, summarizes the gene expression values using RMA, and then uses LIMMA to identify statistically significant differences in gene expression (Smyth, 2004). LIMMA fits a linear model for every gene (like ANOVA or multiple regression analysis), and adjusts P-values for multiple testings (Smyth, 2004). Differentially expressed genes were identified with a B statistic > 0. The B statistic, also known as a likelihood of odds (LOD) score is a moderated t-statistic with posterior residual standard deviations.

DAVID
To uncover enriched processes, data sets were analyzed with DAVID (Database for Annotation, Visualization and Integrated Discovery), a web-based tool that provides statistical methods for identifying over-represented biological themes and pathways within diverse and disparate gene lists (Dennis, Jr., Sherman et al., 2003). DAVID also identifies over-represented biological themes in terms of their Gene Ontology (GO) terms and provides tools to visualize the distribution of genes on BioCarta and KEGG pathway maps. (Ashburner, Ball et al., 2000). GO provides consistent descriptions of genes in terms of biological processes and molecular function. Gene-enrichment analysis computes a modified Fisher exact p-value by comparing the ontological themes identified in our data set to total
possible ontological processes present on the U133A chip. Those ontological processes that had a p-value of less than 0.05 were selected.

Results:

The experimental design consists of direct and indirect comparisons using robust multi-chip average (RMA) algorithm (Figure 1). RMA identifies differentially expressed genes between two samples and performs background correction, quantile normalization, and generates RMA computed expression values (Bolstad, Irizarry et al., 2003; Irizarry, Bolstad et al., 2003; Irizarry, Hobbs et al., 2003). We began our analysis by directly comparing RMA computed values of undifferentiated AFSCs to RMA computed values of AFSCs at day 20 and day 30 of osteogenic differentiation. In the next section, RMA computed values of AFSC were compared to RMA computed values of a reference cell type to generate a reference-define gene expression profile of AFSCs. The last section, we generated a reference-defined global gene expression profile of osteogenic AFSC after 20 and 30 days of differentiation.
Figure 1. Experimental design of microarray experiment. The direct comparison identifies differentially expressed genes between AFSC and their osteogenic derivatives. The indirect comparison compares the AFSC and their osteogenic derivatives to a common reference cell type and generates reference-defined global gene expression profile.
Section 1: Direct comparisons

For the direct comparison, RMA computed values of undifferentiated AFSCs were directly compared to RMA computed values of AFSCs at day 20 and day 30 of osteogenic differentiation. We identified 500 genes that were up-regulated upon 20 days of osteogenic differentiation (Figure 2). To uncover enriched processes, the datasets were analyzed by DAVID, a web-based tool that identified over-represented biological themes based on their Gene Ontology (GO) terms. GO analysis provides consistent descriptions of genes in terms of biological processes and molecular functions. Enriched processes that were up-regulated in this dataset identified skeletal development and muscle development. Specific genes within skeletal development include aggrecan, matrix gla protein, insulin growth factor 2, collagen type 1 alpha 2, osteopontin, and vitamin D receptor. On the other hand, 549 genes were down-regulated in this comparison. Down-regulated genes were clustered with DAVID using their GO terms revealed prominent decreases in biological pathways involved in epidermis development and embryonic development.

The next comparison identified genes that were up-regulated after 30 days of osteogenic differentiation. Gene enrichment analysis of osteogenic differentiation at day 30 was similar to those identified at osteogenic day 20. Gene enrichment analysis of this dataset, consisting of 752 genes, identified processes involved with skeletal development, muscle development and immune
response. Over-represented biological processes of down-regulated genes identified processes involved in epidermis development and embryonic development.

In the direct analysis, we identified processes that were both up-regulated and down-regulated during osteogenic differentiation. Only processes that were unique to one direction are included in Table 1. However, nervous system development and blood vessel development were detected as being both up-and down-regulated upon osteogenic differentiation. Neurogenic genes that were up-regulated included neuropilin 2, EphrinB2, Connexin43. While these genes are expressed during the developing nervous system, these genes have other functions in non-neuronal tissues. While the neurogenic genes that were down-regulated, neuregulin 1, nestin, brain-derived neurotrophic factor, are more specific to neurogenesis. Blood vessel development genes that were up-regulated during osteogenic differentiation include angiogenin and angiopoietin 1 while blood vessel development genes that were down-regulated included vascular growth factor c, nexin, and tumor necrosis factor receptor 12A. Because we identified a down-regulation of processes involved in blood vessel development, neurogenesis, this suggests that undifferentiated AFSC contains these lineage programs. To better understand how these lineage programs change during differentiation, a more detail characterization of undifferentiated AFSCs is discussed in the next section.
Table 1. Ontologies of genes that were up- and down-regulated during osteogenic differentiation of AFSC by direct comparisons. Enrichment analysis of up-regulated genes involved in skeletal and muscle development while a down-regulation of genes involved in embryonic development.

<table>
<thead>
<tr>
<th></th>
<th>Osteogenic Differentiation at Day 20</th>
<th>Osteogenic Differentiation at Day 30</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Up-regulated processes</strong></td>
<td>Skeletal development</td>
<td>Skeletal development</td>
</tr>
<tr>
<td></td>
<td>Muscle development</td>
<td>Muscle development</td>
</tr>
<tr>
<td></td>
<td>Lung development</td>
<td>Immune response</td>
</tr>
<tr>
<td></td>
<td>Immune response</td>
<td></td>
</tr>
<tr>
<td><strong>Down-regulated processes</strong></td>
<td>Epidermis development</td>
<td>Epidermis development</td>
</tr>
<tr>
<td></td>
<td>Embryonic development</td>
<td>embryonic development</td>
</tr>
</tbody>
</table>
Indirect comparison: Defining the AFSC phenotype

The indirect comparison compares RMA computed values of AFSCs to RMA computed values of a reference cell type. Reference cell types were chosen based on their phenotype, type of array, and their public availability and consists of fibroblasts, human embryonic stem cells (ESCs), human umbilical vein endothelial cells (hUVECs), and hematopoietic stem cells (HSCs). This comparison generates a reference-defined global gene expression profile of AFSC. Since genes that are common in both samples are mutually cancelled, we hypothesize that “normalizing” AFSC to a reference cell type can mask specific components of AFSCs. This reference-defined global gene expression profile was divided into two groups. The first group consists of genes with a fold change greater than 2 in AFSC (relative to a reference cell type). The second group consists of the top 100 most up-regulated genes with respect to the reference cell type and is a subset of the first group.

We began by comparing RMA computed values of AFSC to RMA computed values of fibroblast to generate their fibroblast-defined global gene expression profile. Since this comparison identified 82 genes with a fold change greater than 2 in AFSC with respect to fibroblasts, these genes are considered to be highly expressed relative to fibroblasts. Enrichment analysis of these genes identified processes involved in skeletal development, neurogenesis, as well as lung development and epidermal/ectodermal development (Table 2).
To normalize AFSC with respect to ESC, RMA computed values of AFSC were compared to RMA computed values of ESC. This generated an ESC-defined global expression profile of AFSC and contains a diverse set of multilineage programs. Enrichment analysis of the top 100 most up-regulated genes with respect to ESC identified skeletal development, neurogenesis, kidney development and epidermis development. In addition, these data sets was also enriched with mesenchymal processes associated with blood vessel development, muscle development and wound healing. Enrichment analysis of genes with a fold change greater than 2 include processes identified above as well as processes involved in heart development, epidermal development, cartilage development and immune response.

AFSC were then normalized to HSC by comparing the RMA computed values of AFSC to RMA computed values of HSC. The top 100 most up-regulated genes indentified in the HSC-defined global gene expression profile of AFSC were enriched with blood vessel development, skeletal development, neurogenesis, muscle, and lung development. When genes with a fold change greater than 2 were analyzed, pathways involved in heart development and cartilage development were identified. It’s important to note that processes involved in immune response were mutually cancelled in the HSC-defined global gene expression profile of AFSC.
Since we identified processes involved in blood vessel development, AFSC were normalized with respect to hUVECs. The top 100 most up-regulated genes were enriched with programs involved in skeletal development, neurogenesis, muscle development and immune response. However, data sets containing genes with a fold change greater than 2 were enriched with blood vessel development, heart development, kidney/urogenital development and cartilage development. Although this comparison was not able to completely remove the blood vessel development process, we were able to reduce it.

Together, these results provide an overall view of the genetic programs expressed in undifferentiated AFSCs. Skeletal development and nervous system development were identified in all comparisons and thus are considered as major components of AFSC. Processes that are identified in at least two comparisons, such as epidermis development and lung development, might be considered as minor components of AFSCs.
Table 2

<table>
<thead>
<tr>
<th>Fibroblast</th>
<th>hES</th>
<th>HSC</th>
<th>hUVECs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>+++</td>
<td>+</td>
<td>Blood vessel development</td>
</tr>
<tr>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>Skeletal/Limb development</td>
</tr>
<tr>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>Neurogenesis</td>
</tr>
<tr>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td>Muscle development</td>
</tr>
<tr>
<td>+++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Wound healing</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
<td>Heart development</td>
</tr>
<tr>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td></td>
<td>Immune</td>
</tr>
<tr>
<td>+++</td>
<td>+++</td>
<td>+++</td>
<td></td>
<td>Lung</td>
</tr>
<tr>
<td>+++</td>
<td>+</td>
<td></td>
<td></td>
<td>Epidermis/ectoderm</td>
</tr>
<tr>
<td>+++</td>
<td>+</td>
<td></td>
<td></td>
<td>Kidney/urogenital</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>Cartilage</td>
</tr>
</tbody>
</table>

Table 2. Enrichment analysis of multiple reference-define global gene expression profiles of undifferentiated AFSCs. Data sets were divided into two groups. Genes with the highest fold change with respect to each reference file are expressed as (+++). The second data set consists of genes with a fold change greater than two (relative to the reference cell type) are denoted as (+).
Indirect comparisons: Changes in lineage-specific programs during osteogenic differentiation.

In this section, we generated a reference-defined global gene expression profile of osteogenic AFSC after 20 and 30 days of differentiation. Like in the previous section, reference-defined global gene expression profile was divided into two groups for enrichment analysis; the top 100 most up-regulated genes and genes with a fold change greater than 2 in AFSCs (relative to a reference cell type).

We compared the fibroblast-defined global gene expression profiles of undifferentiated AFSC, AFSC after 20 days of osteogenic differentiation, and AFSC after 30 days of osteogenic differentiation. Enrichment analysis was performed at each stage of differentiation using both groups of genes identified in the fibroblast-defined global gene expression profile. The skeletal development process was highly expressed in undifferentiated AFSC and maintained across differentiation. Neurogenesis was highly expressed in AFSC and down-regulated upon differentiation to low levels. We also identified processes involved in immune response as up-regulated upon differentiation. This was unexpected because we do expect AFSC to differentiate into these lineages.

We compared ESC-defined gene expression profile of AFSC to an ESC-defined gene expression profile of AFSCs after 20 and 30 days of differentiation. This comparison also identified a diverse set of tissue-specific lineages. As cells
differentiate, skeletal development is highly enriched at day 20 however it
decreases to low levels of expression at day 30. Cartilage development peaked
at day 20 and returns to low levels of expression. Nervous system development
was enriched at the undifferentiated stage but decreases at day 20 and 30.
Blood vessels development was expressed at low levels throughout
differentiation. Interestingly, we identified an up-regulation of processes
associated with immune response.

We compared HSC-defined global gene expression profile of undifferentiated
AFSCs to an HSC-defined global gene expression profile of AFSC after 20 and
30 days of osteogenic differentiation. Because AFSCs were normalized to
HSCs, the immune-related components expressed in AFSC and their derivatives
were reduced but other programs such as blood vessel development were
identified. In undifferentiated AFSCs, muscle development was highly expressed
but was down-regulated upon differentiation. In contrast, skeletal development
was moderately expressed in AFSC and peaked at Day 20 of osteogenic
differentiation and returned to moderate expression at day 30. Blood vessel
development was moderately expressed in AFSC and up-regulated at day 20
and 30.

We compared the hUVEC-defined global gene expression profiles of AFSC and
osteogenic AFSC. The skeletal development program was highly expressed at
all stages of osteogenic differentiation. The cartilage development program was
moderately expressed in AFSC and peaked at day 20. The neurogenic program was highly expressed in AFSC and down-regulated upon differentiation. Similarly, the blood vessel development program was moderately expressed in AFSC and day 20 but was not detected at day 30. Expectedly, normalizing AFSC to hUVECS reduced the blood vessel development program that was expressed throughout osteogenic differentiation.
Figure 2: Enrichment analysis of reference-define global gene expression profiles during osteogenic differentiation. Fibroblast-define global gene expression profiles during osteogenic differentiation shows skeletal development as highly expressed throughout differentiation while programs involved in nervous system development and lung development were decreased (A). The cartilage program peaked at day 20 but returned to reduced levels at day 30. ESC-defined expression profiles during osteogenic differentiation showed skeletal development was highly expressed in undifferentiated AFSCs but decreased at day 30 of osteogenic differentiation (B). Like the fibroblast comparison, nervous system development program decrease upon differentiation and the cartilage program peaked at day 20. Other lineages, such as heart development and blood vessel development, were down-regulated or maintain low expression throughout differentiation. HSC-defined gene expression profiles during osteogenic differentiation identified skeletal and cartilage programs peaking at day 20 while processes involved in blood vessel development, nervous system development, and immune response were up-regulated at day 30 (C). hUVEC-define global gene expression profiles showed skeletal development highly expressed throughout differentiation, cartilage peaking at day 20, and the down-regulated of programs involved in nervous system development, muscle development, and blood vessel development (D).
Figure 2

Osteogenic differentiation of AFSC with reference to fibroblast

Timepoints during osteogenic differentiation

AFSC Day 20 Day 30

Levels of gene expression (arbitrary units)

Osteogenic differentiation of AFSC with reference to ESC

Timepoints during osteogenic differentiation

AFSC Day 20 Day 30

Levels of gene expression (arbitrary units)

Osteogenic differentiation of AFSC with reference to HSCs

Timepoints during osteogenic differentiation

AFSC Day 20 Day 30

Levels of expression (arbitrary units)

Osteogenic differentiation of AFSC with reference to hUVECs

Timepoints during osteogenic differentiation

AFSC Day 20 Day 30

Levels of gene expression (arbitrary units)
Discussion:

In this report, we examined the global gene expression profile of AFSC and their osteogenic differentiated progeny. Our analysis revealed that undifferentiated AFSCs express multilineage programs. Our *in silico* dissection of the global gene expression profile of AFSCs consistently identified processes involved in skeletal and nervous system development among each comparison and are consider as major components of AFSCs. Processes that were identified in at least two, such as blood vessel development, muscle development, lung development, kidney development and cartilage development, might be considered as minor components of AFSCs. Because it has been shown that AFSC can differentiate into these lineages, the expression of multiple lineage programs prior to differentiation suggests that AFSC are primed to differentiate into these lineages.

A simple microarray experiment looks for changes in gene expression at distinct timepoints of differentiation. The direct comparison identifies the most differentially expressed between two samples however, genes that are expressed but have subtle changes in gene expression are ignored. On the other hand, microarrays can be use to define the molecular phenotype of a cell. Conventional analysis uses MAS5 (Microarray Suite version 5.0) algorithm for generating lists of genes that are present or absent but has been criticized for high false positive rates from exaggerated variances of genes with low levels of
expression (Pepper, Saunders et al., 2007). The reference design, or indirect comparison, is used because gene expression values can be meaningful in a relative sense (Kerr and Churchill, 2001). Comparing AFSC to a reference file identifies genes that are over-expressed in AFSC with reference to a cell type. Genes that are over-expressed in AFSC could be considered as a genetic signature of AFSCs. An alternative to MAS is Robust Multi-chip Average (RMA). Unlike MAS5, RMA identifies differentially expressed genes by comparing two cell types and reduces the variances of low intensity genes. Another feature of RMA is that it identifies genes that are differentially expressed relative to a cell type, meaning genes that are common in both are mutually cancelled. To take advantage of this feature, biologically relevant comparisons were used to purpose subtract out particular phenotypes that are expressed in AFSCs. For example, since AFSC have been shown to differentiate into endothelial cells, we hypothesize that AFSC expressed genes involved in blood vessel development. If we compare the genetic profiles of AFSCs to endothelial cells, we expect to subtract genes involved in blood vessel development. We further hypothesize that by masking particular signatures can identify other signatures.

Since AFSC have the potential to differentiate into multiple cell types, we expect AFSC to express programs for multiple lineages. The identification of programs involved in lung, kidney and cartilage development in our analysis is also consistent with recent publications showing AFSC’s ability to differentiate into lung epithelium, kidney structures, and cartilage cells (Carraro, Perin et al.,
The identification of processes involved in heart development and epidermal development suggest that AFSC have the potential to differentiate into these lineages but further investigations are needed. The identification of multiple lineage-specific transcripts in AFSC is consistent with other stem cells including hES, hES-derived hemangioblast cells, and MSCs (Efroni, Melcer et al., 2009; Golan-Mashiach, Dazard et al., 2005; Lu, Hipp et al., 2007; Tremain, Korkko et al., 2001). Promiscuous genes have also been observed in the hematopoietic system, where expression of genes of multiple lineages was detected prior to commitment (Hu, Krause et al., 1997). It is not likely that the identification of multiple lineage-specific genes in undifferentiated AFSC is due to heterogeneous cell types within the AFSC population. Although the amniotic fluid contains a heterogeneous population of cells, single-cell dilutions and retroviral markings were used to verify that clonal AFSC can give rise to each lineage.

Clearly, gene array analysis is a powerful tool for identifying genes that expressed in stem cells. Considerable efforts have been made to define “stemness,” a set of genes that are commonly expressed in multiple stem cell types, to provide insight underlying self-renewal and the ability to differentiate (Ramalho-Santos, Yoon et al., 2002). However, ascribing stem cell functions from gene lists is analogous to taking a car apart and explaining how it works. This approach focuses on the genes themselves rather than deciphering potential mechanisms of stem cell differentiation.
Microarrays can be used to explore changes in gene expression during stem cell differentiation. Profiling stem cell differentiation in a lineage-specific and temporal-dependent manner may enable one to dissect the genetic wiring of differentiation. This approach has been widely used to characterize the changes in gene expression during embryonic stem cell differentiation and mesenchymal stem cell differentiation (Cao, Wagner et al., 2008; Song, Webb et al., 2006).

However, the experimental design of the microarrays is different between these experiments. In the ESC study, ESC and ESC-derived cardiomyocytes were compared to a pooled reference file while the MSC were compared between undifferentiated and differentiated cells. Interestingly, it is not common to find both direct and indirect comparisons in the same experiment. We believe that both comparisons are valuable because each comparison provides a unique perspective of stem cell differentiation.

The reference design was used to compare ESC profiles to a mixture of RNA obtained from various adult cells (universal RNA) (Bhattacharya, Miura et al., 2004). This reference point is supposed to represent a signature of differentiated adult cells. Since microarrays assess the total RNA with a sample and heterogeneous samples from mixtures may influenced by the predominating cell type, we prefer to use homogeneous reference materials. Furthermore, since reference materials can be designed to have properties in common to the sample in question and RMA ignores genes that are expressed between two samples, we purposely compared AFSC to specific reference files to generate an
expression profile (2003). Another study compared the ES profile to a single differentiated cell type, keratinocytes to identify genes that are unique for ESC (Bhattacharya, Miura et al., 2004; Golan-Mashiach, Dazard et al., 2005).

Likewise, we compared AFSC to a single cell type, fibroblasts. This comparison identified a few developmental processes and we hypothesized that AFSC expressed a signature with multiple lineages because they are multipotent.

To get a better view of the molecular changes that occur during stem cell differentiation, we evaluated the changes of lineage-specific programs in undifferentiated AFSC and their transition into osteogenic cells. Our first analysis directly compared AFSC to their differentiated progeny, providing a kinetic-like relationship of gene expression. When comparing undifferentiated AFSC to AFSC that were differentiated into osteogenic cells for 30 days, enriched processes involved in skeletal development and muscle development was up-regulated. However, the direct analysis identified processes involved in nervous system development and blood vessel development as being both up-and down-regulated upon osteogenic differentiation. This can occur because each up- or down-regulated data set consists of different genes encoding the same developmental process. Neurogenic genes that were up-regulated included neuropilin 2, EphrinB2, Connexin43. These genes are involved in the orientation of projecting axons while connexin 43 is a major gap junction protein present in osteoblast (Giger, Urquhart et al., 1998; Lecanda, Warlow et al., 2000; Zhou, Suh et al., 2001).
When transcriptional profiles were compared to fibroblasts, skeletal development process was highly expressed in undifferentiated AFSC and maintained across differentiation while genes involved in nervous system development was highly expressed in AFSC and down-regulated upon differentiation. Our results suggest that multipotentiality is the expression of genes that represent multiple lineage-specific programs while differentiation is the elimination of other lineages. This dissection method characterized the levels of multiple lineage programs as stem cells transition from undifferentiated cells into an osteogenic lineage. The identification of multi-lineage programs after 30 days of osteogenic differentiation genes may indicate their ability to differentiate into these lineages. It has been recently shown that transdifferentiation can occur after osteogenic differentiation of MSCs (Song and Tuan, 2004). It has also been shown that a mature adipocyte could convert into a fully differentiated osteoblast (Oki, Watanabe et al., 2008). On the other hand, a different interpretation of our results may suggest that differentiated cells expresses genes to recruit neuronal input and blood vessels.

We unexpectedly identified processes involved in immune response as up-regulated upon osteogenic differentiation. Our osteogenic differentiation protocol has been well-characterized by the up-regulation of alkaline phosphatase activity during midpoint of differentiation and by the deposition of calcium (De, Bartsch, Jr. et al., 2007a). Calcium staining also demonstrated that each cell was positive for calcium deposition after 30 days of differentiation (data not shown). This
demonstrates that our differentiation protocol produces homogeneous and functional osteocytes. So it is unlikely that our differentiation protocol resulted in heterogeneous cell types. Another study looked at the gene expression signatures of MLO-Y4, a well-known *in vitro* osteocyte model, and also identified processes involved in acute/defense response pathways and interferon/chemokine pathways when analyzed by DAVID (Yang, Harris *et al.*, 2009). The consistent identification if immune response processes in osteogenic cells may be a limitation of the analysis or a byproduct of the osteogenic differentiation.

Since microarrays assess the RNA in a sample population, it is not clear whether genetic signature of a single cell is similar to that of the entire population. Single cell expression profiles of AFSC may explain whether these genetic signatures derived from single cell differ from those derived from a homogeneous population. Single cell analysis of MSC express mRNA of multiple cell lineages but its not determine whether single cells are express similar profiles to those of clonal derived MSC. It has been shown that even in clonal populations of cells, there is significant phenotypic variation from cell to cell (Chang, Hemberg *et al.*, 2008). Microarray analysis of an “outlier” for sca-1 showed a unique profile that reverted back to original profile of the clonal population. This suggests that the molecular heterogeniety is not due to gene expression noise but due to the weakly stable states of a slowing fluctuating transcriptome.
Our results suggest that multipotentiality is the expression of genes that represent multiple lineage-specific programs while differentiation is the elimination of other lineages. Thus, the molecular basis of stem cells appears to entail a promiscuous gene expression pattern. Understanding the molecular signature of multipotentiality will enable the control and direction of differentiation into particular phenotypes.


Stem cells are an ideal resource for regenerative medicine. Despite the knowledge gained by stem cells’ ability to differentiate into multiple lineages, very little is known about the genes that govern the special properties of stem cells. Stem cells have been previously thought of as a “blank” cell and the activation of lineage-specific programs indicates commitment into a particular lineage. On the other hand, stem cells may express low level of genes associated with multiple lineages to allow a rapid up-regulation of a single lineage program when cells differentiate into a particular lineage (Orkin, 2003). Microarray analysis can provide insights into the genetic programs expressed in stem cells.

While microarrays measure the gene expression of thousands of genes, not all genes are identified due to conservative and stringent statistical methods to analyze microarray data. We described a novel method that harness the power of detect statistically significant by making biologically meaningful comparison. Since RMA algorithm identifies differentially expressed genes by between two samples, genes that are expressed in both samples are mutually cancelled (Irizarry, Bolstad et al., 2003). Thus, masking particular components allows the identification of new genetic signatures. Microarray data of stem cells were compared to microarray data of tissue-specific reference cell types to purposely mask particular features of stem cells. Although gene expression
profiles generated by reference file comparisons do not reflect the entire list of genes expressed in stem cells, multiple comparisons were performed to generate a more comprehensive signature. This approach allowed us to characterize stem cells by masking specific components and revealing others.

Another modification to the reference design is the use of a universal RNA derived from various adult cells (Bhattacharya, Miura et al., 2004). This reference point is supposed to represent a signature of differentiated adult cells. Since microarrays assess the total RNA with a sample and heterogeneous samples from mixtures may influenced by the predominating cell type, we prefer to use homogeneous reference materials. Furthermore, since reference materials can be designed to have properties in common to the sample in question and RMA ignores genes that are expressed between two samples, we purposely compared AFSC to specific reference files to generate an expression profile.

A different approach to understand the genetic programs expressed in stem cells is the identification of “stemness” genes (Ramalho-Santos, Yoon et al., 2002). In these experiments, expression profiles of multiple stem cell types (ES, NSC, and HSC) identified a common set of genes that maybe involved in self-renewal and multipotency. This approach focuses on the genes themselves rather than deciphering potential mechanisms of stem cell differentiation.

Microarrays were used to assess an embryonic stem cell derivative, the hemangioblast. Since ESC differentiation protocols often results in mixed phenotypes, we applied our in silico method to dissect heterogeneous
populations within ESC-derived hemangioblasts. Hemangioblast express genetic signature containing processes involved in blood vessel and hematopoietic development, as well as processes associated with erythropoiesis and myogenesis (Lu, Hipp et al., 2007a). The identification of an erythrocytic signature suggests a strong presence of erythrocytes within the blast population. The results may also represent a fact that BC expansion medium contains several hematopoietic cytokines that pushes the development pathway towards hematopoietic lineage. Optimization of BC differentiation protocol would therefore be valuable to keep these cells bipotential. The myogenic component of hemangioblasts was further investigated after this analysis. Dr. Lu et al demonstrated that hemangioblasts expressed markers of smooth muscle cells and differentiated into smooth muscle-like cells in vivo (Lu, Ivanova et al., 2009). These results confirm the microarray data suggesting that hemangioblasts are tripotent.

To better understand stem cell differentiation, we also applied this method to examine the global gene expression profile of AFSC and their osteogenic differentiated progeny. We began by dissecting the undifferentiated AFSCs and this analysis revealed that AFSCs express multilineage programs. Our in silico dissection of AFSC consistently identified processes involved in skeletal and nervous system development among each comparison and are consider as major components of AFSCs. Processes that were identified in at least two, such as blood vessel development, muscle development, lung development,
kidney development and cartilage development, might be considered as minor components of AFSCs. Because it has been shown that AFSC can differentiate into these lineages, the expression of multiple lineage programs prior to differentiation suggests that AFSC are primed to differentiate into these lineages. The identification of multiple lineage-specific transcripts in AFSC is consistent with other stem cells including hES, hES-derived hemangioblast cells, and MSCs (Efroni, Melcer et al., 2009; Golan-Mashiach, Dazard et al., 2005; Lu, Hipp et al., 2007a; Tremain, Korkko et al., 2001). Promiscuous genes have also been observed in the hematopoietic system, where expression of genes of multiple lineages was detected prior to commitment (Hu, Krause et al., 1997). Thus, the expression of these transcripts may reflect the potential of stem cells to development into these lineages.

To determine how lineage programs change during stem cell differentiation, we performed microarray analysis on AFSC and their osteogenic derivatives. In silico dissection identified the expression of multiple lineage-specific programs, including skeletal development, in undifferentiated AFS cells. Gene expression profiles of AFS cells after 20 and 30 days of differentiation were still enriched genes involved in skeletal development however, processes associated with other lineages were reduced to lower levels. The identification of other lineages in differentiated AFS cells suggests that they still retain the potential to differentiation into other lineages. This means that stem cell derivatives might be more plastic than previously thought (Golan-Mashiach,
A related issue deals with lineage commitment. The expression of lineage-specific genes does not necessarily mean cells are restriction into a particular lineage. Huard et al. has recently showed that muscle-derived stem cells can differentiate into hematopoietic cells, repopulate the bone marrow of lethally irradiated mice and still retain the potential to differentiate into skeletal muscle cells (Cao, Zheng et al., 2003). Another study showed that fully differentiated osteoblasts from hMSCs were capable of differentiating into adipocytes and chondrocytes (Song and Tuan, 2004). More recently, a mature adipocyte-derived preadipocyte can transdifferentiate into osteoblast (Oki, Watanabe et al., 2008). These studies suggest that hierarchial model of stem cell differentiation should be modified to include transdifferentiation pathways. Although these studies did not perform microarray analysis, global gene expression profiles may identify potential mechanism that control transdifferentiation. Exploiting these properties could be of practical value in attempts to control lineage choice in stem cells for therapeutic applications.

The methodology described here can also be applied to complex diseases such as fetal alcohol spectrum disorders. This approach was used to identify an ethanol-responsive genetic signatures using AFSC as a model to study differentiation. This analysis identified an up-regulation of genes involved in skeletal development after ethanol exposure. We further determined that ethanol exposure has a functional effect on osteogenic differentiation by increasing
alkaline phosphatase activity and calcium deposition. Ethanol has been shown to alter the differentiation potential of other stem cells. Ethanol enhanced astrocytic differentiation while decreasing neuronal differentiation of NSC (Tateno, Ukai et al., 2005). In contrast to this study, ethanol inhibits the osteogenic differentiation of adult MSCs while having little or no effect on immortalized human fetal osteoblasts (Gong and Wezeman, 2004; Maran, Zhang et al., 2001). However, the response of these cells may not reflect that of prenatal stem cells. Ethanol has been shown to enhance cartilage differentiation in embryonic limb mesenchyme cultures (Kulyk and Hoffman, 1996). While these cells seem to be lineage restricted because they spontaneously differentiate into chondrocytes, naive AFSCs are not lineage restricted to osteogenesis. Thus, ethanol may act to lineage restrict AFSCs to osteogenesis by elevating the expression of osteogenesis-specific genes. The identification of ethanol-induced osteogenic differentiation in AFS cells provides new information that may be of significance for understanding mechanisms involved in the genesis of FASD. In addition, since these cells can be obtained by non-invasive means early in pregnancy, acquiring a complete picture of the response of AFS cells to ethanol offers promise for identification of a biomarker of prenatal insult.

Another source of cells for regenerative medicine is autologous progenitor cells obtained from a small biopsy of tissue. Patients who need reconstructed organs have “unhealthy” cells and whether these cells can be used for tissue engineering has been a matter of discussion. To provide a molecular evaluation
of cells from diseased organs, transcriptional profiles of bladder smooth muscle
cells derived from patients with myelomeningocele (MM) were compared to
smooth muscle cells from healthy patients. We identified a difference in the
genetic signature between diseased and normal smooth muscle cells. Global
gene expression profile analysis suggests that cells from myelomeningocele
patients are immature and secrete extracellular matrix proteins relative to healthy
cells. Lin et al. reported that these cells have different growth, contraction and
cell-adhesion properties from normal cells and suggests that cells from diseased
bladders might not be a suitable source for tissue engineering (Lin, Cowan et
al., 2004). Dozmorov et al. compared the expression profiles of smooth muscle
cells derived from normal and neuropathic bladders using a cancer specific
array (Dozmorov, Kropp et al., 2007). Their analysis identified 18 differentially
expressed genes and identified three significant pathways, e.g. FGF, PTEN, and
integrin signaling. In contrast, we identified 905 genes that were up-regulated
and 1,447 genes that were down-regulated in MM SMC and pathway-based
analysis showed an expression of genes involved in muscle development, focal
adhesion, and extracellular matrix interaction. These processes are common
pathological features of bladders from myelomeningocele patients. Although the
some of pathways are consistent with Dozmorov’s study, the discrepancy
between our studies is most likely due to the type of array. Dozmorov used an
array containing 1185 genes and was chosen because of its emphasis on cell
cycle, apoptosis, and signaling receptors. In contrast, our chip contained 22,283
genes which may explained why we identified more genes and pathways.
Although smooth muscle cells from diseased bladders can be implanted into patients to improve bladder function, urodynamic studies show that these bladders are not completely normal (Atala, Bauer et al., 2006). Thus, the identification of genetic signatures of diseased tissues and could allow the re-engineering of diseased cells into healthy cells. Modifying pathologically altered gene expression may be necessary to make a normal functioning bladder.

Before stem cells can be used as any type of clinical therapy, strict guidelines must be established to ensure the quality of the cells, the specificity of differentiation, and the assessment of mixed phenotypes. While lineage-specific gene expression and cell surface markers are commonly used to describe a differentiated phenotype, global gene expression profiling is necessary to provide a non-biased evaluation of the quality of cells. Another concern about the clinical potential of stem cells is their tendency to form heterogeneous cell types. Gene arrays are powerful tools to evaluate mixed phenotypes within a sample. The therapeutic potential of stem cells largely relies on efficient and controlled differentiation towards a specific cell type and the generation of homogeneous cell populations.


SCHOLASTIC VITA

Jennifer Anne Hipp
1900 Queen Street #B6, Winston-Salem, NC 27103
(714) 931-3281, jenhipp@wfubmc.edu

EDUCATION
M.D./Ph.D. Program (2004-present)
Wake Forest University School of Medicine,
Winston-Salem NC

Bachelor of Science June 2004
Massachusetts Institute of Technology, Cambridge, MA
Department of Biology

RESEARCH
National Research Service Award 2004-Present
NIAAA, National Institute of Health
MD/PhD Pre-Doctoral Fellow (F30AA16446-01A1)
Graduate Research Assistant, Winston-Salem, NC
August 2006 - Present
Wake Forest Institute of Regenerative Medicine

PUBLICATIONS


ABSTRACTS


HONORS

European Association for Urology, Best Poster in Paediatric Urology (2008)
Research Society on Alcoholism Satellite Symposium Travel Award (2008)
NIH National Graduate Student Research Festival, selected participant (2007)
M.D./Ph.D. Academic Merit Scholarship, WFU School of Medicine (2004-Present)
Intercollegiate Tennis Association Scholar Athlete, MIT Varsity Tennis (2004)
Pre-Cancer Research Training Award, National Cancer Institute (2002)
Pre-Cancer Research Training Award, National Cancer Institute (2001)
Howard Hughes Undergraduate Science Enrichment Program (2000)
PROFESSIONAL MEMBERSHIPS

Genetics and Environmental Mutagenesis Society

North Carolina Medical Society,

American Medical Student Association


MIT Biology Undergraduate Student Association (2001-2004)

SEMINARS
