2016

# Comparative Analysis Of Breast Cancer Stem-Like Cells And The Bulk Tumor Through Whole Transcriptome Sequencing And Expression Analysis

Samuel A. Verga II
*Indiana State University*

# COMPARATIVE ANALYSIS OF BREAST CANCER STEM-LIKE CELLS AND THE BULK TUMOR THROUGH WHOLE TRANSCRIPTOME SEQUENCING AND EXPRESSION ANALYSIS

---

A Thesis

Presented to

The College of Graduate and Professional Studies

Department of Biology

Indiana State University

Terre Haute, Indiana

---

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

---

By

Samuel A. Verga II

December 2016

Keywords: Breast Cancer, Stem cell, Sequencing, RNA, Transcriptome

COMMITTEE MEMBERS

Committee Chair: Catherine E. Steding Ph.D.

Assistant Professor of Biology

Indiana State University

Committee Member: Yongsheng Bai Ph.D.

Assistant Professor of Biology

Indiana State University

Committee Member: Kristopher Schwab Ph.D.

Assistant Professor of Biology

Indiana State University

Committee Member: Roy W. Geib Ph.D.

Alvin S. Levine Professor of Microbiology and Immunology

Professor of Pathology and Laboratory Medicine

Indiana University

Committee Member: Elaina M. Tuttle Ph.D.

Professor of Biology

Indiana State University

## ABSTRACT

Breast cancer is a heterogeneous disease comprised of multiple sub-populations of cells, each with the potential to contribute to the overall cancer in unique ways. Breast cancer stem-like cells (BCSC) are among those unique sub-populations, based on a $CD^{44+}$, $CD^{45+}$ and $CD^{24-/low}$ surface marker profile. This sub-population is responsible for breast cancer propagation and relapse; this makes these stem-like cells essential to study when developing new treatments for breast cancer. Stem cells and stem-like cells are pluripotent, denoting or indicating they have the ability to generate a multitude of cell types and near unlimited replicative potential. A transcriptome analysis of breast cancer cells and the sub-population of breast cancer stem cells will enable us to make a direct comparison of gene activity across the two populations. The analysis of gene activity will give scientists a clearer picture of the genes, the untranslated RNAs and the pathways involved in maintaining breast cancer stem cells. Identifying these pathways and key differences in gene activity will enable us to recognize specific targets for breast cancer stem cells and better understand their behavior in cancer development, progression, chemotherapeutic resistance, and metastasis. The results of this experiment indicate several differentially expressed genes that could significantly affect important pathways. Further investigating of interacting pathways of deferentially, expressed transcripts can determine their specificity to breast cancer stem-like cells.

# PREFACE

This thesis is intended to examine breast cancer stem-like cell gene expression. A greater understanding of BCSCs will open up new avenues to explore in the breast cancer field, especially in terms of acquired resistance and breast cancer therapeutics. Additionally, the use of next generation sequencing will provide a high throughput and accurate analysis of gene expression capable of furthering our knowledge in a variety of areas. This includes; the potential to reveal previously unidentified events, such as altered expression and alternative splice events that might be specific to the activities of the stem-like population versus the bulk tumor cells. This project will provide the foundation to explore a variety of future directions, including the identification of new therapeutic targets for metastatic breast cancer.

TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# CHAPTER 1

## AN INTRODUCTION TO BREAST CANCER

Breast cancer is of significant clinical importance, as it is the most commonly diagnosed cancer among women. Breast cancer statistics show one in eight women in America will develop breast cancer in their lifetime (DeSantis, Ma, Bryan, & Jemal, 2014). Traditionally, white females have a higher rate of breast cancer as compared to black women. However, recently these two groups have converged to demonstrate similar rates (DeSantis et al., 2015). Survivorship following breast cancer treatment is 90% over 5 years (Howlader, Noone, & Krapcho, 2015). There are two common types of breast cancer based on expression of estrogen receptor (ER), epidermal growth factor-2 (ErbB2) gene, human epidermal receptor 2 (HER2), and the progesterone receptor (PR). A third less common type of breast cancer is triple negative breast cancer (TNBC) which lacks expression of ER, HER2, and PR receptors. No effective treatments to TNBC have been developed; therefore TNBC is the most difficult to treat (Wei & Lewis, 2015). Treatments for breast cancer are often hormone-driven treatments that reduce the cancer population by preventing the hormone stimulation required for growth. Due to the lack of receptors, TNBC is more difficult to treat. Treatments for TNBC often use broad therapies like radiation and limited chemotherapy, non-specific targets. Currently, many breast cancer therapies focused on treating TNBC through specific treatments based on a cellular activity unique to this type of cancer.

Breast cancer can be dangerous and life threatening mainly due to ineffective treatment of metastasis. Metastasis involves tumor escape (invasion of surrounding tissue), migration (often through lymph vessel or blood stream), adhesion to a distant site, extravasation at site of tumor initiation, initiation of metastatic growth and maintenance of metastasis (Chambers, Groom, & MacDonald, 2002). Cancer kills at the metastatic site through interfering with normal organ functions.

## Breast Cancer Characteristics and Description

Due to a complex organization of tissue and cellular interactions with the tumor microenvironment, breast cancer exhibits significant heterogeneity. The interactions of both cells and cellular products with the stroma or surrounding breast tissues can affect the cellular sub-populations within the tumor microenvironment and place selective pressure on cancer cells. Differential cellular responses and adaptations to selective pressures therefore not only generate the multiple cellular sub-populations observed in heterogeneous tumors, but also regulate how they will behave during tumor development and progression (Shah & Allegrucci, 2012).

Ductal carcinoma, one of the most common forms of breast cancer, arises from the terminal duct lobular unit after accumulating mutations that deregulate the cell cycle (Chambers et al., 2002). There are many subtypes and categories used to define breast cancers as they often exhibit different characteristics, and often share characteristics from different classifications of breast cancer, highlighting the heterogeneity of breast cancer and the surrounding tissue. Breast carcinoma can be classified into two broad types using molecular and histological methods: luminal, and basal-like. Luminal breast tumors often express high levels of hormone receptor, whereas basal breast tumors typically negatively express hormone receptor, and are known as a triple-negative breast cancer (Makki, 2015). In addition, basal-like breast cancers have been

shown to have similar gene expression signatures to squamous cell lung cancers, which can indicate that these cancers originate from a different cell type (Prat et al., 2013). This suggests a potential difference in origin for basal-like breast cancers, which may alter how this type of breast cancer is treated and studied. The difference in origin of breast cancer could involve very different pathway interactions and function of cells in general. Cancer profiling and analysis are assisted by genomic analysis and is an integral part in better understanding the heterogeneity of breast cancer. Genomic analysis of breast cancer can provide useful information for researchers and medical doctors to improve their approach of how to treat cancer in patients.

## Breast Cancer Stem Cells: Overview and Characteristics

Cancer stem cells were initially identified in acute myeloid leukemia and have now been found in several other cancers including breast cancer (Bonnet & Dick, 1997). Breast cancer stem cells or breast cancer stem-like cells (BCSCs) are a unique sub-population of cells that is responsible for a cancer's invasiveness and its ability to metastasize. These cells were initially identified using the cell surface markers cluster differentiation (CD) $44^{+}$ and CD $24^{low/-}$ and were shown to be capable of initiating tumors at far lower concentrations when enriched for these markers than the bulk tumor cells (Al-Hajj, Wicha, Benito-Hernandez, Morrison, & Clarke, 2003). This sub-population of cells is also considered stem-like progenitor cells or stem-like cancer cells. These cells have stem-like characteristics, which include the ability to produce differentiated cells from an undifferentiated precursor and a more limitless replicative potential (Owens & Naylor, 2013). BCSC have also been observed to be resistant to radiation and chemotherapy and therefore, implicates them in the recurrence of breast cancer after treatment (Mannello, 2013; Phillips, McBride, & Pajonk, 2006). Radiation studies have been used to

further study BCSCs through focusing on the effects of oxidative stress (Ryoo, Choi, & Kwak, 2015; Sai et al., 2015).

**Identifying Breast Cancer Stem Cells**

A sub-population of breast cancer cells capable of initiating tumors when isolated from the bulk tumor were identified and termed "cancer stem cells" (Bonnet & Dick, 1997). When the bulk tumor was depleted of these cells, tumors did not form, nor required a substantially larger number of cells to initiate tumors in mice. The key characteristic of these cells expressed an adhesion molecule $CD^{44}$ and had little to any $CD^{24}$ (Al-Hajj et al., 2003). The prevailing characteristic of tumor-initiating cells or BCSC manage to recapitulate the complex tumor structure and heterogeneity of tumors, which are derived from an unsorted cell line, self-renew, and reconstitute parental cell lines. $CD^{44+}/CD^{24-}$ cells express characteristics of basal cells (Fillmore & Kuperwasser, 2008). BCSCs are characterized several different ways, though mainly through cell receptors and other phenotypic characteristics. BCSCs were identified as being positive for $CD^{44}$, a surface adhesion molecule, and to contain low levels of $CD^{24}$, a marker normally expressed on mature granulocytes responsible for regulating growth and differentiation of B cells (Owens & Naylor, 2013).

Antibodies to the $CD^{44+}$ / $CD^{24-}$ receptors are used to label and isolate the cells based on their surface marker profile. These markers were used for several years. However, recent studies have indicated that these surface marker profiles are not unique to BCSCs, such as BCSCs that have metastasized to the bone marrow (Balic et al., 2006). Initially this phenotype, $CD^{44+}/CD^{24-}$, was thought to not be associated with clinical outcome, invasiveness, or metastasis (Abraham et al., 2005; Sheridan et al., 2006). Further characterizations of the cells were completed, identifying a population of cells that are positive for aldehyde dehydrogenase or $ALDH^{+}$ and

demonstrated a more aggressive population of cancer stem cells (Ginestier et al., 2007). $CD^{44+}/CD^{24-/low}/ALDH^{+}$ cells were shown to have enhanced metastatic behavior (Charafe-Jauffret et al., 2010). ALDH is part of the retinoic acid pathway, which plays a role in cellular differentiation and stem-cell protection (Croker et al., 2009). These markers are not the only markers that can be used to identify this sub-population of cells and can create confusion over the definitions of these isolated cells in each experiment (Hwang-Verslues et al., 2009). There are numerous markers for BCSCs including $Ep\text{-}CAM^{+}/CD49f^{+}$, CD133, epithelial specific antigen (ESA), chemokine receptor type-4 (CXCR4), and protein C receptor (PROCR), with many associated with BCSC functions (Croker et al., 2009; Fillmore & Kuperwasser, 2008; Ghebeh et al., 2013; Hwang-Verslues et al., 2009; Y. Kang et al., 2003; Shipitsin et al., 2007; Wright et al., 2008; F. Ye et al., 2015). Additionally, CD 271 has been suggested as a potential BCSC marker specific to luminal breast cancers without basal-like characteristics (J. Kim et al., 2012). These markers are all considered potential additional markers to identify BCSC. However they are not well established and only associated with increased BCSC characteristics.

**The Origins of Breast Cancer Stem Cells**

Two main theories exist concerning the origin of BCSCs. It is important to determine the origins of these cells to best study the cells and decide on an approach to target them. The competing theories focus on clonal evolution or stem cell instability (Owens & Naylor, 2013). More likely is that these theories are not exclusive, meaning that both activities are occurring within the cell populations. The heterogeneity and selective pressures on cancer cells following treatment has the potential to eliminate cell types resistant to chemotherapy and radiation. This potential is based on the instability of cancer cells in general due to their active growth and lack of DNA repair mechanisms. The other theory suggests that BCSCs arise from long-lived

mesenchymal stem cells and become malignant and is supported by the stem-like characteristics of BCSCs (Shah & Allegrucci, 2012). The focus of many studies is based on either theory and is important in developing treatments, which actively target BCSCs. Properties that are unique to BCSC are used to target the population; this makes the study of BCSCs all the more necessary. Importantly, mammary cancer stem cells were shown to differentiate into several lineages that recapitulate the heterogeneity produced by BCSCs, supporting the stem cell theory (Bao, Cardiff, Steinbach, Messer, & Ellies, 2015).

## What Makes Breast Cancer Stem Cells Unique?

Through genetic analysis, unique attributes have been identified in BCSCs that had better inform what makes this sub-population of cells stem-like. In one such instance, epithelial and mesenchymal-like states of BCSCs have been observed using micro-array. Mesenchymal-like BCSCs were characterized by being $CD^{44+}/CD^{24-/low}$, primarily quiescent, and located near the invading tissue of the tumor, whereas epithelial-like BCSCs are $ALDH^{+}$, actively growing, and are localized in the center of the tumor (S. Liu et al., 2014). Furthermore, cells that were isolated based on the $CD^{44}$ and $CD^{24}$ markers produced what are known as mammospheres (non-adherent spheroid formations) *in vitro* demonstrating for the first time that BCSCs can propagate from existing cell lines (Ponti et al., 2005). This proved to be a viable method to grow BCSCs from established cells lines when previously primary samples were necessary.

### Surface marker expression in breast cancer stem-like cells

Surface markers and expression profiles are important to determine if a cell is indeed a BCSC, however these methods can be costly and time-consuming. One of the often-used measures of BCSCs is through mammosphere formation. Mammosphere formations occurs when cells lose their adhesion molecules and grow in spherical groupings called mammospheres,

which was confirmed as an efficient basis to study BCSCs by comparison to primary cell cultures (R. Wang et al., 2014). The marker $CD^{44}$ and receptor substrate, hyaluronan, are important for drug resistance and colony formation in BCSCs (N. K. Han et al., 2016). Additionally, it was seen that basal-like breast cancers contained a higher percentage of $CD^{44+}/CD^{24-/low}$ and $ALDH^{+}$ cell types, whereas luminal and HER2 over expressing cancers, fail to express these CD surface receptors (Ricardo et al., 2011). Cell lines like MCF7, are luminal-like, contain higher percentages of $ALDH^{+}$ BCSCs, where MDA-MB-231 cells, a triple negative breast cancer, are more basal-like and therefore contain more $CD^{44+}/CD^{24low/-}$ cells. $ALDH^{+}$, CD $133^{+}$, and vascular endothelial growth factor (VEGF) were found to be correlated to each other suggesting a link between ALDH and angiogenesis in BCSCs (Lv, Wang, Song, Pang, & Li, 2016). It was determined that BCSCs can transition between epithelial and mesenchymal-like states. This process is called epithelial-mesenchymal transition or EMT, which suggests this, is responsible for their increased potential for invasion and metastasis. These interesting populations of BCSCs are summarized in Figure 1. This difference between the two groups of characteristics and the ability of these cells to transition from one set to another draws an important perspective on how researchers approach the treatment of BCSCs.

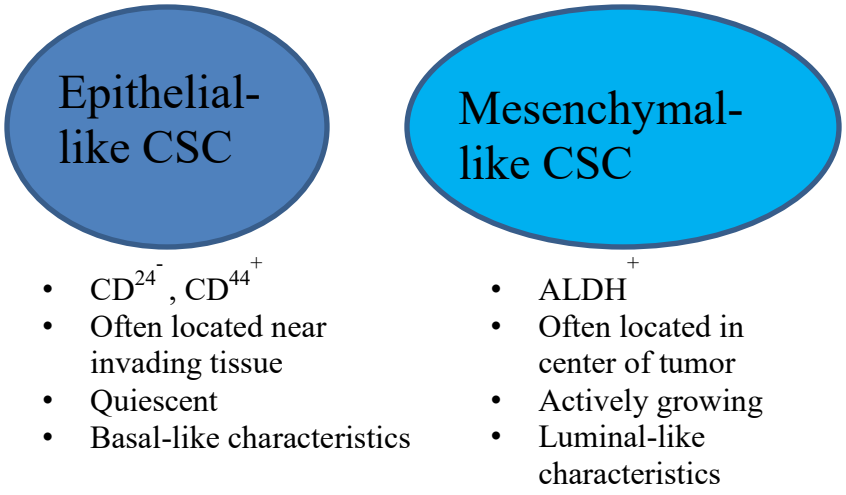



*Figure 1. Comparing the two different characteristics represented by the two different definitions of BCSCs. There are several interchangeable characteristics when looking at stem-like cells. These characteristics are based on the microenvironment in which they reside including, where they are within the tumor, molecules present from chemotherapy, surround cell secretions, and within the blood.*

EMT is necessary for the remodeling of cells and tissue during embryogenesis, as well as wound healing. EMT is directly linked to BCSCs and enhances their stem like characteristics since tumor growth factor-β (TGFβ) is capable of inducing EMT (Shipitsin et al., 2007). PD-L1 has been shown to promote EMT and $CD^{44}$ expression and PF1 is a factor required for maintenance (Alsuliman et al., 2015; Bansal et al., 2015). Additionally, several genes or proteins have been implicated in EMT and BCSC maintenance, including HER, Wnt, Oct4, Sox2, and Nanog, many of which function to activate the Wnt/β-catenin pathways, which regulate stem cell pluripotency (Lawson et al., 2015; C. A. Lee et al., 2015; Martin-Castillo et al., 2015; Niu et al., 2015; Williams, Bundred, Landberg, Clarke, & Farnie, 2015). Cleavage of the intracellular domain of $CD^{44}$ activates Sox2, Oct4, and Nanog, demonstrating a link between the markers and functions of BCSCs (Cho et al., 2015).

**Transcription factors involved in upregulation and downregulation of breast cancer stem-like characteristics, including EMT**

Transcription factors Slug and Snail were implicated in EMT (X. Ye et al., 2015). EMT plays a large role in BCSC activity and involves many pathways, such as RAS-MAPK and Hedgehog (Morel et al., 2008). Combination drug therapy, using paclitaxel and EW-7197, was used to overcome paclitaxel-induced EMT and BCSC properties. Additionally, it suppresses paclitaxel-induced Snail and EMT by attenuating reactive oxygen species (Park et al., 2015). The two different states of this sub-population demonstrate that the quiescent nature of one type may improve the resistance to chemotherapy (Creighton et al., 2009; Farmer et al., 2009; Q.-Q. Li et al., 2009; S. Liu et al., 2014).

EMT is heavily affected by drug treatment and hypoxia. $CD^{44+}/^{24-/low}$ cells are hypoxia-dependent and are limited in the presence ER. The BCSC population is driven by hypoxia, which results from a dedifferentiation process by relying on hypoxia inducible factor-1-α (HIF1α) in $ALDH^{+}$ cells, whereas $CD^{44+}/^{24-/low}$ populations are HIF1α-independent and require prolyl hydroxylase 3 (PHD3) downregulation. This down regulation activates NF-κB signaling and reduces $CD^{24}$ expression (Iriondo et al., 2015). HIF can block apoptosis in cell culture and increase BCSC markers (J. Xie et al., 2016). Additionally, HIF-1-mediated glutathione synthesis plays an important role in the enrichment of BCSC through promoting Nanog and other pluripotency factors. It was also shown that glutathione, MEK1-ERK inhibitor, and copper chelators could promote the BCSCs phenotype through this pathway (Lu et al., 2015). The BCSC marker CD49f is also implicated in HIF function, as its gene is integrin-α-6 and is a direct transcription target of HIF (Brooks et al., 2016).

Transcription factors like, ABCG2, IL-6, Notch, NF-κB, ALDH, JAK2/STAT3, fork-head transcription factors (FOX), or Wnt pathways can induce or otherwise affect EMT and otherwise alter BCSC activity. ATP-binding cassette half transporter (ABCG2) has been shown to be enriched in BCSCs and has been linked to EMT and the enrichment of BCSC markers (J. Y. Jang et al., 2012; Patrawala et al., 2005). Interleukin-6 or IL-6 was shown to be capable of inducing malignant features in the Notch-3 expressing BCSCs by being essential for self-renewal (Sansone et al., 2007). IL-6 is capable of inducing EMT and stem-like characteristics. IL-6 is highly expressed in breast cancer and high levels often correlate with advance tumor stage and metastasis (G. Xie et al., 2012). Autophagy can regulate CD 24 and IL-6 secretion based on the dependency of autophagy in the cell and can either increase or decrease the expression of BCSC markers (Maycotte, Jones, Goodall, Thorburn, & Thorburn, 2015).

The fork head transcription factors have been implicated in some BCSC activities. Identification of the FAK/FAS/FOX03A was found to be specific for BCSCs through the CXCR1, IL-8 receptor (Ginestier et al., 2010). FOXO3A was determined to be a downstream target of Notch and AKT implicated in mammosphere formation and BCSCs marker expression (Smit et al., 2015). FOXA1 was also linked to expression of ALDH and colony formation in BCSCs (Tachi et al., 2015). Furthermore, the FOXC2 transcription factor was found to be highly expressed in BCSC and found to be dependent on cell cycle. This demonstrated the role of the FOX protein in the BCSC cell cycle maintenance (Pietila et al., 2016). Finally, Foxq1 was shown to promote chemo-resistance and stemness traits by acting on the downstream targets of PDGFRα and β (Meng et al., 2015).

DNA regulation and epigenetic changes have been shown to be associated with the BCSC phenotype. Histone deacetylase (HDAC) has been shown to be potentially downregulated

in BCSCs and when inhibited shows an increase in $ALDH^+$ cells and increased mammosphere formation. Additionally, they may be linked to downregulating Wnt reporter activity (Debeb et al., 2016; Debeb et al., 2012). HDAC8 has been shown to activate BCSCs by enhancing the stability of Notch1 outside of its epigenetic functions (Chao et al., 2015). DNA repair was shown to be important for maintenance of BCSCs by showing that inhibition of DNA-PK, a DNA repair enzyme, resulted in the decreased mammosphere formation in MCF7 cells (Lamb et al., 2015). This however may be specific to luminal-like breast cancers. DNA methyltransferase 1 has also been implicated in BCSCs by helping maintain cell memory as they do in normal mammary stem cells (Pathania et al., 2015).

DNA repair mechanisms are important in BCSCs to help maintain genetic stability and maintain stem-like characteristics. Additionally, DNA modifiers help maintain the plasticity of BCSCs and allow them to adjust to their environment, such as establishing a metastatic site or being treated with chemotherapeutics or radiation. The regulation occurs by transcription factors to activate specific genes and functions through DNA modifiers like methyltransferases.

**Effects of mitochondria on breast cancer stem-like cells**

Mitochondrial mass has been associated as a metabolic biomarker for BCSCs by its association with WNT/FGF-driven anabolic signaling (Lamb et al., 2015). Increased mitochondrial mass is seen in BCSC populations and could enhance chemo-resistance and other stem-related traits (Farnie, Sotgia, & Lisanti, 2015). A component of telomerase, hTERT, acts to improve expression of mitochondrial, EMT and glycolytic enzyme genes that facilitate an increased mitochondrial mass and functional activity, which improves mammosphere formation (Lamb et al., 2015).

Mitochondrial inhibitors can be used to reduce BCSC population and inhibit stem-related pathways such as sonic hedgehog, TGFβ, STAT3, and Wnt. They can also function to target FOXM1 and mitochondrial biogenesis in BCSCs to prevent relapse (De Luca et al., 2015). Overall, mitochondrial mass and function are shown to be an important trait in BCSCs and may contribute to additional BCSC functions.

BCSCs perform an important function in the cancer microenvironment and rely more heavily on cellular respiration than other sub-populations. It appears that BCSCs compensate for the increased activity by increasing mitochondrial size and activity. This increased energy output could be directly related to the increased mammosphere formation and maintain the processes needed to express stem-like characteristics.

**Pathways involved in breast cancer stem-like function**

BCSC function has also been linked to Notch function. Increased Notch activity increases mammosphere formation and BCSC marker (D'Angelo et al., 2015). Notch1 can be suppressed by inhibition of integrin-linked kinase (ILK), which suppresses the previously described function of Notch. ILK regulates gamma-secretase complex, which is mediated by Notch 1 activation post-transcriptionally (Hsu et al., 2015). Protein levels of polycomb group protein, B-lymphoma Moloney murine leukemia virus insertion region-1 (Bmi-1) help regulate BCSCs through activating Notch1, Notch2, and Notch4 in luminal-like breast cancers (S. H. Kim & Singh, 2015). Hes1 has also been linked to Notch and BCSC activity (So et al., 2015). Expression of SATB1, special AT-rich sequence binding protein 1, is involved with the Notch signaling pathway to increase expression of the BCSC population (Z. Sun et al., 2015).

The NFκ-B Pathway has also been implicated in BCSC maintenance through use of NFκ-B inhibitors. These inhibitors have been previously shown to inhibit leukemia stem cells and

BCSC proliferation (J. Zhou et al., 2008). NFκ-B is linked to EMT and stem-like characteristics through the ERK/ NFκ-B/Snail pathway and interacting with TGFβ (D. Han et al., 2015).

ALDH was shown to be important to chemotherapy and radiation resistance (Tanei et al., 2009). This is shown through increased effectiveness of chemotherapy and radiation therapy when ALDH is inhibited (Croker & Allan, 2012). JAK2/STAT3 signaling has been shown to increase growth of BCSCs in primary tumors (Marotta et al., 2011). ALDH is a useful marker for identifying and selecting for BCSCs.

The Wnt Pathway is capable of expanding and enhancing the sub-population of BCSCs and the differentiation capacity, leading to recapitulate the heterogeneity of primary tumors (Monteiro et al., 2014). IKKβ, an essential protein in the NFκ-B pathways, regulates a feedback loop with LIN28B, an inhibitor of let7 family microRNAs, by interacting with the Wnt/TCF7L2 signaling pathway to promote stemness and metastasis (Chen et al., 2015). Tumor necrosis factor Alpha-receptor apoptosis inducing ligand (TRAIL) can suppress tumor initiation and clonal expansion of BCSC. cFLIP is a TRAIL inhibitor, which can inhibit cytotoxicity of TRAIL when present in the cytoplasm or promote Wnt-dependent signaling when in the nucleus (French, Hayward, Jones, Yang, & Clarkson, 2015). Wnt/β-catenin signaling is higher in BCSCs, which confers a higher level of therapeutic resistance. An antagonist called CWP232228 binds to β-catenin to T-cell factor (TCF) in the nucleus and inhibits growth of BCSCs and the bulk tumor. It also attenuates insulin growth factor-1 (IGF-1), which has a significantly higher expression in BCSCs, identifying another potential therapeutic target for BCSCs (G. Jang, Hong, et al., 2015; G. Jang, Kim, et al., 2015). Toll-like receptor-3 (TLR3) promotes breast cancer cells toward a BCSC phenotype. TLR3 requires activation from both β-catenin and NF-κB signaling pathways (Jia et al., 2015). Long-term nutrient deprivation induces a Wnt-dependent transition to a stem-

like state caused by chronic metabolic stress using transcriptomic analysis. This enriches for $CD^{44+}/ESA^{+}$ BCSCs (Lee et al., 2015). Mammary stem cells, which may be the origin of BCSCs, may stimulate tumorigenesis through Wnt signaling by affecting self-renewal and differentiation capacity leading to establishment of potential BCSCs (Monteiro et al., 2014).

Additional transcriptome analysis of $CD^{44+}/CD^{24-}$ BCSCs identified genes being overexpressed that are involved in the maintenance of stem like characteristics and PI3K pathways, which indicates a hyperactive endocrine resistance (Hardt et al., 2012). The PI3K/AKT/NF-κB pathway has also been suggested as a possible inhibitor target to eliminate multidrug resistance and BCSC elimination by initiating apoptosis factors, such as caspases (Y. Hu et al., 2015). The PI3K/AKT, MAPK/ERK, and Stat3 pathways have all been implicated in a study using polyphenol-enriched blueberry preparation, which caused inhibition of growth and mammosphere formation (Vuong et al., 2016). This is potentially due to the anti-oxidative influences on these pathways.

In an effort to understand better the underlying mechanisms of cancer, scientists have studied the genome and transcriptome of cancer, specifically breast cancer. Initially, the best ways to study the gene and transcript profiles were through micro-array technology. It was shown that a gene set-based module discovery approach was capable of understanding regulatory programs in cancer cells (Niida et al., 2009). Although this allowed for a basic understanding of what was occurring in breast cancer cells, it was limited by our previous knowledge of the genome and incapable of finding novel genes or transcripts. Genetic analysis of BCSCs was able to determine that transcription does not revert to a Warburg-like state showing a more glycolytic profile (Gordon et al., 2015). The hypoxia resulting from tumor growth does not alter BCSC transcription for metabolism demonstrating a requirement for a more metabolically active cell.

**MircoRNAs and their relation to breast cancer stem-like cells**

Many studies have utilized microarray and gene expression analysis to identify microRNAs (miRNA) which influence BCSC in many ways. Initially miR-200c was identified as a downregulated miRNA in BCSCs, which influences mammary stem cells to prevent duct and tumor formation induction and anti-oncogene roles (Feng et al., 2015; Y. Shimono et al., 2009). Downregulation of miRNAs is not uncommon and was also identified in miR-205-p, miR-100, and miR-34a, which can suppress oncogenes like Zeb1 and ErbB3, involved in EGFR expression and BCSC marker expression (De Cola et al., 2015; Ma, Yang, & Zhang, 2015; Petrelli et al., 2015). Downregulation of miR-141 by progesterone promotes stemness and BCSC marker $CD^{44}$ (Finlay-Schultz et al., 2015).

Pre-adipocytes also play a role in BCSC maintenance by exosome secretion to interact with miR-140, Sox2, and Sox9 to regulate differentiation, stemness, and migration (Gernapudi et al., 2015). Responsiveness to Sox2 reporter is associated with high tolerance to oxidative stress and results in increased mammosphere formation (Gopal et al., 2015). The Notch pathway is also affected by miRNAs such as miR-34a, which is downregulated in breast cancer and can downregulate Notch1 expression (L. Kang et al., 2015). Mammosphere formation can be affected by miRNA through Pleckstrin homology-like domain, family A, member 1 (PHLDA1) expression through ER/NF-κB/miR-181 regulatory axis (Kastrati, Canestrari, & Frasor, 2015). EMT is affected by miR-221, which targets gene ATXN1, which is related to EMT (Ke et al., 2015). Transcription factors like Kruppel-like factor 4 (KLF4) is a pluripotency mediator enriched in BCSCs which is mediated by miR-206 (Lin et al., 2015).

Proteins involved in stem-like characteristics and BCSCs growth, such as Nanog and Oct 4, involved in the Wnt pathway, have also been implicated to have miR-1, miR-204 and miR-221

associated with their expression (T. Liu et al., 2015; Roscigno et al., 2015; L. Wang et al., 2015). A gene panel identified multiple clusters differentially expressed in BCSC including miR-200 clusters, miR0183 clusters, miR-221-222 cluster, let-7, miR-142, and miR-215 to be involved in several pathways involved in BCSCs, such as Notch, Wnt, EMT and Bmi-1 (Y. Shimono, Mukohyama, Nakamura, & Minami, 2015). Let-7a was identified to have an inverse relationship with miR-208a related to ALDH expression and Wnt signaling (X. Sun et al., 2015). The Hippo pathway has also been involved with miRs, such as miR-125a, through leukemia inhibitory factor receptor (LIFR) to promote expression of BCSC markers (Nandy et al., 2015). Overall, miRNAs have been identified to be associated with every pathway involved in resistance, EMT and BCSC maintenance, including some involving long non-coding RNAs (lncRNA) (Bamodu et al., 2016; Boo et al., 2016; Takahashi et al., 2015). Several lncRNAs have been associated with Twist-dependent EMT such as lncRNA-Hedgehog, and may regulate the Wnt signaling pathway (P. Hu et al., 2014; M. Zhou et al., 2016). This demonstrates the importance of miRNAs and using them as therapeutic and inhibitory targets for future research and involvement in the many associated pathways.

## Using Genetic Analysis to Study BCSCs

Several other studies have attempted to study these cells utilizing Next-Generation-Sequencing technology with varying results. This is due to the uncertain nature of their phenotype. This is described by sequencing BCSCs from primary tumors. Negative selection was performed for an additional marker known as $CD^{45}$ to remove any tumor infiltrating leukocytes (Hardt et al., 2012). This demonstrated the ability to sequence rare populations from as few as 500 cells and demonstrated important findings to create an initial profile of BCSCs. Transcriptome analysis of cells overexpressing ErbB2 in mice was able to identify potential

genes driving BCSC populations (Borcherding, Bormann, Kusner, Kolb, & Zhang, 2015). A study by Ghebeh et. al. (2013) has described normal and malignant breast tissue based on the $CD^{44high}/CD^{24low}$ markers in combination with Ep-CAM/CD49f markers to differentiate between luminal and basal cells (Ghebeh et al., 2013). These markers can improve identifications of BCSCs that are ALDH+ within the study as this marker demonstrates a more histological grade cancer (Ginestier et al., 2007). This combined with other varying characteristics suggests that $ALDH^{+}$ BCSCs may be two different populations of BCSCs. The EMT supports a shift to a phenotypic expression of a more BCSC-like state.

Although many researchers have expressed interest in the BCSC transcriptome and other sub-populations, they remain poorly defined. Whole transcriptome analysis of breast cancer cells will produce a comprehensive list of transcribed genes in the breast cancer and cancer stem-like transcriptome. This will demonstrate the ability to accurately sequence established breast cancer cells, which will increase the sample size from previous sequences from only 500 cells. By selecting for the marker $CD44^{+}/CD24^{low/-}/CD45^{-}$ , this will isolate the purest form of BCSCs described by these markers. This approach attempts to reconcile the varying descriptions of BCSCs using the original phenotype described by Al-Hajj et al. (2003) and the use of established cell lines utilizes high sample population numbers for easier access to researchers. The effective use of this methodology can expedite future research of BCSCs. This is because it challenges the previous findings of genes that have been associated with increasing cancer stem-like potential and identifies new ones. Compiling a list of genes being transcribed in breast cancer and the stem-like phenotype population will allow future identification of potential therapeutic targets. Some initial genes of interest are those outlined previously, which have been found to increase the stem-like phenotype and characteristics in the BCSC sub-population. The most compelling

sets of genes include those involved in normal stem cell maintenance and cellular development. These are particularly intriguing because cancer has often been considered a deviation from the development or degradation of stem cell maintenance, which causes cells to act abnormally. Furthermore, by comparative expression analysis of breast cancer cells and BCSCs, changes in the transcriptome can be more easily identified for further study, which can potentially elucidate the trend of stem-related and development genes found to regulate stem-like cells.

## Using Bioinformatics to Analyze Transcript Expression

Bioinformatics allows computerized analysis of genomes using algorithms and automated counting. There are programs developed to analyze transcript expression and can be used to identify novel transcripts, splice junctions, and gene expression. Many analysis programs determine most or all components from a sequenced transcriptome. Initially, the sequenced transcripts need to be aligned to a reference genome. Alignment to the genome is important to understanding sequencing output by identifying where the sequences belong in the genome. Having a reference sequence allows for quick annotation of sequences to identify their place in the genome and the corresponding genes or other genetic information. In studying the human genome, there is significant previous research to appropriately annotate and describe the human genome, allowing researchers to perform complex analysis easily. This can be performed with three different programs: TopHat, Burrow-wheeler aligner (BWA), and Bowtie (D. Kim et al., 2013; Langmead & Salzberg, 2012; H. Li & Durbin, 2009). TopHat is exclusively used for RNA-sequencing data, whereas bowtie and BWA are used for genomic data.

TopHat is a spliced junction mapper alignment program used to analyze transcripts produced during RNA-sequencing. TopHat can be used in conjunction with many other programs such as DeSeq, which can be used for differential expression analysis. To perform

DeSeq, an intermediary program is used to organize the data into the necessary format, including columns with annotation data, to be read by the program, which counts transcripts, called Ht-seq count. TopHat is also used for a program called Cufflinks.

Cufflinks takes aligned reads and assembles them into transcripts that are counted to estimate transcript levels. Cufflinks also has two additional components to allow for differential expression testing. Cuffmerge is used to produce a file with combined transcript files to be used in the differential expression-testing program, CuffDiff. CuffDiff uses the merged transcripts along with the original Cufflinks files to test for significantly expressed transcripts. Each program described has multiple options to customize the algorithms used to count and estimate transcripts or how the transcripts are selected and defined. Several of the customizable options work similarly to DeSeq. This can make it difficult to replicate analysis, but offers a greater ability to customize accordingly to each type of analysis.

Each program requires knowledge of how to write script and use program software. In order to make it more user-friendly, scientists put together a website, usegalaxy.org, to compile programs into a free access and easy to use site. Galaxy allows researchers to find and input data, customize options, and push-button start programs (Afgan et al., 2016). Galaxy reduces the need for programming knowledge and has a community of researchers to help and provide suggestions for analysis. For the following analysis, the Cufflinks programs will be used due to their convenience and easy use. There is no need for additional format conversions or analysis as it is a self contained pipe-line.

# CHAPTER 2

## MATERIALS AND METHODS

Cells- MDA-MB 231 epithelial breast cancer cells are used because they are a well-characterized cell line and can act as a standard for analysis of other cell lines. Cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% Cosmic Calf Serum (CCS) (Fisher Scientific). Every three to four days fresh media was applied and cells were maintained in log phase. Approximately once weekly cells were redistributed or split to lower numbers in new flasks to maintain a less than was ~70% confluency. Cells were redistributed via trypsinization using 0.25% Trypsin-EDTA (Fisher Scientific), inactivation of trypsin with the media supplemented with 10% CCS at a minimum ratio of 3 milliliters of media to 1 milliliter of trypsin-EDTA and selecting an aliquot of cells for placement into a new flask.

Cell Sorting- Cells are removed from plates using 0.25% EDTA Trypsin, collected, and centrifuged. Media containing trypsin is removed and diluted with new media and cells are counted. Cells are allowed to rest for a half hour before beginning cell-sorting protocol. Magnetic activated cell sorting (MACS) (Miltenyi Biotec) is used to separate cells with the biomarkers that identify BCSCs, $CD^{24-}$, $CD^{44-}$, and $CD^{45+}$ (Hardt et al., 2012). A set of cells without the addition of beads is put through the same procedure as a sorting control. The antigens $CD^{24}$ and $CD^{44}$ are separated using negative selection to maintain cell integrity and $CD^{45}$ antigen are positively selected by forcing cells out of the column after remaining cells not

expressing $CD^{45}$ are eluted from the column. A set of cells is put aside as a sort control, which no antigens are selected for. Cells are then flash frozen using liquid nitrogen for downstream analysis.

Cell Staining- A sample of MDA-MB 231, approximately 100,000 cells, of unsorted cells are set aside prior to cell sorting for Staining and Flow Cytometry (BD Biosciences) is performed to confirm presence of BCSCs using fluorescent antibodies for $CD^{44}$, $CD^{24}$, and $CD^{45}$ (BD Biosciences). Cells were collected via trypsinization, re-suspended in 10% CCS DMEM, and incubated at room temperature for 15-30 minutes to allow for cell membrane repair after trypsinization. Cells were then centrifuged at 1000 rpm for 5 minutes to remove media and re-suspended in Stain Buffer (BD Biosciences). All three fluorescent antibodies stains were applied individually then incubated at four at least 30 minutes to as long as overnight. Cells were then centrifuged at 1000 rpm for 10 minutes and re-suspended in stain buffer to wash cells. Finally, cells were centrifuged at 1000 rpm for 10 minutes and re-suspended in Fix Buffer (BD Biosciences), to be ready for flow cytometry.

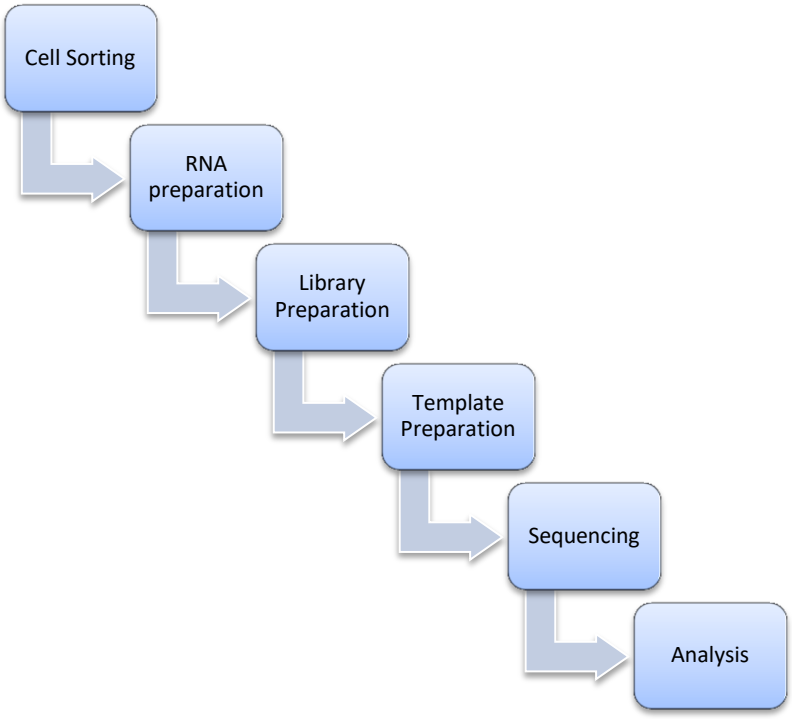



*Figure 2.Outline of methods following cell culture. Three groups of cells are collected following cell sorting, BCSCs, BCSC depleted cells, and Sorted control. RNA preparation, library preparation, template preparation and sequencing are performed on the same day to reduce batch effects per sample group.*

RNA Extraction- PureLink RNA mini-kit and protocol (Life Technologies) is used to obtain pure total RNA. To maintain RNA integrity cells are allowed to rest for thirty minutes after addition of lysis buffer rather than homogenization as suggested in the protocol. This is done to maintain RNA integrity so it does not create very small fragments when treated with RNase III during library preparation. RNA was extracted for further analysis. Using the Qubit fluorimeter it is possible to determine the concentration of RNA obtained. This number is used to confirm that there is a sufficient amount of RNA to be used in sequencing analysis (100ng-1ug of RNA). All RNA samples were diluted to approximately 980ng to be used as input for the Dynabead mRNA Direct kit. The Dynabead mRNA Direct kit was used to isolate only the

mRNA, using positive selection. After sequencing resulted in a small amount of reads, it was decided that a ribosomal RNA (rRNA) depletion kit should be used to preserve the rest of the RNA, which may have been damaged by attaching and removing beads. Analysis of size distribution can only be performed after library completion. It is worth noting that a concentration reading can be taken, but may be detrimental due to the small amounts obtained after isolation. As such, it is not recommended therefore the assessments will be performed via the tape station.

The tape station is similar to gel electrophoresis, miniaturized and automated. Small aliquots of the sample (1 uL) after cDNA transcription and purification are run through a gel and fragments are separated by size and compared to a ladder. The ladder is used to compare to the sample and provides various pieces of information, including size of fragmented cDNA and concentration of the fragments. The size distribution and peak molarity of each size of fragments are used for template preparation. Large quantities of fragments were observed in the 25 base pair range. High peak molarity is observed in the 25 base pair range. This showed that the RNA was too fragmented before library construction, because it is expected that there should be less than 50% of DNA in 50-160bp range. This resulted in the alteration of the RNA extraction protocol. Extraction protocol then changed to allow vortexed cells to sit for 30 minutes. Additionally, the incubation time in RNase III was reduced from 3 minutes to 1 minute.

Ribosomal RNA Depletion- Low Input RiboMinus kit is used to remove rRNA. This enables the analysis to be performed on the entire RNA spectrum and require a positive selection process that may damage the RNA.

RNA sequencing- Ion Total RNA-Seq Kit v2 and protocol for ribosome-depleted RNA will be used for library preparation. This process includes fragmenting the whole transcriptome

RNA and constructing the whole transcriptome library. Template preparation is performed using Ion PI Template OT2 200 Kit v3 and protocol. Ion PI Sequencing 200 Kit v3 and protocol for the Ion Proton is used for chip preparation and RNA sequencing on the Ion Proton system. To prepare the template the Ion One Touch 2 system is used followed by the Ion One Touch ES. Low Input RiboMinus kit and Ion sequencing kits are from Life Technologies. For final analysis, total RNA input was used. Three samples of each total RNA from BCSC, Sorted control, and Depleted control were sequenced.

Down Stream Analysis- The program Cufflinks was used to analyze transcript data, using at least a 0.05 p-value and no minimum number of reads as criteria for considering transcripts as differentially expressed (Trapnell et al., 2010). Cufflinks were merged to use Cuffdiff, a part of cufflinks to identify differentially expressed transcripts. For single replicate analysis, CuffDiff was performed using a blind sampling method appropriate for when only one replicate is used. Gene annotation and functional analysis are done using DAVID (Huang, Sherman, & Lempicki, 2009a, 2009b). Individual analysis was completed to compare individual sequencing results for each pair, BCSC/Sort, BCSC/Depleted, and Sort/Depleted. Combined analysis was also performed using each replicate per condition rather than blind due to the presence of replicates.

# CHAPTER 3

## RESULTS

Cells with the desired profile ($CD^{24low/-}$, $CD^{45-}$, and $CD^{44+}$) were isolated via traditional MACS sorting techniques and used for further analysis. The number of cells obtained varied depending on the number of cells present at the beginning of the sorting procedure. The preferred number of cells to use for the sorting protocol is above forty million cells. This number is based on findings that a sufficient number of stem-like cells are only obtained following the sorting of several million cells (data not shown). Since the focus of this study is on gene activity, overall cellular viability is maintained throughout the procedure. It is important to note that how long the procedure takes, as well as the temperature of the MACS buffer solution both of which may vary during each individual sort can influence the viability of the isolated cells.

Figure 3 demonstrates that within the cell line of interest, MDA-MB231, there are sub-populations of cells that express the three associated cell surface markers of BCSCs. Determination of the presence of these cells was completed via flow cytometry and utilizing three different fluorochrome-conjugated antibodies, one for each surface marker. Figure 3A shows a population of the events for the surface markers $CD^{45}$ and $CD^{24}$, which are compared to each other. Figure 3B demonstrates the number of events from figure 3A that are not included in

the box with a dotted line or gate, which demonstrates an estimate of the number of events which are $CD^{24-/low}$ and $CD^{45-}$. The population of stem cells lies within the region containing 23.5% of events of which 16.5% are breast cancer stem-like cells. This shows that 3.87% of the total identifiable cells analyzed were BCSCs. The upper left quadrant in Figure 3B may also demonstrate a proportion of BCSCs because the $CD^{24}$ surface marker is not always absent in this population and may contain low amounts of $CD^{24}$.

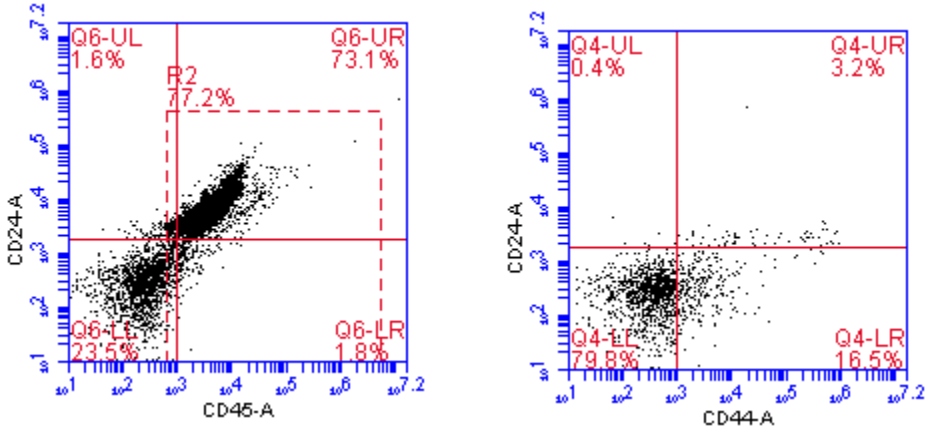



*Figure 3 A. A Cell Flow cytometry stain. Cells that show positive for the stain are in the upper and lower right quadrants or Q6-UR and Q6-LR. The dotted line box is used to gate for all the events or cells that are $CD^{24+}$ and $CD^{44+}$*

*Figure 3 B. A Cell Flow cytometry stain. Cells that show positive for the stain are in the upper and lower right quadrants or Q4-UR and Q4-LR. These are the remaining events after excluding the events from in the gated area from Figure 2. The events in the upper and lower right quadrant represent the sub-population of cells considered Breast Cancer Stem Cells that are $CD^{24-/low}$, $CD^{45-}$, and $CD^{44+}$.*

The results of a sequencing run can be seen in Appendix C. Analysis of the sequencing results show several transcripts being differentially expressed in each pair grouping including, BCSC-to-Sort, BCSC-to-Depleted, and Sort-to-Depleted (see Appendix A). There are several genes of potential interest. Individual results demonstrate a significant difference between each group. Main groups with differential expressed gene consist of membrane, mitochondrial, and metal ion binding proteins. After utilizing three replicates from BCSC and Sort control

conditions no significant differences in transcript expression were identified. Tables 1 through 3 demonstrate the number of genes present in an individual cluster created by DAVID from the lists of differentially expressed genes in each grouping (Huang et al., 2009a, 2009b) (Full list of clusters in Appendix B). The most prominent groups among the BCSC-Sorted control comparison is binding proteins, nucleus and transcript regulation proteins and proteins involved in signaling, disulfide bonds and glycoproteins. The differences in the sort and depleted controls can also give some insight into BCSCs as the sort control contains a small portion of BCSCs. There are a significant number of transcripts represented that function on non-membrane bound organelle and post-translation modification proteins. Finally, the clusters of transcripts from the comparison between Depleted controls and BCSC show only three clusters of transcripts showing transcripts involved in the mitochondria, ion, and membrane bound proteins. As can be seen from the flow cytometry data in Table 2, the BCSC population is a small proportion of the total number of cells present in the sample, which can amplify any inconsistencies in the protocols. These inconsistencies can result in an incomplete or an improperly represented transcriptome, which could have caused the results seen from the differential expression comparison between BCSC and the depleted populations. The single replicate differential expression cluster analysis from DAVID between BCSC to depleted populations does show several clusters of transcripts involved in mitochondria, ion binding and membrane proteins. This may demonstrate some important transcripts involved in BCSC properties.

*Table 1. Differentially expressed transcripts grouped into related categories from BCSC against Sorted control set from single replicate analysis.*

| **Number of genes per annotation cluster** |
|---|
| Endoplasmic reticulum (6) |
| Binding proteins (9) |
| Nucleus and transcript regulation(9) |
| Catabolic processes (5) |
| Signal, disulfide bonds and glycoproteins (10) |
| Organelle and endomembrane (4) |
| Plasma Membrane (4) |
| ion, metal and cation binding (5) |

*Table 2. Differentially expressed transcripts grouped into related categories from Depleted control against Sorted control set from single replicate analysis.*

| **Number of Genes per annotation cluster** |
|---|
| Complex macromolecule assembly (10) |
| Growth regulation (7) |
| Cell proliferation (10) |
| Non-membrane bound organelle (18) |
| Biosynthesis (5) |
| Chromosome (7) |
| Vesicle and protein transport (17) |
| Mitochondria (10) |
| Ribosome (5) |
| Nucleus (12) |
| Response to wound and abiotic stimulus (6) |
| Cell-cell communication (6) |
| DNA repair (6) |
| Post-translational modifications and phosphorylation (18) |
| Organelle envelopes (8) |
| Inflammation response (7) |
| Transcript regulation (16) |
| Cell Cycle (7) |
| Leukocyte activation (4) |
| Metabolic process and apoptosis/programmed cell death regulation (14) |
| Cell movement (3) |
| Positive regulation of biosynthesis and metabolism (5) |
| Cytoskeleton (6) |
| Synapse and neurons (4) |
| adhesion and immune response (13) |
| Metal binding proteins (13) |
| Plasma membrane (5) |
| Transmembrane proteins (24) |

*Table 3. Differentially expressed transcripts grouped into related categories from BCSC against depleted set from single replicate analysis.*

| **Number of Genes per annotation cluster** |
|---|
| Mitochondrial genes (4) |
| Ion binding proteins (3) |
| Membrane protein (4) |

# CHAPTER 4

## DISCUSSION

Flow cytometry results, Figure 3, demonstrate that the initial population of MDA-MB-231 cells contain the sub-population of cells with the desired surface markers, $CD^{44+}/CD^{24-/low}/CD^{45-}$. The flow cytometry data showed only ~4% of the total cells represent the sub-population with the desired surface markers. In a study using flow cytometry to identify stem-like cells, not including $CD^{45}$ in their experimental analysis of MDA-MB-231 cells and similar cell types demonstrate that 90% of the population express the phenotype $CD^{44+}/CD^{24-/low}$ phenotype (Fillmore & Kuperwasser, 2008). Other studies have found similar results looking at the same phenotype. In the current $CD^{45}$ was added to the selection for stem-like cells and to flow cytometry assays. The surface marker, $CD^{45}$, is used to identify tumor infiltrating leukyocytes. This suggests that a large portion of the cells identified in previous studies are leukyocytes rather than a rare sub-population of cancer stem-like cells. If flow cytometry were to be performed following cell sorting, the efficacy of the process could have been measured and determine the purity of the sub-population collected. This would give additional insight into the source of some of the variability within samples.

The initial sequencing results, using whole transcriptome RNA sequencing of the established cell line MDA-MB-231, indicate many potential areas of interest. Sequencing results show twenty differentially expressed genes between BCSCs and their depleted counter-parts.

Several genes are connected with DNA maintenance and other cancers, whereas many other have no relation and could have potentially interesting relations to cancer stem-like phenotypes.

Analysis comparing three replicates showed no significant differences. This data however is only from a single replicate differential expression analysis and a more thorough analysis from more than three samples may prove more enlightening. Additionally, this does not accurately include any sequences shorter than 100 base pairs due to the magnetic beads and ethanol used during library preparation. To obtain an accurate representation of small RNA's special modifications or kits would be required for sequencing. Comparisons between BCSC-to-sort control and Sort-control-to-Depleted have shown a far larger number of differentially expressed genes. This may be due to the lack of stresses on the sort control samples during sorting. However, it is unclear exactly why that happened. Many of the differentially expressed transcripts seen in the comparison between BCSC and Depleted controls are also found in the comparison between BCSC and Sorted control. This is due to the presence of BCSC in the sort control in small numbers. These numbers are not significant enough to interfere with the analysis.

It is important to note that the results of single replicate analysis comparing differentially expressed transcripts demonstrates few significant changes between BCSCs and the bulk tumor. Interestingly, the differences between BCSC and the Sorted control are more pronounced than the comparison with the Depleted controls. These differences may be attributed to batch effect. Batch effect refers to the varying results in experiments often observed when experiments are performed at different times. Example causes often include changes in the environment, such as temperature and humidity. This effect is most pronounced in RNA due to its lack of stability during common sequencing manipulations such as poly-A tail selection or reverse transcription

polymerase chain reaction. To limit batch effect, procedures were performed in sequence until cDNA was produced and samples were frozen only when a stopping point was indicated within the procedure. However, some groups of sequences were performed on separate days within the same fashion as previously described. Additional variability may be a result of varying times within ethanol based buffers during library preparation. When performing protocols with multiple samples, a timer are used to limit variability of the effects of ethanol based solutions. The amount of time that ethanol is present in the samples determines the size of DNA or RNA reads that are attached to the beads; beads collect shorter lengths of DNA/RNA more readily and therefore a longer period of time within ethanol based solutions are required to select for the proper size for sequencing. Varying times within solution can result in varying lengths of reads.

Furthermore, using all three replicates for each condition (BCSC and Depleted control) and using CuffDiff resulted in no significant differences among transcript or gene expression. A significant difference in transcript expression is determined by a p-value of 0.05. The lack of differences among each condition could result from the sorting protocol used or the randomize sampling method for transcript selection during library preparation. As stated previously, sorting was performed using Magnetically-Activated Cell Sorting (MACS). Compared to Flourescence-Activated Cell Sorting (FACS), MACS based sorting have varying results between the two types of sorting (Herrid, Davey, Hutton, Colditz, & Hill, 2009). This is due to the amount of time required to perform MACS which can result in declines in cellular viability. To prevent cell death or decreases in viability due to being outside of growth medium, ice cold phosphate buffered saline (PBS) is used to preserve the cells. If the PBS is allowed to sit long enough, it can return to room temperature and become ineffective. Additonally, the depleted cells need to be removed from the columns while the magnetic beads are still attached to be collected. This

removal can potentially tear open the cell membrane and kill cells surviving from the process. Finally, prior to freezing the cells, they must be allowed to relax in growth medium to allow the cells to chew remove any remaining magnetic beads. Cell death can occur during this process and any samples with media improperly removed prior to freezing can degrade the RNA and DNA contained within the cells.

Just as there can be variability from the initial procedure to select for a sub-population of cells there can be variability during the sequencing preparation procedures. For example, there are a large number of differentially expressed transcripts, there is a significant difference between the Sorted to Depleted controls compared to other comparisons. This may be in part due to the presence of BCSCs and other sub-populations present in the sort control samples, which create a significant difference. Additionally, a large proportion of genes present under one condition are not present in the other, resulting in a significant difference, and a logfold change of infinite by CuffDiff as seen in Tables 5 and 6. More replicates may be needed in order to include many of these missing transcripts, as there is a significant loss of material during library preparation and bead enrichment. During enrichment ~25% of the total beads are enriched with the library to prevent polyclonality during sequencing resulting in a size selection process where the smaller reads have a higher affinity to attach to the beads. If there are a large number of smalled reads then they will more readily be attached to the beads creating a bias for smaller reads. Appendix C shows the sequencing results from each test. An important piece of information shown in Table 10 is the median read size, which is 82 base pairs. This is a little short but is still within what is expected considering the adaptors sequences as seen in Figure 4 being removed from the total size of the read during analysis. The total size of the read cannot

exceed ~200 base pair for the Ion Torrent system when incorporated with the adaptor and primer sequences.

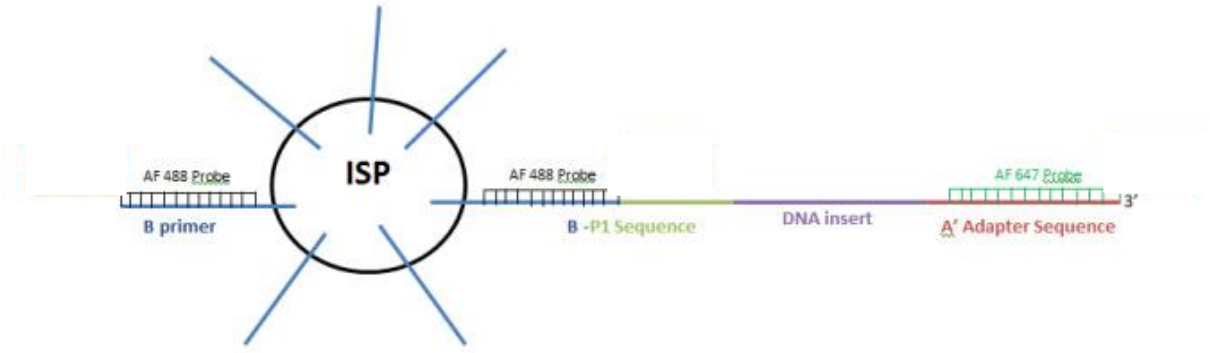



*Figure 4. Example of sequence following enrichment process demonstrating the amount of excess sequence incorporated for each individual read (Ion Pi OT2 200 v2 kit user guide).*

## Differential Expressed Genes Related to BCSC Function and Pathways

The comparison between BCSC and Depleted control is of significant importance to understanding how to develop treatments for BCSCs. Out of 101,289 transcripts identified from single replicate differential expression analysis, 20 transcripts were found to be significantly expressed in either population. Two transcripts for histone cluster 1 and 2, HIST1H2BK and HIST2H2BE respectively, were shown to be downregulated in BCSC populations. HIST1H2BK has previously shown to be overexpressed in micrometastasis, which are cells that maintain a very low population at a distant sight in ER-positive breast cancers. HIST1H2BE has been shown to be downregulated in ER-positive breast cancers (R. S. Kim et al., 2012; Tripathi et al., 2008). Transcripts for the metabolic enzyme and a member acyl-CoA dehydrogenase family, ACADSB, have been shown to be upregulated in ER-postive breast cancers(Rozen et al., 1994; Tozlu et al., 2006). As can be seen in appendix A Table 6, ACADSB is not expressed in BCSCs. If it is truly absent, it may alter the metabolic activity in BCSCs.

Another transcript that may be of interest when studying BCSCs is ST6GALNAC6, which is a sialyltransferase that modifies proteins on the cellular membrane. This transcript is downregulated in BCSCs. It has been shown to be downregulated in renal cancers and is partially responsible for synthesis of a carbohydrate antigen, disiayl Lewis a. Its isomer, sialyl Lewis a (CA19-9), is used as a serum biomarker to diagnose cancers in digestive organs (Kannagi, 2007). Further studies could similarly link this transcript to being a biomarker for breast cancer. Its interactions with the cellular membrane may be important during EMT and metastasis.

In comparison of BCSC to BCSC depleted samples, TRAF5 transcripts are seen to be downregulated in BCSCs. TRAF5 is a signal transducer involved in a large role of physiological functions, including B and T cell activation, inflamation and cell survival. TRAF5 plays a role in NF-κB and TNF pathways which have been shown to have significant roles in regulating BCSC activity and EMT (Au & Yeh, 2007). TRAF5 was also shown to be downregulated in breast cancer tissue compared to breast tissue in insulin-like growth factor-1 (IGF-1) deficient mice (Tang et al., 2008). If TRAF5 is implicated with BCSC this could lead to further investigation of IGF-1 and its implications in BCSC survival.

Only three transcripts were upregulated in BCSCs including, AK124970, visin-like 1 (VSNL1), and interleukin 17 receptor B (IL17RB). AK124970 is a gene with no annotation or function associated with it. IL17B has been shown to mediate the NF-κB pathway and could be potentially implicated in BCSC maintenance and EMT which may support evidence for it being a potential target for therapy (Popova, Kzhyshkowska, Nurgazieva, Goerdt, & Gratchev, 2012).VSNL1 is a tumor suppressor gene which has been shown to be important in neuroblastoma and has been implicated in colorectal cancers. VSNL1 normally regulates the inhibition of rhodospins and almost solely expressed in the brain (Akagi et al., 2012). It has been

show to be overexpressed in metastatic neuroblastomas and has been associated with $CD^{44}$ expression (Y. Xie et al., 2007). These interactions between VSNL1 and $CD^{44}$ may also be functioning in BCSCs to increase metastatic potential.

## Additional Differentially Expressed Genes

Other genes that are not related to BCSC maintenance and growth include; PYCR2, ECHS1, HPS5, COX14, SNBNP25, MVB12A, NAGK, SCG2, TCTA, MRPL1, and SHFM1. PYCR2 is a mitochondrial enzyme at the end of the proline biosythesis pathway. PYCR2 and its partner PYCR1 have been implicated in protecting cancer cells from oxidative stress (Kuo et al., 2016). Although not directly related to mitochondrial function and size, PYCR2 may be a significant player in BCSC maintenance through amino acid production.

ECHS1 (enoyl-CoA hydratase short chain 1) has been linked to breast cancer by using MCF7 cells. 4-amino-5-(4-chlorophenyl)-7-(t-butyl)pyrazolo[3,4-d]pyrimidine (PP2) downregulates expression of ECHS1 and peroxiredoxin 3 (PRDX3) to induce apoptosis in cancer cells. Furthermore, it was observed that bcl-2 family proteins are altered when apoptosis is induced, which suggests mitochondria mediated apoptosis (X. Liu, Feng, & Du, 2010). This study links ECHS1 to the mitochondria which has been demonstrated to be linked to BCSC function and maintenance. Observing the effects of ECHS1 in MCF7 mammospheres and other BCSC assays have the potential to link this protein to BCSC maintenance.

HPS5 is a gene related to Hermansky-Pudlak syndrome, which results in albinism in the eye and platelet deficiency leading to bruising and colitis (Krisp, Hoffman, Happle, König, & Freyschmidt-Paul, 2000). HPS5 also has no direct correlation to BCSC however, it has been identified to have increased expression in deceased patients compared to patients surviving for at least ten years (Jonsson et al., 2015). HPS5 is significantly down regulated in BCSCs which

would be unreliable as a diagnostic marker for BCSCs. SNRNP25, subunit U12 splicesome, splices U12 type introns (Kaida et al., 2010). This gene is completely absent in BCSC and be a potential biomarker for identifying potential BCSCs.

COX14, cytochrome oxidase subunit fourteen, works in conjunction with Coa3 to inhibit translation and prevent cellular respiration of COX1, cytochrome oxidase subunit one (Mick et al., 2010). COX14 is absent in BCSC samples, suggesting the heavy reliance on mitochondrial function in BCSCs. MVB12A is a subunit of ESCRT-1, a mediator of ubiquitinated cargo proteins, and is responsible for regulating EGF receptor bound to ESCRT-1 (Tsunematsu et al., 2010). It is unclear how this may be related to BCSC, as it is completely absent in BCSC samples. Since this is a housekeeping function of the cell, it is unlikely that this gene is truly absent in the sample. More likely, because these results are only from a single replicate; the gene may have not been present in high enough quantities or may be present but not be represented in the sample taken from the BCSCs used. MRPL1 (mitochondrial ribosomal protein L1) is also a mitochondrial protein that lacks expression in the BCSC samples. Decreased expression of MRPL1 was also found in cisplastin resistant ovarian cancers (Cheng et al., 2010). As previously discussed, the presence of chemotherapeutic drugs can alter gene expression and potentially increase stem-like cancer cell expression. The lack of MRPL1 may be a result of increased growth of stem-like cancer cells due to increased mitochondrial activity.

NAGK is another gene that is absent in the BCSC sample set. This protein plays several roles including metabolic pathways and golgi transport. It digests N-acetylglucosamine during lysosomal degradation (Shi, Allewell, & Tuchman, 2015). This means that NAGK is involved in salvaging materials for the cell to reuse and recycle, an important characteristic in cancer cells requiring new materials for growth. As $CD^{44+}/CD^{24-/low}$ cells are primarily quiescent, they are not

as metabolically active as $ALDH^+$ cells. The quiescent nature of this group of BCSCs could explain the lack of NAGK in the sub-population selected for sequencing.

SCG2 (secretogranin II) is a gene absent in BCSC samples. SCG2 functions to help sorting or packaging of peptide hormones and is also considered a suppressor candidate gene which is expressed in all tissues. It has been seen to induce cancer cell proliferation (Utispan et al., 2010). Lacking the SCG2 gene can potentially improve cancer and BCSC growth if it were confirmed to be a tumor suppressor gene.

TCTA (T-cell leukemia translocation associated) gene is also absent from the BCSC sample population. TCTA has been associated with lukemia and lung cancer, where it has been shown to be involved in osteoclastogenesis. As the bone marrow is a common metastatic site for lung cancer, TCTA may be a potential target to treat metastasis of lung cancer (Kotake, Yago, Kawamoto, & Nanke, 2012). Similarly, as bone is a common metastatic site for breast cancer, TCTA could also be a potential target to treat metastasis. MDA-MB-231 cell line is not normally metastatic and this characteristic may be why TCTA is absent in the cell lines (Zhang, Fidler, & Price, 1990). However, expression has been shown to change during EMT and following metastasis. Expression analysis of metastatic BCSC may prove fruitful to multiple gene sets including TCTA. Finally, SHFM1 (split hand foot mutation 1) is associated with BRCA2 and is required for functionality and stability. It has been determined that there is not a direct link between SHFM1, and breast or ovarian cancer (Bonache et al., 2013). In this study it was found to be absent in BCSC sample populations.

# CHAPTER 5

## CONCLUSION AND FUTURE DIRECTIONS

By studying the transcriptome of breast cancer and cancer stem-like cells, exppression differences can be identified in genes that were previously thought to be unrelated. This research has identified many potential candidates for further study. These genes need to be confirmed as active in BCSCs through additional analysis and sequencing. A small panel of custom primers for the candidates genes would further validate the expression profile of these transcripts. Western blots should be used to determine the presence or absence of gene products as proteins. This can be followed by analysis of genes through knockdown or knock-out studies in culture, which can be performed to initially determine the function and requirements of transcripts for BCSC function and maintenance. siRNAs (silencing RNAs) can be used in a knockdown model to prevent translation of mRNA or function of untranslated RNAs. This information can also be used to compare treated and untreated BCSC populations when looking at the drug effectiveness or determining expression changes following treatment.

The next generation sequencing technology being used is highly sensitive, quick and high through-put. Additionally, looking at cancer at the molecular level can be used to inform pathways and other related protein in a multitude of ways. The pathways I would specifically like to look at are pathways involved in telomerase function in elongating telomeres and

additional functions. By looking at proteins like telomerase that affect many types of cancer, this research has the possibility to be applied beyond breast cancer into other types of cancer and informing scientists about telomerase in general. This new data will have the potential to improve future treatments, which can reduce the time for chemotherapy treatments and metastasis rates, further decreasing the risk of damaging the body and mortality in patients.

# REFERENCES


Abraham, B. K., Fritz, P., McClellan, M., Hauptvogel, P., Athelogou, M., & Brauch, H. (2005). Prevalence of CD44+/CD24-/low cells in breast cancer may not be associated with clinical outcome but may favor distant metastasis. *Clin Cancer Res, 11*(3), 1154-1159.

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., . . . Eberhard, C. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research*, gkw343.

Akagi, T., Hijiya, N., Inomata, M., Shiraishi, N., Moriyama, M., & Kitano, S. (2012). Visinin-like protein-1 overexpression is an indicator of lymph node metastasis and poor prognosis in colorectal cancer patients. International Journal of Cancer, 131(6), 1307-1317.

Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., & Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proceedings of the National Academy of Sciences, 100*(7), 3983-3988.

Alspach, E., Fu, Y., & Stewart, S. A. (2013). Senescence and the pro-tumorigenic stroma. *Crit Rev Oncog, 18*(6), 549-558.

Alsuliman, A., Colak, D., Al-Harazi, O., Fitwi, H., Tulbah, A., Al-Tweigeri, T., . . . Ghebeh, H. (2015). Bidirectional crosstalk between PD-L1 expression and epithelial to mesenchymal transition: significance in claudin-low breast cancer cells. *Mol Cancer, 14*, 149. doi:10.1186/s12943-015-0421-2

Au, P. Y., & Yeh, W. C. (2007). Physiological roles and mechanisms of signaling by TRAF2 and TRAF5. Adv Exp Med Biol, 597, 32-47. doi:10.1007/978-0-387-70630-6_3

Balic, M., Lin, H., Young, L., Hawes, D., Giuliano, A., McNamara, G., . . . Cote, R. J. (2006). Most early disseminated cancer cells detected in bone marrow of breast cancer patients have a putative breast cancer stem cell phenotype. *Clinical Cancer Research, 12*(19), 5615-5621.

Bamodu, O. A., Huang, W. C., Lee, W. H., Wu, A., Wang, L. S., Hsiao, M., . . . Chao, T. Y. (2016). Aberrant KDM5B expression promotes aggressive breast cancer through MALAT1 overexpression and downregulation of hsa-miR-448. *BMC Cancer, 16*(1), 160. doi:10.1186/s12885-016-2108-5

Bansal, N., Petrie, K., Christova, R., Chung, C. Y., Leibovitch, B. A., Howell, L., . . . Waxman, S. (2015). Targeting the SIN3A-PF1 interaction inhibits epithelial to mesenchymal transition and maintenance of a stem cell phenotype in triple negative breast cancer. *Oncotarget, 6*(33), 34087-34105. doi:10.18632/oncotarget.6048

Bao, L., Cardiff, R. D., Steinbach, P., Messer, K. S., & Ellies, L. G. (2015). Multipotent luminal mammary cancer stem cells model tumor heterogeneity. *Breast Cancer Res, 17*(1), 137. doi:10.1186/s13058-015-0615-y

Bonache, S., de la Hoya, M., Gutierrez-Enriquez, S., Tenes, A., Masas, M., Balmana, J., & Diez, O. (2013). Mutation analysis of the SHFM1 gene in breast/ovarian cancer families. *J Cancer Res Clin Oncol, 139*(3), 529-532. doi:10.1007/s00432-013-1385-5

Bonnet, D., & Dick, J. E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med, 3*(7), 730-737.

Boo, L., Ho, W. Y., Ali, N. M., Yeap, S. K., Ky, H., Chan, K. G., . . . Cheong, S. K. (2016). MiRNA Transcriptome Profiling of Spheroid-Enriched Cells with Cancer Stem Cell Properties in Human Breast MCF-7 Cell Line. *Int J Biol Sci, 12*(4), 427-445. doi:10.7150/ijbs.12777

Borcherding, N., Bormann, N., Kusner, D., Kolb, R., & Zhang, W. (2015). Transcriptome analysis of basal and luminal tumor-initiating cells in ErbB2-driven breast cancer. *Genom Data, 4*, 119-122. doi:10.1016/j.gdata.2015.04.008

Brooks, D. L., Schwab, L. P., Krutilina, R., Parke, D. N., Sethuraman, A., Hoogewijs, D., . . . Seagroves, T. N. (2016). ITGA6 is directly regulated by hypoxia-inducible factors and enriches for cancer stem cell activity and invasion in metastatic breast cancer models. *Mol Cancer, 15*(1), 26. doi:10.1186/s12943-016-0510-x

Chambers, A., Groom, A., & MacDonald, I. (2002). ā€ Dissemination and Growth of Cancer Cells in Metastatic Sites, ā€ Nat. *Rev. Cancer, 2*, 563ā.

Chao, M. W., Chu, P. C., Chuang, H. C., Shen, F. H., Chou, C. C., Hsu, E. C., . . . Chen, C. S. (2015). Non-epigenetic function of HDAC8 in regulating breast cancer stem cells by maintaining Notch1 protein stability. *Oncotarget*. doi:10.18632/oncotarget.6427

Charafe-Jauffret, E., Ginestier, C., Iovino, F., Tarpin, C., Diebel, M., Esterni, B., . . . Wicha, M. S. (2010). Aldehyde dehydrogenase 1-positive cancer stem cells mediate metastasis and poor clinical outcome in inflammatory breast cancer. *Clin Cancer Res, 16*(1), 45-55. doi:10.1158/1078-0432.CCR-09-1630

Chen, C., Cao, F., Bai, L., Liu, Y., Xie, J., Wang, W., . . . Luo, Y. (2015). IKKbeta Enforces a LIN28B/TCF7L2 Positive Feedback Loop That Promotes Cancer Cell Stemness and Metastasis. *Cancer Res, 75*(8), 1725-1735. doi:10.1158/0008-5472.can-14-2111

Cheng, L., Lu, W., Kulkarni, B., Pejovic, T., Yan, X., Chiang, J. H., . . . Lin, B. (2010). Analysis of chemotherapy response programs in ovarian cancers by the next generation sequencing technologies. *Gynecol Oncol, 117*(2), 159-169. doi:10.1016/j.ygyno.2010.01.041Cho, Y., Lee, H. W., Kang, H. G., Kim, H. Y., Kim, S. J., & Chun, K. H. (2015). Cleaved CD44 intracellular domain supports activation of stemness factors and promotes tumorigenesis of breast cancer. *Oncotarget, 6*(11), 8709-8721. doi:10.18632/oncotarget.3325

Cho, Y., Lee, H. W., Kang, H. G., Kim, H. Y., Kim, S. J., & Chun, K. H. (2015). Cleaved CD44 intracellular domain supports activation of stemness factors and promotes tumorigenesis of breast cancer. Oncotarget, 6(11), 8709-8721. doi:10.18632/oncotarget.3325

Creighton, C. J., Li, X., Landis, M., Dixon, J. M., Neumeister, V. M., Sjolund, A., . . . Herschkowitz, J. I. (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proceedings of the National Academy of Sciences, 106*(33), 13820-13825.

Croker, A. K., & Allan, A. L. (2012). Inhibition of aldehyde dehydrogenase (ALDH) activity reduces chemotherapy and radiation resistance of stem-like ALDHhiCD44+ human breast cancer cells. *Breast Cancer Res Treat, 133*(1), 75-87.

Croker, A. K., Goodale, D., Chu, J., Postenka, C., Hedley, B. D., Hess, D. A., & Allan, A. L. (2009). High aldehyde dehydrogenase and expression of cancer stem cell markers selects for breast cancer cells with enhanced malignant and metastatic ability. *Journal of cellular and molecular medicine, 13*(8b), 2236-2252.

Croker, A. K., Goodale, D., Chu, J., Postenka, C., Hedley, B. D., Hess, D. A., & Allan, A. L. (2009). High aldehyde dehydrogenase and expression of cancer stem cell markers selects for breast cancer cells with enhanced malignant and metastatic ability. *J Cell Mol Med, 13*(8B), 2236-2252. doi:10.1111/j.1582-4934.2008.00455.

D'Angelo, R. C., Ouzounova, M., Davis, A., Choi, D., Tchuenkam, S. M., Kim, G., . . . Korkaya, H. (2015). Notch reporter activity in breast cancer cell lines identifies a subset of cells with stem cell activity. *Mol Cancer Ther, 14*(3), 779-787. doi:10.1158/1535-7163.MCT-14-0228

De Cola, A., Volpe, S., Budani, M. C., Ferracin, M., Lattanzio, R., Turdo, A., . . . De Laurenzi, V. (2015). miR-205-5p-mediated downregulation of ErbB/HER receptors in breast cancer stem cells results in targeted therapy resistance. *Cell Death Dis, 6*, e1823. doi:10.1038/cddis.2015.192

De Luca, A., Fiorillo, M., Peiris-Pages, M., Ozsvari, B., Smith, D. L., Sanchez-Alvarez, R., . . . Sotgia, F. (2015). Mitochondrial biogenesis is required for the anchorage-independent survival and propagation of stem-like cancer cells. *Oncotarget, 6*(17), 14777-14795. doi:10.18632/oncotarget.4401

Debeb, B. G., Lacerda, L., Larson, R., Wolfe, A. R., Krishnamurthy, S., Reuben, J. M., . . . Woodward, W. A. (2016). Histone deacetylase inhibitor-induced cancer stem cells exhibit high pentose phosphate pathway metabolism. *Oncotarget*. doi:10.18632/oncotarget.8631

Debeb, B. G., Lacerda, L., Xu, W., Larson, R., Solley, T., Atkinson, R., . . . Woodward, W. A. (2012). Histone deacetylase inhibitors stimulate dedifferentiation of human breast cancer cells through WNT/beta-catenin signaling. *Stem Cells, 30*(11), 2366-2377. doi:10.1002/stem.1219

DeSantis, C., Ma, J., Bryan, L., & Jemal, A. (2014). Breast cancer statistics, 2013. CA Cancer J Clin, 64(1), 52-62.

DeSantis, C. E., Fedewa, S. A., Goding Sauer, A., Kramer, J. L., Smith, R. A., & Jemal, A. (2015). Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. CA Cancer J Clin, 66(1), 31-42. doi:10.3322/caac.21320

Farmer, P., Bonnefoi, H., Anderle, P., Cameron, D., Wirapati, P., Becette, V., . . . Brain, E. (2009). A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nature medicine, 15*(1), 68-74.

Farnie, G., Sotgia, F., & Lisanti, M. P. (2015). High mitochondrial mass identifies a sub-population of stem-like cancer cells that are chemo-resistant. *Oncotarget, 6*(31), 30472-30486. doi:10.18632/oncotarget.5401

Feng, Z. M., Qiu, J., Chen, X. W., Liao, R. X., Liao, X. Y., Zhang, L. P., . . . Sun, J. G. (2015). Essential role of miR-200c in regulating self-renewal of breast cancer stem cells and their counterparts of mammary epithelium. *BMC Cancer, 15*, 645. doi:10.1186/s12885-015-1655-5

Fillmore, C. M., & Kuperwasser, C. (2008). Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Res, 10*(2), R25. doi:10.1186/bcr1982

Finlay-Schultz, J., Cittelly, D. M., Hendricks, P., Patel, P., Kabos, P., Jacobsen, B. M., . . . Sartorius, C. A. (2015). Progesterone downregulation of miR-141 contributes to expansion of stem-like breast cancer cells through maintenance of progesterone receptor and Stat5a. *Oncogene, 34*(28), 3676-3687. doi:10.1038/onc.2014.298

French, R., Hayward, O., Jones, S., Yang, W., & Clarkson, R. (2015). Cytoplasmic levels of cFLIP determine a broad susceptibility of breast cancer stem/progenitor-like cells to TRAIL. *Mol Cancer, 14*, 209. doi:10.1186/s12943-015-0478-y

Gernapudi, R., Yao, Y., Zhang, Y., Wolfson, B., Roy, S., Duru, N., . . . Zhou, Q. (2015). Targeting exosomes from preadipocytes inhibits preadipocyte to cancer stem cell signaling in early-stage breast cancer. *Breast Cancer Res Treat, 150*(3), 685-695. doi:10.1007/s10549-015-3326-2

Ghebeh, H., Sleiman, G. M., Manogaran, P. S., Al-Mazrou, A., Barhoush, E., Al-Mohanna, F. H., . . . Adra, C. N. (2013). Profiling of normal and malignant breast tissue show CD44high/CD24low phenotype as a predominant stem/progenitor marker when used in combination with Ep-CAM/CD49f markers. *BMC Cancer, 13*, 289. doi:10.1186/1471-2407-13-289

Ginestier, C., Hur, M. H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., . . . Dontu, G. (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell stem cell, 1*(5), 555-567. doi:10.1016/j.stem.2007.08.014

Ginestier, C., Liu, S., Diebel, M. E., Korkaya, H., Luo, M., Brown, M., . . . Birnbaum, D. (2010). CXCR1 blockade selectively targets human breast cancer stem cells in vitro and in xenografts. *The Journal of clinical investigation, 120*(2), 485-497.

Gopal, K., Gupta, N., Zhang, H., Alshareef, A., Alqahtani, H., Bigras, G., . . . Lai, R. (2015). Oxidative stress induces the acquisition of cancer stem-like phenotype in breast cancer detectable by using a Sox2 regulatory region-2 (SRR2) reporter. *Oncotarget*. doi:10.18632/oncotarget.6630

Gordon, N., Skinner, A. M., Pommier, R. F., Schillace, R. V., O'Neill, S., Peckham, J. L., . . . Pommier, S. J. (2015). Gene expression signatures of breast cancer stem and progenitor cells do not exhibit features of Warburg metabolism. *Stem Cell Res Ther, 6*, 157. doi:10.1186/s13287-015-0153-7

Han, D., Wu, G., Chang, C., Zhu, F., Xiao, Y., Li, Q., . . . Zhang, L. (2015). Disulfiram inhibits TGF-beta-induced epithelial-mesenchymal transition and stem-like features in breast cancer via ERK/NF-kappaB/Snail pathway. *Oncotarget, 6*(38), 40907-40919. doi:10.18632/oncotarget.5723

Han, N. K., Shin, D. H., Kim, J. S., Weon, K. Y., Jang, C. Y., & Kim, J. S. (2016). Hyaluronan-conjugated liposomes encapsulating gemcitabine for breast cancer stem cells. *Int J Nanomedicine, 11*, 1413-1425. doi:10.2147/ijn.s95850

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell, 144*(5), 646-674. doi:10.1016/j.cell.2011.02.013

Hardt, O., Wild, S., Oerlecke, I., Hofmann, K., Luo, S., Wiencek, Y., . . . Dittmer, J. (2012). Highly sensitive profiling of CD44+/CD24- breast cancer stem cells by combining global mRNA amplification and next generation sequencing: evidence for a hyperactive PI3K pathway. *Cancer Lett, 325*(2), 165-174. doi:10.1016/j.canlet.2012.06.010

Herbert, B. S., Wright, W. E., & Shay, J. W. (2001). Telomerase and breast cancer. *Breast Cancer Res, 3*(3), 146-149.

Herrid, M., Davey, R. J., Hutton, K., Colditz, I. G., & Hill, J. R. (2009). A comparison of methods for preparing enriched populations of bovine spermatogonia. Reprod Fertil Dev, 21(3), 393-399.

Howlader N, Noone AM, Krapcho M, et al. eds. SEER Cancer Statistics Review, 1975-2012 [seer.cancer.gov/csr.1975_2012/, based on the November 2014 SEER data submission, posted to the SEER website April 2015]. Bethesda, MD: National Cancer Institute; 2015.

Hsu, E. C., Kulp, S. K., Huang, H. L., Tu, H. J., Salunke, S. B., Sullivan, N. J., . . . Chen, C. S. (2015). Function of Integrin-Linked Kinase in Modulating the Stemness of IL-6-Abundant Breast Cancer Cells by Regulating gamma-Secretase-Mediated Notch1 Activation in Caveolae. *Neoplasia, 17*(6), 497-508. doi:10.1016/j.neo.2015.06.001

Hu, P., Yang, J., Hou, Y., Zhang, H., Zeng, Z., Zhao, L., . . . Liu, M. (2014). LncRNA expression signatures of twist-induced epithelial-to-mesenchymal transition in MCF10A cells. *Cell Signal, 26*(1), 83-93. doi:10.1016/j.cellsig.2013.10.001

Hu, Y., Guo, R., Wei, J., Zhou, Y., Ji, W., Liu, J., . . . Zhang, J. (2015). Effects of PI3K inhibitor NVP-BKM120 on overcoming drug resistance and eliminating cancer stem cells in human breast cancer cells. *Cell Death Dis, 6*, e2020. doi:10.1038/cddis.2015.363

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res, 37*(1), 1-13. doi:10.1093/nar/gkn923

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research, 37*(1), 1-13.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols, 4*(1), 44-57.

Hwang-Verslues, W. W., Kuo, W.-H., Chang, P.-H., Pan, C.-C., Wang, H.-H., Tsai, S.-T., . . . Chen, C.-H. (2009). Multiple lineages of human breast cancer stem/progenitor cells identified by profiling with stem cell markers. *PloS one, 4*(12), e8377.

Iriondo, O., Rabano, M., Domenici, G., Carlevaris, O., Lopez-Ruiz, J. A., Zabalza, I., . . . Vivanco, M. (2015). Distinct breast cancer stem/progenitor cell populations require either HIF1alpha or loss of PHD3 to expand under hypoxic conditions. *Oncotarget, 6*(31), 31721-31739. doi:10.18632/oncotarget.5564

Jang, G. B., Hong, I. S., Kim, R. J., Lee, S. Y., Park, S. J., Lee, E. S., . . . Nam, J. S. (2015). Wnt/beta-Catenin Small-Molecule Inhibitor CWP232228 Preferentially Inhibits the Growth of Breast Cancer Stem-like Cells. *Cancer Res, 75*(8), 1691-1702. doi:10.1158/0008-5472.CAN-14-2041

Jang, G. B., Kim, J. Y., Cho, S. D., Park, K. S., Jung, J. Y., Lee, H. Y., . . . Nam, J. S. (2015). Blockade of Wnt/beta-catenin signaling suppresses breast cancer metastasis by inhibiting CSC-like phenotype. *Sci Rep, 5*, 12465. doi:10.1038/srep12465

Jang, J. Y., Kim, M. K., Jeon, Y. K., Joung, Y. K., Park, K. D., & Kim, C. W. (2012). Adenovirus adenine nucleotide translocator-2 shRNA effectively induces apoptosis and enhances chemosensitivity by the down-regulation of ABCG2 in breast cancer stem-like cells. *Exp Mol Med, 44*(4), 251-259. doi:10.3858/emm.2012.44.4.019

Jia, D., Yang, W., Li, L., Liu, H., Tan, Y., Ooi, S., . . . Wang, L. (2015). beta-Catenin and NF-kappaB co-activation triggered by TLR3 stimulation facilitates stem cell-like phenotypes in breast cancer. *Cell Death Differ, 22*(2), 298-310. doi:10.1038/cdd.2014.145

Jonsson, P., Coarfa, C., Mesmar, F., Raz, T., Rajapakshe, K., Thompson, J. F., . . . Williams, C. (2015). Single-molecule sequencing reveals estrogen-regulated clinically relevant lncRNAs in breast cancer. *Molecular Endocrinology, 29*(11), 1634-1645

Kaida, D., Berg, M. G., Younis, I., Kasim, M., Singh, L. N., Wan, L., & Dreyfuss, G. (2010). U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature, 468*(7324), 664-668.

Kang, L., Mao, J., Tao, Y., Song, B., Ma, W., Lu, Y., . . . Li, L. (2015). MicroRNA-34a suppresses the breast cancer stem cell-like characteristics by downregulating Notch1 pathway. *Cancer Sci, 106*(6), 700-708. doi:10.1111/cas.12656

Kang, Y., Siegel, P. M., Shu, W., Drobnjak, M., Kakonen, S. M., Cordón-Cardo, C., . . . Massagué, J. (2003). A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell, 3*(6), 537-549.

Kannagi, R. (2007). Carbohydrate antigen sialyl Lewis a--its pathophysiological significance and induction mechanism in cancer progression. Chang Gung Med J, 30(3), 189-209.

Kastrati, I., Canestrari, E., & Frasor, J. (2015). PHLDA1 expression is controlled by an estrogen receptor-NFkappaB-miR-181 regulatory loop and is essential for formation of ER+ mammospheres. *Oncogene, 34*(18), 2309-2316. doi:10.1038/onc.2014.180.

Ke, J., Zhao, Z., Hong, S. H., Bai, S., He, Z., Malik, F., . . . Liu, S. (2015). Role of microRNA221 in regulating normal mammary epithelial hierarchy and breast cancer stem-like cells. *Oncotarget, 6*(6), 3709-3721. doi:10.18632/oncotarget.2888

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol, 14*(4), R36.

Kim, J., Villadsen, R., Sørlie, T., Fogh, L., Grønlund, S. Z., Fridriksdottir, A. J., . . . Solvang, H. (2012). Tumor initiating but differentiated luminal-like breast cancer cells are highly invasive in the absence of basal-like activity. *Proceedings of the National Academy of Sciences, 109*(16), 6124-6129.

Kim, S. H., & Singh, S. V. (2015). The role of polycomb group protein Bmi-1 and Notch4 in breast cancer stem cell inhibition by benzyl isothiocyanate. *Breast Cancer Res Treat, 149*(3), 681-692. doi:10.1007/s10549-015-3279-5.

Kotake, S., Yago, T., Kawamoto, M., & Nanke, Y. (2012). The role of T-cell leukemia translocation-associated gene protein in human tumorigenesis and osteoclastogenesis. J Biomed Biotechnol, 2012, 675317. doi:10.1155/2012/675317

Krisp, A., Hoffman, R., Happle, R., König, A., & Freyschmidt-Paul, P. (2000). Hermansky-Pudlak syndrome. *European journal of dermatology: EJD, 11*(4), 372-373.

Kuo, M.-L., Lee, M. B.-E., Tang, M., Den Besten, W., Hu, S., Sweredoski, M. J., . . . Su, M. (2016). PYCR1 and PYCR2 interact and collaborate with RRM2B to protect cells from overt oxidative stress. Sci Rep, 6.

Lamb, R., Bonuccelli, G., Ozsvari, B., Peiris-Pages, M., Fiorillo, M., Smith, D. L., . . . Lisanti, M. P. (2015). Mitochondrial mass, a new metabolic biomarker for stem-like cancer cells: Understanding WNT/FGF-driven anabolic signaling. *Oncotarget, 6*(31), 30453-30471. doi:10.18632/oncotarget.5852

Lamb, R., Fiorillo, M., Chadwick, A., Ozsvari, B., Reeves, K. J., Smith, D. L., . . . Lisanti, M. P. (2015). Doxycycline down-regulates DNA-PK and radiosensitizes tumor initiating cells: Implications for more effective radiation therapy. *Oncotarget, 6*(16), 14005-14025. doi:10.18632/oncotarget.4159

Lamb, R., Ozsvari, B., Bonuccelli, G., Smith, D. L., Pestell, R. G., Martinez-Outschoorn, U. E., . . . Lisanti, M. P. (2015). Dissecting tumor metabolic heterogeneity: Telomerase and large cell size metabolically define a sub-population of stem-like, mitochondrial-rich, cancer cells. *Oncotarget, 6*(26), 21892-21905. doi:10.18632/oncotarget.5260

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. Nature methods, 9(4), 357-359.

Lawson, D. A., Bhakta, N. R., Kessenbrock, K., Prummel, K. D., Yu, Y., Takai, K., . . . Werb, Z. (2015). Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature, 526*(7571), 131-135. doi:10.1038/nature15260

Lee, E., Yang, J., Ku, M., Kim, N. H., Park, Y., Park, C. B., . . . Cheong, J. H. (2015). Metabolic stress induces a Wnt-dependent cancer stem cell-like state transition. *Cell Death Dis, 6*, e1805. doi:10.1038/cddis.2015.171

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14), 1754-1760.

Li, Q.-Q., Xu, J.-D., Wang, W.-J., Cao, X.-X., Chen, Q., Tang, F., . . . Xu, Z.-D. (2009). Twist1-mediated adriamycin-induced epithelial-mesenchymal transition relates to multidrug resistance and invasive potential in breast cancer cells. *Clinical Cancer Research, 15*(8), 2657-2665.

Lin, C. C., Sharma, S. B., Farrugia, M. K., McLaughlin, S. L., Ice, R. J., Loskutov, Y. V., . . . Ruppert, J. M. (2015). Kruppel-like factor 4 signals through microRNA-206 to promote tumor initiation and cell survival. *Oncogenesis, 4*, e155. doi:10.1038/oncsis.2015.8

Liu, S., Cong, Y., Wang, D., Sun, Y., Deng, L., Liu, Y., . . . Landis, M. D. (2014). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem cell reports, 2*(1), 78-91.

Liu, T., Hu, K., Zhao, Z., Chen, G., Ou, X., Zhang, H., . . . Liu, C. (2015). MicroRNA-1 down-regulates proliferation and migration of breast cancer stem cells by inhibiting the Wnt/beta-catenin pathway. *Oncotarget, 6*(39), 41638-41649. doi:10.18632/oncotarget.5873

Liu, X., Feng, R., & Du, L. (2010). The role of enoyl-CoA hydratase short chain 1 and peroxiredoxin 3 in PP2-induced apoptosis in human breast cancer MCF-7 cells. *FEBS letters, 584*(14), 3185-3192.

Lu, H., Samanta, D., Xiang, L., Zhang, H., Hu, H., Chen, I., . . . Semenza, G. L. (2015). Chemotherapy triggers HIF-1-dependent glutathione synthesis and copper chelation that induces the breast cancer stem cell phenotype. *Proc Natl Acad Sci U S A, 112*(33), E4600-4609. doi:10.1073/pnas.1513433112

Lv, X., Wang, Y., Song, Y., Pang, X., & Li, H. (2016). Association between ALDH1+/CD133+ stem-like cells and tumor angiogenesis in invasive ductal breast carcinoma. *Oncol Lett, 11*(3), 1750-1756. doi:10.3892/ol.2016.4145

Ma, W., Xiao, G. G., Mao, J., Lu, Y., Song, B., Wang, L., . . . Li, L. (2015). Dysregulation of the miR-34a-SIRT1 axis inhibits breast cancer stemness. *Oncotarget, 6*(12), 10432-10444. doi:10.18632/oncotarget.3394

Makki, J. (2015). Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. *Clin Med Insights Pathol, 8*, 23-31. doi:10.4137/CPath.S31563

Mannello, F. (2013). Understanding breast cancer stem cell heterogeneity: time to move on to a new research paradigm. *BMC Med, 11*, 169. doi:10.1186/1741-7015-11-169

Marotta, L. L., Almendro, V., Marusyk, A., Shipitsin, M., Schemme, J., Walker, S. R., . . . Polyak, K. (2011). The JAK2/STAT3 signaling pathway is required for growth of CD44(+)CD24(-) stem cell-like breast cancer cells in human tumors. *J Clin Invest, 121*(7), 2723-2735. doi:10.1172/JCI44745

Martin-Castillo, B., Lopez-Bonet, E., Cuyas, E., Vinas, G., Pernas, S., Dorca, J., & Menendez, J. A. (2015). Cancer stem cell-driven efficacy of trastuzumab (Herceptin): towards a reclassification of clinically HER2-positive breast carcinomas. *Oncotarget, 6*(32), 32317-32338. doi:10.18632/oncotarget.6094

Maycotte, P., Jones, K. L., Goodall, M. L., Thorburn, J., & Thorburn, A. (2015). Autophagy Supports Breast Cancer Stem Cell Maintenance by Regulating IL6 Secretion. *Mol Cancer Res, 13*(4), 651-658. doi:10.1158/1541-7786.mcr-14-0487

Meng, F., Speyer, C. L., Zhang, B., Zhao, Y., Chen, W., Gorski, D. H., . . . Wu, G. (2015). PDGFRalpha and beta play critical roles in mediating Foxq1-driven breast cancer stemness and chemoresistance. *Cancer Res, 75*(3), 584-593. doi:10.1158/0008-5472.can-13-3029

Mick, D. U., Vukotic, M., Piechura, H., Meyer, H. E., Warscheid, B., Deckers, M., & Rehling, P. (2010). Coa3 and Cox14 are essential for negative feedback regulation of COX1 translation in mitochondria. *The Journal of cell biology, 191*(1), 141-154.

Monteiro, J., Gaspar, C., Richer, W., Franken, P. F., Sacchetti, A., Joosten, R., . . . Fodde, R. (2014). Cancer stemness in Wnt-driven mammary tumorigenesis. *Carcinogenesis, 35*(1), 2-13. doi:10.1093/carcin/bgt279

Morel, A. P., Lievre, M., Thomas, C., Hinkal, G., Ansieau, S., & Puisieux, A. (2008). Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PloS one, 3*(8), e2888. doi:10.1371/journal.pone.0002888

Nandy, S. B., Arumugam, A., Subramani, R., Pedroza, D., Hernandez, K., Saltzstein, E., & Lakshmanaswamy, R. (2015). MicroRNA-125a influences breast cancer stem cells by targeting leukemia inhibitory factor receptor which regulates the Hippo signaling pathway. *Oncotarget, 6*(19), 17366-17378. doi:10.18632/oncotarget.3953

Niida, A., Smith, A. D., Imoto, S., Aburatani, H., Zhang, M. Q., & Akiyama, T. (2009). Gene set-based module discovery in the breast cancer transcriptome. *BMC Bioinformatics, 10*, 71. doi:10.1186/1471-2105-10-71

Niu, D. G., Peng, F., Zhang, W., Guan, Z., Zhao, H. D., Li, J. L., . . . Liu, Q. (2015). Morphine promotes cancer stem cell properties, contributing to chemoresistance in breast cancer. *Oncotarget, 6*(6), 3963-3976. doi:10.18632/oncotarget.2894

Owens, T. W., & Naylor, M. J. (2013). Breast cancer stem cells. *Front Physiol, 4*, 225. doi:10.3389/fphys.2013.00225

Park, S. Y., Kim, M. J., Park, S. A., Kim, J. S., Min, K. N., Kim, D. K., . . . Sheen, Y. Y. (2015). Combinatorial TGF-beta attenuation with paclitaxel inhibits the epithelial-to-mesenchymal transition and breast cancer stem-like cells. *Oncotarget, 6*(35), 37526-37543. doi:10.18632/oncotarget.6063

Pathania, R., Ramachandran, S., Elangovan, S., Padia, R., Yang, P., Cinghu, S., . . . Thangaraju, M. (2015). DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis. *Nat Commun, 6*, 6910. doi:10.1038/ncomms7910

Patrawala, L., Calhoun, T., Schneider-Broussard, R., Zhou, J., Claypool, K., & Tang, D. G. (2005). Side population is enriched in tumorigenic, stem-like cancer cells, whereas ABCG2+ and ABCG2- cancer cells are similarly tumorigenic. *Cancer Res, 65*(14), 6207-6219. doi:10.1158/0008-5472.can-05-0592

Petrelli, A., Carollo, R., Cargnelutti, M., Iovino, F., Callari, M., Cimino, D., . . . Giordano, S. (2015). By promoting cell differentiation, miR-100 sensitizes basal-like breast cancer stem cells to hormonal therapy. *Oncotarget, 6*(4), 2315-2330. doi:10.18632/oncotarget.2962

Phillips, T. M., McBride, W. H., & Pajonk, F. (2006). The response of CD24−/low/CD44+ breast cancer–initiating cells to radiation. *J Natl Cancer Inst, 98*(24), 1777-1785.

Pietila, M., Vijay, G. V., Soundararajan, R., Yu, X., Symmans, W. F., Sphyris, N., & Mani, S. A. (2016). FOXC2 regulates the G2/M transition of stem cell-rich breast cancer cells and sensitizes them to PLK1 inhibition. *Sci Rep, 6*, 23070. doi:10.1038/srep23070

Ponti, D., Costa, A., Zaffaroni, N., Pratesi, G., Petrangolini, G., Coradini, D., . . . Daidone, M. G. (2005). Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. *Cancer Res, 65*(13), 5506-5511. doi:10.1158/0008-5472.CAN-05-0626

Popova, A., Kzhyshkowska, J., Nurgazieva, D., Goerdt, S., & Gratchev, A. (2012). Smurf2 regulates IL17RB by proteasomal degradation of its novel binding partner DAZAP2. Immunobiology, 217(3), 321-328.

Prat, A., Adamo, B., Fan, C., Peg, V., Vidal, M., Galvan, P., . . . Cortes, J. (2013). Genomic analyses across six cancer types identify basal-like breast cancer as a unique molecular entity. *Sci Rep, 3*, 3544. doi:10.1038/srep03544

Quesenberry, P., & Levitt, L. (1979). Hematopoietic stem cells. *New England Journal of Medicine, 301*(14), 755-760.

Ricardo, S., Vieira, A. F., Gerhard, R., Leitão, D., Pinto, R., Cameselle-Teijeiro, J. F., . . . Paredes, J. (2011). Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype. *J Clin Pathol*, jcp. 2011.090456.

Roscigno, G., Quintavalle, C., Donnarumma, E., Puoti, I., Diaz-Lagares, A., Iaboni, M., . . . Condorelli, G. (2015). MiR-221 promotes stemness of breast cancer cells by targeting DNMT3b. *Oncotarget*. doi:10.18632/oncotarget.5979

Rozen, R., Vockley, J., Zhou, L., Milos, R., Willard, J., Fu, K., . . . Fournier, B. (1994). Isolation and expression of a cDNA encoding the precursor for a novel member (ACADSB) of the acyl-CoA dehydrogenase gene family. Genomics, 24(2), 280-287.

Ryoo, I. G., Choi, B. H., & Kwak, M. K. (2015). Activation of NRF2 by p62 and proteasome reduction in sphere-forming breast carcinoma cells. *Oncotarget, 6*(10), 8167-8184. doi:10.18632/oncotarget.3047

Sai, S., Vares, G., Kim, E. H., Karasawa, K., Wang, B., Nenoi, M., . . . Hayashi, M. (2015). Carbon ion beam combined with cisplatin effectively disrupts triple negative breast cancer stem-like cells in vitro. *Mol Cancer, 14*, 166. doi:10.1186/s12943-015-0429-7

Sansone, P., Storci, G., Tavolari, S., Guarnieri, T., Giovannini, C., Taffurelli, M., . . . Marcu, K. B. (2007). IL-6 triggers malignant features in mammospheres from human ductal breast carcinoma and normal mammary gland. *The Journal of clinical investigation, 117*(12), 3988.

Shah, M., & Allegrucci, C. (2012). Keeping an open mind: highlights and controversies of the breast cancer stem cell theory. *Breast Cancer (Dove Med Press), 4*, 155-166. doi:10.2147/BCTT.S26434

Sheridan, C., Kishimoto, H., Fuchs, R. K., Mehrotra, S., Bhat-Nakshatri, P., Turner, C. H., . . . Nakshatri, H. (2006). CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis. *Breast Cancer Res, 8*(5), R59. doi:10.1186/bcr1610

Shi, D., Allewell, N. M., & Tuchman, M. (2015). The N-Acetylglutamate Synthase Family: Structures, Function and Mechanisms. *Int J Mol Sci, 16*(6), 13004-13022. doi:10.3390/ijms160613004

Shimono, Y., Mukohyama, J., Nakamura, S., & Minami, H. (2015). MicroRNA Regulation of Human Breast Cancer Stem Cells. *J Clin Med, 5*(1). doi:10.3390/jcm5010002

Shimono, Y., Zabala, M., Cho, R. W., Lobo, N., Dalerba, P., Qian, D., . . . Chiao, E. (2009). Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell, 138*(3), 592-603.

Shipitsin, M., Campbell, L. L., Argani, P., Weremowicz, S., Bloushtain-Qimron, N., Yao, J., . . . Polyak, K. (2007). Molecular definition of breast tumor heterogeneity. *Cancer Cell, 11*(3), 259-273. doi:10.1016/j.ccr.2007.01.013

Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. *CA Cancer J Clin, 65*(1), 5-29. doi:10.3322/caac.21254

Smit, L., Berns, K., Spence, K., Ryder, W. D., Zeps, N., Madiredjo, M., . . . Clarke, R. B. (2015). An integrated genomic approach identifies that the PI3K/AKT/FOXO pathway is involved in breast cancer tumor initiation. *Oncotarget*. doi:10.18632/oncotarget.6354

So, J. Y., Wahler, J., Das Gupta, S., Salerno, D. M., Maehr, H., Uskokovic, M., & Suh, N. (2015). HES1-mediated inhibition of Notch1 signaling by a Gemini vitamin D analog leads to decreased CD44(+)/CD24(-/low) tumor-initiating sub-population in basal-like breast cancer. *J Steroid Biochem Mol Biol, 148*, 111-121. doi:10.1016/j.jsbmb.2014.12.013

Sun, X., Jiang, S., Liu, J., Wang, H., Zhang, Y., Tang, S. C., . . . Ren, H. (2015). MiR-208a stimulates the cocktail of SOX2 and beta-catenin to inhibit the let-7 induction of self-renewal repression of breast cancer stem cells and formed miR208a/let-7 feedback loop via LIN28 and DICER1. *Oncotarget, 6*(32), 32944-32954. doi:10.18632/oncotarget.5079

Sun, Z., Zhang, C., Zou, X., Jiang, G., Xu, Z., Li, W., & Xie, H. (2015). Special AT-rich sequence-binding protein-1 participates in the maintenance of breast cancer stem cells through regulation of the Notch signaling pathway and expression of Snail1 and Twist1. *Mol Med Rep, 11*(5), 3235-3542. doi:10.3892/mmr.2015.3192

Tachi, K., Shiraishi, A., Bando, H., Yamashita, T., Tsuboi, I., Kato, T., . . . Ohneda, O. (2015). FOXA1 expression affects to the proliferation activity of luminal breast cancer stem cell populations. *Cancer Sci*. doi:10.1111/cas.12870

Takahashi, R. U., Miyazaki, H., Takeshita, F., Yamamoto, Y., Minoura, K., Ono, M., . . . Ochiya, T. (2015). Loss of microRNA-27b contributes to breast cancer stem cell generation by activating ENPP1. *Nat Commun, 6*, 7318. doi:10.1038/ncomms8318

Tanei, T., Morimoto, K., Shimazu, K., Kim, S. J., Tanji, Y., Taguchi, T., . . . Noguchi, S. (2009). Association of breast cancer stem cells identified by aldehyde dehydrogenase 1 expression with resistance to sequential Paclitaxel and epirubicin-based chemotherapy for breast cancers. *Clin Cancer Res, 15*(12), 4234-4241. doi:10.1158/1078-0432.CCR-08-1479

Tang, H. B., Ye, Z. R., Ren, Y. P., Zhang, S., Chen, G. Q., & Wu, Y. P. (2008). [Analysis of apoptosis-related gene expression in different serum level of insulin-like growth factor-1 in mice breast cancer tissue]. Zhonghua Wai Ke Za Zhi, 46(13), 1018-1021.

Tozlu, S., Girault, I., Vacher, S., Vendrell, J., Andrieu, C., Spyratos, F., . . . Bieche, I. (2006). Identification of novel genes that co-cluster with estrogen receptor alpha in breast tumor biopsy specimens, using a large-scale real-time reverse transcription-PCR approach. Endocr Relat Cancer, 13(4), 1109-1120. doi:10.1677/erc.1.01120.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., . . . Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology, 28*(5), 511-515.

Tripathi, A., King, C., de la Morenas, A., Perry, V. K., Burke, B., Antoine, G. A., Hirsch, E. F., Kavanah, M., Mendez, J., Stone, M., Gerry, N. P., Lenburg, M. E. and Rosenberg, C. L. (2008), Gene expression abnormalities in histologically normal breast epithelium of breast cancer patients. Int. J. Cancer, 122: 1557–1566. doi: 10.1002/ijc.23267.

Tsunematsu, T., Yamauchi, E., Shibata, H., Maki, M., Ohta, T., & Konishi, H. (2010). Distinct functions of human MVB12A and MVB12B in the ESCRT-I dependent on their posttranslational modifications. *Biochemical and biophysical research communications, 399*(2), 232-237.

Utispan, K., Thuwajit, P., Abiko, Y., Charngkaew, K., Paupairoj, A., Chau-in, S., & Thuwajit, C. (2010). Gene expression profiling of cholangiocarcinoma-derived fibroblast reveals alterations related to tumor progression and indicates periostin as a poor prognostic marker. Mol Cancer, 9(1), 1.

Vuong, T., Mallet, J. F., Ouzounova, M., Rahbar, S., Hernandez-Vargas, H., Herceg, Z., & Matar, C. (2016). Role of a polyphenol-enriched preparation on chemoprevention of mammary carcinoma through cancer stem cells and inflammatory pathways modulation. *J Transl Med, 14*, 13. doi:10.1186/s12967-016-0770-7.

Wang, L., Tian, H., Yuan, J., Wu, H., Wu, J., & Zhu, X. (2015). CONSORT: Sam68 Is Directly Regulated by MiR-204 and Promotes the Self-Renewal Potential of Breast Cancer Cells by Activating the Wnt/Beta-Catenin Signaling Pathway. *Medicine (Baltimore), 94*(49), e2228. doi:10.1097/MD.0000000000002228.

Wang, R., Lv, Q., Meng, W., Tan, Q., Zhang, S., Mo, X., & Yang, X. (2014). Comparison of mammosphere formation from breast cancer cell lines and primary breast tumors. *J Thorac Dis, 6*(6), 829-837. doi:10.3978/j.issn.2072-1439.2014.03.38.

Wei, W., & Lewis, M. T. (2015). Identifying and targeting tumor-initiating cells in the treatment of breast cancer. Endocr Relat Cancer, 22(3), R135-155. doi:10.1530/erc-14-0447.

Williams, K. E., Bundred, N. J., Landberg, G., Clarke, R. B., & Farnie, G. (2015). Focal adhesion kinase and Wnt signaling regulate human ductal carcinoma in situ stem cell activity and response to radiotherapy. *Stem Cells, 33*(2), 327-341. doi:10.1002/stem.1843.

Wright, M. H., Calcagno, A. M., Salcido, C. D., Carlson, M. D., Ambudkar, S. V., & Varticovski, L. (2008). Brca1 breast tumors contain distinct CD44+/CD24- and CD133+ cells with cancer stem cell characteristics. *Breast Cancer Res, 10*(1), R10. doi:10.1186/bcr1855.

Xie, G., Yao, Q., Liu, Y., Du, S., Liu, A., Guo, Z., . . . Ye, C. (2012). IL-6-induced epithelial-mesenchymal transition promotes the generation of breast cancer stem-like cells analogous to mammosphere cultures. *Int J Oncol, 40*(4), 1171-1179.

Xie, J., Xiao, Y., Zhu, X. Y., Ning, Z. Y., Xu, H. F., & Wu, H. M. (2016). Hypoxia regulates stemness of breast cancer MDA-MB-231 cells. *Med Oncol, 33*(5), 42. doi:10.1007/s12032-016-0755-7.

Xie, Y., Chan, H., Fan, J., Chen, Y., Young, J., Li, W., . . . Tam, P. K. (2007). Involvement of visinin-like protein-1 (VSNL-1) in regulating proliferative and invasive properties of neuroblastoma. Carcinogenesis, 28(10), 2122-2130.

Ye, F., Qiu, Y., Li, L., Yang, L., Cheng, F., Zhang, H., . . . Bu, H. (2015). The Presence of EpCAM(-)/CD49f(+) Cells in Breast Cancer Is Associated with a Poor Clinical Outcome. *J Breast Cancer, 18*(3), 242-248. doi:10.4048/jbc.2015.18.3.242.

Ye, X., Tam, W. L., Shibue, T., Kaygusuz, Y., Reinhardt, F., Ng Eaton, E., & Weinberg, R. A. (2015). Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature, 525*(7568), 256-260. doi:10.1038/nature14897.

Zhang, R., Fidler, I., & Price, J. (1990). Relative malignant potential of human breast carcinoma cell lines established from pleural effusions and a brain metastasis. Invasion Metastasis, 11(4), 204-215.

Zhou, J., Zhang, H., Gu, P., Bai, J., Margolick, J. B., & Zhang, Y. (2008). NF-κB pathway inhibitors preferentially inhibit breast cancer stem-like cells. *Breast Cancer Res Treat, 111*(3), 419-427. doi:10.1007/s10549-007-9798-y.

Zhou, M., Hou, Y., Yang, G., Zhang, H., Tu, G., Du, Y. E., . . . Liu, M. (2016). LncRNA-Hh Strengthen Cancer Stem Cells Generation in Twist-Positive Breast Cancer via Activation of Hedgehog Signaling Pathway. *Stem Cells, 34*(1), 55-66. doi:10.1002/stem.2219.

# APPENDIX A: FULL LIST OF TRANSCRIPTS DIFFERENTIALLY EXPRESSED IN BREAST CANCER STEM CELLS

*Table 4 Sorted control against BCSC differential transcript expression analysis from total RNA MDA-MB-231 samples p-value= .05*

| FPKM BCSC | FPKM Sort Control | log2(fold change) | PValue | Official Gene ID |
|---|---|---|---|---|
| 0.000 | 5.430 | infinite | 5.00E-05 | STX12 |
| 0.000 | 30.901 | infinite | 0.0002 | ADPRHL2 |
| 0.000 | 9.312 | infinite | 0.0002 | ATP6V0B |
| 0.000 | 10.662 | infinite | 0.00015 | SELRC1 |
| 0.000 | 69.818 | infinite | 5.00E-05 | LRIF1 |
| 0.000 | 32.637 | infinite | 5.00E-05 | Unknown |
| 0.000 | 7.712 | infinite | 0.00015 | COX8A |
| 0.000 | 43.427 | infinite | 5.00E-05 | CCS |
| 0.000 | 19.234 | infinite | 5.00E-05 | RHOD |
| 0.000 | 162.038 | infinite | 5.00E-05 | COX14 |
| 0.867 | 37.906 | 5.45028 | 0.0002 | PIP4K2C |
| 0.000 | 8.367 | infinite | 0.0002 | RPLP0 |
| 0.000 | 5.969 | infinite | 0.0001 | ATP5EP2 |
| 0.000 | 13.740 | infinite | 5.00E-05 | RCBTB1 |
| 0.000 | 43.006 | infinite | 5.00E-05 | OXA1L |
| 0.000 | 13.689 | infinite | 5.00E-05 | RHOJ |
| 175.085 | 2349.770 | 3.74639 | 0.0002 | KIAA0247 |
| 0.000 | 152.374 | infinite | 5.00E-05 | MOAP1 |
| 0.000 | 15.745 | infinite | 0.0001 | DNAJA2 |
| 0.000 | 8.470 | infinite | 5.00E-05 | CTCF |
| 0.000 | 8.807 | infinite | 0.0002 | LIG3 |
| 0.000 | 5.611 | infinite | 5.00E-05 | RAB5C |
| 0.000 | 5.177 | infinite | 0.0001 | PCTP |
| 0.000 | 42.997 | infinite | 5.00E-05 | CBX8 |

| FPKM<br><br>BCSC | FPKM Sort Control | log2(fold change) | PValue | Official Gene ID |
|---|---|---|---|---|
| 0.000 | 5.328 | infinite | 5.00E-05 | KEAP1 |
| 0.000 | 9.782 | infinite | 5.00E-05 | Unknown |
| 13.963 | 0.215 | -6.0196 | 0.0002 | ASF1B |
| 0.000 | 54.121 | infinite | 5.00E-05 | NDUFB7 |
| 0.000 | 12.356 | infinite | 5.00E-05 | IMP4 |
| 0.000 | 17.005 | infinite | 0.0002 | GART |
| 0.000 | 19.062 | infinite | 5.00E-05 | MCM5 |
| 0.000 | 5.244 | infinite | 0.0002 | MB |
| 0.000 | 7.863 | infinite | 5.00E-05 | LARS2 |
| 0.000 | 12.517 | infinite | 0.0001 | TMEM115 |
| 0.000 | 18.916 | infinite | 5.00E-05 | SLIT2 |
| 0.000 | 44.109 | infinite | 5.00E-05 | STPG2 |
| 0.000 | 13.312 | infinite | 0.0002 | C4orf33 |
| 0.000 | 12.983 | infinite | 0.0001 | CSF2 |
| 0.000 | 7.666 | infinite | 0.00015 | SLC29A1 |
| 1.595 | 185.430 | 6.86086 | 5.00E-05 | CENPW |
| 0.000 | 43.401 | infinite | 5.00E-05 | EIF4H |
| 0.000 | 6.541 | infinite | 0.0001 | PODXL |
| 14.675 | 0.000 | infinite | 0.0002 | C9orf89 |
| 0.000 | 19.477 | infinite | 5.00E-05 | URM1 |
| 0.000 | 5.140 | infinite | 5.00E-05 | EXO1 |
| 0.000 | 8.926 | infinite | 5.00E-05 | LEO1 |
| 0.000 | 247.470 | infinite | 5.00E-05 | MPP1 |
| 62.469 | 0.000 | infinite | 0.00015 | PITHD1 |
| 0.000 | 18.581 | infinite | 5.00E-05 | CTH |
| 0.000 | 14.587 | infinite | 5.00E-05 | TRAF5 |
| 0.000 | 90.674 | infinite | 0.0002 | VPS11 |
| 1.535 | 26.575 | 4.11381 | 5.00E-05 | SMYD5 |
| 13.230 | 0.000 | infinite | 0.00015 | CCNT2 |
| 16.877 | 0.000 | infinite | 0.0002 | SPDYE5 |
| 0.788 | 17.264 | 4.45317 | 5.00E-05 | LINC00273 |
| 0.000 | 12.373 | infinite | 5.00E-05 | SMIM7 |
| 0.000 | 10.989 | infinite | 5.00E-05 | HOXD-AS1 |
| 47.436 | 0.000 | infinite | 0.0002 | Unknown |
| 0.000 | 23.366 | infinite | 5.00E-05 | LOC100289187 |

*Table 5 Sorted control against Depleted control differential transcript expression analysis from total RNA MDA-MB-231 samples p-value= .05.*

| Sort | Depleted | log2(fold change) | PValue | Official Gene Symbol |
|---|---|---|---|---|
| 0 | 14.5736 | infinite | 5.00E-05 | ATP6V0B |
| 0 | 54.6878 | infinite | 0.0002 | CTH |
| 0 | 22.1541 | infinite | 5.00E-05 | ZNF326 |
| 0 | 120.762 | infinite | 5.00E-05 | LRIF1 |
| 0 | 9.03163 | infinite | 5.00E-05 | DCLRE1B |
| 0 | 5.32466 | infinite | 5.00E-05 | ACP6 |
| 23.7246 | 0.854923 | -4.79444 | 5.00E-05 | HIST2H2BE |
| 903.779 | 206.917 | -2.12692 | 0.00035 | S100A6 |
| 0 | 29.3712 | infinite | 5.00E-05 | hCG_1995134 |
| 149.633 | 29.5554 | -2.33994 | 0.0003 | NUCKS1 |
| 9.97343 | 0 | infinite | 5.00E-05 | SPRTN |
| 0 | 47.1702 | infinite | 5.00E-05 | FAM175B |
| 0.892341 | 24.3483 | 4.77008 | 5.00E-05 | COPB1 |
| 10.7709 | 0 | infinite | 0.00025 | SDHAF2 |
| 0 | 5.42105 | infinite | 5.00E-05 | SSH3 |
| 0 | 3.93678 | infinite | 0.00015 | TSKU |
| 28.3809 | 0 | infinite | 0.0001 | GABARAPL1 |
| 0 | 21.509 | infinite | 5.00E-05 | IL23A |
| 0 | 19.2584 | infinite | 0.0003 | R3HDM2 |
| 0 | 7.25381 | infinite | 0.00025 | METTL21B |
| 0 | 36.9495 | infinite | 5.00E-05 | ANKRD13A |
| 0 | 16.5553 | infinite | 0.0002 | METTL17 |
| 0 | 17.9054 | infinite | 5.00E-05 | BRMS1L |
| 0 | 8.50464 | infinite | 5.00E-05 | ANGEL1 |
| 113.958 | 0 | infinite | 5.00E-05 | NOP10 |
| 14.3732 | 92.6582 | 2.68854 | 5.00E-05 | THBS1 |
| 158.855 | 0 | infinite | 0.00015 | RAD51 |
| 0 | 23.3966 | infinite | 5.00E-05 | SNUPN |
| 0 | 70.6958 | infinite | 0.0002 | POLR3K |
| 0 | 41.4 | infinite | 5.00E-05 | DNAJA2 |
| 0 | 14.0422 | infinite | 5.00E-05 | FAM64A |
| 0 | 12.5733 | infinite | 5.00E-05 | CWC25 |
| 0 | 17.6308 | infinite | 5.00E-05 | RAB5C |
| 0 | 5.03206 | infinite | 5.00E-05 | DBF4B |
| 0 | 33.5352 | infinite | 0.0001 | KIAA1468 |
| 95.4885 | 0 | infinite | 5.00E-05 | LOC284385 |
| 0 | 3.91333 | infinite | 0.00025 | KEAP1 |

| Sort | Depleted | log2(fold change) | PValue | Official Gene Symbol |
|---|---|---|---|---|
| 0 | 12.8283 | infinite | 5.00E-05 | ELOF1 |
| 0 | 9.80126 | infinite | 5.00E-05 | ASF1B |
| 0 | 44.7156 | infinite | 5.00E-05 | SMIM7 |
| 0 | 11.532 | infinite | 5.00E-05 | SERTAD3 |
| 0 | 25.26 | infinite | 5.00E-05 | AP2S1 |
| 22.4294 | 0 | infinite | 5.00E-05 | ZNF749 |
| 0 | 153.222 | infinite | 0.0001 | OST4 |
| 0 | 5.67077 | infinite | 5.00E-05 | NRBP1 |
| 0 | 39.713 | infinite | 5.00E-05 | MTIF2 |
| 0 | 48.973 | infinite | 0.00035 | MRPL53 |
| 0 | 4.78649 | infinite | 5.00E-05 | MTX2 |
| 32.5381 | 0 | infinite | 5.00E-05 | RBM45 |
| 15.4896 | 0 | infinite | 5.00E-05 | NDUFAF5 |
| 0 | 7.92634 | infinite | 0.00015 | SNHG11 |
| 7.4436 | 59.6445 | 3.00232 | 5.00E-05 | YWHAB |
| 99.1651 | 0 | infinite | 5.00E-05 | N6AMT1 |
| 0 | 29.5752 | infinite | 0.0001 | GART |
| 0 | 39.2429 | infinite | 5.00E-05 | CRYBB2P1 |
| 0 | 14.892 | infinite | 5.00E-05 | MCM5 |
| 0 | 10.2945 | infinite | 5.00E-05 | TUSC2 |
| 0 | 7.0496 | infinite | 5.00E-05 | TMEM115 |
| 0 | 44.2322 | infinite | 5.00E-05 | SENP7 |
| 0 | 36.2706 | infinite | 5.00E-05 | EXOC1 |
| 0 | 17.421 | infinite | 5.00E-05 | NEK1 |
| 0 | 63.8687 | infinite | 5.00E-05 | XRCC4 |
| 0 | 57.2516 | infinite | 5.00E-05 | CSF2 |
| 53.3957 | 0 | infinite | 5.00E-05 | CAMLG |
| 69.4105 | 0 | infinite | 5.00E-05 | DQ572964 |
| 0 | 15.3369 | infinite | 5.00E-05 | SLC29A1 |
| 113.692 | 0 | infinite | 5.00E-05 | NDUFAF4 |
| 78.3466 | 0 | infinite | 0.0002 | RPF2 |
| 0 | 4.8442 | infinite | 0.0001 | UST |
| 4.006 | 0 | infinite | 5.00E-05 | NOD1 |
| 0 | 164.676 | infinite | 5.00E-05 | LOC401397 |
| 0.751096 | 13.6474 | 4.18349 | 5.00E-05 | PODXL |
| 0 | 7.91369 | infinite | 5.00E-05 | GTF2E2 |
| 0 | 121.705 | infinite | 5.00E-05 | LACTB2 |
| 0 | 129.045 | infinite | 5.00E-05 | MRPL50 |
| 0 | 37.3246 | infinite | 5.00E-05 | SCML1 |
| 0 | 5.17447 | infinite | 5.00E-05 | EMD |

| Sort | Depleted | log2(fold change) | PValue | Official Gene Symbol |
|---|---|---|---|---|
| 0 | 4428.05 | infinite | 5.00E-05 | RNU11 |
| 219.605 | 0 | infinite | 0.0001 | CDKN3 |
| 0 | 32.308 | infinite | 5.00E-05 | LEO1 |
| 0 | 4.59445 | infinite | 0.0001 | AK307275 |
| 0.422272 | 8.68447 | 4.36219 | 5.00E-05 | LIN54 |
| 28.6631 | 0 | infinite | 5.00E-05 | SLC1A3 |
| 0 | 12.2574 | infinite | 5.00E-05 | HOMER1 |
| 0 | 4.6227 | infinite | 5.00E-05 | DDR1 |
| 0 | 9.93885 | infinite | 0.0001 | STAG3L4 |
| 413.691 | 63.2375 | -2.7097 | 5.00E-05 | SNIP1 |
| 0 | 56.5032 | infinite | 5.00E-05 | ACADM |
| 0 | 13.3562 | infinite | 5.00E-05 | RGL1 |
| 0 | 11.6728 | infinite | 5.00E-05 | VPS11 |
| 0 | 12.5255 | infinite | 0.00025 | CALCOCO1 |
| 66.5706 | 0 | infinite | 5.00E-05 | CRNDE |
| 0 | 40.5877 | infinite | 5.00E-05 | RPS6KB1 |
| 45.5391 | 0 | infinite | 5.00E-05 | THUMPD2 |
| 61.6338 | 0 | infinite | 5.00E-05 | NAPB |
| 41.7459 | 0 | infinite | 5.00E-05 | BC069212 |
| 0 | 33.3518 | infinite | 5.00E-05 | GRAMD1C |
| 11.3745 | 0 | infinite | 5.00E-05 | ST6GALNAC6 |
| 13.1699 | 3.13002 | -2.073 | 5.00E-05 | LINC00273 |
| 1.50742 | 42.9315 | 4.83189 | 5.00E-05 | HOXD-AS1 |
| 0 | 10.0946 | infinite | 5.00E-05 | DCP1A |
| 0 | 97.7524 | infinite | 5.00E-05 | SELK |
| 0 | 15.3071 | infinite | 5.00E-05 | MARCKSL1 |
| 15.187 | 0 | infinite | 0.0002 | PYCR2 |
| 545.344 | 51.977 | -3.39122 | 5.00E-05 | EXTL3 |

*Table 6 BCSC against Depleted control differential transcript expression analysis from total RNA MDA-MB-231 samples p-value= .05.*

| FPKM Depleted | FPKM BCSC | PValue | Official Gene Symbol |
|---|---|---|---|
| 22.9134 | 1.04254 | 5.00E-05 | HIST2H2BE |
| 0 | 77.5764 | 5.00E-05 | AK124970 |
| 7.8649 | 0 | 5.00E-05 | PYCR2 |
| 17.705 | 0 | 5.00E-05 | ECHS1 |
| 51.1197 | 0 | 5.00E-05 | HPS5 |
| 45.703 | 0 | 5.00E-05 | COX14 |
| 26.0536 | 0 | 5.00E-05 | SNRNP25 |
| 6.52765 | 0 | 5.00E-05 | MVB12A |
| 35.6936 | 0 | 5.00E-05 | TMEM18 |
| 0 | 15.2749 | 5.00E-05 | VSNL1 |
| 4.85082 | 0 | 5.00E-05 | NAGK |
| 34.3919 | 0 | 5.00E-05 | SCG2 |
| 24.6532 | 0 | 5.00E-05 | TCTA |
| 0 | 22.7473 | 5.00E-05 | IL17RB |
| 60.9154 | 0 | 5.00E-05 | MRPL1 |
| 60.7 | 1.23797 | 5.00E-05 | HIST1H2BK |
| 66.2796 | 0 | 5.00E-05 | SHFM1 |
| 5.68205 | 0 | 5.00E-05 | ST6GALNAC6 |
| 8.20245 | 0 | 5.00E-05 | TRAF5 |
| 27.1737 | 0 | 5.00E-05 | ACADSB |

# APPENDIX B: GENE CLUSTER RESULTS FROM DAVID OF DIFFERENTIAL EXPRESSED GENES

*Table 7. Annotation clusters developed by DAVID from differentially expressed transcripts in a comparison between BCSC-to-Depleted controls.*

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| Annotation Cluster 1 | Enrichment Score: 1.0185960900191642 | | | | |
| GOTERM_CC_FAT | GO:0005759~mitochondrial matrix | 3 | 1.6 | 0.018 | MRPL1, ACADSB, ECHS1 |
| GOTERM_CC_FAT | GO:0031980~mitochondrial lumen | 3 | 1.6 | 0.018 | MRPL1, ACADSB, ECHS1 |
| UP_SEQ_FEATURE | transit peptide:Mitochondrion | 3 | 1.6 | 0.057 | MRPL1, ACADSB, ECHS1 |
| SP_PIR_KEYWORDS | transit peptide | 3 | 1.6 | 0.058 | MRPL1, ACADSB, ECHS1 |
| GOTERM_CC_FAT | GO:0044429~mitochondrial part | 3 | 1.6 | 0.105 | MRPL1, ACADSB, ECHS1 |
| SP_PIR_KEYWORDS | Mitochondrion | 3 | 1.6 | 0.151 | MRPL1, ACADSB, ECHS1 |
| GOTERM_CC_FAT | GO:0070013~intracellular organelle lumen | 4 | 2.13 | 0.228 | MRPL1, ACADSB, SHFM1, ECHS1 |
| GOTERM_CC_FAT | GO:0043233~organelle lumen | 4 | 2.13 | 0.238 | MRPL1, ACADSB, SHFM1, ECHS1 |
| GOTERM_CC_FAT | GO:0031974~membrane-enclosed lumen | 4 | 2.13 | 0.248 | MRPL1, ACADSB, SHFM1, ECHS1 |
| GOTERM_CC_FAT | GO:0005739~mitochondrion | 3 | 1.6 | 0.272 | MRPL1, ACADSB, ECHS1 |
| Annotation Cluster 2 | Enrichment Score: 0.028152644680335865 | | | | |
| GOTERM_MF_FAT | GO:0046872~metal ion binding | 3 | 1.6 | 0.934 | VSNL1, TRAF5, SCG2 |
| GOTERM_MF_FAT | GO:0043169~cation binding | 3 | 1.6 | 0.937 | VSNL1, TRAF5, SCG2 |
| GOTERM_MF_FAT | GO:0043167~ion binding | 3 | 1.6 | 0.941 | VSNL1, TRAF5, |

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| | | | | | SCG2 |
| Annotation Cluster 3 | Enrichment Score: 0.022997709068004638 | | | | |
| UP_SEQ_FEATURE | topological domain:Cytoplasmic | 3 | 1.6 | 0.802 | ST6GALNAC6, TMEM18, IL17RB |
| UP_SEQ_FEATURE | transmembrane region | 3 | 1.6 | 0.944 | ST6GALNAC6, TMEM18, IL17RB |
| SP_PIR_KEYWORDS | Transmembrane | 3 | 1.6 | 0.945 | ST6GALNAC6, TMEM18, IL17RB |
| SP_PIR_KEYWORDS | alternative splicing | 4 | 2.13 | 0.978 | ST6GALNAC6, TMEM18, HPS5, IL17RB |
| UP_SEQ_FEATURE | splice variant | 4 | 2.13 | 0.978 | ST6GALNAC6, TMEM18, HPS5, IL17RB |
| SP_PIR_KEYWORDS | Membrane | 3 | 1.6 | 0.984 | ST6GALNAC6, TMEM18, IL17RB |
| GOTERM_CC_FAT | GO:0016021~integral to membrane | 3 | 1.6 | 0.985 | ST6GALNAC6, TMEM18, IL17RB |
| GOTERM_CC_FAT | GO:0031224~intrinsic to membrane | 3 | 1.6 | 0.988 | ST6GALNAC6, TMEM18, IL17RB |
| Not Clustered | | | | | |
| UP_SEQ_FEATURE | region of interest:Substrate binding | 3 | 1.6 | 0.005 | ACADSB, ECHS1, NAGK |
| GOTERM_BP_FAT | GO:0006952~defense response | 4 | 2.13 | 0.019 | HIST1H2BK, HIST2H2BE, IL17RB, SCG2 |
| SP_PIR_KEYWORDS | Acetylation | 5 | 2.66 | 0.166 | PYCR2, ACADSB, HIST1H2BK, HIST2H2BE, ECHS1 |
| GOTERM_CC_FAT | GO:0043228~non-membrane-bounded organelle | 4 | 2.13 | 0.453 | MRPL1, HIST1H2BK, HIST2H2BE, TRAF5 |
| GOTERM_CC_FAT | GO:0043232~intracellular non-membrane-bounded organelle | 4 | 2.13 | 0.453 | MRPL1, HIST1H2BK, HIST2H2BE, TRAF5 |
| SP_PIR_KEYWORDS | Polymorphism | 10 | 5.32 | 0.717 | MRPL1, ACADSB, HPS5, VSNL1, SHFM1, ECHS1, NAGK, TRAF5, IL17RB, SCG2 |

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| SP_PIR_KEYWORDS | Phosphoprotein | 6 | 3.19 | 0.784 | PYCR2, HPS5, HIST1H2BK, HIST2H2BE, NAGK, SCG2 |
| UP_SEQ_FEATURE | sequence variant | 10 | 5.32 | 0.789 | MRPL1, ACADSB, HPS5, VSNL1, SHFM1, ECHS1, NAGK, TRAF5, IL17RB, SCG2 |
| SP_PIR_KEYWORDS | Nucleus | 3 | 1.6 | 0.901 | HIST1H2BK, HIST2H2BE, SNRNP25 |

*Table 8. Annotation clusters developed by DAVID from differentially expressed transcripts in a comparison between BCSC-to-sort controls.*

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| Annotation Cluster 1 | Enrichment Score: 1.8009982082793852 | | | | |
| GOTERM_MF_FAT | GO:0022890~inorganic cation transmembrane transporter activity | 4 | 0.741 | 0.008 | ATP5EP2, COX8A, CCS, ATP6V0B |
| GOTERM_BP_FAT | GO:0006091~generation of precursor metabolites and energy | 5 | 0.926 | 0.012 | OXA1L, ATP5EP2, NDUFB7, COX8A, ATP6V0B |
| KEGG_PATHWAY | hsa00190:Oxidative phosphorylation | 4 | 0.741 | 0.014 | ATP5EP2, NDUFB7, COX8A, ATP6V0B |
| GOTERM_MF_FAT | GO:0015078~hydrogen ion transmembrane transporter activity | 3 | 0.556 | 0.024 | ATP5EP2, COX8A, ATP6V0B |
| GOTERM_MF_FAT | GO:0015077~monovalent inorganic cation transmembrane transporter activity | 3 | 0.556 | 0.032 | ATP5EP2, COX8A, ATP6V0B |
| Annotation Cluster 2 | Enrichment Score: 1.4488225878843253 | | | | |
| UP_SEQ_FEATURE | metal ion-binding site: Magnesium 2 | 3 | 0.556 | 0.004 | EXO1, ADPRHL2, LIG3 |
| UP_SEQ_FEATURE | metal ion-binding site: Magnesium 1 | 3 | 0.556 | 0.004 | EXO1, ADPRHL2, LIG3 |

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| SP_PIR_KEYWORDS | magnesium | 3 | 0.556 | 0.277 | EXO1, ADPRHL2, LIG3 |
| GOTERM_MF_FAT | GO:0000287~magnesium ion binding | 3 | 0.556 | 0.346 | EXO1, ADPRHL2, LIG3 |
| Annotation Cluster 3 | Enrichment Score: 1.1341952970900713 | | | | |
| GOTERM_BP_FAT | GO:0009152~purine ribonucleotide biosynthetic process | 3 | 0.556 | 0.045 | ATP5EP2, ATP6V0B, GART |
| GOTERM_BP_FAT | GO:0009260~ribonucleotide biosynthetic process | 3 | 0.556 | 0.050 | ATP5EP2, ATP6V0B, GART |
| GOTERM_BP_FAT | GO:0009150~purine ribonucleotide metabolic process | 3 | 0.556 | 0.060 | ATP5EP2, ATP6V0B, GART |
| GOTERM_BP_FAT | GO:0044271~nitrogen compound biosynthetic process | 4 | 0.741 | 0.067 | CTH, ATP5EP2, ATP6V0B, GART |
| GOTERM_BP_FAT | GO:0009259~ribonucleotide metabolic process | 3 | 0.556 | 0.067 | ATP5EP2, ATP6V0B, GART |
| GOTERM_BP_FAT | GO:0006164~purine nucleotide biosynthetic process | 3 | 0.556 | 0.068 | ATP5EP2, ATP6V0B, GART |
| GOTERM_BP_FAT | GO:0006163~purine nucleotide metabolic process | 3 | 0.556 | 0.100 | ATP5EP2, ATP6V0B, GART |
| GOTERM_BP_FAT | GO:0009165~nucleotide biosynthetic process | 3 | 0.556 | 0.100 | ATP5EP2, ATP6V0B, GART |
| GOTERM_BP_FAT | GO:0034404~nucleobase, nucleoside and nucleotide biosynthetic process | 3 | 0.556 | 0.107 | ATP5EP2, ATP6V0B, GART |
| GOTERM_BP_FAT | GO:0034654~nucleobase, nucleoside, nucleotide and nucleic acid biosynthetic process | 3 | 0.556 | 0.107 | ATP5EP2, ATP6V0B, GART |
| Annotation Cluster 4 | Enrichment Score: 1.0433001461402358 | | | | |
| SP_PIR_KEYWORDS | mitochondrion inner membrane | 4 | 0.741 | 0.010 | OXA1L, ATP5EP2, NDUFB7, COX8A |
| GOTERM_BP_FAT | GO:0006091~generation of precursor metabolites and energy | 5 | 0.926 | 0.012 | OXA1L, ATP5EP2, NDUFB7, COX8A, ATP6V0B |

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| KEGG_PATHWAY | hsa00190:Oxidative phosphorylation | 4 | 0.741 | 0.014 | ATP5EP2, NDUFB7, COX8A, ATP6V0B |
| GOTERM_CC_FAT | GO:0044455~mitochondrial membrane part | 3 | 0.556 | 0.043 | OXA1L, ATP5EP2, NDUFB7 |
| GOTERM_CC_FAT | GO:0005743~mitochondrial inner membrane | 4 | 0.741 | 0.047 | OXA1L, ATP5EP2, NDUFB7, COX8A |
| GOTERM_CC_FAT | GO:0019866~organelle inner membrane | 4 | 0.741 | 0.056 | OXA1L, ATP5EP2, NDUFB7, COX8A |
| GOTERM_CC_FAT | GO:0044429~mitochondrial part | 5 | 0.926 | 0.072 | OXA1L, ATP5EP2, NDUFB7, COX8A, LARS2 |
| GOTERM_CC_FAT | GO:0031966~mitochondrial membrane | 4 | 0.741 | 0.086 | OXA1L, ATP5EP2, NDUFB7, COX8A |
| KEGG_PATHWAY | hsa05012:Parkinson's disease | 3 | 0.556 | 0.089 | ATP5EP2, NDUFB7, COX8A |
| GOTERM_CC_FAT | GO:0005740~mitochondrial envelope | 4 | 0.741 | 0.099 | OXA1L, ATP5EP2, NDUFB7, COX8A |
| SP_PIR_KEYWORDS | mitochondrion | 5 | 0.926 | 0.129 | OXA1L, ATP5EP2, NDUFB7, COX8A, LARS2 |
| KEGG_PATHWAY | hsa05010:Alzheimer's disease | 3 | 0.556 | 0.133 | ATP5EP2, NDUFB7, COX8A |
| KEGG_PATHWAY | hsa05016:Huntington's disease | 3 | 0.556 | 0.156 | ATP5EP2, NDUFB7, COX8A |
| GOTERM_CC_FAT | GO:0031967~organelle envelope | 4 | 0.741 | 0.227 | OXA1L, ATP5EP2, NDUFB7, COX8A |
| GOTERM_CC_FAT | GO:0031975~envelope | 4 | 0.741 | 0.228 | OXA1L, ATP5EP2, NDUFB7, COX8A |
| UP_SEQ_FEATURE | transit peptide:Mitochondrion | 3 | 0.556 | 0.301 | OXA1L, COX8A, LARS2 |
| SP_PIR_KEYWORDS | transit peptide | 3 | 0.556 | 0.306 | OXA1L, COX8A, LARS2 |
| GOTERM_CC_FAT | GO:0005739~mitochondrion | 5 | 0.926 | 0.327 | OXA1L, ATP5EP2, NDUFB7, COX8A, LARS2 |
| GOTERM_CC_FAT | GO:0031090~organelle membrane | 5 | 0.926 | 0.333 | OXA1L, ATP5EP2, NDUFB7, COX8A, VPS11 |

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| Annotation Cluster 5 | Enrichment Score: 0.9420850726739483 | | | | |
| GOTERM_BP_FAT | GO:0006091~generation of precursor metabolites and energy | 5 | 0.926 | 0.012 | OXA1L, ATP5EP2, NDUFB7, COX8A, ATP6V0B |
| KEGG_PATHWAY | hsa00190:Oxidative phosphorylation | 4 | 0.741 | 0.014 | ATP5EP2, NDUFB7, COX8A, ATP6V0B |
| GOTERM_BP_FAT | GO:0006119~oxidative phosphorylation | 3 | 0.556 | 0.032 | ATP5EP2, NDUFB7, ATP6V0B |
| GOTERM_BP_FAT | GO:0016310~phosphorylation | 3 | 0.556 | 0.680 | ATP5EP2, NDUFB7, ATP6V0B |
| GOTERM_BP_FAT | GO:0006793~phosphorus metabolic process | 3 | 0.556 | 0.782 | ATP5EP2, NDUFB7, ATP6V0B |
| GOTERM_BP_FAT | GO:0006796~phosphate metabolic process | 3 | 0.556 | 0.782 | ATP5EP2, NDUFB7, ATP6V0B |
| Annotation Cluster 6 | Enrichment Score: 0.9157387050560352 | | | | |
| GOTERM_BP_FAT | GO:0016568~chromatin modification | 5 | 0.926 | 0.008 | RCBTB1, LEO1, CTCF, ASF1B, CBX8 |
| SP_PIR_KEYWORDS | chromatin regulator | 4 | 0.741 | 0.014 | RCBTB1, CTCF, ASF1B, CBX8 |
| GOTERM_BP_FAT | GO:0006325~chromatin organization | 5 | 0.926 | 0.023 | RCBTB1, LEO1, CTCF, ASF1B, CBX8 |
| GOTERM_BP_FAT | GO:0051276~chromosome organization | 5 | 0.926 | 0.050 | RCBTB1, LEO1, CTCF, ASF1B, CBX8 |
| GOTERM_BP_FAT | GO:0006333~chromatin assembly or disassembly | 3 | 0.556 | 0.052 | CTCF, ASF1B, CBX8 |
| GOTERM_CC_FAT | GO:0070013~intracellular organelle lumen | 9 | 1.667 | 0.091 | CCNT2, ATP5EP2, LIG3, LEO1, KEAP1, CTCF, LARS2, IMP4, MCM5 |
| GOTERM_CC_FAT | GO:0043233~organelle lumen | 9 | 1.667 | 0.100 | CCNT2, ATP5EP2, LIG3, LEO1, KEAP1, CTCF, LARS2, IMP4, MCM5 |
| GOTERM_CC_FAT | GO:0031974~membrane-enclosed lumen | 9 | 1.667 | 0.110 | CCNT2, ATP5EP2, LIG3, LEO1, KEAP1, CTCF, LARS2, IMP4, MCM5 |
| SP_PIR_KEYWORDS | cell division | 3 | 0.556 | 0.127 | CCNT2, LIG3, MCM5 |

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| GOTERM_CC_FAT | GO:0031981~nuclear lumen | 7 | 1.296 | 0.182 | CCNT2, LIG3, LEO1, KEAP1, CTCF, IMP4, MCM5 |
| SP_PIR_KEYWORDS | transcription regulation | 8 | 1.481 | 0.191 | RCBTB1, CCNT2, LEO1, KEAP1, CTCF, ASF1B, CBX8, MCM5 |
| GOTERM_CC_FAT | GO:0005654~nucleoplasm | 5 | 0.926 | 0.205 | CCNT2, LIG3, LEO1, CTCF, MCM5 |
| SP_PIR_KEYWORDS | Transcription | 8 | 1.481 | 0.206 | RCBTB1, CCNT2, LEO1, KEAP1, CTCF, ASF1B, CBX8, MCM5 |
| GOTERM_BP_FAT | GO:0051301~cell division | 3 | 0.556 | 0.209 | CCNT2, LIG3, MCM5 |
| GOTERM_CC_FAT | GO:0044427~chromosomal part | 3 | 0.556 | 0.274 | CTCF, ASF1B, CBX8 |
| GOTERM_CC_FAT | GO:0005730~nucleolus | 4 | 0.741 | 0.283 | CCNT2, KEAP1, CTCF, IMP4 |
| SP_PIR_KEYWORDS | nucleus | 13 | 2.407 | 0.290 | EXO1, CCNT2, C9ORF89, LIG3, CTCF, KEAP1, CBX8, MCM5, RCBTB1, ADPRHL2, LEO1, ASF1B, IMP4 |
| GOTERM_CC_FAT | GO:0005694~chromosome | 3 | 0.556 | 0.347 | CTCF, ASF1B, CBX8 |
| GOTERM_BP_FAT | GO:0006350~transcription | 8 | 1.481 | 0.403 | RCBTB1, CCNT2, LEO1, KEAP1, CTCF, ASF1B, CBX8, MCM5 |
| GOTERM_BP_FAT | GO:0045449~regulation of transcription | 8 | 1.481 | 0.644 | RCBTB1, CCNT2, LEO1, KEAP1, CTCF, ASF1B, CBX8, MCM5 |
| Annotation Cluster 7 | Enrichment Score: 0.9122037737039032 | | | | |
| SP_PIR_KEYWORDS | prenylation | 4 | 0.741 | 0.006 | RHOJ, RAB5C, RHOD, DNAJA2 |
| SP_PIR_KEYWORDS | lipoprotein | 6 | 1.111 | 0.017 | RHOJ, MPP1, NDUFB7, RAB5C, RHOD, DNAJA2 |
| UP_SEQ_FEATURE | short sequence motif:Effector region | 3 | 0.556 | 0.021 | RHOJ, RAB5C, RHOD |
| INTERPRO | IPR013753:Ras | 3 | 0.556 | 0.038 | RHOJ, RAB5C, RHOD |
| INTERPRO | IPR001806:Ras GTPase | 3 | 0.556 | 0.044 | RHOJ, RAB5C, RHOD |

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| GOTERM_CC_FAT | GO:0009898~internal side of plasma membrane | 4 | 0.741 | 0.051 | RHOJ, RAB5C, RHOD, VPS11 |
| INTERPRO | IPR005225:Small GTP-binding protein | 3 | 0.556 | 0.064 | RHOJ, RAB5C, RHOD |
| GOTERM_MF_FAT | GO:0003924~GTPase activity | 3 | 0.556 | 0.110 | RHOJ, RAB5C, RHOD |
| UP_SEQ_FEATURE | nucleotide phosphate-binding region:GTP | 3 | 0.556 | 0.153 | RHOJ, RAB5C, RHOD |
| SP_PIR_KEYWORDS | gtp-binding | 3 | 0.556 | 0.180 | RHOJ, RAB5C, RHOD |
| GOTERM_BP_FAT | GO:0007264~small GTPase mediated signal transduction | 3 | 0.556 | 0.219 | RHOJ, RAB5C, RHOD |
| GOTERM_MF_FAT | GO:0005525~GTP binding | 3 | 0.556 | 0.265 | RHOJ, RAB5C, RHOD |
| GOTERM_MF_FAT | GO:0032561~guanyl ribonucleotide binding | 3 | 0.556 | 0.275 | RHOJ, RAB5C, RHOD |
| GOTERM_MF_FAT | GO:0019001~guanyl nucleotide binding | 3 | 0.556 | 0.275 | RHOJ, RAB5C, RHOD |
| GOTERM_CC_FAT | GO:0044459~plasma membrane part | 8 | 1.481 | 0.368 | RHOJ, SLC29A1, MPP1, RAB5C, PODXL, KEAP1, RHOD, VPS11 |
| GOTERM_CC_FAT | GO:0005886~plasma membrane | 9 | 1.667 | 0.831 | RHOJ, SLC29A1, MPP1, RAB5C, PODXL, KEAP1, RHOD, VPS11, SLIT2 |
| GOTERM_BP_FAT | GO:0007242~intracellular signaling cascade | 3 | 0.556 | 0.889 | RHOJ, RAB5C, RHOD |
| SP_PIR_KEYWORDS | cell membrane | 3 | 0.556 | 0.971 | RHOJ, RAB5C, RHOD |
| Annotation Cluster 8 | Enrichment Score: 0.8723876533331111 | | | | |
| GOTERM_MF_FAT | GO:0022890~inorganic cation transmembrane transporter activity | 4 | 0.741 | 0.008 | ATP5EP2, COX8A, CCS, ATP6V0B |
| GOTERM_BP_FAT | GO:0006812~cation transport | 3 | 0.556 | 0.477 | ATP5EP2, CCS, ATP6V0B |
| GOTERM_BP_FAT | GO:0006811~ion transport | 3 | 0.556 | 0.658 | ATP5EP2, CCS, ATP6V0B |
| Annotation Cluster 9 | Enrichment Score: 0.7392411811789346 | | | | |
| SP_PIR_KEYWORDS | nucleotide-binding | 9 | 1.667 | 0.040 | RHOJ, RAB5C, LIG3, RHOD, LARS2, VPS11, PIP4K2C, MCM5, GART |

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| GOTERM_MF_FAT | GO:0000166~nucleotide binding | 11 | 2.037 | 0.068 | RHOJ, RAB5C, EIF4H, LIG3, RHOD, LARS2, VPS11, PIP4K2C, ATP6V0B, MCM5, GART |
| SP_PIR_KEYWORDS | acetylation | 11 | 2.037 | 0.080 | EXO1, ATP5EP2, RAB5C, EIF4H, RPLP0, LIG3, LARS2, CBX8, PIP4K2C, MCM5, GART |
| GOTERM_MF_FAT | GO:0032555~purine ribonucleotide binding | 9 | 1.667 | 0.111 | RHOJ, RAB5C, LIG3, RHOD, LARS2, PIP4K2C, ATP6V0B, MCM5, GART |
| GOTERM_MF_FAT | GO:0032553~ribonucleotide binding | 9 | 1.667 | 0.111 | RHOJ, RAB5C, LIG3, RHOD, LARS2, PIP4K2C, ATP6V0B, MCM5, GART |
| GOTERM_MF_FAT | GO:0017076~purine nucleotide binding | 9 | 1.667 | 0.135 | RHOJ, RAB5C, LIG3, RHOD, LARS2, PIP4K2C, ATP6V0B, MCM5, GART |
| SP_PIR_KEYWORDS | ligase | 3 | 0.556 | 0.160 | LIG3, LARS2, GART |
| GOTERM_MF_FAT | GO:0005524~ATP binding | 6 | 1.111 | 0.367 | LIG3, LARS2, PIP4K2C, ATP6V0B, MCM5, GART |
| SP_PIR_KEYWORDS | atp-binding | 5 | 0.926 | 0.376 | LIG3, LARS2, PIP4K2C, MCM5, GART |
| GOTERM_MF_FAT | GO:0032559~adenyl ribonucleotide binding | 6 | 1.111 | 0.378 | LIG3, LARS2, PIP4K2C, ATP6V0B, MCM5, GART |
| GOTERM_MF_FAT | GO:0030554~adenyl nucleotide binding | 6 | 1.111 | 0.423 | LIG3, LARS2, PIP4K2C, ATP6V0B, MCM5, GART |
| GOTERM_MF_FAT | GO:0001883~purine nucleoside binding | 6 | 1.111 | 0.437 | LIG3, LARS2, PIP4K2C, ATP6V0B, MCM5, GART |
| GOTERM_MF_FAT | GO:0001882~nucleoside binding | 6 | 1.111 | 0.443 | LIG3, LARS2, PIP4K2C, ATP6V0B, MCM5, GART |
| Annotation Cluster 10 | Enrichment Score: 0.48057878868833986 | | | | |
| GOTERM_CC_FAT | GO:0043232~intracellular non-membrane-bounded organelle | 10 | 1.852 | 0.241 | CCNT2, MPP1, RPLP0, KEAP1, CTCF, ASF1B, VPS11, CBX8, IMP4, TRAF5 |
| GOTERM_CC_FAT | GO:0043228~non-membrane-bounded organelle | 10 | 1.852 | 0.241 | CCNT2, MPP1, RPLP0, KEAP1, CTCF, ASF1B, VPS11, CBX8, IMP4, TRAF5 |
| GOTERM_CC_FAT | GO:0005730~nucleolus | 4 | 0.741 | 0.283 | CCNT2, KEAP1, CTCF, IMP4 |

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| GOTERM_CC_FAT | GO:0005856~cytoskeleton | 4 | 0.741 | 0.727 | MPP1, KEAP1, VPS11, TRAF5 |
| Annotation Cluster 11 | Enrichment Score: 0.3834531138201595 | | | | |
| SP_PIR_KEYWORDS | metal-binding | 11 | 2.037 | 0.147 | EXO1, ADPRHL2, SMYD5, LIG3, CTCF, CCS, VPS11, TRAF5, DNAJA2, MB, GART |
| SP_PIR_KEYWORDS | zinc-finger | 6 | 1.111 | 0.375 | SMYD5, LIG3, CTCF, VPS11, TRAF5, DNAJA2 |
| SP_PIR_KEYWORDS | zinc | 7 | 1.296 | 0.406 | SMYD5, LIG3, CTCF, CCS, VPS11, TRAF5, DNAJA2 |
| GOTERM_MF_FAT | GO:0043169~cation binding | 13 | 2.407 | 0.458 | EXO1, PCTP, LIG3, CTCF, SLIT2, GART, ADPRHL2, SMYD5, CCS, VPS11, TRAF5, MB, DNAJA2 |
| GOTERM_MF_FAT | GO:0043167~ion binding | 13 | 2.407 | 0.482 | EXO1, PCTP, LIG3, CTCF, SLIT2, GART, ADPRHL2, SMYD5, CCS, VPS11, TRAF5, MB, DNAJA2 |
| GOTERM_MF_FAT | GO:0046914~transition metal ion binding | 9 | 1.667 | 0.485 | SMYD5, LIG3, CTCF, CCS, VPS11, TRAF5, DNAJA2, MB, GART |
| GOTERM_MF_FAT | GO:0046872~metal ion binding | 12 | 2.222 | 0.586 | EXO1, ADPRHL2, SMYD5, LIG3, CTCF, CCS, VPS11, TRAF5, SLIT2, DNAJA2, MB, GART |
| GOTERM_MF_FAT | GO:0008270~zinc ion binding | 7 | 1.296 | 0.610 | SMYD5, LIG3, CTCF, CCS, VPS11, TRAF5, DNAJA2 |
| Annotation Cluster 12 | Enrichment Score: 0.3700685424689228 | | | | |
| GOTERM_BP_FAT | GO:0006259~DNA metabolic process | 4 | 0.741 | 0.178 | EXO1, LIG3, CTCF, MCM5 |
| GOTERM_CC_FAT | GO:0005654~nucleoplasm | 5 | 0.926 | 0.205 | CCNT2, LIG3, LEO1, CTCF, MCM5 |
| SP_PIR_KEYWORDS | dna-binding | 3 | 0.556 | 0.941 | EXO1, CTCF, MCM5 |
| GOTERM_MF_FAT | GO:0003677~DNA binding | 4 | 0.741 | 0.964 | EXO1, LIG3, CTCF, MCM5 |
| Annotation Cluster 13 | Enrichment Score: 0.2987365248102637 | | | | |

| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| SP_PIR_KEYWORDS | protein transport | 3 | 0.556 | 0.314 | STX12, RAB5C, VPS11 |
| GOTERM_BP_FAT | GO:0008104~protein localization | 4 | 0.741 | 0.472 | OXA1L, STX12, RAB5C, VPS11 |
| GOTERM_BP_FAT | GO:0015031~protein transport | 3 | 0.556 | 0.654 | STX12, RAB5C, VPS11 |
| GOTERM_BP_FAT | GO:0045184~establishment of protein localization | 3 | 0.556 | 0.659 | STX12, RAB5C, VPS11 |
| Annotation Cluster 14 | Enrichment Score: 0.19418904990903646 | | | | |
| GOTERM_CC_FAT | GO:0044459~plasma membrane part | 8 | 1.481 | 0.368 | RHOJ, SLC29A1, MPP1, RAB5C, PODXL, KEAP1, RHOD, VPS11 |
| GOTERM_CC_FAT | GO:0005887~integral to plasma membrane | 3 | 0.556 | 0.838 | SLC29A1, MPP1, PODXL |
| GOTERM_CC_FAT | GO:0031226~intrinsic to plasma membrane | 3 | 0.556 | 0.847 | SLC29A1, MPP1, PODXL |
| Annotation Cluster 15 | Enrichment Score: 0.16298372853376233 | | | | |
| GOTERM_BP_FAT | GO:0042981~regulation of apoptosis | 3 | 0.556 | 0.683 | CSF2, MOAP1, TRAF5 |
| GOTERM_BP_FAT | GO:0043067~regulation of programmed cell death | 3 | 0.556 | 0.688 | CSF2, MOAP1, TRAF5 |
| GOTERM_BP_FAT | GO:0010941~regulation of cell death | 3 | 0.556 | 0.690 | CSF2, MOAP1, TRAF5 |
| Annotation Cluster 16 | Enrichment Score: 0.06893264590802266 | | | | |
| SP_PIR_KEYWORDS | membrane | 17 | 3.148 | 0.385 | RHOJ, OXA1L, MPP1, NDUFB7, RAB5C, PODXL, COX8A, ATP6V0B, TMEM115, SLC29A1, STX12, ATP5EP2, RHOD, VPS11, KIAA0247, PIP4K2C, DNAJA2 |
| GOTERM_CC_FAT | GO:0031224~intrinsic to membrane | 11 | 2.037 | 0.964 | TMEM115, SLC29A1, OXA1L, STX12, MPP1, RAB5C, PODXL, COX8A, KIAA0247, ATP6V0B, DNAJA2 |

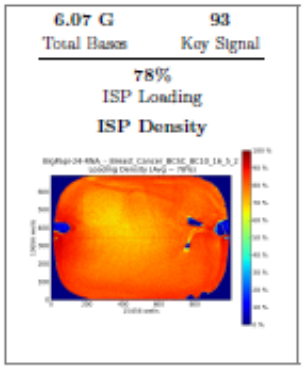
| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| UP_SEQ_FEATURE | transmembrane region | 8 | 1.481 | 0.964 | TMEM115, SLC29A1, OXA1L, STX12, PODXL, COX8A, KIAA0247, ATP6V0B |
| SP_PIR_KEYWORDS | transmembrane | 8 | 1.481 | 0.966 | TMEM115, SLC29A1, OXA1L, STX12, PODXL, COX8A, KIAA0247, ATP6V0B |
| UP_SEQ_FEATURE | topological domain:Cytoplasmic | 5 | 0.926 | 0.969 | SLC29A1, STX12, PODXL, KIAA0247, ATP6V0B |
| GOTERM_CC_FAT | GO:0016021~integral to membrane | 9 | 1.667 | 0.991 | TMEM115, SLC29A1, OXA1L, STX12, MPP1, PODXL, COX8A, KIAA0247, ATP6V0B |
| UP_SEQ_FEATURE | topological domain:Extracellular | 3 | 0.556 | 0.992 | SLC29A1, PODXL, KIAA0247 |
| Annotation Cluster 17 | Enrichment Score: 0.005856957988287524 | | | | |
| UP_SEQ_FEATURE | disulfide bond | 4 | 0.741 | 0.971 | CSF2, CCS, KIAA0247, SLIT2 |
| SP_PIR_KEYWORDS | disulfide bond | 4 | 0.741 | 0.976 | CSF2, CCS, KIAA0247, SLIT2 |
| SP_PIR_KEYWORDS | signal | 4 | 0.741 | 0.988 | CSF2, PODXL, KIAA0247, SLIT2 |
| UP_SEQ_FEATURE | signal peptide | 4 | 0.741 | 0.988 | CSF2, PODXL, KIAA0247, SLIT2 |
| UP_SEQ_FEATURE | Glycosylation site: N-linked (GlcNAc...) | 4 | 0.741 | 0.998 | SLC29A1, CSF2, PODXL, SLIT2 |
| SP_PIR_KEYWORDS | glycoprotein | 4 | 0.741 | 0.999 | SLC29A1, CSF2, PODXL, SLIT2 |
| Not Clustered | | | | | |
| UP_SEQ_FEATURE | region of interest:Substrate binding | 3 | 1.596 | 0.005 | ACADSB, ECHS1, NAGK |
| GOTERM_BP_FAT | GO:0006952~defense response | 4 | 2.128 | 0.019 | HIST1H2BK, HIST2H2BE, IL17RB, SCG2 |
| SP_PIR_KEYWORDS | acetylation | 5 | 2.660 | 0.166 | PYCR2, ACADSB, HIST1H2BK, HIST2H2BE, ECHS1 |

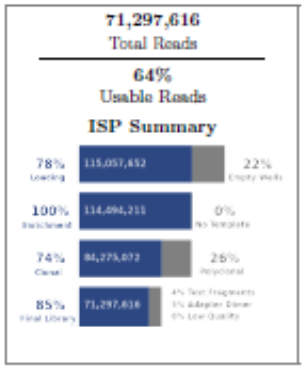
| Category | Term | Count | % | PValue | Genes |
|---|---|---|---|---|---|
| GOTERM_CC_FAT | GO:0043228~non-membrane-bounded organelle | 4 | 2.128 | 0.453 | MRPL1, HIST1H2BK, HIST2H2BE, TRAF5 |
| GOTERM_CC_FAT | GO:0043232~intracellular non-membrane-bounded organelle | 4 | 2.128 | 0.453 | MRPL1, HIST1H2BK, HIST2H2BE, TRAF5 |
| SP_PIR_KEYWORDS | polymorphism | 10 | 5.319 | 0.717 | MRPL1, ACADSB, HPS5, VSNL1, SHFM1, ECHS1, NAGK, TRAF5, IL17RB, SCG2 |
| SP_PIR_KEYWORDS | phosphoprotein | 6 | 3.191 | 0.784 | PYCR2, HPS5, HIST1H2BK, HIST2H2BE, NAGK, SCG2 |
| UP_SEQ_FEATURE | sequence variant | 10 | 5.319 | 0.789 | MRPL1, ACADSB, HPS5, VSNL1, SHFM1, ECHS1, NAGK, TRAF5, IL17RB, SCG2 |
| SP_PIR_KEYWORDS | nucleus | 3 | 1.596 | 0.901 | HIST1H2BK, HIST2H2BE, SNRNP25 |

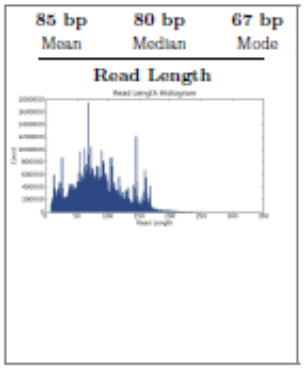
APPENDIX C: SEQUENCING RESULTS FROM ION TORRENT PROTON SEQUENCER

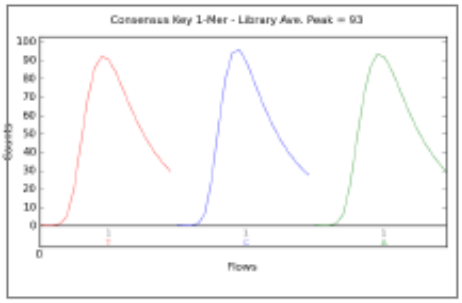
*Table 9. Sequencing Run Containing two replicates of Sorted control and BCSC and three replicates of BCSC depleted samples, all samples were used in analysis.*

**Run Summary**



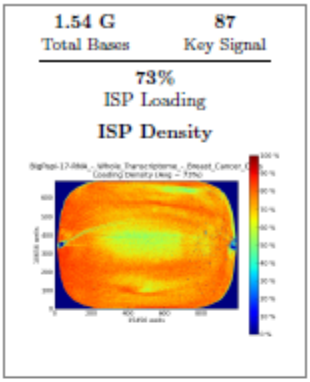

| Addressable Wells | 148,155,732 | |
|---|---|---|
| With ISPs | 115,057,652 | 77.7% |
| Live | 114,494,211 | 99.5% |
| Test Fragment | 3,569,424 | 03.1% |
| Library | 110,924,787 | 96.9% |
| **Library ISPs** | **110,924,787** | |
| Filtered: Polyclonal | 30,219,139 | 27.2% |
| Filtered: Low Quality | 5,440,421 | 04.9% |
| Filtered: Adapter Dimer | 4,075,285 | 03.7% |
| **Final Library ISPs** | **71,297,616** | **64.3%** |

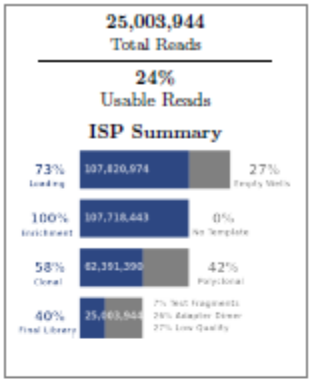
| Barcode Name | Sample | Bases | ≥ Q20 | Reads | Mean Read Length |
|---|---|---|---|---|---|
| No barcode | none | 149,067,394 | 103,092,572 | 2,468,720 | 60 bp |
| IonXpressRNA_010 | BCSC 1 p27 | 91,639,822 | 81,043,269 | 1,538,607 | 60 bp |
| IonXpressRNA_011 | BCSC 2 p27 | 1,073,575,275 | 936,008,208 | 11,922,476 | 90 bp |
| IonXpressRNA_012 | Sort Control p27 | 1,411,550,642 | 1,237,467,260 | 15,250,697 | 93 bp |
| IonXpressRNA_013 | Sorth Contorl p27 | 1,189,574,812 | 1,051,048,153 | 15,640,146 | 76 bp |
| IonXpressRNA_014 | Depleted p27 | 174,104,301 | 149,843,248 | 1,401,692 | 124 bp |
| IonXpressRNA_015 | Depleted p27 | 16,166,806 | 14,067,138 | 145,859 | 111 bp |
| IonXpressRNA_016 | Delpleted p27 | 1,965,822,199 | 1,717,982,184 | 22,915,492 | 86 bp |

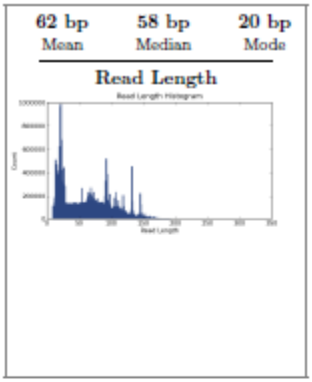
| Test Fragment | Reads | Percent 50AQ17 | Read Length Histogram |
|---|---|---|---|
| **TF_C** | **2,783,480** | **73%** | |

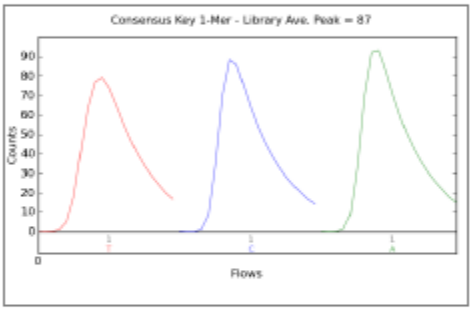
*Table 10. Sequencing results for total RNA and rRNA depleted samples. Only Total RNA samples used for analysis, Bar Codes 007 and 008.*

**Run Summary**




| | | |
|---|---|---|
| **Addressable Wells** | **148,155,732** | |
| With ISPs | 107,820,974 | 72.8% |
| Live | 107,718,443 | 99.9% |
| Test Fragment | 4,351,053 | 04.0% |
| Library | 103,367,390 | 96.0% |
| **Library ISPs** | **103,367,390** | |
| Filtered: Polyclonal | 45,327,053 | 43.9% |
| Filtered: Low Quality | 17,036,112 | 16.5% |
| Filtered: Adapter Dimer | 16,171,855 | 15.6% |
| **Final Library ISPs** | **25,003,944** | **24.2%** |

| Barcode Name | Sample | Bases | ≥ *Q*20 | Reads | Mean Read Length |
|---|---|---|---|---|---|
| No barcode | none | 158,279,515 | 63,949,003 | 6,517,627 | 24 bp |
| IonXpressRNA_002 | BCSC 2 1-21 rRNA depleted | 110,270,308 | 93,952,965 | 1,916,240 | 58 bp |
| IonXpressRNA_003 | Sort 1 -21 rRNA depleted | 11,178,101 | 9,678,736 | 312,664 | 36 bp |
| IonXpressRNA_007 | BCSC 2 2-24 Total RNA | 753,020,850 | 643,025,480 | 9,139,079 | 82 bp |
| IonXpressRNA_008 | Sort 2 2-4 Total RNA | 510,653,634 | 426,614,361 | 7,114,650 | 72 bp |

| Test Fragment | Reads | Percent 50AQ17 | Read Length Histogram |
|---|---|---|---|
| **TF_C** | **3,985,510** | **75%** | |