INTEGRATIVE IMMUNOTRANSCRIPTOMIC ANALYSIS OF LONG-LIVED PLASMA CELLS

A Dissertation
Presented to
The Academic Faculty

by

Swetha Garimalla

In Partial Fulfillment
of the Requirements for the Degree
Doctorate of Philosophy in the
School of Biological Sciences

Georgia Institute of Technology
November 2018

INTEGRATIVE IMMUNOTRANSCRIPTOMIC ANALYSIS OF LONG-LIVED PLASMA CELLS

Approved by:

Dr. Greg Gibson, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Soojin Yi
School of Biological Sciences
Georgia Institute of Technology

Dr. Frances Eun-Hyung Lee
Department of Medicine
Emory Univerisity

Dr. Peng Qiu
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Rabind Tirouvanziam
Department of Pediatric Infectious Diseases
Emory University

Date Approved: November 02, 2018

To my family, Nagaraj, Satish, and Nisha – none of this would have been possible without you.

ACKNOWLEDGEMENTS

I would like to thank Dr. Greg Gibson for all of the support, guidance, and patience over the past few years as well as giving me an opportunity to be in his lab. I would like to thank him for reminding me to have faith in myself, always expecting more of me, and teaching me to approach life with humor.

I would also like to thank Dr. Rabin Tirouvanziam, Dr. Soojin Yi, and Dr. Peng Qiu for their willingness to be part of my committee and their comments and suggestions.

I would like to thank Dr. Frances Eun-Hyung Lee for her collaboration and developing the experiments and studies to obtain the samples for this research. I would especially like to thank her for her commitment to the science, attention to detail, and unwavering excitement for the research. We would also like to thank all of the volunteers, patients, and families for participating in these studies.

I would like to thank all of my colleagues in the Gibson Lab, especially Monica, Urko, Biao, and Hari for being my sounding boards, co-conspirators, supports, and friends in this research process.

I would also like to thank my friends all over Atlanta and the world, especially Toyya, Blake, Brandon, Dejah, Sheila, Monique, Sonja, Max, Erin, David, Emily, Anar, and Nick. Every text, every late-night phone call, every cup of coffee, every dinner, every word of comfort helped me to remind me that I always have someone in my corner.

Finally, I would like to thank my family: Satish, Nagaraj, & Nisha. Thank you to my father, Nagaraj, for teaching me passion and persistence. Thank you to my mother,

Satish, for teaching me boldness and kindness. Thank you to my little sister, Nisha, for being my therapist, protector, and partner-in-crime. Most of all, thank you to my entire family for showing me that failure and success are just passing moments in a day and that all that matters, in the end, is whether you gave your life to the pursuit of what you love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

ASC	Antibody-Secreting Cells
PC	Plasma Cells
BM	Bone Marrow
PB	Peripheral Blood
MSC	Mesenchymal Stromal Cells
SLPC	Short-Lived Plasma Cells
LLPC	Long-Lived Plasma Cells
PMB	Plasmablasts

SUMMARY

Long-lived plasma cells are a key component of serological memory encoded by the adaptive immune response. To date, prior studies of these cell types have largely assessed plasma cells as one homogenous population. The Lee Lab at Emory University provided strong evidence for the presence of subpopulations of plasma cells in the bone marrow and peripheral blood. In this work, I analyze six of these populations: pops 2, 3, and 5 in the peripheral blood and pops A, B, and D in the bone marrow sampled from live human donors. Whole-transcriptome analysis of these six flow-sorted plasma cell populations is compared within and across tissue within this study. Further, I present 4 possible models of development of these plasma cells based on the results. Finally, I performed a single-cell RNA sequencing analysis of the long-lived plasma cells (popD) and present evidence for the possibility of subpopulations of this cell type. This work describes many mechanisms of molecular development of long-lived plasma cells, some of which have been experimentally validated by Dr. Doan Nguyen as a result of these analyses, however, further experimental validation is required to validate all findings in this work.

INTRODUCTION

1.1 Immune System

Our primary defense against foreign agents is the immune system. This protection is enacted by the two arms of the immune system, innate and adaptive, through the elimination of pathogens and immune surveillance. While the innate immune response is immediate and responsive to a wide variety of antigens, the adaptive arm provides an antigen-specific response and long-lasting immunity. Plasma cells and memory B-cells play key roles in the program of long-lived immunity. {Shi et al (2015); Jourdan et al (2011); Halliley et al (2015)}

As illustrated in Figure 1.1, after differentiation from hematopoietic stem cells in the bone marrow, naïve B-cells travel through the peripheral blood into secondary lymph organs (such as the lymph node and spleen) where, upon binding to antigen, they are activated. Activated B-cells then either form short-lived plasmablasts, for an immediate (though weak) immune response, or proliferate in a lymphoid germinal center to form plasma cells and memory B-cells. The plasma cells secrete copious amounts of antibody, either while remaining in the secondary lymph organ or after migrating back to the bone marrow. Antibody production in the marrow by these long-lived plasma cells continues well after the initial infection is cleared, allowing for rapid detection of previously encountered antigens, thereby conferring long-lived immunity.{Halliley et al (2015)}

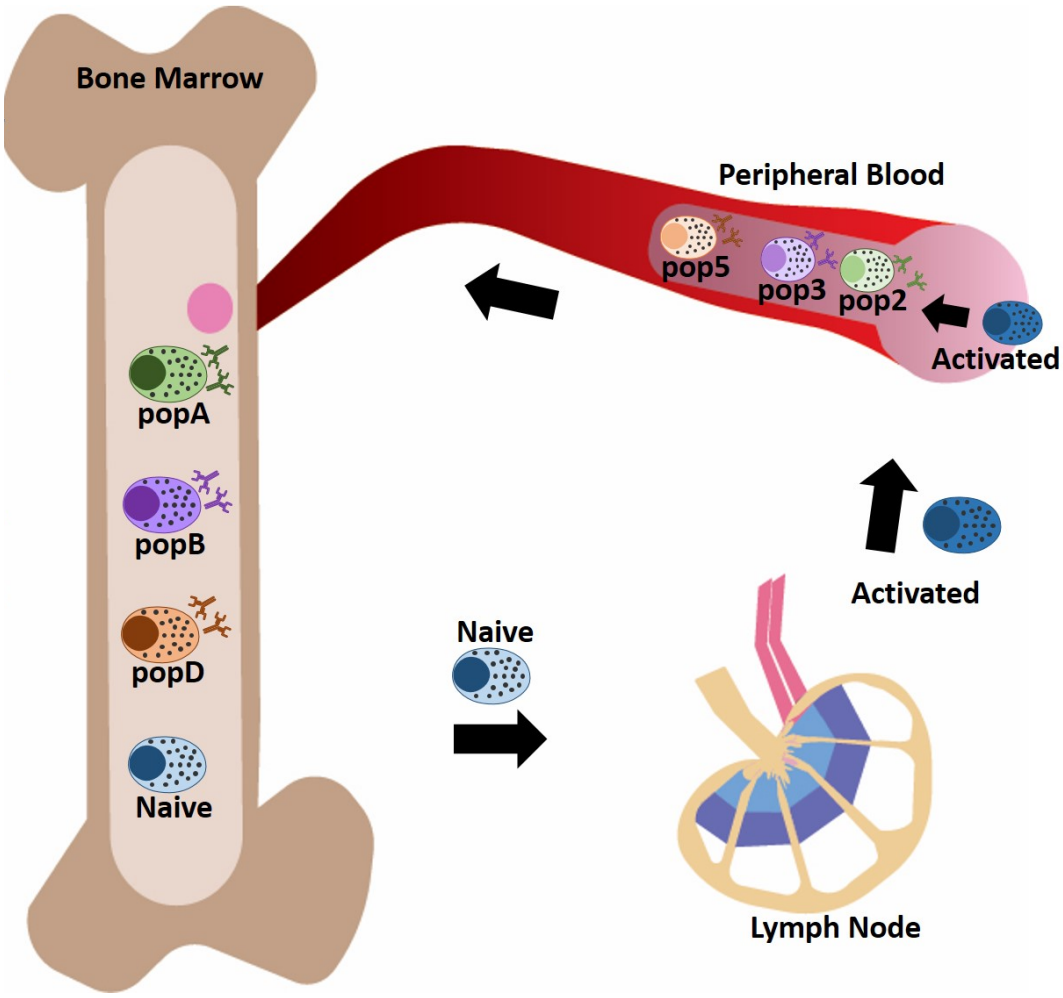


Figure 1.1. Progression from Naïve B-cells to Long-Lived Plasma Cells. Naïve B-cells leave the marrow and move to the germinal center where they are exposed to antigen and become activated. These activated B-cells develop into memory B-cells and plasmablasts, the latter of which is a precursor population to pops2, 3, and 5. In this study, I propose that these cell populations enter the marrow and develop into bone marrow plasma cell populations popA, popB, and popD (long-lived plasma cells).

1.2 Adaptive vs. Innate

Upon infection, the innate immune system of an individual is activated. Unlike the adaptive immune system, the innate immune system is fully developed and present since birth. The effector molecules of the innate immune system target a range of highly-conserved motifs presented by many pathogenic microbes. This non-specific recognition of one of these motifs by an innate immune cell triggers a series of signaling pathways, including inflammatory pathways. This plays a dual role: 1) to keep the infection in check until the generation of the adaptive immune system, 2) priming the activation of the adaptive immune system. {Trezise et al (2018); Tellier et al (2018); Tellier et al (2017); Shi et al (2015); Nutt et al (2015)}

Unlike the innate immune system, a key characteristic of the adaptive immune system is antigen-specific recognition. B- and T-lymphocytes, the major classes of adaptive immune cells, recognize antigens through functional receptors on their surface. The receptors on T-cells are non-soluble. Innate immune cells, known as antigen-presenting cells, present foreign antigens to T-cells. The antigen-specific immune response is manifested through a lock-and-key process of receptors seeking matching antigens. This is made possible through the high diversity of B- and T-cell receptors which occurs prior to exposure to antigen. Once a T-cell recognizes an antigen, it becomes activated and undergoes clonal selection and expansion. {Trezise et al (2018);Tellier et al (2018);Tellier et al (2017);Shi et al (2015);Nutt et al (2015)}

A similar process occurs for B-cells. Unlike T-cells, naïve B-cells transverse the peripheral blood system from the bone marrow to various primary lymphoid organs where

they are exposed to antigen. Once a B-cell receptor matches antigen, like T-cells, clonal selection and expansion occurs. During this process, further mutation to the receptor occurs, known as somatic hypermutation, which allows for increased binding affinity to the antigen after clonal selection. Additionally, once a naïve B-cell is activated and develops into a plasma cell, unlike T-cells, it can secrete soluble antibodies which play a key role in immune surveillance. Due to the adaptive immune system's ability to distinguish between self- and non-self, adaptation to infection (specificity), and immune memory, manipulation of this arm of the system has been the subject of many studies.{Trezise et al (2018);Tellier et al (2018);Tellier et al (2017);Shi et al (2015);Nutt et al (2015)}

1.3 Development of B-cells

The cellular components of both the innate and adaptive immune systems derive from the same hematopoietic stem cells (HSCs). This progenitor cell-type differentiates into myeloid cells, which further develop into innate immune cells, and into lymphoid cells, which differentiate into the adaptive immune cells including T-cells and B-cells. Both of these cell types originate from HSCs in the bone marrow. T-cells then migrate to the thymus to differentiate and mature, whereas B-cells engage in antigen-independent differentiation into naïve B-cells. They then migrate to a secondary lymphoid organ where they engage in antigen-dependent differentiation once they recognize antigen presented to them in germinal vesicles.{Trezise et al (2018);Tellier et al (2018);Tellier et al (2017);Shi et al (2015);Nutt et al (2015)}

1.4 Immunological Memory

The resolution of an infection marks the clearance of the pathogens that triggered the immune response. This occurs along with death of the majority of effector cells involved in the immune response, through neglect and apoptosis. The remnants of these apoptosed immune cells are cleared by phagocytosis. While most effector cells are disposed of in this manner, some of these cells are the precursors for plasma cells and memory cells. {Tellier et al (2018); Tellier et al (2017); Shi et al (2015); Peperzak et al (2013); Nutt et al (2015); Kallies et al (2010); Corcoran et al (2016); Corcoran et al (2015)}

Memory response in the immune system is characterized by the activation of memory cells that formed during the primary response, to aid in the resolution of reinfection by a previously encountered pathogen. Memory responses encapsulate secondary responses, tertiary responses, and so forth, based on the number of exposures to the antigen. The property of the memory response that enables this is the B-lymphocyte's ability to persist in the absence of the antigen that triggered the primary response. While most of these memory cells are in a resting state, typically a small percentage of these cells are dividing. The benefit of these memory cells over the naïve response is that, due to clonal expansion and affinity maturation during the primary response, the antibodies produced by the memory response have had time to develop a much higher affinity for the antigen, rendering future memory responses much more effective.{Tellier et al (2018);Tellier et al (2017);Shi et al (2015); Peperzak et al (2013);Nutt et al (2015);Kallies et al (2010);Corcoran et al (2016); Corcoran et al (2015)}

1.5 Plasma Cells

1.5.1 Overview of previous plasma cell gene expression analyses

In addition to memory B cells, plasmablasts can also differentiate into plasma cells, also known as antibody secreting cells (ASC). Unlike memory B-cells, plasma cells are terminally differentiated and hence they have ceased proliferation. These plasma cells migrate to the bone marrow where they increase antibody synthesis and immunoglobulin secretion. The differentiation of mature B-cells into antibody secreting plasma cells is known to be governed by a set of transcription factors. Amongst these are PAX5, BACH2, BCL6, PU.1, SPIB, ETS1, IRF8, IRF4, BLIMP1, and XBP1. The expression of PAX5, BACH2, and BCL6 plays a role throughout B-cell development, however in all three cases, expression in mature B-cells prevents some of the ASC differentiation programs. BACH2 is regulated by PAX5 and is not expressed in ASCs. Further, BACH2 targets BLIMP1 which also plays a significant role in ASC differentiation. BCL6 also enforces repression of BLIMP1, so, like BACH2 and PAX5, down-regulation of BCL6 is required for ASC differentiation. PU.1, IRF8, and SPIB are also downregulated in ASCs due to their contribution in maintaining the B-cell state. ETS1 must also be downregulated in order to allow for the development of ASCs because it plays a role in terminal differentiation, PAX5 expression, and inhibition of BLIMP1. BLIMP1 is one of the main drivers of ASC differentiation which promotes the B-cell terminal differentiation and downregulates other promoters of B-cell programs. IRF4 plays a role both in the maintenance the B-cell programs as well as the promotion of ASC differentiation. While it interacts with ASC differentiation repressors such as the IRF4-ETS complex and PU.1 in low concentrations, it promotes ASC differentiation when it is highly expressed. XBP1, in contrast, does not

affect differentiation into ASC but plays a key role in the unfolded protein response and ER stress in the context of promotion of ER remodeling. Additionally, mTOR pathways have been shown to compensate for partial loss of XBP1.{ Zhang et al (2009); Tellier et al (2018); Shi et al (2015); Nutt et al (2015); Klein et al (2008); Kassambara et al (2015); Jourdan et al (2011); Chatterjee et al (2007); Rawstron et al (2006); Klein et al (2003); Good et al (2009) ; Mei et al (2015)}

There have been numerous studies exploring differential gene expression analysis between the various stages of development from B-cells to plasma cells. One particular direction of our transcriptomic analysis that has not previously been pursued is our focus on antibody secreting cells (pops2, 3, 5), short-lived plasma cells (popsA and B), and popD plasma cells which have be previously proven to present a long-lived phenotype. This contrasts with most previous studies that consider plasma cells as a single cell type. Many of these studies, additionally, focus on microarray analysis rather than RNASeq as with my work. {Halliley et al (2015)}

Goode 2009 and Jourdan 2011 both pursued microarray analysis in order to profile gene expression during the progression of B-cell differentiation. Good 2009 focused on Naïve B-cells, IgM memory cells, isotype-switched memory, and switched memory B-cells harvested from the spleens of human cadavers. Additionally, this study focused on the differential gene expression of TNSF, SLAM, TLF, and TACI genes across those cell types. Jourdan 2011 focused on human cell lines using microarray analysis to compare preplasmablasts, B-cells, plasmablasts, and plasma cells. They also utilized cultured media to determine factors that promote differentiation in the context of cell cycle and Ig secretion. The cells from our study came from the peripheral blood and bone marrow of

living human donors. As previously stated, we used whole transcriptome RNASeq analysis to perform an unsupervised gene expression analysis across peripheral blood and bone marrow plasma cell subtypes. While we also used cultured media to determine which features promote survival and differentiation into the long-lived phenotype, unlike Jourdan, our focus was not specifically on cell cycle and number of cell divisions. Rather, the focus of my thesis is on characterizing differentially regulated pathways across plasma cell types. {Jourdan et al (2011);Good et al (2009) }

Klein 2013 and Mei 2015 studied lymph nodes, tonsils, peripheral blood, and bone marrow using microarray analysis. Both of these studies considered subsets of cells defined by specific cell surface markers. Klein 2013 uses CD20, CD38, and CD138 expression to define their cell types and determine the differentiation patterns and transcription factors that contribute from preplasmablasts to memory cells to plasma cells, including splice variants. Further, they used cultured media to activate various cell types. Mei 2015, on the other hand, used the presence of CD19 to distinguish plasma cell types in the marrow. They noted that since CD19+ plasma cells (which would include our popA and popB cell types) have PB ASCs as their precursors and that, due to upregulation of gene expression of survival signatures, CD19- could be the long-lived phenotype for plasma cells; however, they did not explicitly prove that CD19- is the long-lived plasma cell as in Halliley 2015. Additionally because of the use of this single cell marker, they lost the granularity and nuance of the gene expression that I will describe in our comparative analysis of long-lived PCs with SLPCs and ASCs.{Klein et al (2003); Mei et al (2015)}

Kassambara 2015 attempted to address cross-study discrepancies in documenting the genomic atlas of B-cell differentiation. Their study focused on human bone marrow

microarray analysis of the different stages of differentiation of B-cells from GC B-cells to bone marrow plasma cells, via a meta-analysis of publicly available gene sets. However, they based their entire analysis on 10,000 genes and attempted to determine the specific gene signatures that characterize these cells types within the context of only 9303of the genes that appeared to be differentially expressed. {Kassambara et al (2015)}

Nutt 2015 performed a whole transcriptome RNASeq analysis of murine splenic follicular B cells, marginal zone B cells, splenic peripheral blood cells, splenic plasma cells, B1 cells, and bone marrow plasma cells. They reported on exon level gene expression. Additionally, they attempted to determine developmental distance between cell types using "leading fold change," a metric which utilizes only the top 500 differentially expressed genes between two populations. They attempted to correlate these differences in expression with the number of cell divisions between each stage and their relative epigenetic differences. {Tellier et al (2017);Nutt et al (2015)}

Our study reports primarily on gene-level expression in order to avoid inaccuracies in inference of isoform abundance from exon-level analysis that inevitably arise with current analytical tools. Further, rather than using only the top 500 differentially expressed genes, I considered all significantly differentially expressed genes in my analysis. My differential expression analysis is primarily focused on differences between cell types rather than number of cell divisions. A future direction will be to include epigenetic modifications into our analysis in determining differences between cell types.

Previous studies have been extensively conducted on the transcriptomic analysis of long-lived plasma cells and their precursors. The primary limitations of these previous

studies include the focus on specific pre-determined genes in a supervised bottom-up processing of the transcriptome, focus on cell types that were not clearly determined to be long-lived plasma cells, use of mouse models rather than living human donors, as well as other more nuanced discrepancies. In this study, my collaborators in the Lee laboratory at Emory University used CD19, CD39, and CD138, cell surface markers that were previously experimentally proven to demarcate the long-lived plasma cell phenotype, to cell sort peripheral blood ASCs and bone marrow plasma cells from living, healthy, human donors into six subsets: pop2, pop3, pop5, and popA, popB, and popD. I describe the use of unsupervised, top-down processing to perform differential gene expression analysis and gene set enrichment analysis on whole transcriptome RNASeq data to determine the cell-type specific signatures of each cell type, the ontological relationships between these PB ASCs and BM LLPCs, and functional inferences of the different biophysical and biochemical capacities of the different plasma cell populations.

1.5.2 Peripheral Blood Antibody Secreting Cells

In 2016, the F.E. Lee lab at Emory University characterized five sub-populations of antibody-secreting cells (ASCs) that circulate in the peripheral blood (pop1 - pop5), and four that are resident in the bone marrow (popA-popD). The three cell surface markers that identify these nine subsets in the blood and marrow are CD19, CD38, and CD138 and are all implicated in functional differentiation of plasma cells as well. CD19 is a differentiation marker for B-cells generally{Ishikawa et al (2002)}, whereas CD38 and CD138 are plasma-cell specific.{Rawstron et al (2006); Medina et al (2002); Mei et al (2015)}. CD138 encodes Syndecan, a cell-cell adhesion protein that is required to maintain the localization of terminally differentiated cells within the marrow. {Aref et al (2003);Bayer-Garner et al

(2001);Halliley et al (2015);Rawstron et al (2006)}. CD38 encodes a cell surface glycoprotein with roles in cell adhesion, signal transduction, and calcium signaling. {Rawstron et al (2006);Medina et al (2002);Halliley et al (2015); Mei et al (2015) }

Halliley 2015 described a preliminary transcriptomic signature of these cell-sorted (on CD19, CD38, and CD138) cell types in the marrow. This thesis builds on their foundation by generating functional inferences of the different biophysical and biochemical capacities of the different plasma cell populations. {Halliley et al (2015)}

1.5.3 Bone Marrow Plasma Cells

Of the four bone marrow populations, popA, popB, popC, and popD, Halliley et al (2015) characterized 3 sub-populations with respect to their long-lived PC phenotype. The results demonstrated that popD showed specificity to viral antigens (such as measles) to which the individual had not been exposed in multiple decades, suggesting that some cells in this population had existed in the marrow for that extended period of time. This was not found to be true for pops A and B. Additionally, the VH repertoire of this cell population was not dominated by any particular clone, nor were there strong clonal ties to any of the other populations, which also suggests that these popD cells had not been generated recently. These pieces of evidence led to the conclusion that popD are long-lived plasma cells. {Halliley et al (2015)}

1.5.4 Connections between Antibody-Secreting Cells and Plasma Cells

In Nguyen 2017, circulating antibody-secreting cell (ASC) plasma cell populations (pop2, pop3, and pop5) were further investigated for their role in vaccine responses.

Peripheral blood ASCs were identified based on the levels of CD38, CD27, and CD19 detected on particular cell types. Discussing the heterogeneity of post-vaccine ASC populations, they argued that, counter to established models, both CD19+ and CD19- ASCs contribute to the vaccine response. This suggests that the CD19- populations could either be newly created ASCs, or long-lived plasma cells (LLPCs) released from the marrow. While peripheral blood ASCs have been proposed as possible precursors to LLPCs in the bone marrow, due to lack of functional characterization of these cell types in both the blood and marrow, the ontological relationships between these PB ASCs and BM LLPCs was unknown. {Halliley et al (2015);Nguyen et al (2018)}

1.5.5 Four Models

The parallels in the cell surface markers used to define the BM and PB subsets gives rise to the parsimonious hypothesis that ASCs in the peripheral blood are the precursors of their cell surface marker homologues in the BM. Alternatively, these plasma cells subsets could have quite different relationships across peripheral blood and bone marrow.

Tissue	Population	CD19	CD38	CD138	Population	Tissue
Blood	pop2	+	+	-	popA	Marrow
	pop3	+	+	+	popB	
	pop5	-	+	+	popD	

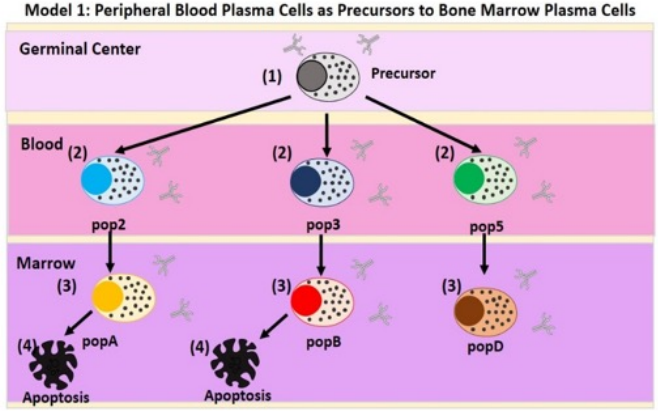
Table 1.1. Six Populations of Peripheral Blood ASCs and Bone Marrow Plasma Cells and their Parallel Cell Surface Marker Presentation. Table 1 shows the six blood and marrow populations of plasma cells and their cell surface marker expression: CD19, CD38, CD138. Each of the pairs have parallel patterns of these three markers: (1) pops 2 and A, (2) pops 3 and B, and (3) pops 5 and D.

Three alternative models are represented in Figure 1.2. Models 2 and 3 are incorporate the inference that pop3 appear to be derived from pop2 in the blood (Nguyen et al, 2018). Model 4 allows for the possibility that the PB and BM populations are independently derived in the two compartments. In Chapter 3, I compare the relative gene expression across the different types of plasma cells and different tissues. Similarity of profiles or pathway regulation across tissues types or based on cell surface markers illuminates the likely ontological relationship between the cell types.

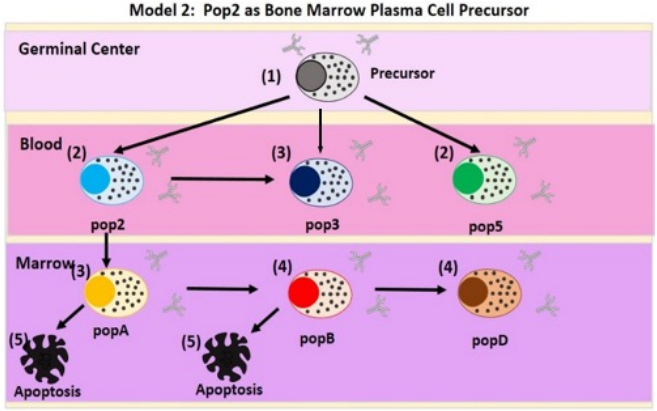
1.6 Transcriptomic Analysis

Transcriptomic analysis by bulk RNASeq is commonly used in gene expression analysis, where mRNA is sequenced from a very large number ("bulk") of cells. The sequenced RNA is then mapped back to the coding regions within the genome by a process known as short read alignment. The relative levels of reads across samples are normalized

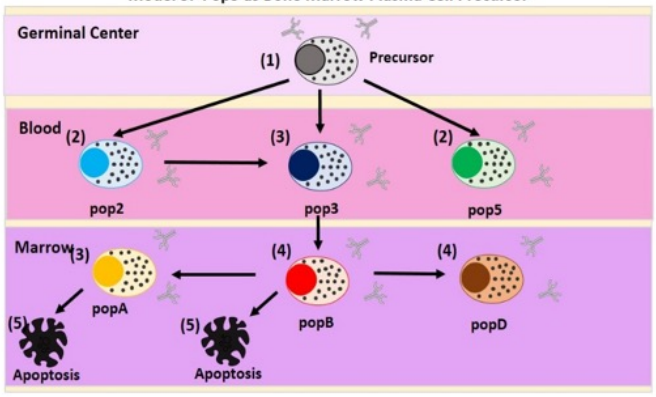
Model 1: Peripheral Blood Plasma Cells as Precursors to Bone Marrow Plasma Cells



Model 2: Pop2 as Bone Marrow Plasma Cell Precursor



Model 3: Pop3 as Bone Marrow Plasma Cell Precursor



Model 4: Independent Precursor Model

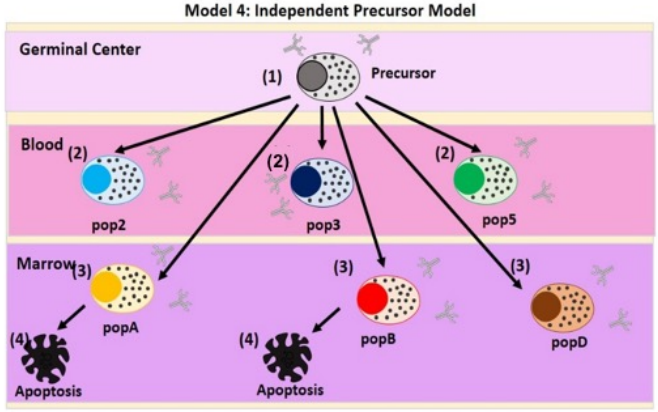


Figure 1.2. Four Models of Development of Long-Lived Plasma Cells and their Precursors. These are four potential models of development of long-lived plasma cells from peripheral blood and germinal center precursors. (1) Model 1 presents a germinal center precursor that eventually develops into peripheral blood antibody secreting cells (pops 2, 3, and 5), which subsequently enter the marrow and become their counterpart (based on parallel cell surface markers, Table 1) bone marrow plasma cell populations (pops A, B, and D, respectively). (2) Model 2 shows the GC precursor population develops into pops 2, 3, or 5. Of these three populations, pop2 enters the marrow and becomes popA or remains in the peripheral blood and develops into pop3. In the marrow, popA either apoptoses or develops into popB. PopB, as a transitional population, also either apoptoses or develops into popD. (3) Model 3 is similar to Model 2, however, it is pop3 as popB enters the marrow rather than pop2. At this stage, popB either apoptoses, or develops into popA or popD. (4) Model 4 is the independent precursor model which presents the null hypothesis that these 6 populations arise independently from each other.

to their library sizes using software such as EdgeR, and then converted to the log2 scale to facilitate statistical assessment of differential gene expression across experimental conditions. Additionally, gene expression can be ranked and assessed for gene set enrichment within particular pathway sets that are relatively upregulated or downregulated across experimental conditions and cell types.

1.7 Normalization and Differential Expression Analysis

Further normalization to remove or adjust for batch or experimental conditions can be performed using supervised normalization of microarray data (SNM). {Mecham et al (2010)}. This method utilizes modelling of genomic data in the context of study-specific biological variables and adjustment variables. The biological variables describe the variability in the data that is the target of the study and should be preserved (i.e. cell type, cell population, etc.). The adjustment variables are the variables that capture variability due the “experimental or biological setting” and can be removed, usually if they are technical such as a batch effect, or fit as a covariate, for example if they are a known biological effect

such as gender that influence the data but are not of primary interest. Normalization in this way uses information across all genes to make the adjustments, so is more accurate than simply including the factors in each individual gene model. {Law et al (2016);Nikolayeva et al (2014);Robinson et al (2010);Varet et al (2016)}

1.8 Gene Set Enrichment Analysis

With increasing amounts of gene expression data, an ongoing challenge is the interpretation of this data in a biological context. Gene set enrichment analysis is a tool developed by the Broad Institute that provides one method for the analysis and interpretation of gene expression data within the context of pre-defined gene sets. These gene sets were created based on similar biological meaning, function, chromosomal location, or regulation of molecular processes. The principle behind GSEA is to determine whether the genes in a particular pre-defined gene set are mostly upregulated or downregulated or if the members of this gene set are interspersed throughout the ordered expression profile of the study. The former would indicate that the gene set (or pathway) shows a coordinated pattern of expression that differentiates the samples of interest. {Mootha et al (2003);Subramanian et al (2005)}.

In order to evaluate the significance of the enrichment, GSEA calculates an enrichment score that determines how many genes in the pre-defined pathway are overrepresented in the ranked expression profiles at the top or the bottom (most expressed or least expressed genes). This method is executed by iterating through the ranked gene list and increasing the value of a statistic when a gene in the pathway of interest is encountered and decreasing the value of the statistic when a gene that is not in the pathway is

encountered. The enrichment score returned is based on the maximum value of the statistic over the course of the entire "walk". This statistic is based on the weighted Kolmogorov–Smirnov-like statistic. {Mootha et al (2003);Subramanian et al (2005)}

Once this statistic is determined, the statistical significance of the enrichment score is determined by the use of multiple permutations after randomization of the data and recalculating the enrichment score for the randomized data. This permutation and recalculation is performed 1000 times. The resultant p-value is based on the distribution of enrichment scores observed in the permutations. Finally, this set of significance values is adjusted for multiple hypothesis testing by 1) normalizing the enrichment score for each pre-defined gene set to account for the varying number of genes in each set, 2) determining the false discovery rate for each of the normalized enrichment scores by comparing the observed and null distributions for the normalized enrichment scores. These two values, the normalized enrichment score and the calculated false discovery rate, are both used to determine which pathways are included as significantly differentially regulated across conditions. {Mootha et al (2003);Subramanian et al (2005)}

In this thesis, analyses with purified populations of each of the six cell types (pop2, pop3, pop5, popA, popB, and popD) was performed, including differential gene expression analysis across the BM and PB tissues as well as between the cell types. Gene set enrichment analysis (GSEA) was performed to determine which pathways are up- and down- regulated across these cell types. The pathways of interest were drawn from public knowledge databases such as the MSigDB and BioCarta, PantherDB, and KEGG. {Mootha et al (2003);Subramanian et al (2005)}

1.9 Proteomics

The environmental microniche of the LLPC is hypothesized to play a significant role in the differentiation of these plasma cells via signaling from the cell types in the bone marrow, including mesenchymal stromal cells (MSCs), osteoclasts, and osteoblasts. The Lee group at Emory have created an *in vitro* system that partially replicates this BM environment and supports the long-lived phenotype using the secretome created by mesenchymal stromal cells. Irradiated MSCs have been shown to generate greater amounts of exosomes than non-irradiated MSC and appear to support higher survival of co-cultured plasma cells. In order to assess which proteins directly secreted by the MSCs (i.e. the MSC secretome) as well as the proteins output by MSCs via their exosomes, Ronghu Wu's group at Georgia Tech used mass spectrometry to assess the presence of the proteins in each of these fractions. In Chapter 2, I describe how the proteomic profiles were matched to transcriptomic data of genes with which they potentially interact, allowing for further interpretation. {Nguyen et al (2018)}

In short, mass spectrometry ionizes the components of a sample (such as the proteins or peptides in a secretome) and sorts these by their mass-to-charge ratio (m/z). The individual signatures of these molecules are then identified by correlating their spectra to those of known molecules. {Aebersold et al (2003)}. A limitation of proteomics is that it is not comprehensive, so the absence of detection does not necessarily imply absence of a protein from the sample. That is, some proteins could be present at such low levels that given the amount of sample, they would be undetected by the mass spectrometer.

However, when a protein is present in the secretome or exosome samples, particularly when its receptor is significantly differentially expressed between different populations of plasma cells (SLPC vs. LLPC), this lends itself as stronger evidence of the role of that protein and its candidate pathways, thereby elucidating the role of the extracellular environment in the development and differentiation of LLPC. {Halliley et al (2015);Nguyen et al (2018);Nguyen and Lewis et al (2018)}

1.10 Single-Cell RNA Sequencing

Nguyen 2017 provided a cell-surface marker-based assessment of heterogeneity in the plasma cell populations in the blood. Further investigation into the heterogeneity of these plasma cells at the whole transcriptome level in both the PB and BM bears consideration and can be assessed using single cell RNASeq (scRNASeq). Unlike bulk RNAseq, scRNASeq allows a more discerning analysis of the cell-to-cell differences in gene expression, since bulk RNASeq analysis is unable to resolve contributions of multiple sub-cell types to the signal present in the "bulk" of cells. In Chapter 4, I characterize transcriptomes generated using the SmartSeq2 method to generate full-length single cell cDNA libraries. The single-cell data was clustered using canonical correlation analysis which determines association between two variables by measuring the canonical covariates that capture the variability within and across both variables. After clustering, the gene expression data was pooled across cells, to minimize variability of lower-expressed genes. Further, pooling allowed us to analyze the scRNASeq data using standard bulk RNASeq

methods, overcoming over-dispersion and drop-out of low-abundance transcripts. {Chambers et al (2018);Sharma et al (2018);Wallrapp et al (2018)}

1.11 Scope of Thesis

Lack of granular understanding of the biochemical properties of the plasma cells (PC) in peripheral blood and bone marrow has impeded the study of diseases characterized by plasma cell dyscrasia. Since PC are responsible for the maintenance of serological memory, the recent insight that they cannot be understood as a single population has implications for analysis of their contribution to disease as well as healthy immune function. We consequently here describe comparative whole-transcriptome analysis of three populations (pop2, pop3, and pop5) from flow-sorted PC preparations from peripheral blood of five donors, and for three populations (popA, popB, and popD) of flow-sorted PC preparations from bone marrow of four donors.

1.11.1 Peripheral Blood Antibody Secreting Cells

Whole transcriptomic analysis of peripheral blood Pops 2, 3, and 5 highlights differences among these three sub-populations across multiple pathways. Especially of interest is that pops 2 and 3 are more similar to each other to the exclusion of pop5 to the extent that pop2 and pop3 cluster within individual. Further, Pops 2 and 3 show some genes that are upregulated in BM pops, again to the exclusion of pop5.

1.11.2 Comparison of BM Plasma Cells to PB Antibody Secreting Cells

Comprehensive assessment of the transcriptional capacities of the PC populations highlights differences among the sub-populations across multiple pathways, and allows us to contrast models for the order of differentiation of the PC sub-populations with popB and pop3 as precursors, or as transitional cell types. Although popA and pop2 may be considered to be short-lived (SLPC) and popD and pop5 as long-lived (LLPC), the transcriptomes of these peripheral blood and bone marrow cells are notably different, implying independent differentiation rather than the blood as the source of the differentiated bone marrow PC sub-types.

Further, LLPC showed a strong hypoxic signature and its metabolic signatures differed from SLPC and ASC. LLPC also showed differential expression of androgen and estrogen response pathways. This analysis showed that ontologically PB and BM are different. We have developed 4 models of development for plasmablast precursor to peripheral blood populations (2, 3, and 5) to bone marrow populations (A, B, and D).

Research reported in this Chapter has been published in Nguyen, D.C., Garimalla, S., et al. (2018) Factors of the bone marrow microniche that support human plasma cell survival and immunoglobulin secretion. *Nature Commumications* **9**: 3698, with the first two authors making joint contributions.

1.11.3 Single Cell RNASeq Analysis of Lupus and Healthy B-cells

I performed single-cell RNASeq analysis of two individuals (one healthy, and one lupus) of popD and naïve B-cells. The 246 popD cells and 36 naïve B-cells from each individual were sequenced using the SmartSeq protocol at ~1 million reads per cell, allowing for great depth of sequenced reads. The data was provided by Dr Steven Bossinger

at the Yerkes Genome Core, in collaboration with the Lee lab at Emory. I created a modified method of analysis through use of Seurat and *scran* to pool the reads within Seurat-generated clusters based on canonical correlation analysis, and then conducted differential gene expression analysis on these pooled sets. I describe specific pathways, including oxidative phosphorylation, antigen processing and presentation, and specific immunoglobulins that are significantly differentially expressed between lupus and healthy cells. Due to lack of power (2 individuals), it is difficult to determine if the results of the differential expression between individuals or due to disease.

CHAPTER 2

PERIPHERAL BLOOD ANTIBODY SECRETING CELLS

The primary goal of this study is to characterize the transcriptomes of long-lived plasma cells (LLPCs) and their potential precursors to ascertain the molecular basis of their physiological differences. The discovery of this long-lived cell type in the bone marrow and its role in long-lived immunity occurred in 2015. In conjunction with this discovery, three populations of peripheral blood (PB) antibody-secreting cells (ASC) were isolated and determined to be potential precursor populations based on parallel cell surface markers (Table 1). While Halliley 2015 conducted a gene expression analysis of bone marrow plasma cells, thus far, an in-depth analysis of the transcriptional properties of the PB-specific ASCs as LLPC precursors has not been performed. {Tipton et al (2018);Halliley et al (2015)}

2.1 Introduction

Both ASCs and plasmablasts (PMBs) contribute to significant antibody production during infection, despite both of these cell types being relatively rare. While PMBs are especially prominent in numbers during the initial stages of primary or secondary immune response (arising from memory B-cells in the case of the latter), they are relatively short-lived cell types. ASCs, on the other hand, are plasma cells that arise from activated B-cells.{Tellier et al (2018);Tellier et al (2017);Nguyen and Lewis et al (2018);Nguyen et al (2018);Halliley et al (2015)}

While some cursory analyses of ASC transcriptional programs have been conducted, they have largely performed on murine models. Even those studies that have been

considered within the context of naturally occurring ASCs from the human immune system did not assess ASCs as individual populations (such as pops1-5) but rather as a bulk population of cells. { Zhang et al (2009); Tellier et al (2018); Shi et al (2015); Nutt et al (2015); Klein et al (2008); Kassambara et al (2015); Jourdan et al (2011); Chatterjee et al (2007); Rawstron et al (2006); Klein et al (2003); Good et al (2009) ; Mei et al (2015)}. Subsequently, the results of these studies have largely been in determining ASC signatures that separate them from other B-cell subsets.{Tellier et al (2018);Tellier et al (2017);Shi et al (2015);Nguyen and Lewis et al (2018);Nguyen et al (2018);Kassambara et al (2015);Halliley et al (2015)}

One such tool that has been developed for this purpose is GenomicScape that provides a unique transcriptomic signature for various stages of development and differentiation of B-cells from naïve B-cells to terminally differentiated plasma cells. {Kassambara et al (2015)} These studies have converged on ASC-specific transcription factors that are associated with the differentiation process of ASCs: IRF4, Blimp-1, and XBP1. All three of these factors have been implicated in the differentiation of B-cells into ASCs. IRF4 has been determined to be necessary for the survival of ASCs. Blimp1 has been suggested to play a role in the migration of ASCs from the lymph organs where they develop. XBP1 plays a key role in ASC secretion. Blimp1, in conjunction with IRF4, has been shown to be essential for regulation of ASC metabolism. In addition to XBP1, Blimp1 is also key in the unfolded protein response in ASCs. {Halliley et al (2015);Shi et al (2015);Chatterjee et al (2007);Kassambara et al (2015);Klein et al (2008);Rawstron et al (2006);Klein et al (2003); Good et al (2009); Mei et al (2015) }

Integrins (especially ITGß7), sphingosine-1-phosphate-receptor-1 (S1PR1), CXCR4, and VLA4, are all required, at various stages, for the egress of ASCs from the secondary lymph organs in which they arise, as well as for homing to the bone marrow.{Nutt et al (2015);Tellier et al (2017);Tellier et al (2018)}

As the primary role of the ASCs is to produce copious amounts of antibody, much of its cell organization and energy are devoted to this function. The ER, as well as the golgi apparatus, increase dramatically to accommodate this high protein production. The unfolded protein response (UPR) and autophagy are both key in the management of physiological stress. The UPR manages endoplasmic reticulum (ER) stress during phases of high protein production. Without effective curtailing of this physiological stress by the UPR, the cell may apoptose. {Shi et al (2015);Nutt et al (2015);Tellier et al (2017);Tellier et al (2018)}

Autophagy, especially via mTOR signaling, is also involved in the management of ER stress by restraining ER growth and providing further energy to the cell for protein production. Energy for protein production in ASCs is derived largely from glycolysis (as evidenced by increase in GLUT1 in ASCs) and fatty acid metabolism, which the shift in preferred metabolism thought to be a hallmark of the long-lived phenotype. {Whitehead et al (2016);Patergnani et al (2013)}

The survival of ASCs has been associated with genes such as BCMA, APRIL, and Mcl-1, and SDC1 (or CD138). However, a clear, granular understanding of the mechanisms involved in the longevity of ASCs has been elusive, thus far. A large reason for this could be the basic assumption that these cell types should be considered as one

homogenous group during analysis. In doing so, the transcriptomic signal associated with subsets of ASCs would have been undetectable. This has prevented a more thorough assessment and understanding of the development of ASCs and their potential role as precursors of long-lived plasma cells. {Shi et al (2015);Nutt et al (2015);Tellier et al (2018);Klein et al (2003); Good et al (2009); Mei et al (2015); Nguyen et al (2018);Nguyen and Lewis et al (2018)}

2.2 Materials and Methods

FastQ files representing peripheral blood samples were retrieved from Doan Nguyen in the Lee lab at Emory. {Nguyen et al (2018)}. Paired end 100bp reads were mapped to the reference human genome hg38, using STAR {Dobin et al (2013)} to generate bam files. HTseq {Anders et al (2015)} was then used to assess read counts at the gene level. The peripheral blood samples were analyzed as single-end reads with approximately 25 million reads. The mapped libraries were then normalized accounting for differences in size and dispersion using the default parameters of EdgeR for gene abundance levels as TMM. {Robinson et al (2010);Nikolayeva et al (2014)}. Read counts were converted to the log base 2 scale and principal component variance analysis (PCVA) in JMP-Genomics (version 8.0) was used to assess the contributions of Individual, Batch, Cell Population, and Tissue to the weighted average variance explained by the principal components. Batch was removed in order to maximize the contrasts across the populations, and I elected to use remove the batch effect using the SNM (supervised normalization of microarray) method with Cell Population as the Biological variable and Batch as the adjustment variable with the Rm=True option. {Mecham et al (2010)}

For differential expression analysis, I then set a lower threshold for inclusion of genes in the analysis of 3 log2 units, selected by plotting the coefficient of variance against average abundance. Lower-abundance features were removed for all downstream analyses as the high variance and small sample size is likely to lead to false positives. Differences among cell populations were assessed by general linear modeling on a gene by gene basis using JMP v8.0. Gene expression was visualized by hierarchical clustering using standardization of each gene across individuals and Ward's method to weight the correlations, using JMP Genomics.

Gene set enrichment analysis was conducted using the Broad Institute's GSEA pre-ranked program. {Subramanian et al (2005); Liberzon et al (2011); Liberzon et al (2015)}. The t-statistic was used as the user-input rank for this analysis and it was run against the hallmark pathways of GSEA. The GSEA pre-ranked program analyzes this user input list by organizing the genes by rank. The algorithm then iterates through the list, adding to the enrichment score of a gene set when a gene within that gene set is encountered; and subtracting from the enrichment score whenever a gene that is not within the gene set is encountered. The enrichment score that is reported by GSEA is the maximum value of the enrichment score across the entire iteration. Since the enrichment score is affected by the gene set size, the normalized enrichment score is calculated by dividing the calculated enrichment score by the mean of the enrichment scores of multiple permutations of the dataset (in this study, 1000 permutations were used). {Subramanian et al (2005);Liberzon et al (2011); Liberzon et al (2015)}.

The GSEA was run against Hallmark gene sets in the MSigDB resource. These gene sets are generated by computing the overlaps of other gene sets in the MSigDB gene sets

to define biological states that have been well-characterized. In using these gene sets, redundancy and noise is minimized in the GSEA analysis. Further, these Hallmark gene sets provide the "founder" gene sets from which they were compiled for further analysis or individual pathways. The radar plot analysis in both Chapters 2 and 3 is based on these founder gene sets from which the Hallmark gene set analysis is derived. These founder pathways were further highlighted within the Hallmark gene sets that were determined to be significantly differentially regulated across subsets of cell types. {Subramanian et al (2005);Liberzon et al (2011);Liberzon et al (2015)}.

The first principal component of the genes in each pathway was used to visualize the overall regulation of the gene set. The sign of the PC does not necessarily signify that the associated pathway is up- or down-regulated as principal component analysis (PCA) simply evaluates covariance, and hence is blind to whether genes are up- or down-regulated. Furthermore, activators and repressors can be regulated in either direction, so systems modeling must be used to infer the likely consequence of differential expression in any given pathway – but in most cases appropriate mathematical models are not yet available. Three further potential issues arise in use of this method of portraying the regulation of these gene sets using the first PC: (1) the PC1 captures less than 20% of the variance of the gene sets, (2) the sign of the PC1 arbitrarily flips when computed in R, (3) the PC1 is dominated by the signal from a subset the genes and thus does not accurately represent the overall gene expression of the gene set. To address these issues, we only presented gene sets for which the first principal component captures at least 45% of the variance, verified the direction of the PC, and compared the most highly contributing genes

of each PC1. Nevertheless, PCA is a powerful approach to representing coordinated regulation of genes in pathways.

2.3 Results

While the Lee Lab identified five populations of PB ASCs (pops1-5), pops 1 and 4 are extremely rare and transitional and we were not able to isolate adequate numbers of these cells to further analyze them. The remaining three populations of interest in the blood included pop2, pop3, and pop5.

Analysis of the transcript abundance measures of 674 genes that are significantly differentially expressed between these three populations is shown in Figure 2.1A. The overall structure of the dendrogram shows pop2 (light blue) and 3 (dark blue) clustering together (Fig 2.1A, Box1) within individual to the exclusion of pop5 (Fig 2.1A, Box2). There are 50 genes that are significantly ($p < 0.05$, FDR < 0.05) differentially expressed between pops 2 and 3. Between pops 2 and 5, 529 genes are significantly differentially expressed. Between pops 5 and 3, there are 432 significantly differentially expressed genes. These relationships are confirmed by the principal components analysis in Fig 2.1B. The first principal component explains 56.61% of the variance which separates pop5 from pops 2 and 3. The second principal component, explaining approximately 7% of the variance, captures the differences between pops 2 and 3.

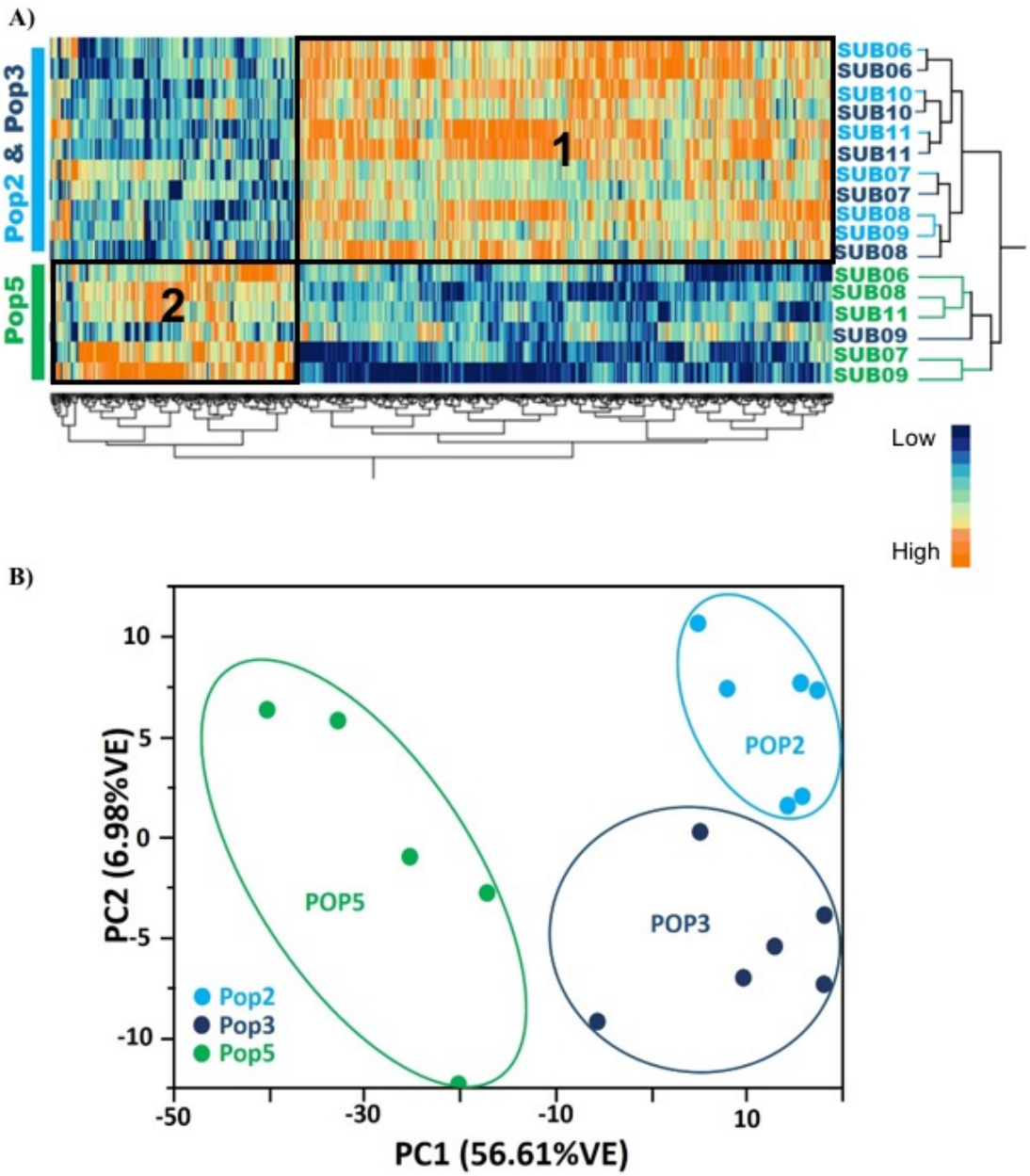


Figure 2.1. Analysis of Differentially Expressed Genes between Pop2, Pop3, and Pop5. (A) Two-way hierarchical clustering of 674 genes that are significantly (FDR < 0.05, p < 0.05) differentially expressed genes between any two of the three antibody-secreting cell populations (pop2, pop3, pop5). Box 1 encompasses the genes that are upregulated in pop2 and pop3 in comparison with pop5. Box 2 encompasses the genes that are upregulated in pop5 compared to pops 2 and 3. Pop2 and pop3 are more similar to each other than to pop5, clustering mostly within individual. (B) Principal component analysis of these differentially expressed genes shows that the first principal component captures 56.61% of variation. PC1 separates out pops 2 and 3 from pop5. PC2, which captures 6.98% of the variation, separates out pop2 and pop3.

Further analysis of these genes was conducted through gene set enrichment analysis (GSEA). GSEA (Figure 2.2) highlights the molecular differences between the three populations from the peripheral blood. The abbreviations for these gene sets and their full names are displayed in Table 1.1. Figure 2.2A shows the two-way hierarchical clustering of the first principal components of each of these significantly differentially expressed (nominal p-value < 0.05, FDR < 0.05) gene sets between pops 2, 3, and 5. The PC1 of each of the 30 pathways (x-axis) is displayed for each of the three ASC populations (y-axis). In concordance with Fig 2.2A, the gene set enrichment patterns of pops 2 and 3 are extremely similar to one another to the exclusion of pop5. However, enrichment in a particular gene set does not necessarily equate to upregulation of the pathway due to the fact that inhibitors and activators are not differentiated from one another in the enrichment analysis.

These same 30 gene sets are shown in Figure 2.2B as bubble plots that compare the normalized enrichment score of each pathway across all pairwise comparisons of the three peripheral blood ASC populations. Enrichment scores provide a representation of the bias toward up- or down-regulation of all of the genes in the pathway. They can be used to polarize the PC signs, which are arbitrarily assigned wherever there are similar numbers of genes in each direction. In these plots, the enrichment score is on the x-axis with the size of the point being directly associated with the negative log p-value of each gene set. The gene sets that show enrichment in pop2 in comparison to pop3 and pop5 falls broadly under cell cycle and proliferation (E2F Targets, G2M Checkpoint, Mitotic Spindle, Myc Targets V1 and V2). While no pathways are uniquely enriched in pop3, there are two gene sets that are upregulated in pop3 as compared to pop2: Hypoxia, and TNFA Signaling.

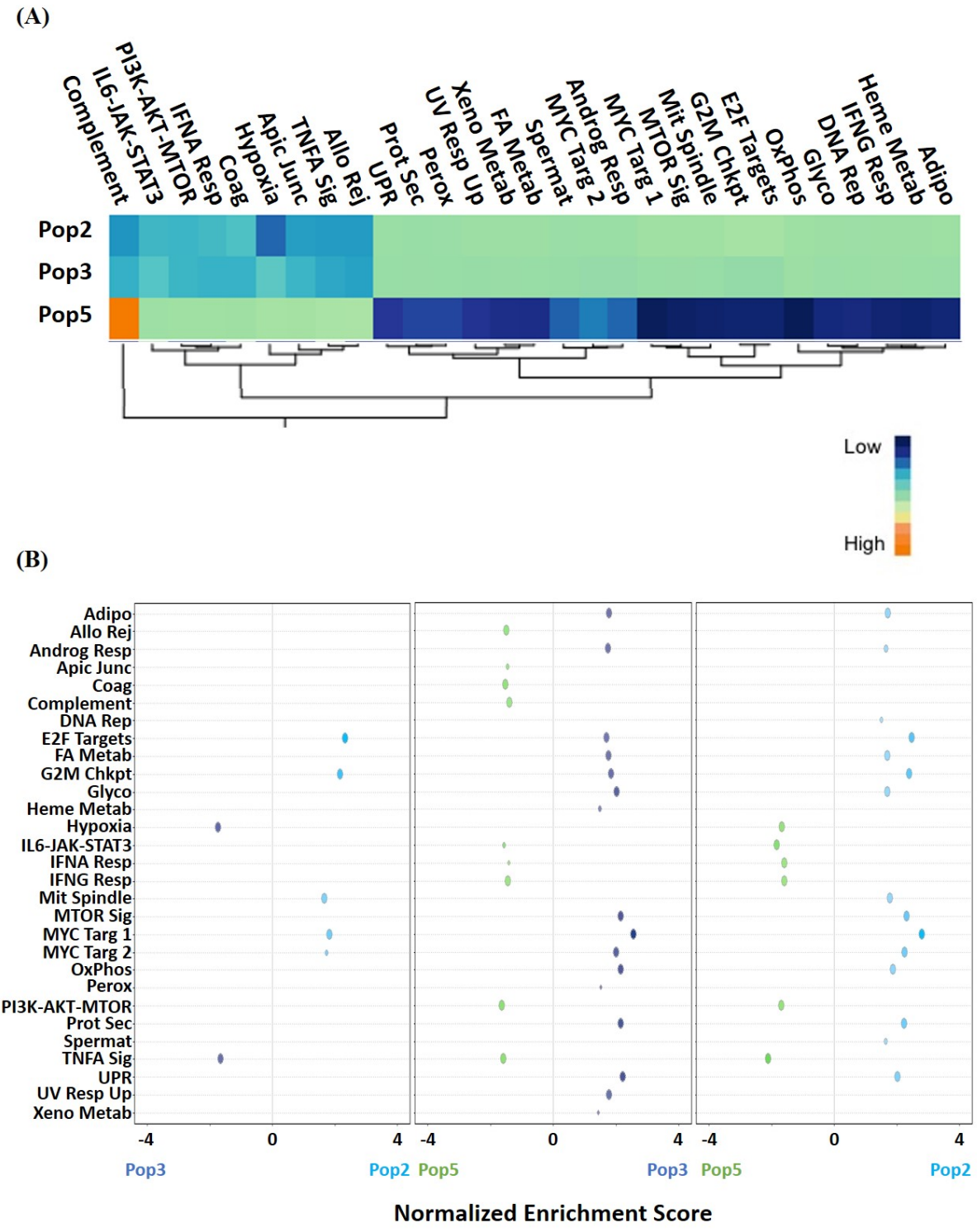

(A)
PI3K-AKT-MTOR
IL6-JAK-STAT3
Complement
IFNA Resp
Coag
Hypoxia
Apic Junc
TNFA Sig
Allo Rej
UPR
Prot Sec
Perox
UV Resp Up
Xeno Metab
FA Metab
Spermat
MYC Targ 2
Androg Resp
MYC Targ 1
MTOR Sig
Mit Spindle
G2M Chkpt
E2F Targets
OxPhos
Glyco
DNA Rep
IFNG Resp
Heme Metab
Adipo
Pop2
Pop3
Pop5
Low
High
(B)
Adipo
Allo Rej
Androg Resp
Apic Junc
Coag
Complement
DNA Rep
E2F Targets
FA Metab
G2M Chkpt
Glyco
Heme Metab
Hypoxia
IL6-JAK-STAT3
IFNA Resp
IFNG Resp
Mit Spindle
MTOR Sig
MYC Targ 1
MYC Targ 2
OxPhos
Perox
PI3K-AKT-MTOR
Prot Sec
Spermat
TNFA Sig
UPR
UV Resp Up
Xeno Metab
-4
0
4
Pop3
Pop2
Pop5
Pop3
Pop5
Pop2
Normalized Enrichment Score

Figure 2.2. Significantly Differentially Regulated Pathways between Pop2, Pop3, and Pop5. (A) This is the two-way hierarchical cluster of the first principal component (%VE > 45) of 30 pathways (nominal p-value < 0.05, FDR < 0.05) that were significantly differentially regulated between pops 2, 3, and 5 based on the Broad Institute preranked gene set enrichment analysis. (B) This bubble plot of the normalized enrichment score of each of the 30 pathways (from 3A) show the normalized enrichment score of each pathway within each of the two-way comparisons of pops 2, 3, and 5. The size of each bubble is directly correlated with the negative log nominal p-value of each pathway. The x-axis shows the normalized enrichment score of the each of two-way comparisons of pops 2, 3, and 5. The color the bubbles is a spectrum between each of the two populations, trending closer to the color of denoting the population as the enrichment score increases. The abbreviations for these gene sets and their full names are displayed in Table 2.

Pop5 shows enrichment in comparison with pop2 and pop2 in gene sets that fall broadly under the category of inflammation (IL6-JAK-STAT3, IFNα Response, IFNG Response, and TNFα Signaling) as well as PI3K-AKT-MTOR pathway.

There is some discordance between the Figure 2.2A and 2.2B for 12 pathways: Allograft Rejection, Coagulation, Complement, DNA Repair, Heme Metabolism, IFNG Response, Mitotic Spindle, MTORC1 Signaling, Peroxisome, Spermatogenesis, UV Response Up, and Xenobiotic Metabolism. Potential reasons for this discrepancy is described in the previous section (Chapter2, Section 2.2: Materials and Methods). Further analysis of these pathways presents the expression of the genes driving the first principal component of these pathways (shown in Figure 2.2A) in contrast with the accompanying gene expression for each population for all genes within the gene set. The latter of these was used in the analysis of the weighted enrichment score for each pathway (Figure 2.2B). While the hierarchical clusters of the genes driving PC1 show very clear patterns across populations within each gene set, it is apparent from the expression patterns of all of the genes within each gene set that differential regulation and enrichment of genesets is not as straightforward as the PC1 expression presents which leads to discrepancies between the two presentations of gene set enrichment (Appendix Figures 1 and 2). In such cases of

discrepancy, normalized enrichment score is used for interpretation of the enrichment of these pathways.

Of these 30 pathways, six pathways were noteworthy due to their purported role in ASC survival: Apoptosis, ECM, Hypoxia, Cell Cycle, TNFα Signaling, and UPR. In order to explore these pathways further, I plot the normalized expression values of differentially expressed genes in the form of radar plots in Figure 2.3.

In the apoptosis pathway (Figure 2.3A), pop5 shows the highest expression for five of the seven differentially expressed genes as compared to pop2 and pop3 which each present the highest expression for one gene (*CASP3* and *PARP2*, respectively). Pop5 shows the highest expression for *TNFRSF1A, FAS, MAPK3, PI3KCD*, and *NFKBIA* within the apoptosis pathway (Figure 2.3A).

The Extracellular Membrane Interaction pathway (ECM, Fig 2.3B), shows pops 3 and 5 presenting the highest expression of *SDC1*, a known survival factor in ASCs, to the exclusion of pop2. Pop2 shows highest expression in the *CHAD* gene, involved in cell adhesion, and *ITGB1*, an ASC migration marker, as compared to pops 3 and 5. Pop3 presents highest expression in *LAMC1*, which has been implicated in a variety of biological roles including migration, differentiation, and adhesion, as compared to pops2 and 5.

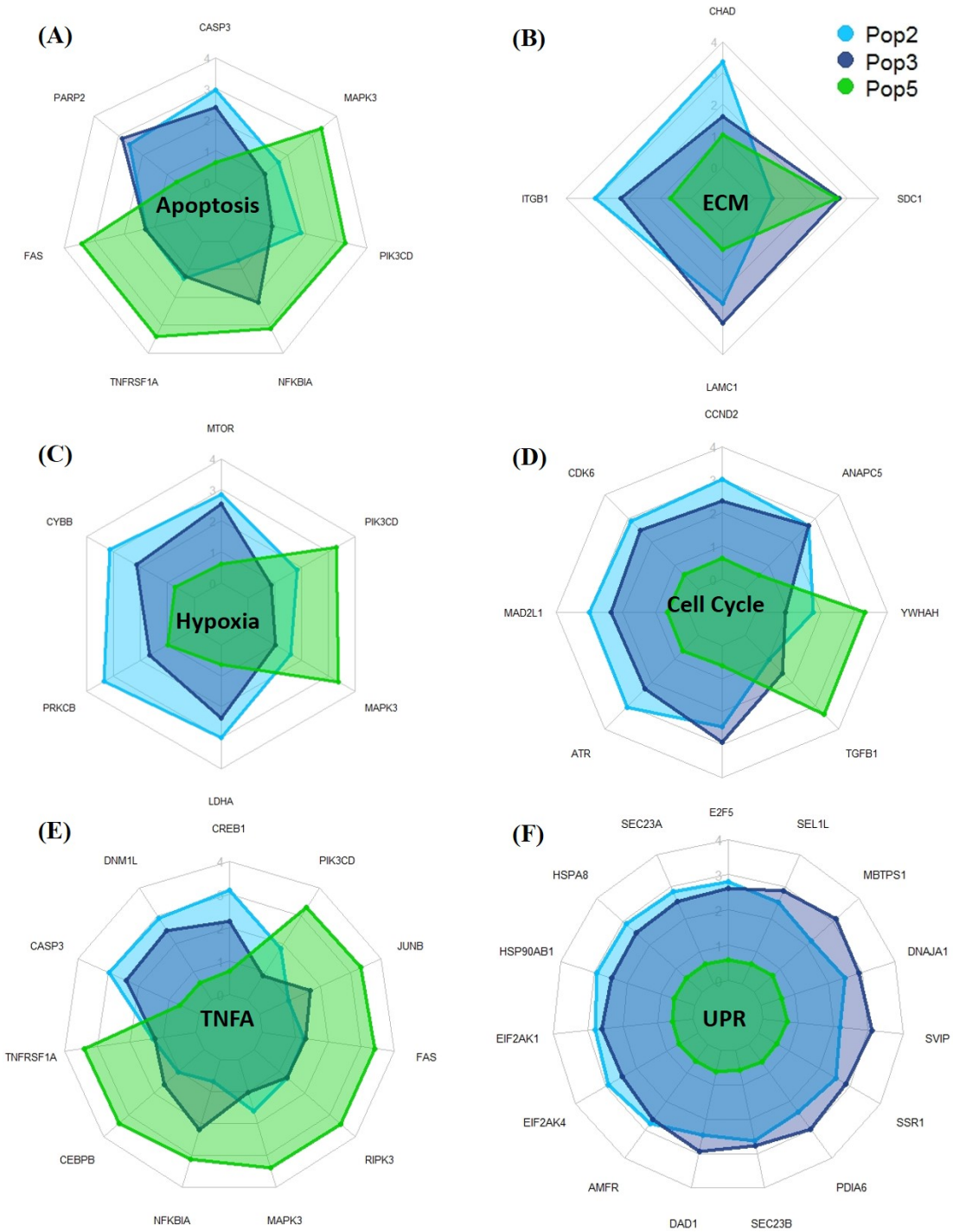


Figure 2.3. Differential Gene Expressed for Apoptosis, ECM, Hypoxia, Cell Cycle, TNFA, and UPR Pathways for Pop2, Pop3, and Pop5. These are radar plots of selected pathways (apoptosis (A), extracellular matrix interactions (B), hypoxia (C), cell cycle (D), TNF alpha (E), and unfolded protein response (F)) and the standardized least square means for the significantly ($p < 0.05$, FDR < 0.05) differentially expressed genes within them across populations 2, 3, and 5. Pop2 is shown in light blue, pop3 is shown in dark blue, and pop5 is shown in green.

Figure 2.3C also shows that hypoxia was most enriched in pop5. In this plot, pop5 has the highest expression of the three populations in *MAPK3*, which plays a role in inhibition of autophagy, and *PIK3CD*, which is necessary for authophagy. Pop2 presents the highest expression of the remaining four DEGs*: LDHA, PRKCB, CYBB*, and *MTOR*; the former three of which all are implicated in the inhibition of autophagy. *MTOR*, however, is part of the mTORC signaling pathway which is used to modulate ER stress during protein production. Pop3 does not significantly differentially express these four genes as compared to pop2.

Within the cell cycle pathway (Figure 2.3D), popD shows the highest expression for *TGFB1* and *YWAH*. Pop2 and 3 have comparable expression for the remaining six genes: *E2F5, CDK6, CCND2, ATR, MAD2L1*, and *ANAPC5*. Pop5 is significantly lower than pop2 and 3 for these remaining six genes. In the TNFA pathway (Fig 2.3E), popD is significantly higher than pops2 and 3 for eight genes: *MAPK3, NFKBIA, CEBPB, TNFRSF1A, PIK3CD, JUNB, FAS*, and *RIPK3*. Pop2 present the highest expression of the three populations in three genes: *CREB1, DNM1L*, and *CASP3*; the latter two of these is associated with apoptosis.

As previously described, UPR is necessary for the management of physiological ER stress in the ASC to accommodate its primary function of high antibody production. In the absence or ineffectiveness of the UPR, the cell apoptoses. Figure 2.3F shows the differentially expressed genes (DEG) in UPR across all pairwise comparisons of the three ASC populations. While pops2 and 3 have nearly comparable levels of most of the differentially expressed genes, pop5 has conspicuously low expression of all of the DEGs within this pathway.

2.4 Discussion

ASCs have been characterized in numerous studies in both human and murine models. All of these analyses, however, have treated ASCs as one homogenous population and were unable to establish granular understanding of functional mechanisms and ontogenic relationships of these cells. In this study, I have conducted a transcriptomic analysis of three subsets of these cell types to further elucidate their role in the rise of long-lived plasma cells.

While cell-sorted whole RNASeq profiling is insufficient to fully determine the relationships between these three cell types and their functional consequences therein, I discuss specific aspects of the processes relevant for cell survival, ER stress, and autophagy in the context of each cell type. Further I infer relationships between pops 2, 3, and 5 based on this analysis.

When ASCs are analysed as a group, the transcriptomic signal of individual populations becomes impossible to determine. In this work, we see that there are significant differences between three of, at least, five subpopulations of ASCs in the peripheral blood. Figure 2.1 shows that two of these populations (pop2 and pop3) are more similar to each other, at least transcriptomically, than to the third population (pop5). While it is possible that transcriptomic profiling of pops 1 and 4 might have shed further light on the relationship between these populations, their rare and transient nature made their collection impossible for the purposes of this study.

Figure 2.2 further establishes the similarity of pops 2 and 3 to the exclusion of pop5. Principal component analysis of the Hallmark pathways especially shows the stark

differences in function that is likely to derive from differential gene expression. Interestingly, pop5 is especially enriched for many inflammatory pathways. This could be indicative of preparation of this cell type for apoptosis, for example as is suggested by in-depth individual pathway analysis of the apoptosis and UPR pathways (Fig 2.3A and 2.3B). As previously discussed, the primary function of ASCs is antibody production. As compared to its precursors, ASCs have enlarged ER and Golgi to accommodate this demand. This massive demand for protein production places significant stress on the ER and the cell which is modulated by the UPR and apoptosis pathways. Both the gene set enrichment analysis and individual gene-level analyses show that the UPR for pop5 is comparatively downregulated, which presumably means that there is reduced requirement for protection against ER stress in this population.

These results are contrary to the Model 1 which posits that each ASC cell type migrates to the marrow to give rise to the bone marrow plasma cell that presents the parallel pattern of cell surface markers (Table 1). Since pop5 has the same cell surface marker presentation as popD which has been determined to present the long-lived phenotype in the marrow, it might have been expected to have the highest relative demand for protein production. One possibility is that pop5 is completely independent of pops2 and 3 (Figure 1.2, Model 4). Another possibility is that while pop2 and pop3 still maintain the potential to develop into the long-lived phenotype, pop5 is a subset of cells that are in a pre-apoptotic stage due to their inability to tolerate the ER stress.

Another interesting finding from the GSEA is that the hypoxia and the TNFA signalling gene sets are both similarly enriched in pops 3 and 5 in comparison with pop2. This is unusual considering how similarly pops2 and 3 are to one another in other gene sets

as compared with pop5. Long-lived plasma cells in the marrow have to subsist in the hypoxic marrow environment. Under the assumption that ASCs are precursors to bone marrow plasma cells, enrichment of the hypoxia gene set is expected. Similarly, TNFA signalling is associated with upregulation of other plasma cell survival pathways including MAPK signalling, NFKB signalling, and PI3K-AKT signalling. This would be consistent with pop2 being the precursor population to pop3. Both of these populations are still healthy ASCs with managed-ER stress and an active UPR; pop5, in contrast, does not. Another possibility is as pop3 begins differential regulation of certain pathways to prepare it for entry into the bone marrow, the ER stress for some of these pop3 cells might exceed management by the unfolded protein response resulting in impairment of several gene sets, including UPR; subsequently, these pop3 that are unable to manage ER stress assume the pop5 identity by these cells as a pre-apoptotic stage.{Patergnani et al (2013);Liu et al (2012);Klein et al (2003)}.

Many of the pathways that are enriched in pop2 compared to pops 3 and 5 are specific for cell cycle and proliferation. Antibody-secreting cells, like other plasma cells, are known to be terminally differentiated cells. Consequently, the enrichment of these pathways in pop2 could be an artefact of transition from a proliferating precursor into the ASC phenotype, further suggesting that pop2 is the precursor population to pop3.

2.5 Conclusion

The findings, in conjunction, suggest that pops2 and 3 are closely related cell types that represent healthy, viable ASCs that continue to maintain the potential for the long-lived phenotype. The enrichment of proliferation pathways in pop2 suggest that it is a precursor to pop3 which begins to downregulate these pathways and upregulate hypoxia and TNFA signalling, in preparation for entry into the marrow space. However, during this time, if the autophagic and UPR stress response are unable to manage the additional stress of these transcriptional shifts, the ASC transitions into pop5, a pre-apoptotic stage. However, this hypothesis cannot be fully verified through whole transcriptome profiling alone. Further experiments will have to be performed to verify these conclusions.

CHAPTER 3

COMPARISON OF BONE MARROW PLASMA CELLS TO PERIPHERAL BLOOD ANTIBODY SECRETING CELLS

3.1 Introduction

Understanding of the pathological mechanisms responsible for diseases characterized by plasma cell dyscrasia has been hampered by lack of a granular understanding of the biochemical properties of the contributing cells. Multiple studies have suggested widely varying behavior of plasma cells, and Halliley 2015 provided strong evidence that bone marrow plasma cells (PC) should not be understood to be a single population. Since plasma cells are responsible for the maintenance of serological memory, this insight has implications for analysis of the contributions of PC to disease as well as to healthy immune function. {Halliley et al (2015);Bayer-Garner et al (2001);Aref et al (2003)}.

Four sub-populations of bone marrow plasma cells can be distinguished by the cellular differentiation surface markers CD19, CD38, and CD138, and are referred to as popA, popB, popC, and popD. {Halliley et al (2015)}. Similarly defined sub-populations are observed in peripheral blood (discussed in chapter 2), where they are referred to as pop1, pop2, pop3, pop4 and pop5. Of these, we have been able to assess the cell-sorted whole transcriptomes by RNASeq for the six most abundant populations: popA, popB, popD, pop2, pop3, and pop5, based on the CD antigens shown in Table 1. These plasma cell subsets were previously distinguished using analysis of their immunoglobulin VH repertoires, additional flow cytometry, and preliminary transcriptome analysis. {Halliley

et al (2015)}. However, the potential contributions of specific aspects of gene expression to functional differences among these subpopulations was not described.

The three CD markers themselves are implicated in functional differentiation as well. CD38, like CD138, is a plasma cell marker which distinguishes PC from other leukocytes. {Rawstron et al (2002);Medina et al (2002);Medina et al (2002)}. CD19 is a differentiation marker for B-cells and its differential expression suggests a role in differentiation of pop2, pop3, popA, and popB. {Ishikawa et al (2002)}. CD138 is known to be a cell-cell adhesion marker, which is required to maintain localization of cells within the bone marrow.{Moreaux et al (2009);Jourdan et al (1998);Gokden et al (2006);Bayer-Garner et al (2001);Bayer-Garner et al (2002);Bayer-Garner et al (1999);Bayer-Garner et al (2001);Bayer-Garner et al (2004);Bayer-Garner et al (2004);Bayer-Garner et al (2001);Aref et al (2003)}.

There has been extensive research conducted on plasma cell development and differentiation. Prior studies often focused on plasma cells in the spleen, gut, lymph nodes and bone marrow. Additionally, a majority of these studies focused on specific processes of the differentiation program. {Mei et al (2015); Zhang et al (2009);Rawstron et al (2006);Klein et al (2008);Kassambara et al (2015); Good et al (2009);Klein et al (2003);Chatterjee et al (2007);Jourdan et al (2011);Nguyen et al (2018);Mecham et al (2010);Robinson et al (2010);Shi et al (2015)}

The follow-up analysis of RNASeq data described in this Chapter provides a more comprehensive assessment of the transcriptional capacities of the plasma cell populations, not only in bone marrow but also in peripheral blood. The objectives are first to

characterize the differences in physiological capacities of the different PC sub-types, particularly as they relate to short-and long- lifespans and to their survival and differentiation in the hypoxic environment in bone marrow; second to evaluate whether sub-types defined by the same cell surface markers in the two compartments are homologous or more likely to be independently differentiated, and third to contrast four models for the differentiation of PC, with PopB/Pop3 as precursors of the other sub-types, or as transitional states between PopA and PopD or Pop2 and Pop5. It is concluded that the short-and long-lived subtypes are probably independently generated in the two compartments, with a precursor relationship as the more likely model.

3.2 Materials and Methods

3.2.1 Short Read Mapping and Gene Expression Quantification

FastQ files samples of peripheral blood samples and bone marrow samples were retrieved from Jessica Halliley and Doan Nguyen of the Lee lab at Emory. {Nguyen and Lewis et al (2018);Nguyen et al (2018);Halliley et al (2015)}. Paired end 100bp reads were mapped to the reference human genome hg38, using STAR {Dobin et al (2013)} to generate bam files. HTseq {Anders et al (2015)}was then used to assess read counts at the gene level. The bone marrow samples were paired-end and contained 80 millions reads. For cross-tissue analyses, the bone marrow samples were reduced to just the forward read to facilitate more direct comparison with the single-end peripheral blood samples which has approximately 25 million reads each. The mapped libraries were then normalized accounting for differences in size and dispersion using the default parameters of EdgeR for gene abundance levels as TMM. {Robinson et al (2010);Nikolayeva et al (2014)}. Read counts

were converted to the log base 2 scale and principal component variance analysis (PCVA) in JMP-Genomics (version 8.0) was used to assess the contributions of Individual, Batch, Cell Population, and Tissue to the weighted average variance explained by the first 3 PCs. Since Batch explained 53% of the variance, in order to maximize the contrasts across the populations, I elected to remove the batch effect using the SNM (supervised normalization of microarray) method {Mecham et al (2010)} with Cell Population as the Biological variable and Batch as the adjustment variable with the Rm=True option.

For differential expression analysis, I then set a lower threshold for inclusion of genes in the analysis of 3 log2 units, selected by plotting the coefficient of variance against average abundance as described in chapter 2. Differences among cell populations were assessed by general linear modeling on a gene by gene basis using JMP v8.0. Gene expression was visualized by hierarchical clustering using standardization of each gene across individuals and Ward's method to weight the correlations, using JMP Genomics. GSEA was performed as described in Chapter 2.

3.2.2 Proteomics Data Analysis

Fastq sequence files for the cell-sorted populations of 4 to 6 subjects were aligned to hg38 using STAR and HTSeq was used to generate gene-level counts which where normalized for library size and outlier transcript abundance with the TMM procedure in EdgeR as in Chapter 2. Resultant TMM were log2-transformed for each sample. Analysis of variance (ANOVA) was performed on these transcript abundance measures, contrasting the three ASC populations and three PC populations, retaining all genes at FDR < 0.05. From the original 17,640 genes, 2,558 were found to be significantly differentially

expressed between blood ASC and BM LLPC. Of these 2,558 genes, 556 were shown to be potential interacting partners for the 91 BM MSC secretome proteins using HIPPIE. {Alanis-Lobato et al (2017)}. These 556 genes were then standardized to account for difference in levels of expression across genes and hierarchically clustered using Ward's method in JMP Genomics (version 8.0). Furthermore, all potential PPI (protein-protein interactors) of YWHAZ and FN-1 (436 and 430, respectively) resulted in 126 overlapping potential PPI (of which 31 potential PPI are differentially expressed between blood ASC and BM LLPC). For both the set of 556 DEG and 126 overlapping genes, gene set enrichment analysis (GSEA) was performed using the t-statistic (derived from comparison of each gene between each pair of ASC populations) to rank genes, and pre-ranked gene set sets significantly enriched for high or low expression with FDR < 0.05 were deemed to be dysregulated between the two tissues. The expression of the standardized least-square-means (SLSM) for each gene in each gene set was then used to compute PC1 for the gene set, and these values were hierarchically clustered again using Ward's method in JMP Genomics.

I used the JMP software version 8 to conduct an analysis of variance (ANOVA) using a 5% false discovery rate to consider the interactions effects of treatment to analyze the data. Models were fit with day and treatment as fixed effects and significance of full model was reported.

3.3 Results

3.3.1 Differential Gene Expression Analysis of Bone Marrow Plasma Cells

Beginning with the bone marrow compartment, analysis of the transcript abundance measures of 2266 genes that are significantly differentially expressed between the three bone marrow sub-types (popA, popB, and popD) is shown in Figure 3.1A. The overall structure of the dendrogram shows each of the three populations forming a distinct cluster in which popB appears to be intermediate. These relationships are confirmed by the principal components analysis in Fig 3.1B. Approximately 923 genes are shared as differentially expressed by pop B and popD as compared to 359 genes shared by popB and popA. These patterns are consistent with popB being a precursor that differentiated into the other two types, or with popB being transitional between popA and popD, though it is possible that all 3 pops have independent origins with shared genes in common with a progenitor cell type (Figure 2, Models 1 and 4).

GSEA (Figure 3.2) provides insight into the molecular differences between all three populations, remarkably highlighting most of the same gene sets as in the peripheral blood, though with different patterns with respect to the PC populations. However, enrichment in a particular gene set does not necessarily equate to upregulation of the pathway, due to the fact that inhibition and activation is not differentiated from one another in the enrichment analysis. Figure 3.2A shows the two-way hierarchical clustering of the first principal components of each of these significantly differentially expressed (nominal p-value < 0.05, FDR < 0.05) gene sets between these three populations. The PC1 of each of the 32 pathways (x-axis) is displayed for each of the three populations (y-axis), pops A-D. In

concordance with analyses presented in Fig 3.1A, the gene set enrichment patterns of popA and popD are opposite to each other with popB appearing transitional between the two populations.

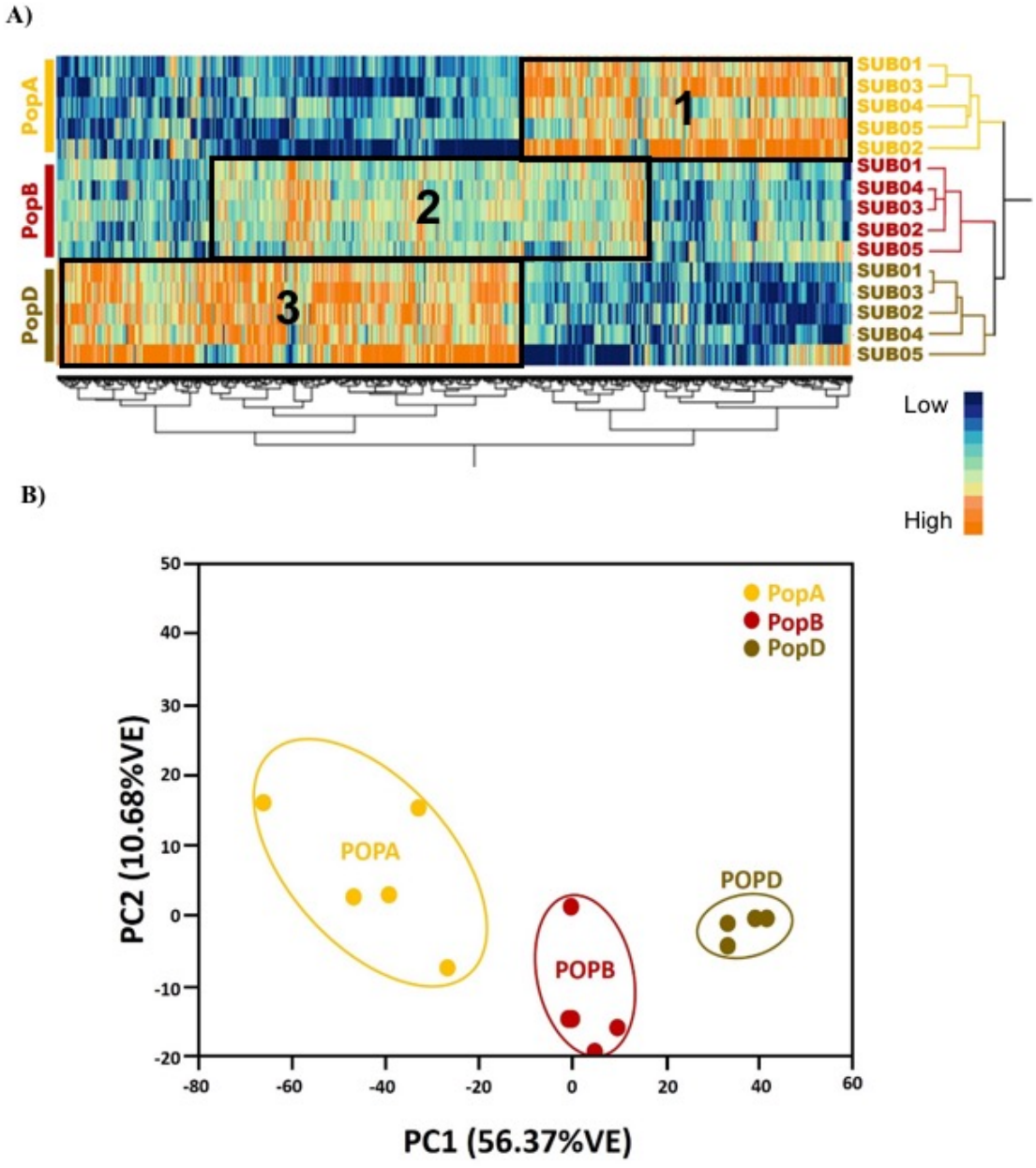


Figure 3.1. Analysis of Differentially Expressed Genes between PopA, PopB, and PopD. (A) Two-way hierarchical clustering of 2265 genes that are significantly (FDR < 0.05, p < 0.05) differentially expressed genes between any two of the three bone marrow plasma cell populations (popA, popB, popD). The samples cluster within population. Box 1 encompasses the genes that are upregulated in popA (yellow). Box 2 encompasses the genes that are upregulated in popB (red). Box 3 encompasses the genes that are upregulated in popD (brown). PopA and popD present opposite expression profiles with popB presenting an intermediate expression profile between pops A and D. (B) Principal component analysis of these differentially expressed genes shows that the first principal component captures 56.37% of variation. PC1 separates out all three bone marrow populations from one another.

To further elucidate the enrichment of these gene sets in Figure 3.2A, I recapitulated them in Figure 3.2B, by displaying enrichment scores on the x-axis with size of the point being directly associated with the negative log p-value of each gene set. The gene sets that show enrichment in popA in comparison to popB and popD falls broadly under inflammation (Allograft Rejection, Interferon Alpha Response, Interferon Gamma Response, and Reactive Oxygen Species), cell cycle and proliferation (DNA Repair, G2M Checkpoint, MYC Targets V1, and Mitotic Spindle), energy metabolism (Fatty Acid Metabolism, and Oxidative Phosphorylation), and MTOR pathways (MTORC1 Signaling and PI3K-AKT-MTOR). Note that In some cases, the enrichment shown in figure 3.2B does not match up with the PC1 patterns shown in 3.2A (for example with inflammatory response) because the genes that most contribute to PC1 are a subset of the genes that contribute to the normalized enrichment score and present a slightly different regulation pattern across the three populations in PC1.

As in Chapter 2, some pathways show discrepancy between the Figure 3.2A and 3.2B for 8 pathways: Apoptosis, Fatty Acid Metabolism, Hypoxia, IL2-STAT5, IL6-JAK-STAT3, P53, Reactive Oxygen Species, and Unfolded Protein Response. Some possible reasons for this discrepancy are described in Chapter2, Section 2.2: Materials and Methods. Additional analysis of these 8 pathways is presented in Appendix Figures 3 and 4. As is

discussed in Chapter 2, Section 2.3: Results, while the hierarchical clusters of the genes driving PC1 show very clear patterns across populations within each gene set, it is apparent from the expression patterns of all of the genes within each gene set that differential regulation and enrichment of gene sets has more nuance based on the weighted ranked enrichment score which presents the discrepancies between the two presentations of gene set enrichment (Appendix Figures 3 and 4). As previously discussed, in such cases of discrepancy, normalized enrichment score is used for interpretation of the enrichment of these pathways.

PopD shows enrichment in comparison with popB and popA in gene sets that fall broadly under the category of cell-cell adhesion and survival (Apical Surface, Epithelial Mesenchymal Transition, Hypoxia, and Myogenesis), sex hormones (Androgen Respose, Estrogen Response Early, and Estrogen Response Late), and signaling (TGF-β Signaling, TNF-α Signaling, and UV Response Down). There are no pathways that are uniquely enriched in popB in comparison to pops A and D.

Out of these 32 gene sets, I focused on six pathways of interest to further elucidate some of these functions (Figure 3.3). Each plot shows the significantly differentially expressed genes between any two of the three bone marrow populations. The least-squared means of each significant DEG is plotted for each population (popA in yellow, popB in red, and popD in brown) for each of these pathways in a radar plot.

Klein 2003 suggested a role for IL6, SDC1 (CD138), and IGF1 in the survival of plasma cells in the context of multiple myeloma; consequently, I selected the IL6 pathway (Fig 3.3A), Extracellular Membrane Interactions pathway (Fig 3.3B), IGF1 pathway (Fig

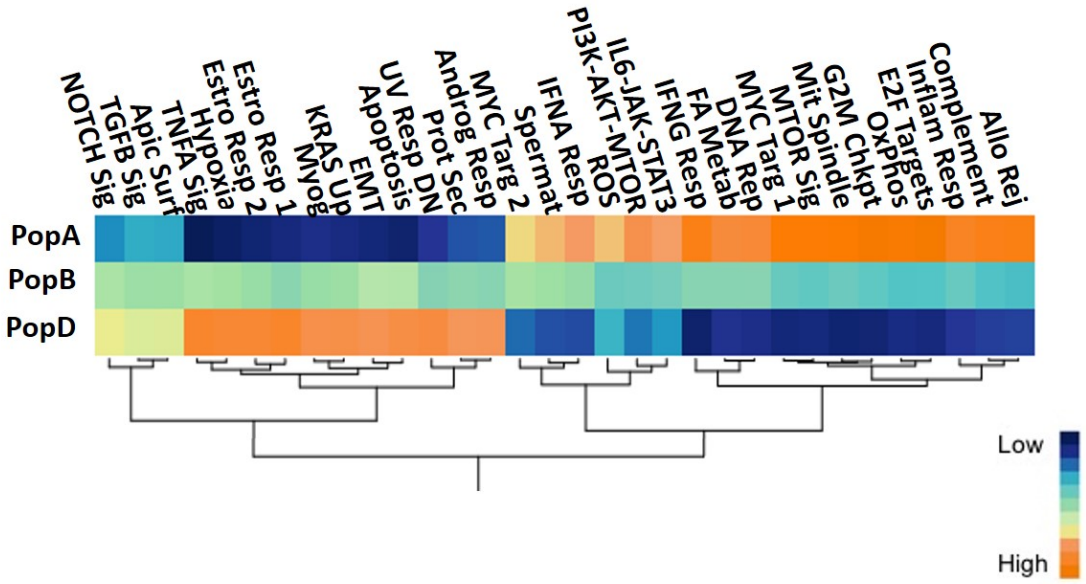
3.3C), and Apoptosis pathway (Fig 3.3D) for further analysis. TNFA Signaling (Fig 3.3E) and Hypoxia (Fig 3.3F) were selected as well because they were enriched in popD compared to the remaining two BMPC populations. { Klein et al (2003)}

Within the IL6 pathway (Figure 3.3A), *IL6ST*, *GAB1*, *TIMP1*, *PRKCD*, *NLK*, *NCOA1*, and *JUNB* are more highly expressed in popD than pops B or A. Of these, IL6ST binds to free IL6 and forms a signaling complex with IL6R within the context of plasma cells, which then upregulates the IL6-JAK-STAT3, MAPK, Ras/Raf, PI3K, and NFκB signaling pathways. *NCOA1* has been found to be a multiple myeloma susceptibility gene through variant analysis. *GAB1* is also involved in protein-protein interactions resulting in PI3K activation. *TIMP1* has been known to play a role in modulation of inflammation. *PRKCD* has been shown to play a role in survival of plasma cells. *NLK* also plays a role in signaling pathways. {Rodriguez-Bayona et al (2013); Wang et al (1995); Cokic et al (2015); Jourdan et al (2005); Klein et al (2003)}

Within the ECM (Figure 3.3B), PopD has increased expression of *SDC1, CD44, ITGB8, ITGB7, ITGB1, ITGAV, ITGA8, HSPG2, TNXB, RELN, LAMC1, LAMB2, LAMA2,* and *COL6A2* compared to pops A and B. SDC1 (or CD138) is a cell surface marker of these plasma cells. The integrins have been detected in multiple myeloma and have been shown to play a role in cell survival. HSPG2 (or perlecan) plays a role in cell-adhesion. In contrast to this, TNXB has an anti-adhesive effect. RELN has been noted to be upregulated in the presence of CD138 and plays a role in survival and adhesion in multiple myeloma. {Moreaux et al (2009);Jourdan et al (1998);Gokden et al (2006);Bayer-Garner et al (2001);Bayer-Garner et al (2002);Bayer-Garner et al (1999);Bayer-Garner et al (2001);Bayer-Garner et al (2004);Bayer-Garner et al (2004);Bayer-Garner et al

(2001);Aref et al (2003)}.. Finally, the laminins play a role in the chemotaxis of these LLPC cells {Ota et al (2006)}, though they are generally known for their contribution to attachment of cells to extracellular epithelial matrices.

(A)


NOTCH Sig
TGFB Sig
Apic Surf
TNFA Sig
Hypoxia
Estro Resp 2
Estro Resp 1
Myog
KRAS Up
EMT
Apoptosis
UV Resp DN
Prot Sec
Androg Resp
MYC Targ 2
Spermat
IFNA Resp
ROS
PI3K-AKT-MTOR
IL6-JAK-STAT3
IFNG Resp
FA Metab
DNA Rep
MYC Targ 1
MTOR Sig
Mit Spindle
G2M Chkpt
OxPhos
E2F Targets
Inflam Resp
Complement
Allo Rej
PopA
PopB
PopD
Low
High

(B)

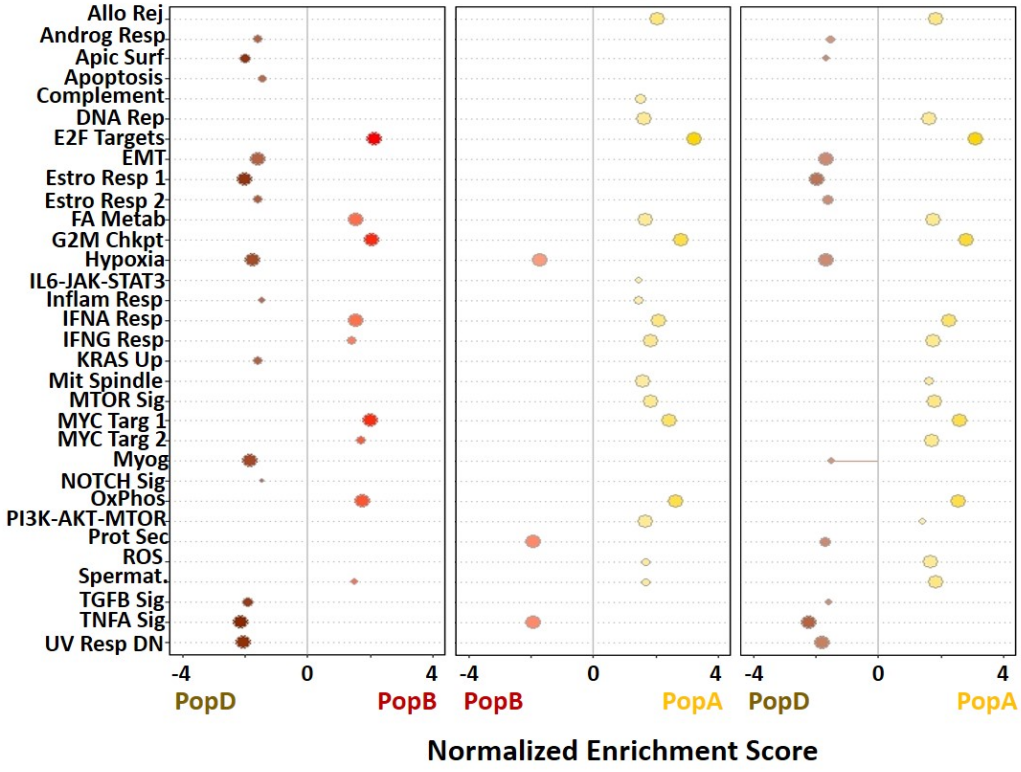

Allo Rej
Androg Resp
Apic Surf
Apoptosis
Complement
DNA Rep
E2F Targets
EMT
Estro Resp 1
Estro Resp 2
FA Metab
G2M Chkpt
Hypoxia
IL6-JAK-STAT3
Inflam Resp
IFNA Resp
IFNG Resp
KRAS Up
Mit Spindle
MTOR Sig
MYC Targ 1
MYC Targ 2
Myog
NOTCH Sig
OxPhos
PI3K-AKT-MTOR
Prot Sec
ROS
Spermat.
TGFB Sig
TNFA Sig
UV Resp DN
-4
0
4
PopD
PopB
PopB
PopA
PopD
PopA
Normalized Enrichment Score

Figure 3.2. Significantly Differentially Regulated Pathways between PopA, PopB, and PopD. (A) Two-way hierarchical cluster of the first principal component (%VE > 45) of 32 pathways (nominal p-value < 0.05, FDR < 0.05) that were significantly differentially regulated between pops A, B, and D based on GSEA. (B) Bubble plot of the normalized enrichment score of each of the 32 pathways (from A) show the normalized enrichment score of each pathway within each of the two-way comparisons of pops A, B, and D. The size of each bubble is directly correlated with the p-value of each pathway. The x-axis shows the normalized enrichment score of the each of two-way comparisons of pops 2, 3, and 5. Bubbles are colored on a spectrum between each of the two populations, trending closer to the color of denoting the population as the enrichment score increases.

Within the Apoptosis pathway (Figure 3.3C), PopD has increased expression of *GADD45B, GADD45A, NFKBIA, FOS, JUN, MAP3K5, BCL2, BIRC3, CTSO, ITPR3, LMNA, MCL1, TUBA3D*, and *CFLAR* compared to pops A and B. Many of these genes overlap with the TNFA (Figure 3.4D) pathway genes, including *CLAR, JUN, BIRC3, FOS, NFKBIA*, and *MAP3K5*. In addition to these, the TNFA profile also includes *VEGFC*, *AKT3, PTGS2, JUNB, TNFAIP3, CRB3L2*, and *CEBPB* which encode inflammatory and survival factors.

Within the IGF1 pathway (Figure 3.3E), PopD has increased expression of *IGF1, PDK1*, and *SMAD3* compared to pops A and B. IGF1 is involved in survival and proliferation through the activation of the PI3K-AKT pathway and MAPK pathways. PDK1 has been shown to play a role in survival in multiple myeloma. SMAD3 has been shown to play a role in survival via interaction with TGFβ. {Klein et al (1999); Klein et al (2007); Jundt et al (2004)}.

Regarding hypoxia (Figure 3.3F), several genes were highly expressed in popD as compared to pops A and B: *IGF1R, INSR, IGF1, IFNG, CDKN1A, BCL2, PDK1, TIMP1, CAMK2D, PRKCA, CYBB*, and *AKT3*. Among these genes are apoptotic genes as well as genes that play a role in the activation of PI3K-AKT, MAPK, and NFκB pathways.

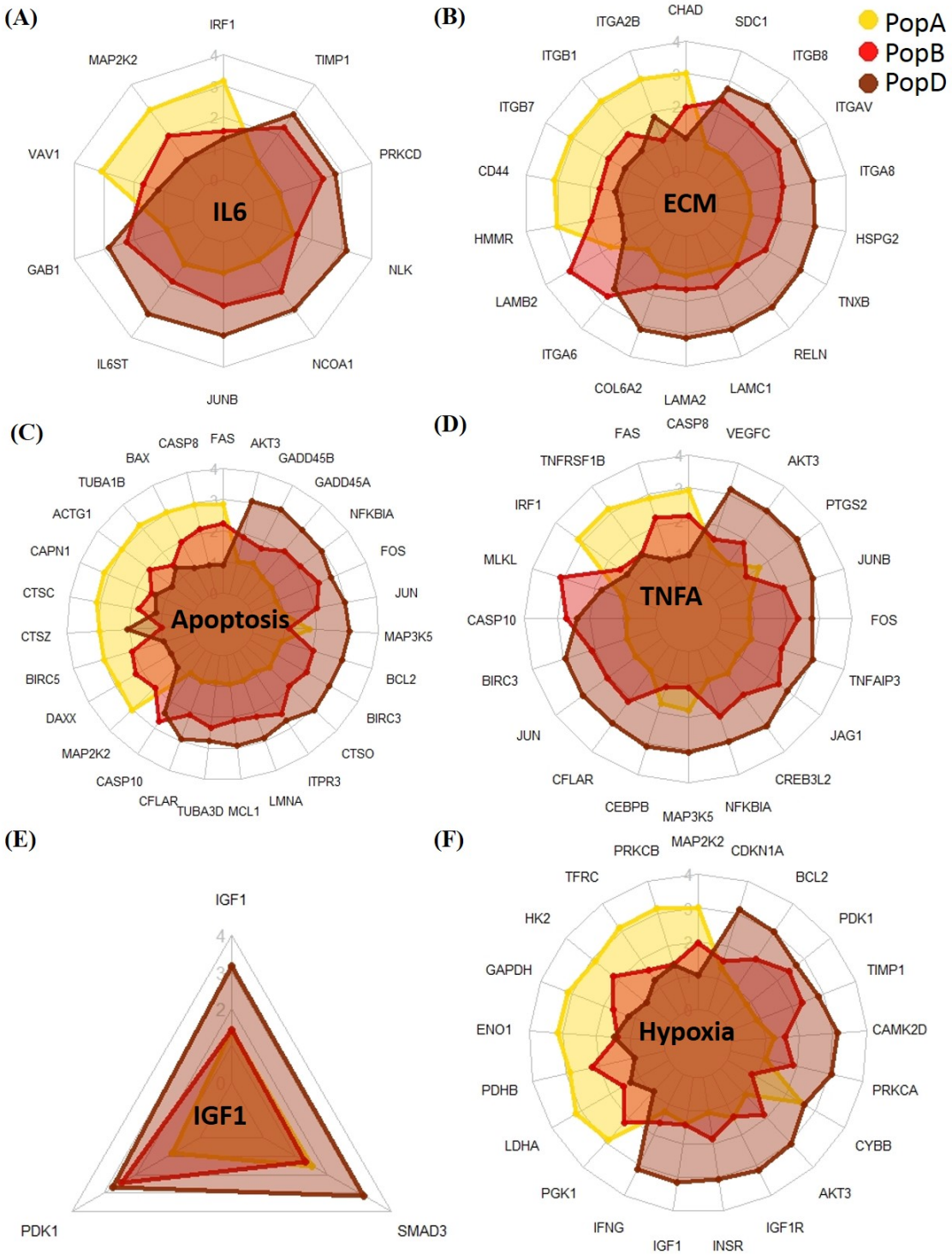

(A)
IL6
IRF1
TIMP1
PRKCD
NLK
NCOA1
JUNB
IL6ST
GAB1
VAV1
MAP2K2
(B)
ECM
CHAD
SDC1
ITGB8
ITGAV
ITGA8
HSPG2
TNXB
RELN
LAMC1
LAMA2
COL6A2
ITGA6
LAMB2
HMMR
CD44
ITGB7
ITGB1
ITGA2B
PopA
PopB
PopD
(C)
Apoptosis
FAS
AKT3
GADD45B
GADD45A
NFKBIA
FOS
JUN
MAP3K5
BCL2
BIRC3
CTSO
ITPR3
LMNA
MCL1
TUBA3D
CFLAR
CASP10
MAP2K2
DAXX
BIRC5
CTSZ
CTSC
CAPN1
ACTG1
TUBA1B
BAX
CASP8
(D)
TNFA
CASP8
VEGFC
AKT3
PTGS2
JUNB
FOS
TNFAIP3
JAG1
CREB3L2
NFKBIA
MAP3K5
CEBPB
CFLAR
JUN
BIRC3
CASP10
MLKL
IRF1
TNFRSF1B
FAS
(E)
IGF1
IGF1
SMAD3
PDK1
(F)
Hypoxia
MAP2K2
CDKN1A
BCL2
PDK1
TIMP1
CAMK2D
PRKCA
CYBB
AKT3
IGF1R
INSR
IGF1
IFNG
PGK1
LDHA
PDHB
ENO1
GAPDH
HK2
TFRC
PRKCB

Figure 3.3. Differential Gene Expression for the IL6, ECM, Apoptosis, TNFA, IGF1, and Hypoxia Pathways for PopA, PopB, and PopD. Radar plots show the standardized least square means for significantly ($p < 0.05$, FDR < 0.05) differentially expressed genes across plasma cell populations A, B, and D for the indicated pathways. PopA is shown in yellow, popB is shown in red, and popD is shown in brown.

3.3.2 Comparison of Bone Marrow Plasma Cells to Peripheral Blood Antibody-Secreting Cells

RNASeq was performed on pop2, pop3, and pop5 flow-sorted preparations from peripheral blood of five donors with one missing pop5 sample; and on popA, popB, and popD flow-sorted preparations from bone marrow of four different donors. After normalization as described in the Methods, analysis of variance was performed to identify 5684 genes that are significantly differentially expressed at a false discovery rate of 5% between any two sub-populations from either of the compartments analyzed together (Figure 3.4A). Clusters of high expression within the two-way hierarchical expression of these DEGs indicate signatures of various cell types. These clusters are delineated by black numbered boxes. Box 1 (2403 genes) contains the set of genes that are uniquely upregulated in the bone marrow populations. Box 2 (1583 genes) contains the set of genes upregulated in the peripheral blood populations. Boxes 3 (167 genes) and 5 (177 genes) show signatures that are shared by pop2, pop3, and the bone marrow populations, to the exclusion of pop5. Box 4 (254 genes) is a signature that is unique to pop2 & pop3.

Principal components analysis confirms the hierarchical clustering implication that the transcriptomes of the bone marrow and peripheral blood cell populations are notably divergent. The first principal component captures 34.04% of the variance and separates bone marrow from peripheral blood sets. The second principal component captures 10.4% of the variance and separates pops2 and 3 from pop5. Since cell proliferation is elevated

in ASCs, we were concerned that expression of proliferation-related genes might be driving the separation between the compartments. However, after removing the signature of such genes, the global difference in gene expression between ASCs, SLPCs, and LLPCs remained present in the resulting data. Given this high level divergence in the transcriptomes, all subsequent analyses consider the two cell compartments separately.

In order to characterize the biochemical functions associated with the sub-population-specific gene expression, gene-set enrichment analysis was performed. {Subramanian et al (2005);Liberzon et al (2011);Liberzon et al (2015)}. For each gene set identified, the first principal component was generated as a score summarizing gene expression in the pathway, though exceptions were discussed above.

The results of this gene set analysis for these six populations are summarized in Figure 3.5A which shows the PC1 of each of these significantly differentially expressed pathways similar to figure 3.2A. To further analyze the enrichment of these gene sets, I recapitulated them in Figure 3.5B, which as before displays the enrichment score on the x-axis with size of the point being directly associated with the negative log p-value of each gene set. Discordance between the two panels is attributable to the fact that the PC1 scores are driven by a subset of genes in each pathway that are more closely co-regulated between compartments than among populations within a compartment. There are four such pathways including Apoptosis, Complement, IL6-JAK-STAT3, and Inflammatory Response. As previously discussed, the genes driving PC1 and the genes that contribute to the normalized enrichment score (shown in greater detail in Appendix Figure 5) present a discrepancy for these pathways. In such cases of discrepancy, normalized enrichment score is used for interpretation of the enrichment of these pathways.

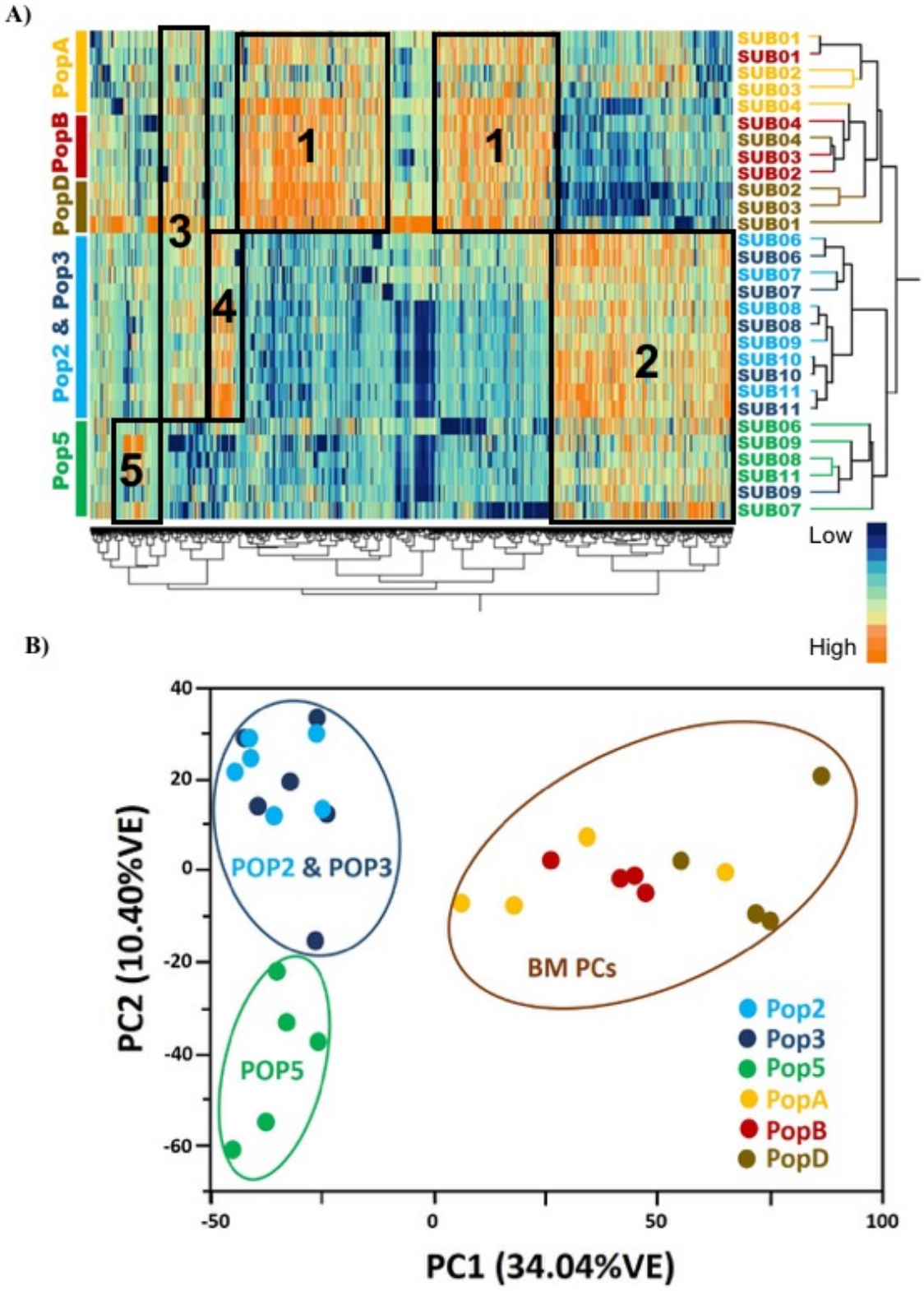

A)
PopA
PopB
PopD
Pop2 & Pop3
Pop5
1
1
2
3
4
5
SUB01
SUB01
SUB02
SUB03
SUB04
SUB04
SUB04
SUB03
SUB02
SUB02
SUB03
SUB01
SUB06
SUB06
SUB07
SUB07
SUB08
SUB08
SUB09
SUB10
SUB10
SUB11
SUB11
SUB06
SUB09
SUB08
SUB11
SUB09
SUB07
Low
High
B)
PC2 (10.40%VE)
PC1 (34.04%VE)
POP2 & POP3
POP5
BM PCs
Pop2
Pop3
Pop5
PopA
PopB
PopD
40
20
0
-20
-40
-60
-50
0
50
100

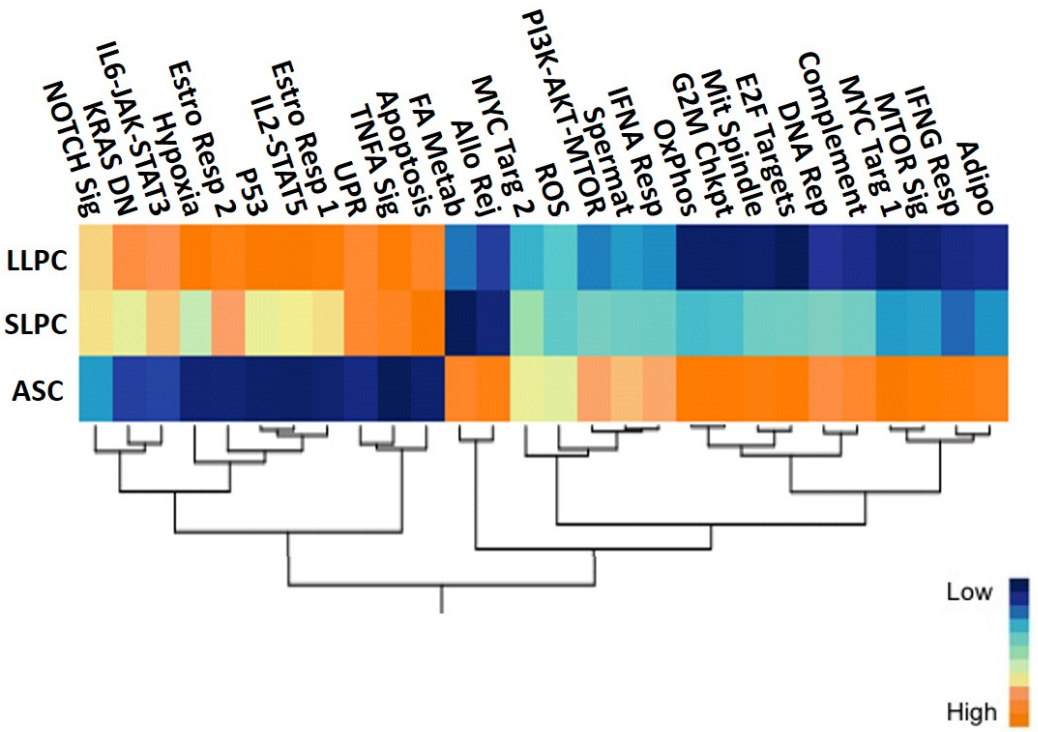
Figure 3.4. Analysis of Differentially Expressed Genes between ASC, SLPC, and LLPC. (A) Two-way hierarchical clustering of 5684 genes that are significantly (FDR < 0.05, p < 0.05) differentially expressed genes between any two of the six populations of antibody secreting cells or plasma cells populations (pop2, pop3, pop5, popA, popB, and popD). The samples cluster within population. Box 1 encompasses genes that are upregulated in bone marrow plasma cells (pops A, B, and D); Box 2 genes that are upregulated in antibody secreting cells (pops 2, 3, and 5); Box 3 genes that are upregulated cross-tissue (pop2, pop3, popA, popB, and popD); Box 4 genes that are upregulated in antibody-secreting cell populations, pops 2 and 3; Box 5 genes that are upregulated in antibody-secreting cell population, pop5. (B) Principal component analysis of these differentially expressed genes shows that the first principal component captures 34.04% of variation. PC1 separates the bone marrow populations (pops A, B, and D) from antibody secreting cell populations (pops 2, 3, and 5). PC 2 captures 10.4% of the variance and separates pops 2 and 3 from pop5.

The peripheral blood Pops 2, 3, and 5 (all Antibody Secreting Cells, ASCs) show coordinated enrichment of proliferation and cell cycle gene sets as compared to the bone marrow compartment SLPCs (pops A and B) and LLPC (popD). These proliferation gene sets include MYC Targets, E2F Targets, G2M Checkpoint, and DNA Repair. Enrichment in ASCs as compared to SLPC and LLPC is also observed for the metabolism gene sets, oxidative phosphorylation, and MTOR signaling. ASCs are also enriched for some inflammatory pathways (IFNG response and allograft rejection).

In contrast, SLPC and ASC are not significantly differentially regulated for specific metabolism gene sets (adipogenesis and fatty acid metabolism), cell cycle gene sets (DNA repair and mitotic spindle), inflammation pathways (complement, IL6-JAK-STAT3, IFNA response, reactive oxygen species), MTOR signaling (PI3K-AKT-MTOR), as well as apoptosis, NOTCH signaling, and late estrogen response.

Gene sets that are enriched in LLPC as compared to both SLPC and ASC include early estrogen response and TNFA signaling. Conversely, SLPCs are enriched for early estrogen response, hypoxia, P53, and TNFA signaling as compared to ASCs.

(A)


NOTCH Sig
KRAS DN
IL6-JAK-STAT3
Hypoxia
Estro Resp 2
P53
IL2-STAT5
Estro Resp 1
UPR
TNFA Sig
Apoptosis
FA Metab
Allo Rej
MYC Targ 2
ROS
PI3K-AKT-MTOR
Spermat
IFNA Resp
OxPhos
G2M Chkpt
Mit Spindle
E2F Targets
DNA Rep
Complement
MYC Targ 1
MTOR Sig
IFNG Resp
Adipo
LLPC
SLPC
ASC
Low
High

(B)

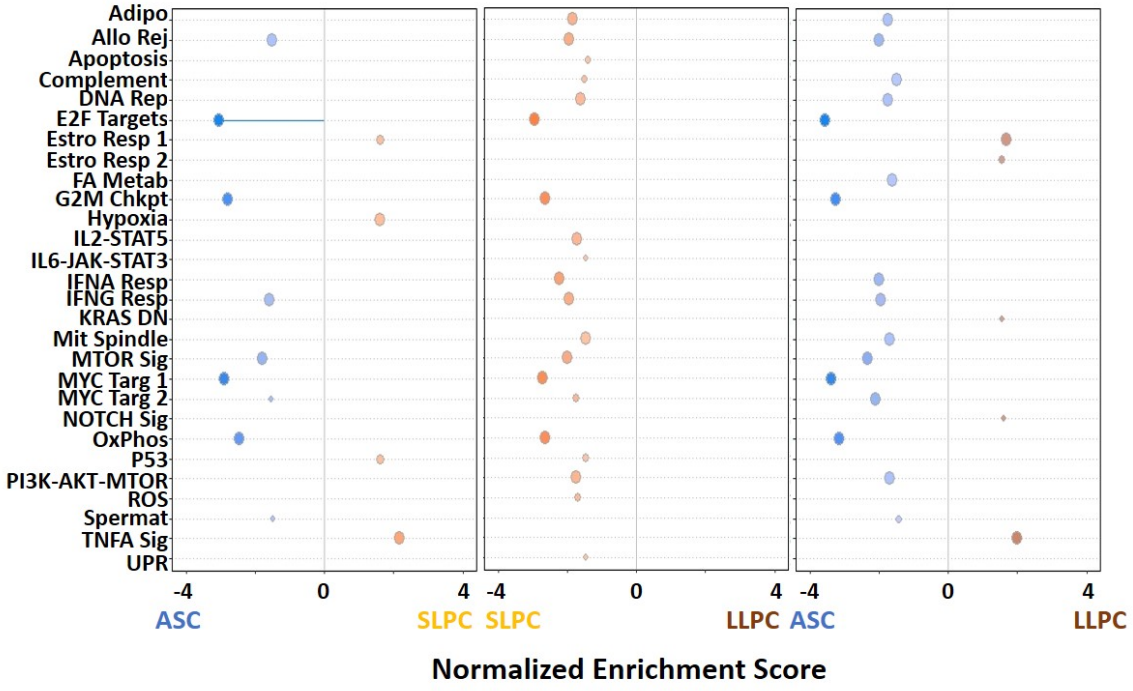

Adipo
Allo Rej
Apoptosis
Complement
DNA Rep
E2F Targets
Estro Resp 1
Estro Resp 2
FA Metab
G2M Chkpt
Hypoxia
IL2-STAT5
IL6-JAK-STAT3
IFNA Resp
IFNG Resp
KRAS DN
Mit Spindle
MTOR Sig
MYC Targ 1
MYC Targ 2
NOTCH Sig
OxPhos
P53
PI3K-AKT-MTOR
ROS
Spermat
TNFA Sig
UPR
-4
0
4
ASC
SLPC
SLPC
LLPC
ASC
LLPC
Normalized Enrichment Score

Figure 3.5. Significantly Differentially Regulated Pathways between ASC, SLPC, and LLPC. (A) Two way hierarchical cluster of the first principal component (%VE > 45) of 28 pathways (nominal p-value < 0.05, FDR < 0.05) that were significantly differentially regulated between ASCs, SLPCs, and LLPCs based on GSEA. (B) Bubble plot of the normalized enrichment score of each of the 28 pathways (from A) show the normalized enrichment score of each pathway within each of the two-way comparisons of ASCs, SLPCs, and LLPCs. The size of each bubble is directly correlated with the p-value of each pathway. The x-axis shows the normalized enrichment score of the each of two-way comparisons of ASCs, SLPCs, and LLPCs. The bubbles are colored on a spectrum between each of the two populations, trending closer to the color denoting the population as the enrichment score increases.

Of these 28 pathways, I have chosen to pursue in depth analysis of six pathways due to their purported roles in ASC, SLPC, or LLPC development and survival: IL6, Apoptosis, IL2, ECM, TNFα Signaling, IGF1, and OxPhos. Klein 2003 implicated IL6 (Fig 3.6A), IL2 (Fig 3.6C), and IGF1 (Fig 3.6E) pathways as major contributors to plasma differentiation and survival. The ECM pathway (Fig 3.6B) provides greater insight into the proteins that are interacting with the extracellular membrane, including SDC1 (or CD138). The TNFA Signaling pathway (Fig 3.6D) has been shown to be key in upregulation of other plasma cell pathways including MAPK, NFKB, and PI3K-AKT. Finally, oxidative phosphorylation is an energy metabolism pathway that is differentially regulated across all pairwise comparisons of ASCs, SLPCs, and LLPCs. {Klein et al (2003)}.

The IL6 pathway (Fig 3.6A) shows that *IL6R* is slightly upregulated for ASC as are *JAK*, which is involved in plasma cell differentiation programs, and *VAV1*, a regulator of Blimp-1. SLPC and LLPC strongly upregulate *JUNB*, *NCOA1*, *SHC1*, *PRKCD*, *TIMP1*, and *BCL2L1*, many of which are expected from bone marrow plasma cell population analysis.

Figure 3.6. Differential Gene Expressed for IL6, ECM, IL2, TNFA, IGF1, and Oxidative Phosphorylation Pathways for LLPC, SLPC, and ASC. Radar plots of selected pathways show the standardized least square means for the significantly ($p < 0.05$, FDR < 0.05) differentially expressed genes across ASCs (pops 2, 3, and 5), SLPCs (pops A and B), and LLPCs (popD). LLPC is shown in maroon, SLPC in yellow-orange, and ASC in light blue.

In the ECM pathway (Fig 3.6B), ASCs have the highest expression for several integrins (*ITGB7* and *ITGB1*), *COL4A4*, and *CD44*. The remaining DEGs are more highly expressed in the bone marrow plasma cells (*COL1A2*, *ITGA8*, *ITGA3*, *SDC1*, *ITGA6*, *LAMA5*, *COL9A3*, *AGRN*, and *TNR*.

The IL2-STAT5 (Fig 3.6C) pathway is essential to the commitment of the B-cell differentiation program. Within this pathway, ASCs most highly express *SYK*, *CBL*, *IL2RB*, *NMI*, and *CCND2*. SLPC highly express *NMI* and *CCND2*, as well. LLPC especially upregulate *CISH*. SLPC and LLPC upregulate *JUN*, *BCL2*, *FOS*, *MYC*, *SHC1*, and *STAT5A*.

The IGF1 pathway shows ASCs as having the highest expression of *TNF* and *PTEN*. LLPC has especially high expression of *PLD1*. The remaining genes are upregulated for both LLPC and SLPC: *FOXO3*, *PDK1*, *PPARGC1A*, *FBXO32*, *ACVR2B*, and *IGF1R*.

In the TNFA pathway (Fig 3.6D), *MAP2K4*, *MAP3K8*, and *TRADD* are most highly expressed in ASCs. *JAG1*, *DAB2P*, *PTGS2*, and *LTA* are most highly expressed in popD. The remaining DEGs are most highly expressed in both SLPC and LLPC.

Finally, oxidative phosphorylation (Fig 12E) shows highest expression for ASC as compared to the other two subsets, with the exception of *GMPPB,* suggesting that oxidative phosphorylation is uniquely upregulated in ASC as compared to SLPC and LLPC.

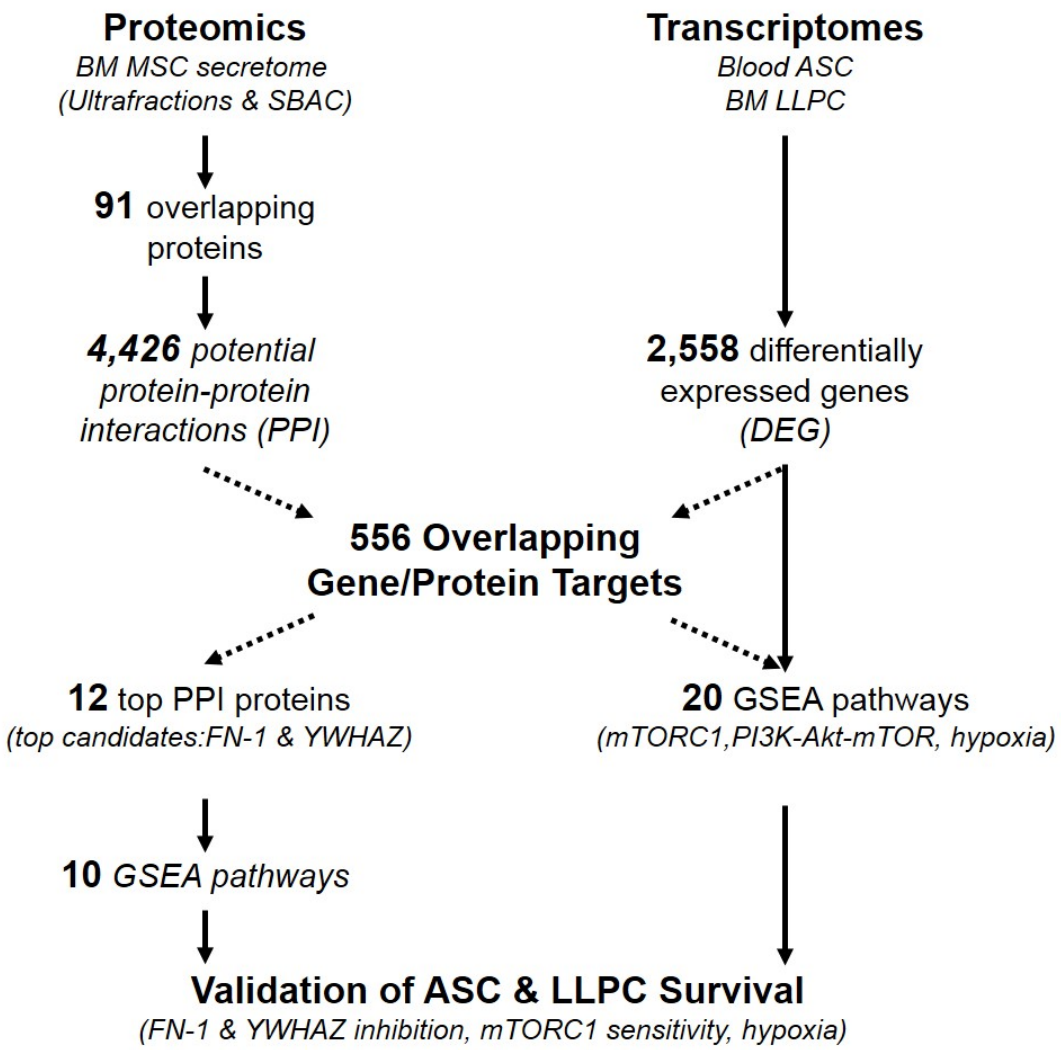


Figure 3.7. Proteomic-Transcriptomic Analysis of the Role of the Bone Marrow Microniche on Long-Lived Plasma Cells. Using LC-MS/MS, 91 proteins were identified to be present in the MSC secretome. All 4,426 potential protein interactors of these 91 proteins were compared against the 2,558 genes that were significantly ($p < 0.05$, FDR < 0.05) differentially expressed between long-lived plasma cells and peripheral blood ASCs. The result was a list of 556 potential protein interactors that were significantly differentially expressed between LLPCs and ASCs. Gene set enrichment analysis of these 556 genes produced 20 differentially regulated pathways.

3.3.3 Proteomic identification of likely regulators of LLPC survival

Mesenchymal stromal cells (MSCs) are known to play a key role in LLPC survival. Nguyen et al (2018) identified novel MSC plasma plasma cell survival factors using LC-MS/MS in the MSC secretome. I utilized a bioinformatics approach to integrate this proteomic analysis of the MSC proteins with the ASC and LLPC transcriptomic analysis (Fig 3.7). Proteomic analysis of the MSC secretome produced 91 proteins. Using HIPPIE, which contains a database of potential protein-protein interactions of ≥ 0.6 score, these 91 proteins were determined to have 4,426 potential protein-protein interactions (PPI) using HIPPIE. {Alanis-Lobato et al (2017)}.

To determine which of these PPI played essential roles in the LLPC survival process, differential gene expression analysis was performed on LLPC samples (cells that had been exposed to the MSC secretome) as compared to ASC samples (cells that were naïve to the MSC secretome).

A total of 2,558 genes were found to be significantly differentially expressed ($p < 0.05$, FDR < 0.05) between blood ASC and BM LLPC. The overlap between these DEGs and 4426 PPI from the proteomic analysis identified 556 PPI that were significantly differentially expressed between LLPC and ASC. Gene set enrichment analysis revealed 20 statistically significant Hallmark pathways (Fig 3.7). Highlighted pathways included mTORC1 signaling, PI3K-Akt-mTOR signaling, TNFA signaling, and hypoxia. Furthermore, twelve of the 91 proteins had more than 45 PPI that overlapped with the ASC-LLPC DEGs. These top hits included fibronectin (FN-1), YWHAZ (also known as 14-3-3 zeta/delta), heat shock proteins, endolase, glyceraldehyde 3-phosphate

dehydrogenase (GAPDH), eukaryotic elongation factor (EEF2), and various cytoskeletal proteins (actin, tubulin, and vimentin).

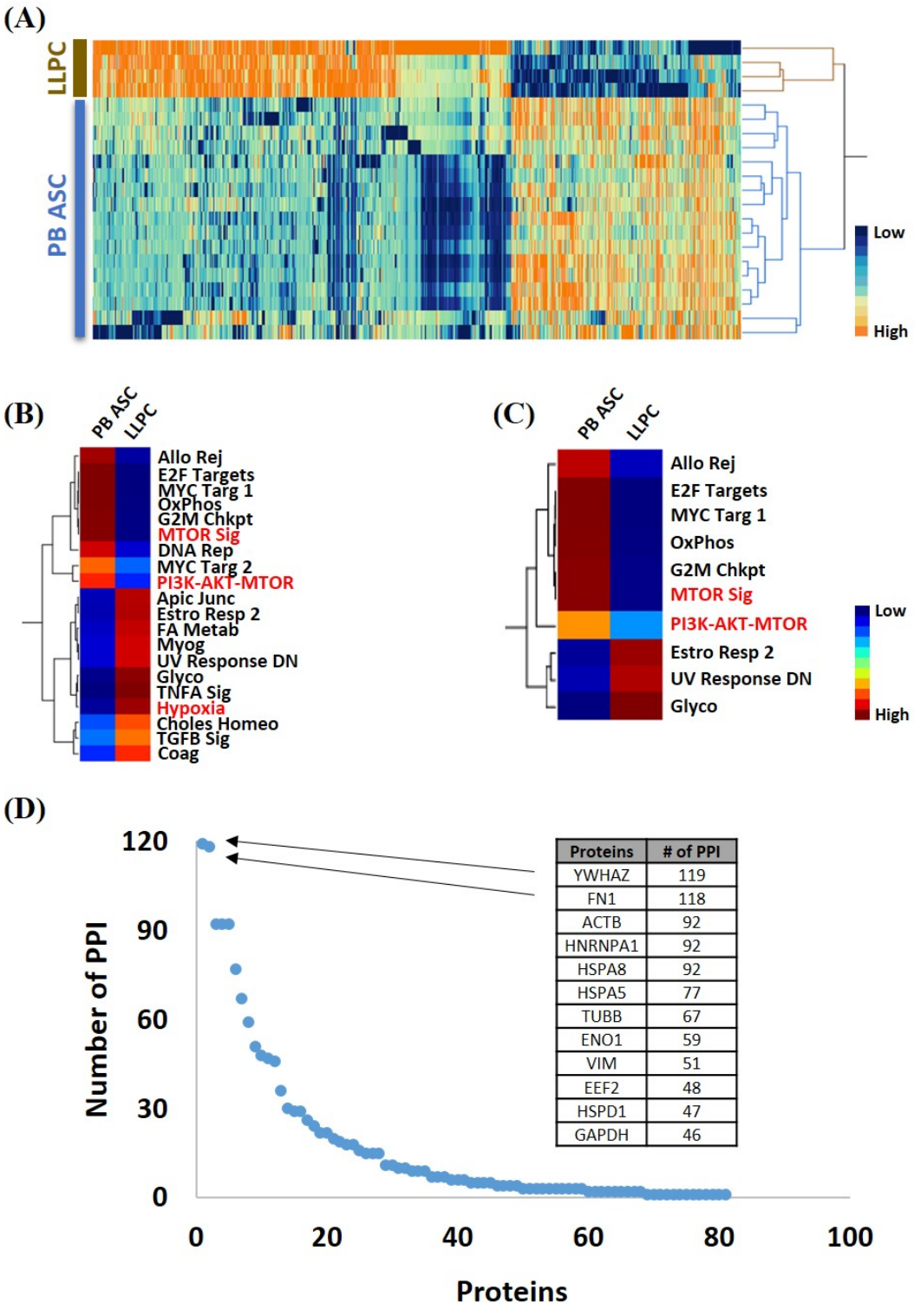

(A)
LLPC
PB ASC
Low
High
(B)
PB ASC
LLPC
Allo Rej
E2F Targets
MYC Targ 1
OxPhos
G2M Chkpt
MTOR Sig
DNA Rep
MYC Targ 2
PI3K-AKT-MTOR
Apic Junc
Estro Resp 2
FA Metab
Myog
UV Response DN
Glyco
TNFA Sig
Hypoxia
Choles Homeo
TGFB Sig
Coag
(C)
PB ASC
LLPC
Allo Rej
E2F Targets
MYC Targ 1
OxPhos
G2M Chkpt
MTOR Sig
PI3K-AKT-MTOR
Estro Resp 2
UV Response DN
Glyco
Low
High
(D)
Number of PPI
120
90
60
30
0
0
20
40
60
80
100
Proteins
Proteins # of PPI
YWHAZ 119
FN1 118
ACTB 92
HNRNPA1 92
HSPA8 92
HSPA5 77
TUBB 67
ENO1 59
VIM 51
EEF2 48
HSPD1 47
GAPDH 46

Figure 3.8 Integrated bioinformatics of MSC secretome proteomics with transcriptomics of blood ASC and BM LLPC. (A) A heat map of the 2,558 DEG between 17 blood ASC (PB ASC) obtained from 7 healthy subjects and BM LLPC (LLPC) obtained from 4 adult subjects. **(B)** HIPPIE analysis of 91 MSC protein revealed 4,429 potential protein partners (PPI). Overlap of the 4,429 PPI with the 2,558 DEG uncovered 556 overlapping gene/protein targets and led to 20 statistically significant GSEA hallmark pathways. **(C)** Of the aforementioned 91 MSC secretome proteins, FN-1 and YWHAZ had the highest number (118 and 119, respectively) of potential interacting partners. Of these potential partners, 31 were shared between both FN-1 and YWHAZ. **(D)** From the 20 GSEA pathways, FN-1 and YWHAZ were found to be involved in these 10 potential GSEA hallmark pathways.

The top two proteins, FN-1 and YWHAZ, had 436 and 430 PPI, respectively, of which 118 and 119 overlapped with the DEG in blood ASC and LLPC (FDR < 0.05; p-value <0.05). FN-1 and YWHAZ were only involved in 10 of the 20 GSEA hallmark pathways. These included proliferation signatures i.e. E2F targets and G2M checkpoints, which is consistent with evidence that most blood ASC have undergone recent proliferation in contrast to the BM LLPC . However, the downregulation of Myc targets, mTORC1 signaling, and PI3K-Akt-MTOR signaling suggested novel pathways that the MSC secretome may activate for maintaining ASC survival. While YWHAZ and FN-1 have been experimentally as essential for ASC and LLPC survival, whether the other 89 proteins have a direct role in LLPC maturation will require further investigation.{Nguyen and Lewis et al (2018);Nguyen et al (2018);Halliley et al (2015)}. Further experiments have shown a dose-dependent effect of GAPDH, HSPD1, and hnRNPA1 when comparing ASCs to PCs, further validating several of these proteins.

3.4 Discussion

My transcriptomic profiling characterizes differential expression of thousands of genes in various gene sets of development and differentiation across two compartments

and six cell types. Cell-sorted whole RNASeq was insufficient to resolve the full functional consequences of these transcript abundance differences for the development of long-lived plasma cells. However, I discuss some aspects of the processes relevant to each cell type and compartment before turning to inferences as to relationships between these cell types in the context of order of maturation within and across compartment.

3.4.1 Bone Marrow Plasma Cells: PopA, PopB, and PopD

This Chapter started with a detailed analysis of pops A, B, and D independently in the bone marrow. Many of the gene sets previously discussed in the peripheral blood compartment appear again in this discussion, including proliferation, inflammation, MTOR, androgen-estrogen response, cell mobility, tumor suppression, metabolism, stress response, and survival (Figure 3.2B).

Figure 1.2 provided a model for the relationship between pops A, B, and D. In addition to the possibility that they arise independently from each other (Figure 1.2, Models 1 and 4), the differential gene expression clustering in Fig 3.2A suggests that popB could be an intermediate population between popA and popD. This could be explained in one of two ways: 1) popA is the precursor population and transitions into popB which in turn becomes popD (Fig 1.2, Model 2), or 2) popB is the precursor population and from it, either popA or popD arise. Bayer Garner 2003 suggested the possibility of three bone marrow plasma cell populations that they studied in the murine model, which they differentiated based on the strength of CD138 association (strongly (popD), weakly (popB), or absent (popA)) wherein the weakly expressed CD138+ cell was the precursor population and either more strongly expressed or shed CD138. Unfortunately, transcriptomic analysis

alone cannot conclusively prove which, if any, of these models is the true model of development of LLPC. Here, I will discuss further the findings of the transcriptomic analysis of these three cell types, followed by inference of their relationship with one another. {Bayer-Garner et al (2001)}

Fig 3.2 presents popA as having enrichment of inflammation gene sets, cell cycle and proliferation, energy metabolism gene sets (including oxidative phosphorylation), and MTOR pathways. In contrast, popD was enriched in cell-cell adhesion, signaling, and survival pathways. PopB, as could be surmised from the DEG (Fig 3.5), did not uniquely upregulate any pathway in comparison with both popA and popD. The pathways upregulated in popA are more similar to the signatures found in the ASC (Chapter 2), especially in relation to oxidative phosphorylation and cell cycle and proliferation. These pathways are not relevant for the hypoxic environment of the terminally differentiated long-lived phenotype. This suggests the possibility that differential expression of these pathways is an artefact of recent transitioning of ASC into BMPC. This lends itself to the model that popA is the precursor population that develops into popB and then popD (Figure 1.2, Model 2). {Tellier et al (2018);Tellier et al (2017);Nutt et al (2015);Klein et al (2003)}

These pathways of cell-cell adhesion and survival being enriched in popD are recapitulated in the individual pathway analysis (Figure 3.6). Across all 6 pathways considered in depth, genes associated with survival and adhesion are mainly highly expressed by popD as compared to popA and popB.

3.4.2 Comparison of Bone Marrow Plasma Cells to Peripheral Blood Antibody-Secreting Cells

Initially, BM and PB sets were analyzed together (Figure 3.4) through clustering of all significantly differentially expressed genes (DEGs) across all six populations (Figure 3.4A). The gene expression separated the two compartments with pops A, B, and D clustering together relative to the marrow compartment (Fig 3.4, Box 1), and pops 2, 3, and 5 clustering together representing the distinct blood compartment (Fig 3.4, Box 2). Consequently, I analyzed peripheral blood and bone marrow subsets separately by considering all three PB populations as one set of cells (ASCs: pop2, pop3, and pop5), short-lived plasma cells as a second set (SLPCs: popA and popB), and long-lived plasma cells (LLPC: popD).

As expected from the peripheral blood analysis, proliferation gene sets were enriched in ASCs as compared to LLPCs and SLPCs. Some cell cycle gene sets were enriched in both SLPC and ASC consistent with the model that ASCs are precursors to SLPCs. Several enrichment pathways were similarly enriched in both ASC and SLPC as well as MTOR pathways. Nguyen et al (2018) showed through experimental manipulations that MTOR is essential for SLPC survival, though not for ASC or LLPC survival. One potential explanation for the upregulation of MTOR in ASC, as compared to SLPC and LLPC, could be that its upregulation is in preparation for entrance into the marrow space, establishing a need for MTOR at the SLPC stage. {Nguyen et al (2018)}.

Of the significantly differentially regulated gene sets between ASC, SLPC, and LLPC, the estrogen response gene sets were unexpected. However, Fortini et al [2017]

showed that the upregulation of estrogen provides a protective function for cells against TNFα-induced apoptosis by activating Notch signaling, which is involved in cell-cell interactions and cell-fate decisions. {Fortini et al (2017);Jundt et al (2004)}. Estrogen has also been noted to induce B-cell hyperactivity. {Ahmed et al (1999)}. In fact, the shift in cholesterol conversion from androgen to estrogen has been known to upregulate B-cell development.{Olsen and Kovacs (2001)}. Hence, the enrichment of androgen gene sets may indicate suppression of B-cell development. Whether or not differences in androgen and estrogen levels in men and women influence gender-biases in vaccine efficacy by virtue of the documented differences in sex-hormone response gene expression remains to be evaluated.

The individual pathway analysis of ASC, SLPC, and LLPC (Figure 3.6) focused on six specific pathways. Amongst these, the IL2-STAT5 signaling gene set has been noted to commit B-cells to the PC fate. {Tellier et al (2018);Le Gallou et al (2012);Klein et al (2003);Hipp et al (2017)}. The remaining gene sets have more complicated roles in the development of PCs. TNFα signaling plays a dual role in the context of PCs. Depending on its downstream targets (either the caspase cascade or NFκB signaling), it either induces apoptosis or proliferation. {Jourdan et al (1999)}. As previously discussed, estrogen response gene sets upregulated in PCs might modulate the apoptotic effects of TNFα such that NFκB-promoted "proliferation" allows for further development into the long-lived phenotype. Additionally TNFα, as well as IFNα signaling, has been a key player in the shift from generation of LLPCs to SLPCs and the mobilization of PCs from the marrow during infection.{Slocombe et al (2013);Mathian et al (2011)}. The IL6-JAK-STAT3 gene set, in contrast, plays a role in survival and differentiation of plasma cells. {Liu et al (2012);J

ourdan et al (1991); Jourdan et al (2003); Jourdan et al (2005); Cokic et al (2015)}. The P53 survival gene set also modulate development and survival of plasma cells in the context of tumor suppression. {Shi et al (2017); Ma et al (2014); Vicente-Duenas et al (2012); Ackermann et al (1998); Wang et al (1995)}.

These gene sets of development, differentiation, and survival produce a stress response in PCs that is also reflected in the enrichment of the Unfolded Protein Response, and UV Response. These gene sets are known to be crucial for significant changes in demands for antibody production and the changes in cell structure to compensate for these changes. {Ma et al (2014)}. However, it is important to note that the differential expression at the level of individual genes implies a much more nuanced picture than suggested by considering the pathway as a whole. Experimental manipulation following approaches described n Nguyen, Garimalla et al (2018) is critical to validate specific inferences.{Tellier et al (2018);Tellier et al (2017);Nutt et al (2015);Nguyen et al (2018);Klein et al (2003);Good et al (2009); Jourdan et al (2011)}

In conclusion, while the exact mechanisms in which these and other pathways engage to promote survival and development into the long-lived phenotype is unresolved, what is clear is that all of these pathways play a significant role in all three subsets of plasma cells (ASCs, SLPCs, and LLPCs); and it is possible that different portions of these pathways are upregulated and downregulated through different stages of this development process. Furthermore, current gene set enrichment analysis is inadequate to fully determine the extent of interaction within and across pathways and genes in the development of novel cell types, however it can serve to suggest a direction of more in-depth analysis and focus within each functional group of pathways and gene sets.

3.4.3 Four Models of Development of Long-Lived Plasma Cells

We have recently demonstrated that rather than being a homogeneous population of cells, plasma cells actually consist of multiple subpopulations based on key cell surface markers CD138, CD38, and CD19, both in bone marrow and peripheral blood. {Nguyen and Lewis et al (2018);Nguyen et al (2018);Halliley et al (2015)}. Various models for the ontological relationships between the six predominant PC populations, three in each compartment, can be considered, as schematized in Figure 1.2. Each model assumes that PC first arise in germinal centers, and that pops A and B are short lived, undergoing high rates of apoptosis if they do not further differentiate. The first is that peripheral blood plasma cells are precursors to bone marrow plasma cells of the same cell surface marker pattern (Model 1). In this case, each PB ASC would arise independently from a germinal center precursor and differentiate into one of the bone marrow plasma cells, retaining their matching cell surface markers. Second, pop2 could be the precursor of all bone marrow plasma cells (Model 2). In this model, some unidentified precursor from the germinal center gives rise to pops 2, 3, and 5. Then pop2 enters the marrow as popA; popA then differentiates popB or undergoes apoptosis. PopB then differentiates into popD or apoptoses. Subsequently, popD, migrates further into the hypoxic space of the bone marrow as an LLPC. A third possibility is that pop3 is the precursor of all bone marrow plasma cells (Model 3). That is, some unidentified precursor from the germinal center gives rise to pops 2, 3, and 5 and it is pop3 that enters the marrow as popB; popB then differentiates either into popA (which eventually apoptosis) or differentiates into popD upon migration further into the hypoxic space of the bone marrow as a LLPC. In the fourth

model, each cell type arises independently in both the peripheral blood and bone marrow compartments from a common precursor.

Model 1 is unlikely due to the lack of similarity between pops 5 and D. Since pops 2 and 3 show such high similarity that they are paired within individual, whereas their bone marrow counterparts, pops A and B, are quite different from each other, it is difficult to distinguish between Models 2 and 3 which postulate that either pop2 or pop3 respectively are the precursors to the bone marrow cells. Model 4 is that all 6 cell types derive independently, such that the CD marker similarities between the matching populations represent convergent gene expression, rather than retention from peripheral blood precursors.

Models 2 & 3 encapsulate the concept of pop 2 and 3 being closely linked and also as the BM precursors. Model 2 suggests that popA differentiates into popB. This differentiation would result from significantly altered expression of cell cycle, inflammation, and stress gene sets. A failure or inability to proceed with this transition may result in apoptosis. Subsequently, popB would differentiates into popD, accompanied by significantly altered expression of hypoxia, androgen and estrogen response, and proliferation related gene sets. This model is favored by the fact that there are no gene sets that are upregulated in B alone. Furthermore, numerous gene sets that are high in popA are also characteristic of pops2 and 3.

Model 3 suggests that at the popB stage of the differentiation, a decision on cell fate is made and the plasma cell either differentiates into popA, which significantly altered expression of cell cycle, inflammation, and stress gene sets; or popB differentiates into

popD, which causes significantly altered expression of hypoxia, androgen and estrogen response, and proliferation related gene sets, leading to the long-lived phenotype. This model is supported by the observation that pop3 begins to upregulate BMPC relevant pathways such as hypoxia and TNFA signaling (Chapter 2) which suggests that it is the BM PC precursor population.

3.5 Conclusion

While it is clear that we cannot explain the origin of the BM populations without postulating independent gain and loss of expression of genes in multiple gene sets on either the precursor or transitional models, a mixture of the two is possible as well. Resolution of these relationships is important because it will allow for greater insight into the developmental ontogeny of LLPC which could then, ideally, be manipulated for autoimmune preventative therapeutics or increased vaccine efficacy, amongst other uses. With currently available experimental methods, further understanding of the ontology and function of the plasma cell types will be reached by extensive experimental validation of in-depth bioinformatic analysis of each individual gene set and its relevant interactors.

CHAPTER 4

SINGLE CELL RNASEQ ANALYSIS OF LUPUS AND HEALTHY NAÏVE B-CELLS AND POPD PLASMA CELLS

4.1 Introduction

This study has largely been based on the premise that neither bone marrow plasma cells nor peripheral blood antibody secreting cells should be considered to be a homogenous population; that is, by sorting these cells into subtypes, we are able to better determine development and survival mechanisms for the long-lived plasma cell and its potential precursors. Where the line should be drawn for the subsetting of cell-types is a question under ongoing contention, especially with the advent and perpetual improvement of single-cell RNASeq (scRNASeq) technology. Traditionally, fluorescence-activated cell-sorting (FACS) based on diagnostic cell-surface markers has been used to determine subpopulations *a priori*. Now, scRNASeq and algorithms of analysis of these new datasets allow the possibility of using gene expression of the cells themselves to determine how many subtypes are present in a given sample, and how they differ.

Here, I ask the question if there are subpopulations of LLPCs: does the development of LLPC happen on a continuum or is it a binary switch? To this end, we were able to procure bone marrow samples for two individuals, one healthy and one with systemic lupus erythematosus (hereafter, 'lupus'). Within each of the samples, naïve and popD cells were cell-sorted and sequenced using single-cell SmartSeq2. In this chapter I discuss the analysis.

4.1.1 Systemic Lupus Erythematosus (SLE)

Lupus is a plasma cell dyscrasia that predominantly affects women, especially of African descent. Both genetics and environment play a role in SLE. Lupus is caused by the production of autoantibodies. Oftentimes, these antibodies can bind to each other and form immune complexes which transverse through the bloodstream to various areas of the body and lodge in various tissues causing, at times, severe inflammation. {Tipton et al (2018); Ching et al (2012); Banchereau et al (2016)}

As previously mentioned, there is a predominance of lupus in women. Sex hormones have been shown to play a role in lupus, wherein women who were given estrogen replacement therapy were 1.3x more likely to have an SLE flare-up as compared to women who did not. {Buyon et al (2005);Askanase et al (2004);Askanase et al (2012)}.

While there have been significant studies and gains in understanding of the disease over the last decade that have led to improved patient care, a granular understanding of mechanisms behind the production of these aberrant plasma cells remains elusive. Through single-cell RNASeq analysis, determination of subpopulations of cell types including naïve B-cells and popD cells could provide some insight into the molecular mechanisms behind this disorder.

4.1.2 Naïve B-cells

I was provided with data for two cell types: naïve and popD. Until now, the focus of this thesis has been primarily on plasma cells rather than the naïve B-cell precursor.

All B-cells arise from a multipotent precursor in the bone marrow. This precursor then differentiates into a common lymphoid progenitor. From here, the B-cell develops through many stages of development from early pro-B-cell to immature B-cell with rearranged heavy and light chains and IgM expression. These cells then leave the bone marrow for the circulation as transitional B-cells and eventually become mature B-cells. These naive B-cells eventually encounter antigen and become activated and differentiate into plasma cells and memory cells. During the naïve stage, however, they are not producing the copious amounts of antibody that their later counterparts do. {Askanase et al (2012);Bordon et al (2015); Liu et al (2011); Taylor et al (2015);Tipton et al (2015);Tipton et al (2018);Tsokos et al (2016)}

While the role of plasma cells is established in lupus, only recently has the role of naïve autoreactive B-cells been considered. Tipton et al 2015 determined that significant numbers of ASCs in lupus patients arise from recently activated naïve cells and are an important contributor to the disease. {Tipton et al (2015)}

4.2 Materials and Methods

4.2.1 Short Read Mapping and Gene Expression Quantification

Bone marrow aspirate was collected from one healthy patient and one lupus patient. From each of these, 36 Naïve B-cells and 246 popD cells were sorted and collected. They were sequenced using the SmartSeq2 protocol {Picelli et al (2014)} by Dr Stephen Bossinger at the Yerkes genome core. One million single end 50bp reads were mapped to the reference human genome hg38, using STAR {Dobin et al (2013)} to generate bam files. HTseq {Anders et al (2015)} was then used to assess read counts at the gene level. The mapped

libraries were then normalized accounting for differences in size and dispersion using the default parameters of EdgeR for gene abundance levels as TMM {Robinson et al (2010);Nikolayeva et al (2014)}. Subsequent analysis was performed using Seurat v2.0 {Butler et al (2018)}, including normalization and differential expression analysis.

Gene set enrichment analysis was conducted using the Broad Institute's GSEA pre-ranked program as in Chapters 2 and 3. The average difference between conditions was used as the rank for this analysis and it was run against the hallmark pathways of GSEA.

4.2.2 Seurat

In this study, I use Seurat v. 2.0 as an R package to analyse the SmartSeq2 single-cell data. Because I have two datasets (one from each individual), I performed a canonical correlation analysis (CCA) within the Seurat package to determine which genes provide common sources of variation between the healthy individual and the lupus individual. To determine the number of canonical components to be used in this analysis, a metagene bicorrelation plot was used to determine the saturation point between the two datasets. This was further verified using the CC hierarchical clustering plots. Once the number of CCs was determined, Seurat then used to align the CC subspaces and cluster the results in one single integrated analysis. Further, tobit differential expression analysis implemented also in Seurat was used to identify genes differentially expressed across conditions, individuals, and clusters. {Butler et al (2018)}

4.2.3 tSNE

Visualization of clusters determined through use of canonical components was performed using the tSNE dimensionality reduction procedure. Though principal component analysis, a linear and parametric method, has been used in the past for dimension reduction and analysis of genomic data, for visualization of scRNASeq data, it has been supplanted by t-distributed Stochastic Neighbor Embedding, or tSNE, a non-linear and non-parametric dimension reduction method. Many commonly used scRNASeq analysis pipelines automatically have tSNE included within the clustering and visualization of single-cell data, including Seurat. {Butler et al (2018); Maaten and Hinton, 2008; Wattenberg, 2016}

The origin of tSNE is the SNE (or Stochastic Neighbor Embedding) algorithm, which uses Euclidean distances and measures similarity between two data points, *i* and *j*, as the probability that *i* would have *j* as its neighbour under a Gaussian distribution that is centered on *i.* Further, SNE requires the use of variance in the calculations of these probabilities. Since the dense regions require a smaller variance than the less dense regions, SNE performs a binary search for the optimal variance to be used for calculation based on perplexity, which is user-defined. Perplexity is effectively a measurement of number of neighbors and accounts for entropy of the cluster. Because of these inputs, often the computation time for SNE is high, and multiple runs are required to optimize the user-selected perplexity score, which can result in crowding of datapoints. {Butler et al (2018); Maaten and Hinton, 2008; Wattenberg, 2016}

Unlike SNE, tSNE utilizes a heavy-tailed Student's t-distribution rather than the Gaussian distribution which minimizes crowding. Additionally, tSNE uses a symmetrized SNE cost function to lower computational time. Further, it is more robust to changes in perplexity which minimizes the need for multiple runs to optimize the clustering of datapoints. The presentation of single-cell clusters is shown as a tSNE plot. This has become common practice within the context of single-cell RNASeq analysis. {Butler et al (2018); Maaten and Hinton, 2008; Wattenberg, 2016}

4.3 Results

Due to limitations of this experiment containing only two individuals, joint analysis of these cells was confounded by the fact that there was just a single individual within each disease state (healthy or lupus) and cell type (naïve, or popD). Nevertheless, I used Seurat's canonical correlation (CC) analysis pipeline to identify common cell types across both individuals, by identifying common sources of variation between both sets of data. Given these I then quantified differential expression between the two samples. Any inference that the difference is a mark of disease must be tempered by the lack of repetition, though comparison with knowledge from existing bulk transcriptomic data can be used as a guide.

Figure 4.1 provides a bicorrelation plot of the healthy and lupus samples for all cells. A "saturation" point for CCs is at approximately the first 10 CCs after which there is a drop-off in signal. To ascertain the genes that drive each of these CCs, Figure 4.2 displays heatmaps of the top genes that drive each CC for CC1-12. Cells are on the x-axis and genes are shown on the y-axis. As can be seen, there is some structure to the first 10 CCs which drops off at CCs 11 and 12. Following the selection of CCs to include, these 10 CCs were

aligned between lupus and healthy samples, allowing discrimination of cell types by integrated cluster analysis in Seurat.

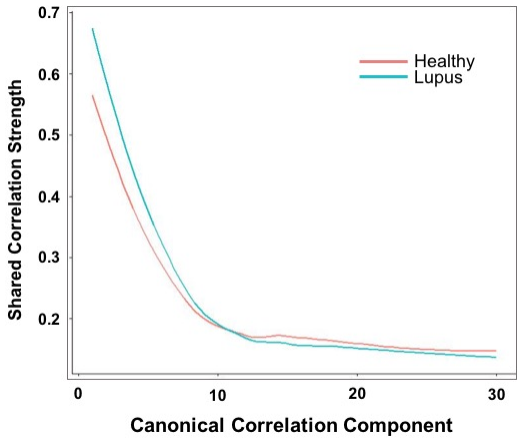


Figure 4.1. Bicorrelation Plot of Healthy and Lupus Samples. This plot compares the shared correlation strength of the Healthy and Lupus samples on y-axis of each canonical correlation component (on the x-axis). The drop-off in correlation strength begins at after CC10 so CC1-10 was used for this analysis.

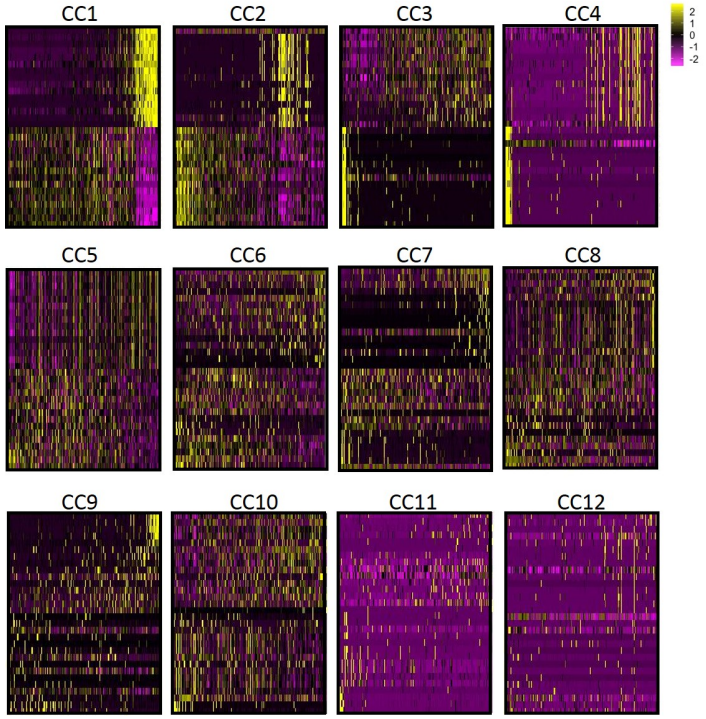


Figure 3.2. Heatmaps of First 6 CCs. This is a heatmap to visualize the first 6CCs and their top driving genes. The genes are represented on the rows and the cells are represented as columns. CC1-6 shows structure in driving gene expression.

The result of this clustering is displayed in Figure 4.3. Four identical tSNE plots are shown, varying only in their color schemes. Figure 4.3A shows the three clusters (clusters 0, 1, and 2) that result from the integrated CCA clustering. Figure 4.3B shows the distribution of healthy cells and lupus cells within these clusters. As can be seen, there is no special clustering of lupus or healthy samples within each cluster. Figure 4.3C shows the distribution of naïve and popD cells within each cluster. Cluster 1 is dominated by naïve B-cells with the exception of 3 cells. Clusters 0 and 2 are dominated by popD cells, suggesting that there might be a subpopulation of popD cells. Figure 4.3D shows the distribution of disease and population within the same tSNE plot, effectively recapitulating Figs 4.3B and 4.3C in one plot.

Differential expression analysis was performed using the tobit method in Seurat across disease state, cell population, cluster, and combinations of these for genes that were present in greater than 10% of the cells in both conditions. Figure 4.4A shows the differential expression analysis of healthy and lupus popD cells. The 312 differentially expressed genes are on the x-axis and disease state is displayed on the y-axis. Gene set enrichment analysis was performed on these 337 genes ($p < 0.05$, FDR < 0.05) and resulted in two significantly ($p < 0.05$, FDR < 0.05) dysregulated gene sets (Fig 4.4B): TNFA signaling (enriched in healthy cells) and UPR (enriched in lupus cells).

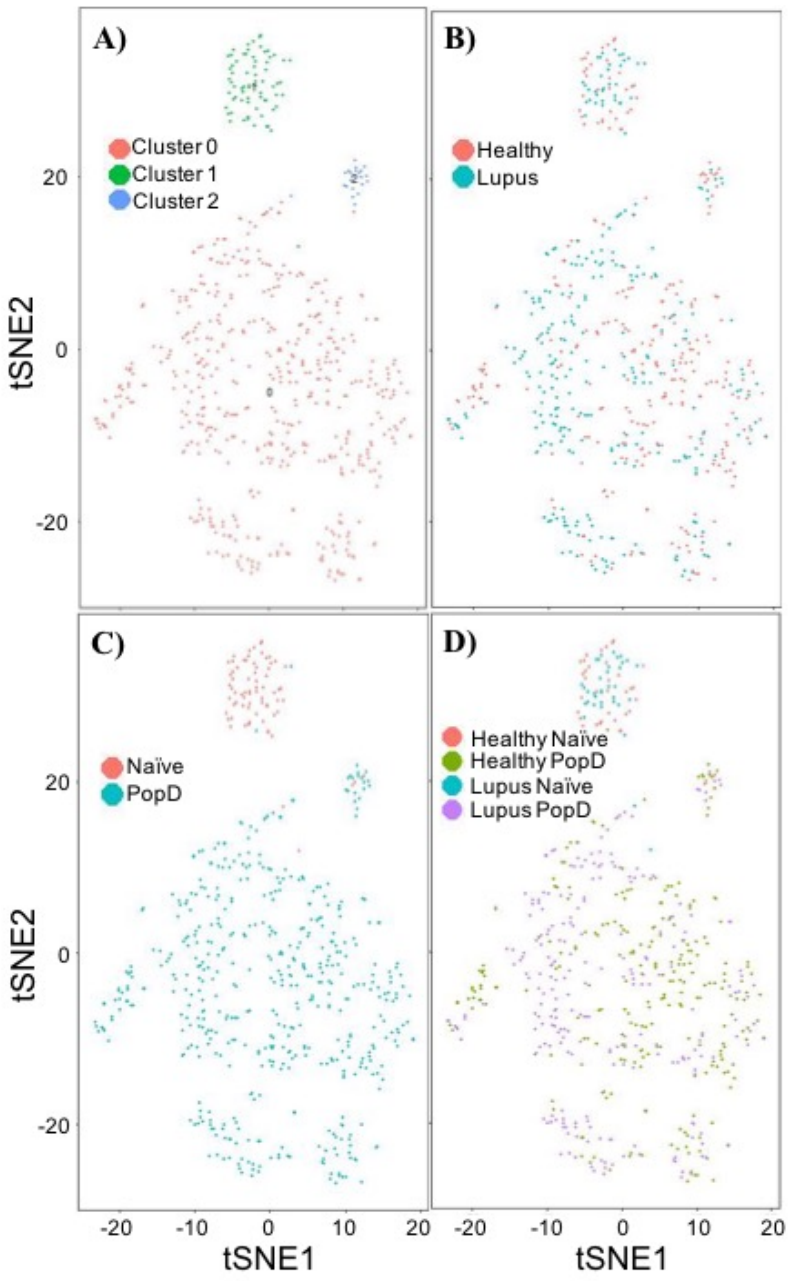


Figure 4.3. tSNE plots of CC-based Clusters. These four graphs present (A) the three clusters (cluster 0, 1, and 2) that were determined using the first 10 CCs. (B) These clusters do not separate out by disease state (healthy or lupus). (C) These clusters do separate by cell population (Naïve and popD) where clusters 0 and 2 are comprised nearly entirely of popD cells and cluster 1 is comprised of nearly entirely naïve cells. (D) This is an amalgamation of (A-C) to show all four combinations of disease and cell populations: Healthy Naïve, Healthy PopD, Lupus Naïve, Lupus PopD.

A similar analysis of differential gene expression analysis followed by GSEA was performed on naïve cells in comparison with popD cells (Fig 4.4C). This resulted in 471 significantly differentially expressed genes ($p < 0.05$, FDR < 0.05) with 10 significantly differentially regulated pathways ($p < 0.05$, FDR < 0.05) between popD and naïve, nine of which are enriched in popD (Fig 4.4D). These pathways fall under the broad categories of inflammation (allograft rejection, inflammatory response, and IFNG response). Both KRAS Up and KRAS Down gene sets are enriched. While both can be enriched due to the fact that both activators and repressors are treated equally in these analyses, this also suggests that there may be coregulation of the KRAS signaling pathway in opposite directions. Further IL2-STAT5, PI3K-AKT-MTOR, and apical junction are enriched in popD as well. The only pathway that is enriched in naïve cells, interestingly, is UPR. This is counterintuitive due to the fact that UPR is commonly associated in B-cells with high rates of antibody production which is not characteristic of naïve B-cells in the marrow.

Figure 4.4E compares cluster 0 with cluster 2, effectively comparing the bulk of the popD cells with the novel popD cells. The result is 163 significantly differentially expressed genes. Gene set enrichment analysis of these genes highlights Allograft Rejection, Apical Junction, Apoptosis, Coagulation, Complement, EMT, Glycolysis, Heme Metabolism, Hypoxia, IL2-STAT5, Inflammatory Response, KRAS Up, PI3K-AKT-MTOR, Reactive Oxygen Species, and TNFA Signaling pathways previously discussed in Chapters 2 and 3, while adding Cholesterol Homeostasis as well. All of these pathways are upregulated in cluster 2, which is the cluster of novel popD cells.

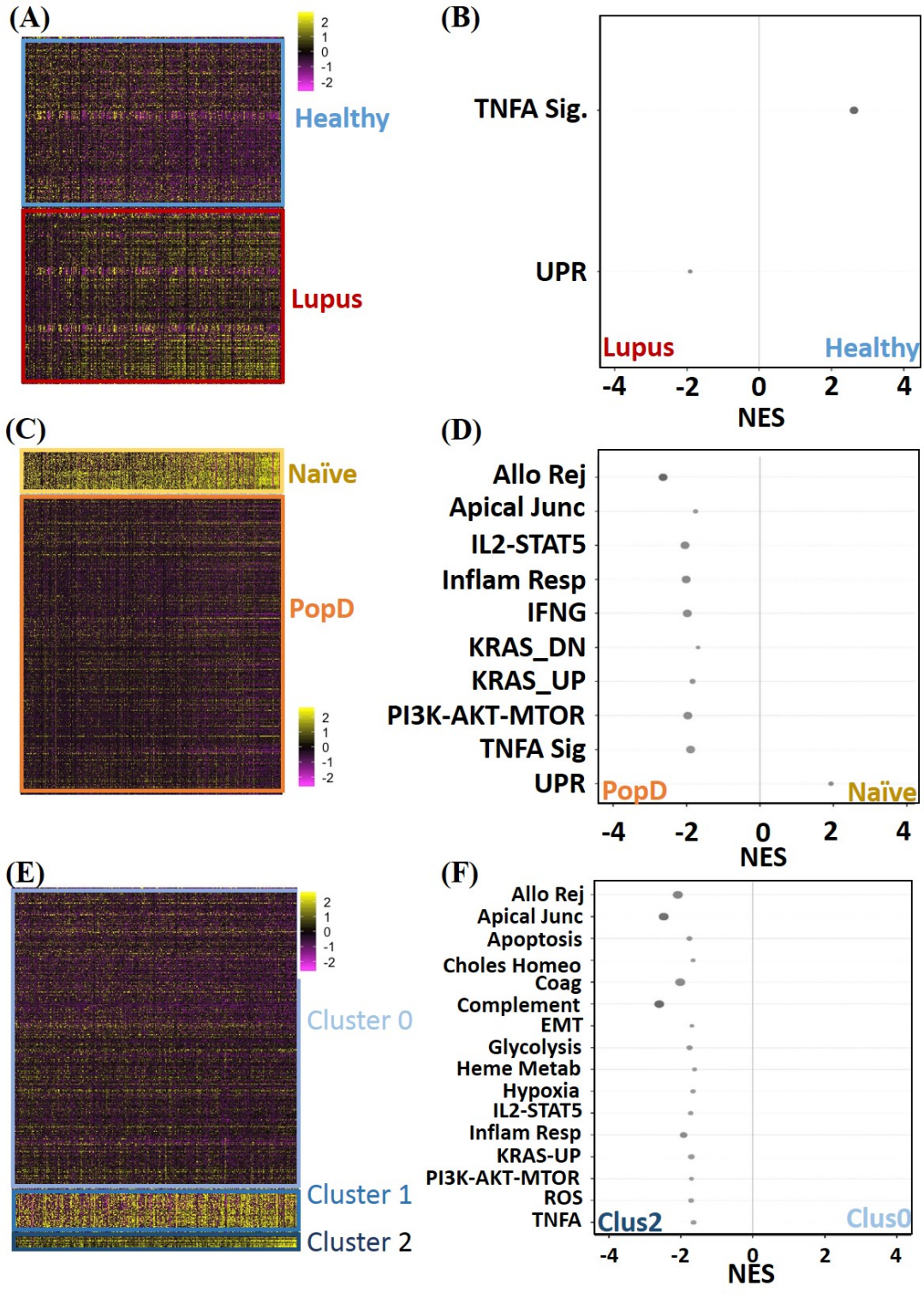

(A)
2
1
0
-1
-2
Healthy
Lupus
(B)
TNFA Sig.
UPR
Lupus
Healthy
-4
-2
0
2
4
NES
(C)
Naïve
PopD
2
1
0
-1
-2
(D)
Allo Rej
Apical Junc
IL2-STAT5
Inflam Resp
IFNG
KRAS_DN
KRAS_UP
PI3K-AKT-MTOR
TNFA Sig
UPR
PopD
Naïve
-4
-2
0
2
4
NES
(E)
2
1
0
-1
-2
Cluster 0
Cluster 1
Cluster 2
(F)
Allo Rej
Apical Junc
Apoptosis
Choles Homeo
Coag
Complement
EMT
Glycolysis
Heme Metab
Hypoxia
IL2-STAT5
Inflam Resp
KRAS-UP
PI3K-AKT-MTOR
ROS
TNFA
Clus2
Clus0
-4
-2
0
2
4
NES

Figure 4.4. Differential Expression analysis of Markers of Disease States and Cell Populations. (A) Heatmap representing the gene markers that differentiate healthy donor cells from lupus donor cells. The cells are represented on the y-axis and genes are represented on on the x-axis. (B) Bubble plot of the gene set enrichment analysis of the genes that are significantly ($p < 0.05$, FDR < 0.05) differentially expressed between healthy and lupus cells The size of each bubble is directly correlated with the negative log nominal p-value of each pathway. The x-axis shows the normalized enrichment score of the comparison between Healthy and Lupus. The color the bubbles is a spectrum between black (towards -4 or 4) to white (trending closer to zero). The bubble becomes darker as the absolute value of the NES. (C) Heatmap representing gene markers that differentiate naïve B cells from popD cells. (D) The results of the GSEA for this comparison. (E) Heatmap representing the gene markers that differentiate cells in Cluster 0 (the main cluster for PopD) from Cluster 2 (the smaller popD cluster). The cells are represented on the y-axis and genes are represented on on the x-axis. (F) The results of the gene set enrichment analysis of the genes that are significantly ($p < 0.05$, FDR < 0.05) differentially expressed between Cluster 0 and Cluster 2 are shown in this bubble plot of the normalized enrichment scores.

4.4 Discussion

Using the Seurat package, CCA analysis of samples obtained from lupus and healthy donors allowed for determination of cell subpopulations independent of disease state (and, consequently, individual). Subsequent analysis showed the presence of three clusters (Fig 4.2A). None of these three clusters was specific to a disease state (Fig 4.2B). While the presence of a naïve cluster separate from popD was expected (Fig.4.2C), in addition to a naïve cluster and a large popD cluster, a third cluster also emerged that was also predominantly filled with an alternate set of popD cells.

Differential expression analysis and gene set enrichment analysis resulted in the recapitulation of many of the pathways determined to be significant in the comparison of LLPC to SLPC and ASC (Chapter 3). When comparing healthy with lupus cells, the UPR was enriched in the lupus popD cells (Fig 4.4A). Pending reanalysis with more samples,

this could suggest a greater amount of ER stress and protein production demand in the lupus cells.

In the comparison of naïve B to PopD cells, inflammatory pathways and survival and signaling pathways were enriched in popD. The only pathway upregulated in naïve B cells is the unfolded protein response pathway. This is unexpected due to the fact that this pathway is associated with high antibody production rates that are more typical of plasma cells rather than naïve B-cells. It is possible that the UPR gene set, which is noted to be associated with stress response in the B-cell, is accommodating a different stress that is unrelated to high antibody production rates.

I further compared differential expression across the three clusters, focusing on the naïve B and novel popD clusters in Figure 4.5. The DE analysis and GSEA between clusters 0 and 1, not shown, effectively recapitulated the analysis between naïve and popD cells as cluster 0 is effectively the majority of popD cells and cluster 1 is primarily naïve cells. When comparing cluster 0 with cluster2 (comparing the two populations of popD), however, the results included a greater number of pathways that were upregulated in cluster 2. Under the assumption that this is representative of true signal, this suggests a potential subset of long-lived plasma cells. Whether or not it is a transitional cell type or novel type of mature popD cell remains to be elucidated through further single-cell samples and experiments as well as potential lineage tracing experiments.

4.5 Conclusion

Because of the low power of this experiment, only preliminary conclusions can be drawn from this dataset. Despite this, my analysis of single cell popD and naïve cells does suggest the possibility of additional sub-populations of plasma cells, in addition to the discrete states defined by the three cell surface markers CD19, CD38 and CD138 as analyzed in Chapters 2 and 3.

CONCLUSIONS

5.1 Introduction

Mammalian immune systems are characterized by two major components: the innate and adaptive immune systems. A powerful key feature of the adaptive immune system is its ability to provide long-lasting immunity. This feature is largely composed of plasma cells and memory B-cells. {Shi et al (2015);Jourdan et al (2011);Halliley et al (2015)}

Plasmablasts are precursors to both memory B-cells and terminally differentiated plasma cells. Amongst other locations, plasma cells migrate to the bone marrow and synthesize copious amounts of antibody. { Zhang et al (2009);Tellier et al (2018);Shi et al (2015);Nutt et al (2015);Klein et al (2008);Kassambara et al (2015);Jourdan et al (2011);Chatterjee et al (2007);Rawstron et al (2006);Klein et al (2003);Good et al (2009); Mei et al (2015) }

While numerous studies have explored gene expression of plasma cells, none have yet focused on the subsets of plasma cells in both the peripheral blood and bone marrow compartments: antibody secreting cells (pops2, 3, 5) in the peripheral blood, and short-lived (popsA and B) and long-lived (popD) plasma cells in the marrow. {Halliley et al (2015)} Since these previous studies have primarily addressed these cell types as a homogenous population, analysis of the biochemical properties of the plasma cells (PC) in peripheral blood and bone marrow have not proved fruitful in the context of study of diseases characterized by plasma cell dyscrasia. In this work, I described comparative whole-transcriptome analysis of three peripheral blood populations (pop2, pop3, and pop5)

from flow-sorted PC preparations and for three bone marrow samples, all of which came from live human donors.

5.2 Peripheral Blood Antibody Secreting Cells

During the initial discovery of the long-lived phenotype, the Lee Lab at Emory also documented the existence of three populations of peripheral blood antibody-secreting cells that were determined to be potential precursor populations to the three populations in the marrow due to their similarity in cell surface marker presentation (Table 1.1). Until this study, no in-depth analysis of the transcriptional and molecular properties of the PB-specific ASCs had been performed. {Tipton et al (2018);Halliley et al (2015)}

While it is known that ASCs contribute to antibody production during infection despite the rarity of the cell type, most analyses of ASCs have addressed them as a homogenous cell type and were primarily performed within the context of the murine model. Subsequently, these studies have largely only been able to determine ASC signatures that identify them in comparison to other B-cell subsets rather than as subsets themselves. {Tellier et al (2018);Tellier et al (2017);Shi et al (2015);Nguyen and Lewis et al (2018);Nguyen et al (2018);Kassambara et al (2015);Halliley et al (2015)}.

In my work, I analyzed pop2, pop3, and pop5 transcriptionally. My findings showed a marked similarity between pops2 and 3 to the exclusion of pop5. Further, these finding suggest that pops2 and 3 maintain the potential to develop into the long-lived phenotype. Proliferation pathway enrichment in pop2 suggest that it is the precursor to pop3 which begins to downregulate these pathways and upregulate hypoxia and TNFA signalling, both of which are upregulated in the SLPC and LLPC (discussed in Chapter 3). The

downregulation of UPR in pop5 suggest that it is autophagic and that the UPR stress response may be unable to manage the stress of ER development and transcriptional shifts, possibly resulting in a pre-apoptotic stage. However, further validation experiments must be performed to verify these hypotheses as transcriptional analysis alone is insufficient to draw any conclusions.

5.3 Bone Marrow Plasma Cells and Peripheral Blood Antibody Secreting Cells

Understanding and analysis of the role of plasma cells in plasma cell dyscrasias has largely been hampered by the consideration of these cell types as a single population. In 2015, Halliley et al presented clear evidence that bone marrow plasma cells are not homogenous. Further, they went on to characterize a subset of plasma cells that present the long-lived phenotype that is responsible for the maintenance of serological memory. {Halliley et al (2015);Bayer-Garner et al (2001);Aref et al (2003)}.

As discussed in Chapter 2, five sub-populations are observed in peripheral blood based on the cell surface markers CD19, CD38, and CD138: pop1, pop2, pop3, pop4 and pop5. Based on parallel cell surface marker presentation, four bone marrow populations have been observed: popA, popB, popC, and popD. {Halliley et al (2015)}. Due to the parallel cell surface marker presentation, the peripheral blood populations were hypothesized to be precursors of the bone marrow populations. Of these, we have been able to assess the cell-sorted whole transcriptomes by RNASeq for the six most abundant populations: popA, popB, popD, pop2, pop3, and pop5. {Halliley et al (2015)}. Until this

study, contributions of specific aspects of gene expression to functional differences among these subpopulations across tissues was not described.

I found that peripheral blood and bone marrow population gene expression cluster within tissue. I put forth 4 models (Figure 1.2) that were more thoroughly described in Chapter 3. My analysis showed that the origin of the BM populations cannot be explained without further experimentation including single cell analysis and lineage tracing, while a mixture of the precursor or transitional models is possible. By resolving these relationships through future studies, more granular and clearer insight into the development of long-lived plasma cells can be determined which would have implications in the context of health and disease. Further, integration of proteomic and transcriptomic analyses put forth several hypotheses of the role of numerous proteins on the development and survival of LLPC that were then validated experimentally by Doan Nguyen of the Lee Lab in Emory, including: TNFA, GAPDH, MTOR, HSPD1, FN1, YWHAS, hnRPA1, and Hypoxia.

5.4 Single Cell RNASeq of PopD and Naïve B-cells

This work has been primarily focused on granular understanding of plasma cells which have, until recently, been understood to be a homogenous cell type. In Chapters 2 and 3 I considered 6 subsets of the plasma cells populations: pop2, pop3, pop5, popA, popB, and popD. In this work, I increased the granularity of assessment of popD cells through single-cell RNASeq analysis. In this chapter, the question was posed as to whether LLPCs contain within them, subpopulations of long-lived plasma cells.

A major implication of this work is that we were only able to procure bone marrow samples for two individuals, one healthy and one with systemic lupus erythematosus (hereafter, 'lupus'). For each of these individuals, we procured naïve B-cells and LLPC and performed a clustering analysis using a canonical component analysis within the Seurat program and tSNE to cluster the cells. The results were three clusters; one contained Naïve cells, and two were popD populations.

As previously stated, the low power of this experiment only allows for preliminary conclusions to be drawn. Despite this, my analysis of single cell popD and naïve cells does suggest the possibility of additional sub-populations of plasma cells, in addition to the discrete states defined by the three cell surface markers CD19, CD38 and CD138 as analyzed in Chapters 2 and 3.

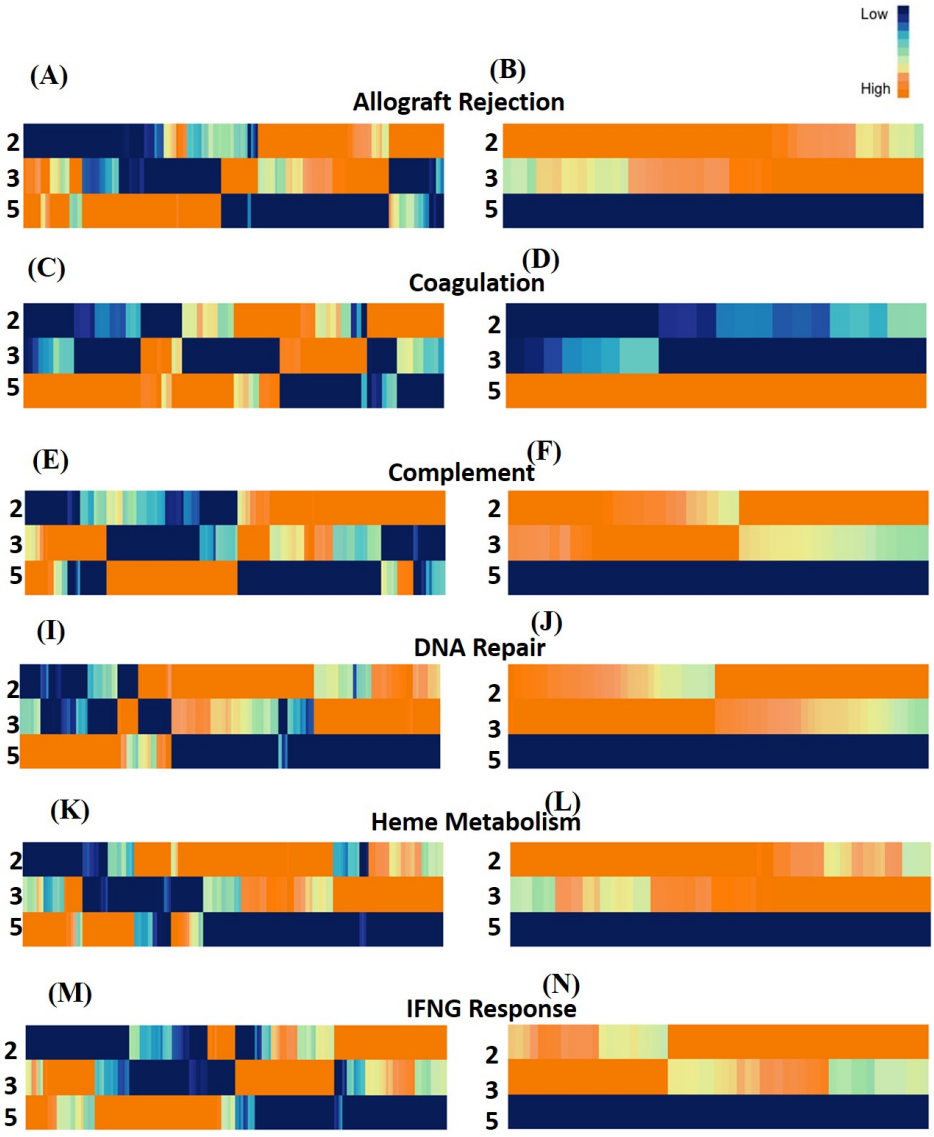
5.5 Conclusions

In this work, I transcriptionally characterized six populations of plasma cells; three from the bone marrow (pops A, B, and D), and three populations from the peripheral blood (pops 2, 3, and 5). I created four potential models of development of long-lived plasma cells (Figure 1.2) as a result of these analyses. While this study resulted in the determination of many components that contribute to long-lived development and survival were determined and experimentally validated across these cell types, further experimental validation, single-cell RNASeq, and lineage tracing will be required to more fully elucidate the granular mechanisms of development of long-lived plasma cells.

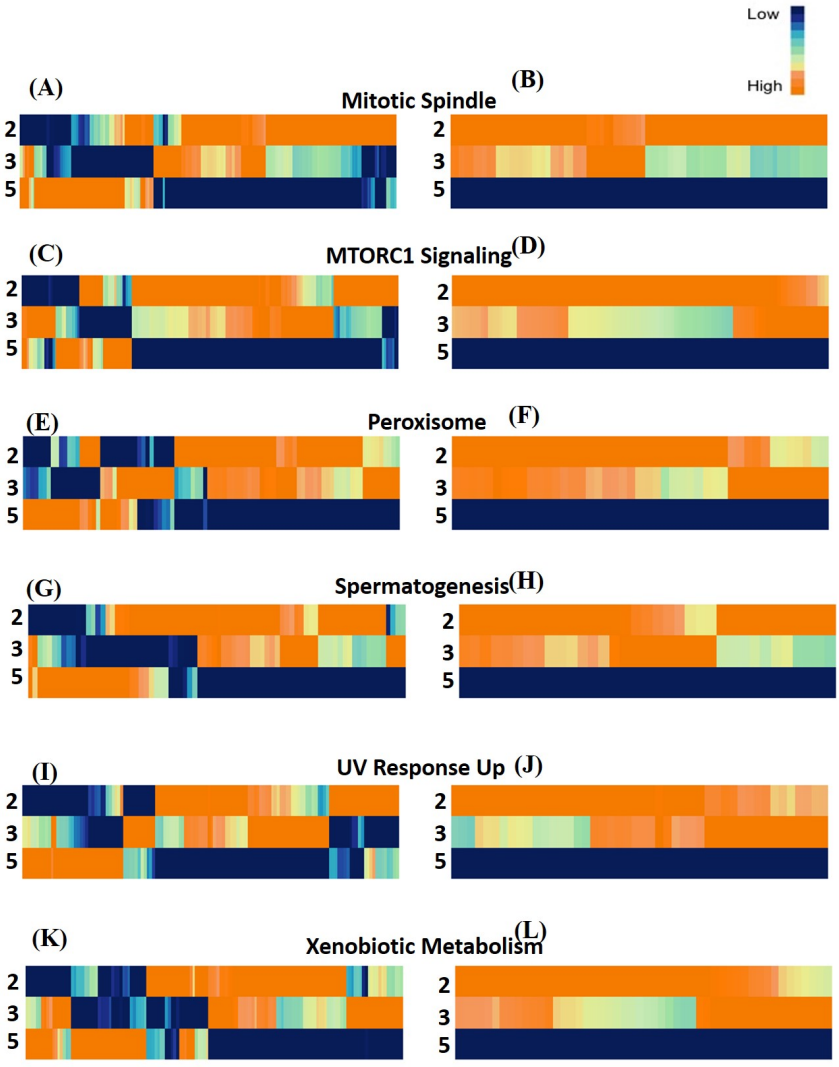
APPENDIX A.

Abbreviation	Pathway Name
Adipo	Hallmark Adipogenesis
Allo Rej	Hallmark Allograft Rejection
Androg Resp	Hallmark Androgen Response
Apic Junc	Hallmark Apical Junction
Apic Surf	Hallmark Apical Surface
Apoptosis	Hallmark Apoptosis
Choles Homeo	Hallmark Cholesterol Homeostasis
Coag	Hallmark Coagulation
Complement	Hallmark Complement
DNA Resp	Hallmark DNA Repair
E2F Targets	Hallmark E2F Targets
EMT	Hallmark Epithelial Mesencymal Transition
Estro Resp 1	Hallmark Estrogen Response Early
Estro Resp 2	Hallmark Estrogen Response Late
FA Metab	Hallmark Fatty Acid Metabolism
G2M Chkpt	Hallmark G2M Checkpoint
Glyco	Hallmark Glycolysis
Heme Metab	Hallmark Heme Metabolism
Hypoxia	Hallmark Hypoxia
IL2-STAT5	Hallmark IL2-STAT5 Signaling
IL6-JAK-STAT3	Hallmark IL6-JAK-STAT3 Signaling
Inflam Resp	Hallmark Inflammatory Response
IFNA Resp	Hallmark Interferon Alpha
IFNG Resp	Hallmark Interferon Gamma
KRAS Up	Hallmark KRAS Signalling Up
Mit Spindle	Hallmark Mitotic Spindle
MTOR Sig	Hallmark MTORC1 Signaling
MYC Targ 1	Hallmark MYC Targets V1
MYC Targ 2	Hallmark MYC Targets V2
Myog	Hallmark Myogenesis
NOTCH Sig	Hallmark NOTCH Signaling
OxPhos	Hallmark Oxidative Phosphorylation
P53	Hallmark P53 Pathway
Perox	Hallmark Peroxisome
PI3K-AKT-MTOR	Hallmark PI3K-AKT-MTOR
Prot Sec	Hallmark Protein Secretion
ROS	Hallmark Reactive Oxigen Species
Spermat	Hallmark Spermatogenesis
TGFB Sig	Hallmark TGF Beta Signalling via NFKB
TNFA Sig	Hallmark TNFA Signalling
UPR	Hallmark Unfolded Protein Response
UV Response DN	Hallmark UV Response DN
UV Response Up	Hallmark UV Response Up
Xeno Metab	Hallmark Xenobiotic Metabolism

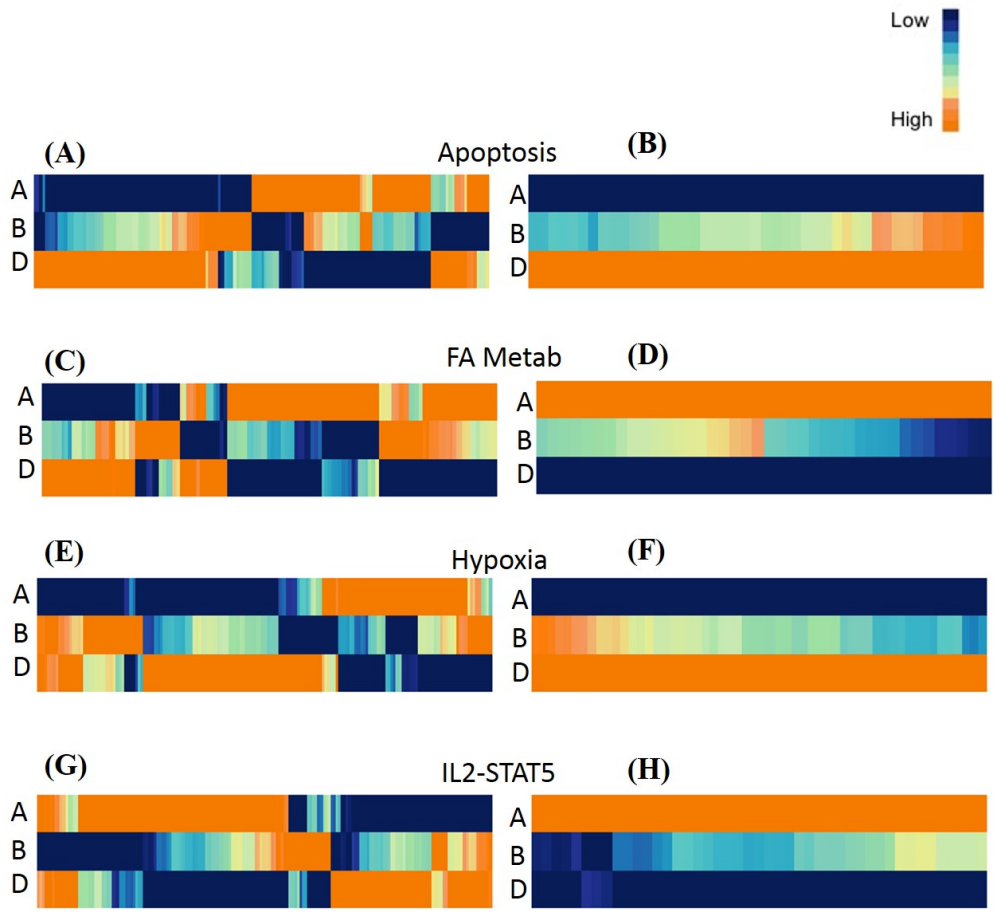
Appendix Table 2. Pathway Abbreviations and Associated Pathway Names. This table summarizes the abbreviations of pathway names in this study and their associated full name of each pathway. The column on the left contains the pathway abbreviation. The column on the right contains the full pathway name from the Broad Institute's Gene Set Enrichment Analysis Hallmark pathways.



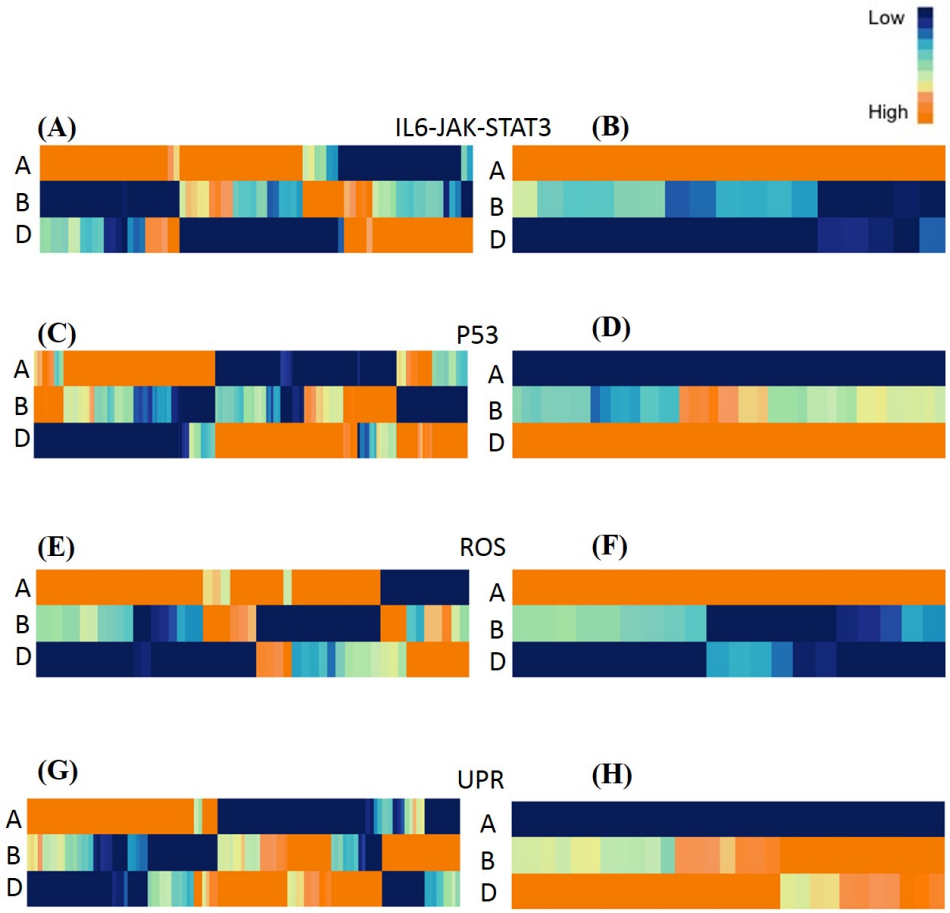
Appendix Figure 1. Two-Way Hierarchical Clusters of Incongruous Hallmark Gene Sets in the GSEA of pops 2, 3, and 5 and the Top Loading Genes of the First Principal Component for Each Pathway. These are two-way hierarchical clusters of 6 Hallmark gene sets. On the x-axis are genes within the denoted gene set and on the y-axis are the three populations of peripheral blood ASCs; where pop2 is denoted by the number 2, pop3 is denoted by the number 3, and pop5 is denoted by the number 5. (A) and (B) both describe genes within the Hallmark Allograft Rejection gene set. (C) and (D) describe genes within the Hallmark Coagulation gene set. (E) and (F) denote the Hallmark Complement gene set. (I) and (J) display Hallmark DNA Repair; (K) and (L) denote genes in the Hallmark Heme Metabolism gene set; and (M) and (N) show gene expression for Hallmark IFNG Response gene set. All 6 hierarchical clusters on the left (A, C, E, G, I, K, and M) show the expression of the all of the genes in the gene set for these 3 populations. The 6 hierarchical clusters on the right (B, D, F, H, J, L, and N) provide the gene expression for the subset of the genes in their respective gene sets with a loading of greater than or equal to 0.9.



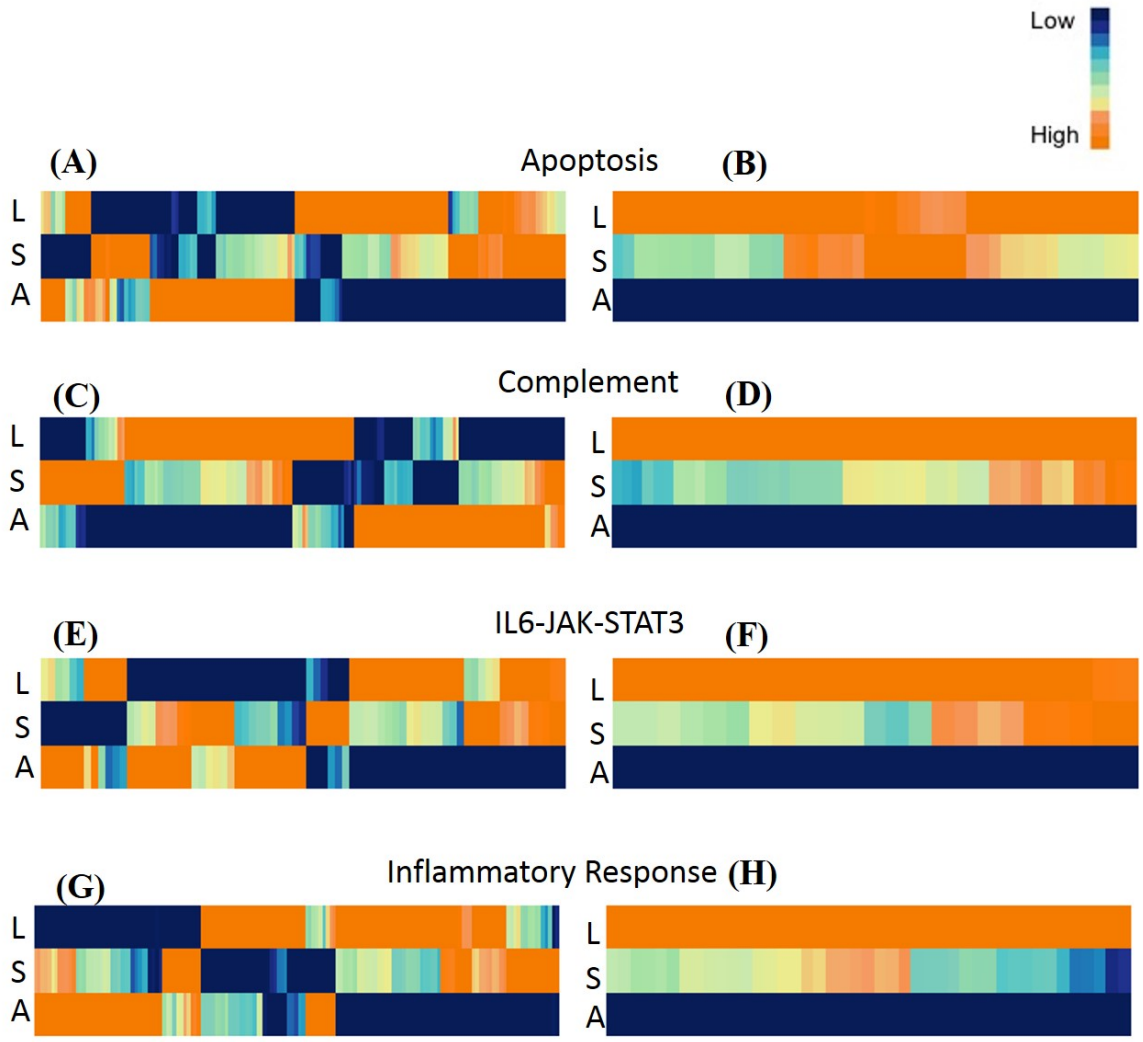
Appendix Figure 2. Two-Way Hierarchical Clusters of Incongruous Hallmark Gene Sets in the GSEA of pops 2, 3, and 5 and the Top Loading Genes of the First Principal Component for Each Pathway. These are two-way hierarchical clusters of 6 Hallmark gene sets. On the x-axis are genes within the denoted gene set and on the y-axis are the three populations of peripheral blood ASCs; where pop2 is denoted by the number 2, pop3 is denoted by the number 3, and pop5 is denoted by the number 5. (A) and (B) both describe genes within the Hallmark Mitotic Spindle gene set. (C) and (D) describe genes within the Hallmark MTORC1 Signaling gene set. (E) and (F) denote the Hallmark Peroxisome gene set. (I) and (J) display Hallmark Spermatogenesis; (K) and (L) denote genes in the Hallmark UV Response Up gene set; and (M) and (N) show gene expression for Hallmark Xenobiotic Metabolism gene set. All 6 hierarchical clusters on the left (A, C, E, G, I, K, and M) show the expression of the all of the genes in the gene set for these 3 populations. The 6 hierarchical clusters on the right (B, D, F, H, J, L, and N) provide the gene expression for the subset of the genes in their respective gene sets with a loading of greater than or equal to 0.9.



Appendix Figure 3. Two-Way Hierarchical Clusters of Incongruous Hallmark Gene Sets in the GSEA of pops A, B, and D and the Top Loading Genes of the First Principal Component for Each Pathway. These are two-way hierarchical clusters of 4 Hallmark gene sets. On the x-axis are genes within the denoted gene set and on the y-axis are the three populations of bone marrow plasma cells; where popA is denoted by the letter A, popB is denoted by the letter B, and popD is denoted by the letterD. (A) and (B) both describe genes within the Hallmark Apoptosis gene set. (C) and (D) describe genes within the Hallmark Fatty Acid Metabolism gene set. (E) and (F) denote the Hallmark Hypoxia gene set. (I) and (J) display Hallmark IL2-STAT5 Signaling gene set. All 4 hierarchical clusters on the left (A, C, E, and G) show the expression of the all of the genes in the gene set for these 3 populations. The 4 hierarchical clusters on the right (B, D, F, and H) provide the gene expression for the subset of the genes in their respective gene sets with a loading of greater than or equal to 0.9.



Appendix Figure 4. Two-Way Hierarchical Clusters of Incongruous Hallmark Gene Sets in the GSEA of pops A, B, and D and the Top Loading Genes of the First Principal Component for Each Pathway. These are two-way hierarchical clusters of 4 Hallmark gene sets. On the x-axis are genes within the denoted gene set and on the y-axis are the three populations of bone marrow plasma cells; where popA is denoted by the letter A, popB is denoted by the letter B, and popD is denoted by the letterD. (A) and (B) both describe genes within the Hallmark IL6-JAK-STAT3 gene set. (C) and (D) describe genes within the Hallmark P53 gene set. (E) and (F) denote the Hallmark Reactive Oxygen Species gene set. (I) and (J) display Hallmark Unfolded Protein Response gene set. All 4 hierarchical clusters on the left (A, C, E, and G) show the expression of the all of the genes in the gene set for these 3 populations. The 4 hierarchical clusters on the right (B, D, F, and H) provide the gene expression for the subset of the genes in their respective gene sets with a loading of greater than or equal to 0.9.



Appendix Figure 5. Two-Way Hierarchical Clusters of Incongruous Hallmark Gene Sets in the GSEA of LLPC, SLPC, and ASC and the Top Loading Genes of the First Principal Component for Each Pathway. These are two-way hierarchical clusters of 4 Hallmark gene sets. On the x-axis are genes within the denoted gene set and on the y-axis are the three populations of bone marrow plasma cells; where LLPC is denoted by the letter L, SLPC is denoted by the letter S, and ASC is denoted by the letter A. (A) and (B) both describe genes within the Hallmark Apoptosis gene set. (C) and (D) describe genes within the Hallmark Complement gene set. (E) and (F) denote the Hallmark IL6-JAK-STAT3 gene set. (I) and (J) display Hallmark Inflammatory Response gene set. All 4 hierarchical clusters on the left (A, C, E, and G) show the expression of the all of the genes in the gene set for these 3 populations. The 4 hierarchical clusters on the right (B, D, F, and H) provide the gene expression for the subset of the genes in their respective gene sets with a loading of greater than or equal to 0.9.

REFERENCES

1. Ackermann, J., et al., *Absence of p53 deletions in bone marrow plasma cells of patients with monoclonal gammopathy of undetermined significance.* Br J Haematol, 1998. **103**(4): p. 1161-3.
2. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
3. Ahmed, S.A., et al., *Gender and risk of autoimmune diseases: possible role of estrogenic compounds.* Environ Health Perspect, 1999. **107 Suppl 5**: p. 681-6.
4. Alanis-Lobato, G., M.A. Andrade-Navarro, and M.H. Schaefer, *HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks.* Nucleic Acids Res, 2017. **45**(D1): p. D408-D414.
5. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.
6. Aref, S., et al., *c-Myc oncogene and Cdc25A cell activating phosphatase expression in non-Hodgkin's lymphoma.* Hematology, 2003. **8**(3): p. 183-90.
7. Aref, S., T. Goda, and M. El-Sherbiny, *Syndecan-1 in multiple myeloma: relationship to conventional prognostic factors.* Hematology, 2003. **8**(4): p. 221-8.
8. Askanase, A., K. Shum, and H. Mitnick, *Systemic lupus erythematosus: an overview.* Soc Work Health Care, 2012. **51**(7): p. 576-86.
9. Askanase, A.D., *Estrogen therapy in systemic lupus erythematosus.* Treat Endocrinol, 2004. **3**(1): p. 19-26.
10. Banchereau, R., et al., *Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients.* Cell, 2016. **165**: p. 551-565.
11. Bayer-Garner, I.B., et al., *Syndecan-1 (CD138) immunoreactivity in bone marrow biopsies of multiple myeloma: shed syndecan-1 accumulates in fibrotic regions.* Mod Pathol, 2001. **14**(10): p. 1052-8.
12. Bordon, Y., *B cells: Whatever will B cell be?* Nat Rev Immunol, 2015. **15**(3): p. 132.
13. Bortnick, A. and D. Allman, *What is and what should always have been: long-lived plasma cells induced by T cell-independent antigens.* J Immunol, 2013. **190**(12): p. 5913-8.
14. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions, technologies, and species.* Nat Biotechnol, 2018. **36**(5): p. 411-420.
15. Buyon, J.P., et al., *The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: a randomized trial.* Ann Intern Med, 2005. **142**(12 Pt 1): p. 953-62.
16. Chambers, D.C., et al., *Transcriptomics and single-cell RNA-sequencing.* Respirology, 2018.
17. Chatterjee, M., et al., *STAT3 and MAPK signaling maintain overexpression of heat shock proteins 90alpha and beta in multiple myeloma cells, which critically contribute to tumor-cell survival.* Blood, 2007. **109**(2): p. 720-8.
18. Ching, K.H., et al., *Two major autoantibody clusters in systemic lupus erythematosus.* PLoS One, 2012. **7**(2): p. e32001.
19. Cokic, V.P., et al., *Proinflammatory Cytokine IL-6 and JAK-STAT Signaling Pathway in Myeloproliferative Neoplasms.* Mediators Inflamm, 2015. **2015**: p. 453020.

20. Corcoran, L.M. and S.L. Nutt, *The flavors of plasma cells.* Oncotarget, 2015. **6**(32): p. 32305-6.
21. Corcoran, L.M. and S.L. Nutt, *Long-Lived Plasma Cells Have a Sweet Tooth.* Immunity, 2016. **45**(1): p. 3-5.
22. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
23. Elman, S.A., et al., *Developing classification criteria for discoid lupus erythematosus: an update from the World Congress of Dermatology 2015 meeting.* Int J Womens Dermatol, 2016. **2**(2): p. 44-45.
24. Fortini, F., et al., *Estrogen receptor beta-dependent Notch1 activation protects vascular endothelium against tumor necrosis factor alpha (TNFalpha)-induced apoptosis.* J Biol Chem, 2017. **292**(44): p. 18178-18191.
25. H, M., *A unique population of IgG-expressing plasma cells lacking CD19 is enriched in human bone marrow.* Blood, 2015.
26. Halliley, J.L., et al., *Long-Lived Plasma Cells Are Contained within the CD19(-)CD38(hi)CD138(+) Subset in Human Bone Marrow.* Immunity, 2015. **43**(1): p. 132-45.
27. Hinton, M.a., *Visualizing Data Using t-SNE.* Journal of Machine Learning Research, 2008. **9**: p. 26.
28. Hipp, N., et al., *IL-2 imprints human naive B cell fate towards plasma cell through ERK/ELK1-mediated BACH2 repression.* Nat Commun, 2017. **8**(1): p. 1443.
29. Ishikawa, H., et al., *CD19 expression and growth inhibition of tumours in human multiple myeloma.* Leuk Lymphoma, 2002. **43**(3): p. 613-6.
30. Jenks, S.A., et al., *Distinct Effector B Cells Induced by Unregulated Toll-like Receptor 7 Contribute to Pathogenic Responses in Systemic Lupus Erythematosus.* Immunity, 2018.
31. Jourdan, M., et al., *Characterization of a transitional preplasmablast population in the process of human B cell to plasma cell differentiation.* J Immunol, 2011. **187**(8): p. 3931-41.
32. Jourdan, M., et al., *The myeloma cell antigen syndecan-1 is lost by apoptotic myeloma cells.* Br J Haematol, 1998. **100**(4): p. 637-46.
33. Jourdan, M., et al., *Delineation of the roles of paracrine and autocrine interleukin-6 (IL-6) in myeloma cell lines in survival versus cell cycle. A possible model for the cooperation of myeloma cell growth factors.* Eur Cytokine Netw, 2005. **16**(1): p. 57-64.
34. Jourdan, M., et al., *Tumor necrosis factor is a survival and proliferation factor for human myeloma cells.* Eur Cytokine Netw, 1999. **10**(1): p. 65-70.
35. Jourdan, M., et al., *A major role for Mcl-1 antiapoptotic protein in the IL-6-induced survival of human myeloma cells.* Oncogene, 2003. **22**(19): p. 2950-9.
36. Jourdan, M., et al., *IFN-alpha induces autocrine production of IL-6 in myeloma cell lines.* J Immunol, 1991. **147**(12): p. 4402-7.
37. Jundt, F., et al., *Jagged1-induced Notch signaling drives proliferation of multiple myeloma cells.* Blood, 2004. **103**(9): p. 3511-5.
38. Kallies, A. and S.L. Nutt, *Bach2: plasma-cell differentiation takes a break.* EMBO J, 2010. **29**(23): p. 3896-7.
39. Kassambara, A., et al., *GenomicScape: an easy-to-use web tool for gene expression data analysis. Application to investigate the molecular events in the differentiation of B cells into plasma cells.* PLoS Comput Biol, 2015. **11**(1): p. e1004077.

40. KL, G., *Resting Human Memory B Cells Are Intrinsically Programmed for Enhanced Survival and Responsiveness to Diverse Stimuli Compared to Naive B Cells.* Journal of Immunology 2009.
41. Klein, B., *Survival and Proliferation Factors of Normal and Malignant Plasma Cells.* International Journal of Hematology, 2003.
42. Klein, B., et al., *Production of growth factors by human myeloma cells.* Cancer Res, 1987. **47**(18): p. 4856-60.
43. Klein, B., et al., *Activation molecules on human myeloma cells.* Curr Top Microbiol Immunol, 1999. **246**: p. 335-41.
44. Klein, B., et al., *Interleukin-6 is the central tumor growth factor in vitro and in vivo in multiple myeloma.* Eur Cytokine Netw, 1990. **1**(4): p. 193-201.
45. Klein, B., et al., *Interleukin-6 is a major myeloma cell growth factor in vitro and in vivo especially in patients with terminal disease.* Curr Top Microbiol Immunol, 1990. **166**: p. 23-31.
46. Klein, U. and R. Dalla-Favera, *Germinal centres: role in B-cell physiology and malignancy.* Nat Rev Immunol, 2008. **8**(1): p. 22-33.
47. Kuhn, A., et al., *S2k guideline for treatment of cutaneous lupus erythematosus - guided by the European Dermatology Forum (EDF) in cooperation with the European Academy of Dermatology and Venereology (EADV).* J Eur Acad Dermatol Venereol, 2017. **31**(3): p. 389-404.
48. Law, C.W., et al., *RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR.* F1000Res, 2016. **5**: p. 1408.
49. Le Gallou, S., et al., *IL-2 requirement for human plasma cell generation: coupling differentiation and proliferation by enhancing MAPK-ERK signaling.* J Immunol, 2012. **189**(1): p. 161-73.
50. Liberzon, A., et al., *The Molecular Signatures Database (MSigDB) hallmark gene set collection.* Cell Syst, 2015. **1**(6): p. 417-425.
51. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0.* Bioinformatics, 2011. **27**(12): p. 1739-40.
52. Liu, Z., Y. Zou, and A. Davidson, *Plasma cells in systemic lupus erythematosus: the long and short of it all.* Eur J Immunol, 2011. **41**(3): p. 588-91.
53. Ma, Y., et al., *Non-thermal atmospheric pressure plasma preferentially induces apoptosis in p53-mutated cancer cells by activating ROS stress-response pathways.* PLoS One, 2014. **9**(4): p. e91947.
54. Mathian, A., et al., *Interferon-alpha induces unabated production of short-lived plasma cells in pre-autoimmune lupus-prone (NZBxNZW)F1 mice but not in BALB/c mice.* Eur J Immunol, 2011. **41**(3): p. 863-72.
55. Mecham, B.H., P.S. Nelson, and J.D. Storey, *Supervised normalization of microarrays.* Bioinformatics, 2010. **26**: p. 1308-1315.
56. Medina, F., et al., *The heterogeneity shown by human plasma cells from tonsil, blood, and bone marrow reveals graded stages of increasing maturity, but local profiles of adhesion molecule expression.* Blood, 2002. **99**(6): p. 2154-61.
57. Mootha, V.K., et al., *PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.* Nat Genet, 2003. **34**(3): p. 267-73.

58. Nguyen, D.C., et al., *Factors of the bone marrow microniche that support human plasma cell survival and immunoglobulin secretion.* Nat Commun, 2018. **9**(1): p. 3698.
59. Nguyen, D.C., et al., *Extracellular vesicles from bone marrow-derived mesenchymal stromal cells support ex vivo survival of human antibody secreting cells.* J Extracell Vesicles, 2018. **7**(1): p. 1463778.
60. Nikolayeva, O. and M.D. Robinson, *edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology.* Methods Mol Biol, 2014. **1150**: p. 45-79.
61. Nutt, S.L., et al., *The generation of antibody-secreting plasma cells.* Nat Rev Immunol, 2015. **15**(3): p. 160-71.
62. Oleksyn, D., et al., *Protein kinase Cbeta is required for lupus development in Sle mice.* Arthritis Rheum, 2013. **65**(4): p. 1022-31.
63. Patergnani, S., et al., *PRKCB/protein kinase C, beta and the mitochondrial axis as key regulators of autophagy.* Autophagy, 2013. **9**(9): p. 1367-85.
64. Peperzak, V., et al., *Mcl-1 is essential for the survival of plasma cells.* Nat Immunol, 2013. **14**(3): p. 290-7.
65. Picelli, S., et al., *Full-length RNA-seq from single cells using Smart-seq2.* Nat Protoc, 2014. **9**(1): p. 171-81.
66. Pracht, K., et al., *A new staining protocol for detection of murine antibody-secreting plasma cell subsets by flow cytometry.* Eur J Immunol, 2017. **47**(8): p. 1389-1392.
67. Rawston, A.F., JA; et al, *The interleukin-6 receptor alpha-chain (CD126) is expressed by neoplastic but not normal plasma cells.* Blood, 2006.
68. Richardson, C., et al., *Molecular basis of 9G4 B cell autoreactivity in human systemic lupus erythematosus.* J Immunol, 2013. **191**(10): p. 4926-39.
69. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
70. Rodriguez, V.E., E.N. Gonzalez-Pares, and C. Rivera, *Clinical manifestations and vascular events in patients with lupus erythematosus anticardiolipin antibodies and raynaud's phenomenon.* P R Health Sci J, 2006. **25**(4): p. 307-13.
71. Rodriguez-Bayona, B., et al., *STAT-3 activation by differential cytokines is critical for human in vivo-generated plasma cell survival and Ig secretion.* J Immunol, 2013. **191**(10): p. 4996-5004.
72. Sharma, S., et al., *Using single cell analysis for translational studies in immune mediated diseases: Opportunities and challenges.* Mol Immunol, 2018. **103**: p. 191-199.
73. Shi, L., et al., *Gene expression profiling and functional analysis reveals that p53 pathway-related gene expression is highly activated in cancer cells treated by cold atmospheric plasma-activated medium.* PeerJ, 2017. **5**: p. e3751.
74. Shi, W., et al., *Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells.* Nat Immunol, 2015. **16**(6): p. 663-73.
75. Slocombe, T., et al., *Plasma cell homeostasis: the effects of chronic antigen stimulation and inflammation.* J Immunol, 2013. **191**(6): p. 3128-38.
76. Stephenson, L.M., et al., *Vav proteins regulate the plasma cell program and secretory Ig production.* J Immunol, 2006. **177**(12): p. 8620-5.
77. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.

78. Taylor, J.J., et al., *Humoral immunity. Apoptosis and antigen affinity limit effector cell differentiation of a single naive B cell.* Science, 2015. **347**(6223): p. 784-7.
79. Tellier, J. and S.L. Nutt, *Standing out from the crowd: How to identify plasma cells.* Eur J Immunol, 2017. **47**(8): p. 1276-1279.
80. Tellier, J. and S.L. Nutt, *Plasma cells: the programming of an antibody-secreting machine.* Eur J Immunol, 2018.
81. Tellier, J., et al., *Blimp-1 controls plasma cell function through the regulation of immunoglobulin secretion and the unfolded protein response.* Nat Immunol, 2016. **17**(3): p. 323-30.
82. Tipton, C.M., et al., *Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus.* Nat Immunol, 2015. **16**(7): p. 755-65.
83. Tipton, C.M., et al., *Understanding B-cell activation and autoantibody repertoire selection in systemic lupus erythematosus: A B-cell immunomics approach.* Immunol Rev, 2018. **284**(1): p. 120-131.
84. Trezise, S., et al., *Mining the Plasma Cell Transcriptome for Novel Cell Surface Proteins.* Int J Mol Sci, 2018. **19**(8).
85. Tsokos, G.C., et al., *New insights into the immunopathogenesis of systemic lupus erythematosus.* Nat Rev Rheumatol, 2016. **12**(12): p. 716-730.
86. van der Maaten L, Hinton G (2008) Visualizing Data using t-SNE. Journal of Machine Learning Research 9: 2579–2605.
87. Varet, H., et al., *SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data.* PLoS One, 2016. **11**(6): p. e0157022.
88. Vicente-Duenas, C., et al., *Loss of p53 exacerbates multiple myeloma phenotype by facilitating the reprogramming of hematopoietic stem/progenitor cells to malignant plasma cells by MafB.* Cell Cycle, 2012. **11**(20): p. 3896-900.
89. Wallrapp, A., et al., *Type 2 innate lymphoid cells in the induction and resolution of tissue inflammation.* Immunol Rev, 2018. **286**(1): p. 53-73.
90. Wang, L., et al., *Modulation of interleukin-6-induced plasma protein secretion in hepatoma cells by p53 species.* J Biol Chem, 1995. **270**(39): p. 23159-65.
91. Wattenberg, et al., "How to Use t-SNE Effectively", Distill, 2016. http://doi.org/10.23915/distill.00002
92. Whitehead, N.P., *Enhanced autophagy as a potential mechanism for the improved physiological function by simvastatin in muscular dystrophy.* Autophagy, 2016. **12**(4): p. 705-6.
93. Wilmore, J.R. and D. Allman, *Here, There, and Anywhere? Arguments for and against the Physical Plasma Cell Survival Niche.* J Immunol, 2017. **199**(3): p. 839-845.
94. Wilmore, J.R., D.D. Jones, and D. Allman, *Protocol for improved resolution of plasma cell subpopulations by flow cytometry.* Eur J Immunol, 2017. **47**(8): p. 1386-1388.
95. WJ, O.N.K., *Effects Of Androgens On T And B Lymphocyte Development.* Immunologic Research, 2001.
96. Zhang, J., et al., *Patterns of microRNA expression characterize stages of human B-cell differentiation.* Blood, 2009. **113**(19): p. 4586-94.