Gene expression differences within an *in vitro* germinal center model

Heather Rene Schiffhauer

Thesis for the degree of Master of Science
University of Iceland
Faculty of Medicine
School of Health Sciences

HÁSKÓLI ÍSLANDS

Breytingar á genatjáningu í ræktunarlíkani af kímstöðvarhvarfi

Heather Rene Schiffhauer

Ritgerð til meistaragráðu í Líf- og læknavísindum

Umsjónarkennari: Helga M. Ögmundsdóttir

Meistaranámsnefnd: Hlíf Steingrímsdóttir og Magnús Karl Magnússon

Læknadeild

Heilbrigðisvísindasvið Háskóla Íslands

Júní 2012

Gene expression differences within an *in vitro* germinal center model

Heather Rene Schiffhauer

Thesis for the degree of Master of Science

Supervisor: Helga M. Ögmundsdóttir

Masters committee: Hlíf Steingrímsdóttir og Magnús Karl Magnússon

Faculty of Medicine

School of Health Sciences

June 2012

Prentun: Háskólaprent

Reykjavík, Ísland 2012

ÁGRIP

Átta fjölskyldur hafa verið skilgreindar á Íslandi með mörg tilfelli af einstofna mótefnahækkun (monoclonal gammopathies (MG)), þar á meðal „monoclonal gammopathy of undetermined significance“ (MGUS), mergæxli (multiple myeloma (MM)) og Waldenströms makróglobulinemíu (WM). Einnig hefur verið sýnt fram á umbreytingu á MGUS í MM í nokkrum skrefum. Með örvunarprófi í frumuræktun með poke-weed mitogeni (PWM) fundust 13 heilbrigðir ættingjar sjúklinga í fjórum af þessum átta fjölskyldum með ofursvarandi (hyper-responding (HR)) B-frumur sem framleiddu immúnóglóbúlín í magni sem var > 2 SD yfir meðaltali viðmiða. Við komum upp frumuræktunarlíkani af kímstöðvarhvarfinu (germinal center (GC)) og framkvæmdum genatjáningarranssókn á óörvuðum sýnum og sýnum sem safnað var eftir 14 daga í GC rækt til að kanna hvort GC líkanið gefur rétta mynd af kímstövarhvarfi í líkamanum. Við skimuðum einnig fyrir mun milli viðmiða (skyldra og óskyldra) og HR hópsins. B frumur voru einangraðar úr blóðsýnum frá 11 ofursvörum, 11 óskyldum viðmiðum og 9 skyldum viðmiðum. mRNA var einangrað strax úr B frumum á degi 0 og síðan úr B frumum eftir 14 daga örvun í kímstöðvarlíkaninu og notað til þess að framleiða cDNA. Cy-3 merkt sýni voru sett á fákirna örflögur (oligonucleotide microarrays). Genatjáning var normalíseruð og viðeigandi ANOVA líkön notuð á undirbúin gögn til greiningar milli hópa og daga. Beitt var prófun fyrir mestu líkum (maximum likelihood) og líkindahlutfalli (likelihood ratio) og leiðréttingu samkvæmt Benjamini og Hochberg. Við greiningu var m.a. beitt „unsupervised hierarchical cluster analysis“ og prófun fyrir „ontology“ hópum með hugbúnaði sem mælt er með hjá Gene Ontology gagnagrunninum. Niðurstöður genatjáningar sýndu ólíka flokkun á hitakorti (heatmap) milli daga 0 og 14. Um það bil þriðjungur (14,277) genaumrita var marktækt breyttur eftir 14 daga í kímstöðvarræktuninni ($p_{adjusted} < 0.001$). Genin flokkuðust í marga flokka eins og við var að búast. Mest margfaldur munur (fold change (FC) differences (FC > 2.5)) kom fram fyrir fjölmörg gen og flokka, svo sem frumufjölgun, DNA pökkun og viðgerð, stýrðan frumudauða og frumulifun, og samrýmdist vel þeim breytingum sem fyrri rannsókn hefur lýst sem sértækum fyrir kímstöðvarfrumur í líkamanum. Enginn marktækur munur kom fram milli ofursvara (HR hóps) og viðmiðunarhópa þegar leiðrétt var fyrir endurteknum prófunum en ef ekki var leiðrétt sýndu 44 gen ólíka tjáningu á degi 14 ($p_{raw} < 0.001$). Niðurstöður genatjáningarrannsóknarinnar gefa til kynna að ræktunarlíkanið gefi raunsanna mynd af kímstöðvarhvarfi í líkamanum. Ofursvararnir (HR hópurinn) sem heild sýndi ekki marktækt frábrugðið genatjáningarmynstur. Unnið er að frekari greiningum.

ABSTRACT

Eight families in Iceland have been identified with multiple cases of monoclonal gammopathies (MG), including monoclonal gammopathy of undetermined significance (MGUS), multiple myeloma (MM) and Waldenström's macroglobulinemia (WM). It has also been shown that there is a multi-step transformation from MGUS to MM. An *in vitro* poke-weed mitogen (PWM) stimulation assay identified 13 disease-free relatives of patients within four of the eight families as having hyper-responding (HR) B-lymphocytes (B cells) producing Ig at > 2 Standard Deviations (SD) above controls. We established an *in vitro* model of the germinal center (GC) reaction and performed gene expression (GE) analysis on unstimulated samples and those collected after 14 days in the GC culture to investigate if the GC model faithfully mimics *in vivo* GC. We also screened for differences between controls (related and unrelated) and the HR group. B cells were isolated from peripheral blood (PB) samples from each HR (n=11), unrelated control (UC; n=11), and related (RC; n=9). mRNA was extracted immediately from B cells at day 0 and then from B cells after 14 days of stimulation in the *in vitro* GC culture and used to synthesize cDNA. Cy-3 labeled samples were loaded onto oligonucleotide microarrays. GE was normalized and appropriate linear-mixed effects models were applied to prepared data for analysis between study groups and stimulation days. Maximum likelihood and likelihood ratio tests and Benjamini and Hochberg correction were applied. Analysis methods included unsupervised hierarchical cluster analysis and ontology group testing using software programs recommended by the GO database. Day 0 and day 14 GE results clustered separately on a heatmap. Approximately one-third (14,277) of gene transcripts were significantly different after 14 days in the GC culture ($p_{adjusted} < 0.001$). Significant genes were ranked by t-statistic to identify the most up- and down- regulated genes. These genes grouped into multiple ontology classes as expected. The most up-regulated genes were involved in ontology groups, such as cell proliferation, DNA packaging and repair, apoptosis and cell survival, which mimicked those described by previous studies as being specific to the GC compartment *in vivo*. No significant differences were detected between the HR study group and the control groups when adjusted for multiple testing, but data which were unadjusted for multiple testing revealed different expression of 45 gene transcripts on day 14 and 198 gene transcripts on day 0 ($p_{raw} < 0.001$). The GE results suggest that the culture model faithfully mimics an *in vivo* GC response but some improvements could be fitted to the model. The HR group as a whole does not display a significantly different gene expression pattern (GEP) but an additional, later time point for sample collection might give insights or more significant results in the HR group.

ACKNOWLEDGEMENTS

Special thanks must go to my supervisor, Helga M. Ögmundsdóttir. Without her thought-provoking discussions, guidance, vast knowledge and support, I would not have been able to experience this exciting study abroad opportunity. I greatly appreciate her commitment to my success, quick email responses and good spirits throughout this study. She has not only been a great supervisor but also a friend.

I would like to thank my M.Sc. committee Hlíf Steingrímsdóttir and Magnús Karl Magnússon for many stimulating discussions and taking the time to contribute to this study. They served an influential role in the process of this study.

I would like to express my gratitude to the staff of Roche NimbleGen Iceland, Llc., especially Freyja Eiríksdóttir, Elsa Eysteinsdóttir, Elisabet Gudmundsdóttir, and Sigridur Valgeirsdóttir. Their advice, guidance, use of materials and the laboratory, and microarray processing are all greatly appreciated.

I would like to acknowledge the other M.Sc. students involved with this study, Sóley Valgeirsdóttir and Lóa Björk Óskarsdóttir. I appreciate the time, knowledge of laboratory protocols, and isolated samples which Sóley shared with me and the help of Lóa in completing ELISA measurements and for many interesting discussions.

Special thanks must go to Jóhannes Guðbrandsson. Without the extensive knowledge of computer programming and statistics of Jóhannes, I would have been lost. The statistical analysis of gene expression data was successful thanks to him. In conjunction, I would like to thank his supervisor Arnar Pálsson for encouraging his collaboration with this study.

I would like to express my appreciation to Kristín Bergsteinsdóttir for allowing me to share the laboratory and its equipment, as well as her advice with interpreting results.

I would like to thank Vilhelmína Haraldsdóttir for her participation in lively discussions and friendly encouragement during my time working on this project.

I would also like to express my deepest gratitude to the following people: my parents, Jerry L. and Rene Schiffhauer, for their love and support of my decision to move to Iceland; my boyfriend Finnur Atli Magnússon for his encouragement, the patience he had with me during the writing process of this thesis and for giving me the opportunity to experience Iceland´s beauty; my boyfriend´s parents and family for giving me a home and support system here in Iceland.

I would like to acknowledge the value of the participants who agreed to be in this study because without them none of this would be possible.

The work described in this thesis was performed at the Faculty of Medicine, University of Iceland, Reykjavík and Roche NimbleGen, Iceland Llc., Reykjavík.

This study was supported by funding from The Icelandic Research Fund (RANNIS) and Roche NimbleGen, Iceland Llc.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AID	Activation-Induced cytidine Deaminase
BcR	B-cell Receptor
BM	Bone Marrow
CC	Chemokine (C-C motif)
CHO	Chinese Hamster Ovarian
CLL	Chronic Lymphocytic Leukemia
CSR	Class Switch Recombination
CXC	Chemokine (C-X-C motif)
DC	Dendritic Cell
ELISA	Enzyme-Linked ImmunoSorbent Assay
FADD	Fas-Associated protein with Death Domain
FC	Fold Change
FISH	Fluorescence In-Situ Hybridization
FL	Follicular Lymphoma
FLC	Free Light Chain
GC	Germinal Center
GE	Gene Expression
GEP	Gene Expression Pattern
GO	Gene Ontology
HR	Hyper-responder
Ig	Immunoglobulin
IL	Interleukin
M-protein	Monoclonal Ig protein
MG	Monoclonal Gammopathy
MGUS	Monoclonal Gammopathy of Undetermined Significance
MGZ	Marginal Zone
MHC	Major Histocompatibility Complex
MM	Multiple Myeloma
MNZ	Mantle Zone
NHEJ	Non-Homologous End Joining
NHL	Non-Hodgkin’s Lymphoma
PB	Peripheral Blood
PBMN	Peripheral Blood MonoNuclear cells

PBS	Phosphate Buffered Saline
PC	Plasma Cell
PWM	Poke-Weed Mitogen
RT	Reverse Transcriptase
RIN	RNA Integrity Number
RMA	Robust Multichip Average
SD	Standard Deviation
SHM	Somatic HyperMutation
SMM	Smoldering Multiple Myeloma
TGF	Transforming Growth Factor
TF	Transcription Factor
TNF	Tumor Necrosis Factor
V(D)J	Variable (Diversity) Joining regions
WM	Waldenström's Macroglobulinaemia

1 INTRODUCTION

1.1 B cells; Development/function

1.1.1 Origin and function

B-lymphocytes (B cells) originate in the bone marrow (BM) from pluripotent hematopoietic stem cells. From those stem cells the common lymphoid progenitors, precursors to antigen-specific lymphocytes, are derived. They remain in the BM through many developmental processes before entering into the vascular and lymphoid system as mature naïve B cells. Typical development happens by interaction with BM stromal cells through cell adhesion molecules, interleukins (IL) and cytokines (Vangsted, Klausen, & Vogel, 2012). The first step in B cell development is the generation of antigen receptors which possess heavy (H) and light (L) chains. Each chain type contains a constant (C) region, variable (V) region with a joining (J) region and in the immunoglobulin heavy chain gene (*IgH*) a diversity (D) region. This requires a somatic recombination of these different segments, known as V(D)J recombination, that compose genes coding for the B cell receptor (BcR). This recombination allows for a vast number of antigen receptor possibilities. The *IgH* is produced first, followed by the immunoglobulin light chain gene (*IgL*) with rearrangement of the segments kappa (κ) and lambda (λ) (Figure 1). Allelic exclusion will result in only one of the two alleles being functional of the recombined *IgH* and *IgL*, which is all that is necessary, as reviewed by Maddaly et al (Maddaly et al., 2010).

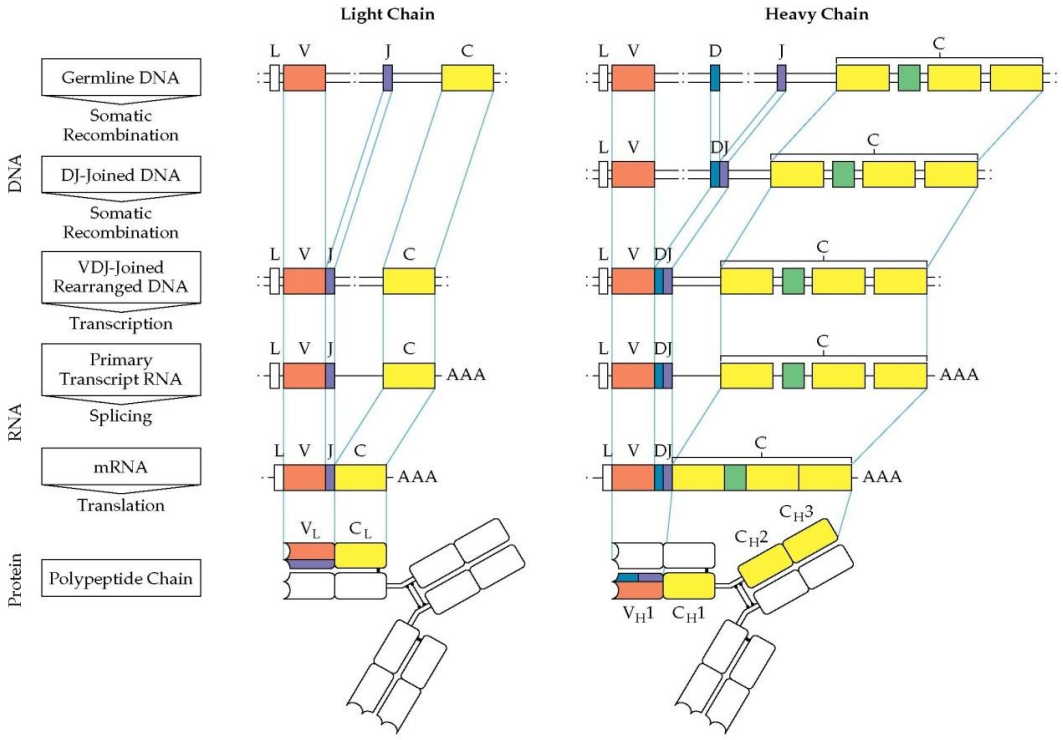


Figure 1: Immunoglobulin V-region gene segments.

Genes coding for the Ig V-regions are constructed from somatic recombination of gene segments. L-chain V-regions are composed of two gene segments—Variable (V) and Joining (J). H-chain V-regions are composed of three gene segments--Diversity, (J), and (V)) (Travers, Murphy, & Walport, 2007).

Successful gene recombination is important for progression from stage to stage during B cell development. Any cells which deviate from the normal, such as B cells that fail to receive signals from its antigen receptor or those which recognize self-antigens are driven into apoptosis and removed through negative selection. Once antigen receptors are developed and B cells express IgM and IgD on their surface, they are considered mature naïve B cells prepared to enter the vascular and lymphoid systems (Vangsted et al., 2012).

1.1.2 Maturation

In BM, antigen stimulation and B cell development are independent. Once developmental stages are finished and a mature naïve B cell is released from the BM and is carried to the peripheral lymphoid organs, which includes the spleen, lymph nodes and mucosal lymphoid tissues. Typically, a B cell will recirculate via blood and lymph and repetitively reenter the lymphoid tissues until it meets its corresponding antigen. Once B cells have left the BM, antigen exposure is obligatory for maturation but T lymphocytes (T cell) help is not always necessary. In the vascular or lymphoid system, once a B cell encounters the antigen which has been presented by dendritic cells (DC) and with the assistance of T cells, although this is not always necessary, that cell is activated. This activation occurs when the membrane-bound immunoglobulin´s (Ig) V-region bonds to the epitope of the antigen. Once activated, the density of chromatin in the nucleus decreases, nucleoli become visible, new mRNAs and proteins are produced and the cell is enlarged. Then some cells begin to proliferate rapidly in the splenic red pulp and switch from IgM membrane-bound to short-lived plasma cells (PC) which are produced for primary immune response and typically secrete IgM (Radbruch et al., 2006). During an immune response, some activated B cells migrate to form an area called the germinal center (GC) which is where dramatic proliferation, somatic hypermutation (SHM) and affinity maturation occurs. A rigorous selection process is in place to make certain that only B cells which express the highest affinity survive. Surviving cells will then go through *IgH* class switching recombination (CSR), a process that produces PCs which secrete antibodies of a different class (IgG, IgA, or IgE) with different effector functions yet still retain their antibody specificity (Travers et al., 2007). Signals (cytokines, co-stimulatory signals) received by the cells during class switching will determine the type of antibody secreted by the PC. Long-lived PCs will migrate to BM as a supply of longer-lasting antibody. Other surviving B cells will differentiate into memory B cells as long-lived cells which will allow for a quick response to repeated exposure to its specific antigen (M. G. McHeyzer-Williams & Ahmed, 1999). Figure 2 summarizes the pathway through B cell stages and possible differentiation results (Kuehl & Bergsagel, 2002).

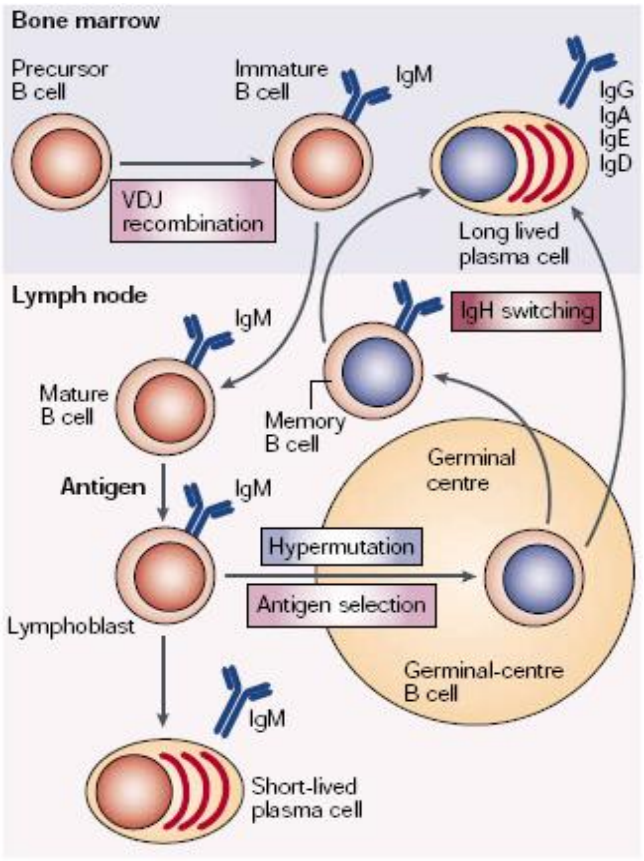


Figure 2: Stages of B cells.

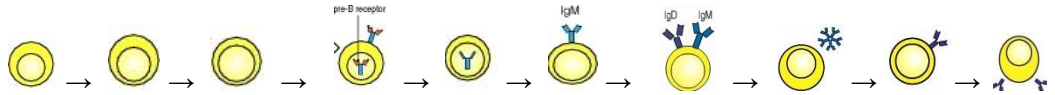
Stages of B cell development and the mechanisms involved. DNA remodeling mechanisms include V(D)J recombination, SHM, IgH CSR. Precursor B cells undergo V(D)J recombination to become immature naïve B cells that will exit the BM and home to peripheral lymphoid tissues. B cells interaction with its antigen produces proliferation and differentiation. Pre-GC short-lived PCs are produced for primary immune response and typically secrete IgM. Activated B cells that enter the GC undergo multiple rounds of SHM and antigen selection and differentiation into memory B cells or plasmablast. Plasmablasts then undergo CSR to become long-lived PCs that will reside in the BM (Kuehl & Bergsagel, 2002).

1.1.3 Developmental stages of B cells

Expression of surface markers, transcription factors (TF) and genetic events change throughout stages of development or maturation (Table 1). Many reviews have reported expression patterns for many of those stages (Klein & Dalla-Favera, 2008; Kuehl & Bergsagel, 2002; Perez-Andres et al., 2010; Vangsted et al., 2012). For example, pro-B cells typically express CD19+, CD10+ and CD34+, whereas pre-B cells express CD79A, CD19+, CD10+ and CD34-. Table 1 is not a complete list of CD molecules involved throughout development and maturation but includes those important to and exemplifies general development. Lymphoma classification is mainly based on different developmental stages defined by expression of surface markers and proteins as well as considering the genetic events that have occurred. For instance, a diagnosis of Mantle Cell Lymphoma (MCL) will occur at an earlier stage of maturation (naïve B cells) than Multiple Myeloma (MM) which involves malignant PCs (Klein & Dalla-Favera, 2008).

Table 1: Surface markers, Bcl-2, and Blimp expression along the B-lineage.

The different stages of B cell development and the typical surface marker expression and BCL-2 and BLIMP1 transcription factor expression (Travers et al., 2007).

	Stem cell	Common lymphoid progenitor	Pro-B cell	Large pre-B cell	Small pre-B cell	Immature B cell	Mature naïve B cell	Activated B cell	Memory cell	Plasma Cell
										
CD19										
CD20										
CD21										
CD27										
CD38										
CD40										
CD45R										
CD86										
CD95										
Bcl-2								???		
Blimp1										
Surface IgD										
Surface IgM										
MHC II										

Naïve B cells are B cells that have matured in the BM and now circulate through the vascular and peripheral lymphoid organs and tissues, such as the spleen, lymph nodes and mucosal lymphoid tissues, but have not been presented with their specific antigen. If never activated these cells die or a small number will make residence in the marginal zone of the spleen and are considered a separate subtype of B cells—marginal zone (MGZ) B cells. MGZ B cells are the first to respond to foreign antigen and differentiate into PCs. Naïve B cells express high levels of IgM and low levels of IgD. They also co-express important surface markers, such as CD19, CD20, CD21 and CD40 which allow activation when presented with their specific antigen. Of the total B cell population, the daily yield of new B cells from bone marrow is around 5-10% with a half-life of 3-8 weeks (Travers et al., 2007). Once activated, these B cells begin to express CD86, which is a ligand for CD28 that delivers stimulating signal to B cells and promotes cell survival (Podojil & Sanders, 2005). Typically, the switch from membrane-bound antibody expression to antibody secretion signifies the differentiation of B cells into plasmablasts which become short-lived PCs that will serve as an immediate response while the slower GC reaction produces other differentiated B cells (MacLennan et al., 2003).

Resulting short-lived PCs, which secrete antibodies, will help defend the body in an immune defense. These cells differentiate in the lymph nodes, rapidly produce antibodies and then die within a few days. Differentiation into PCs relies on the regulation of the transcription factor BLIMP1 (B-lymphocyte-induced maturation protein 1). In B cells where BLIMP1 is induced, PCs differentiation occurs, Ig secretion is increased and cell-surface properties change. BLIMP1 acts as a transcriptional repressor in B cells which turns off genes required for B cell proliferation in GC which allows for B cells to become PC, and for class switching and affinity maturation (Andrea Cerutti, 1998). PCs lose certain receptors which are no longer needed, for example—CD19, CD22, CD40, CD72 and Major Histocompatibility Complex (MHC) class II which assist in signal transduction in the development and assisting activation in B cells. However, CD38 is re-expressed on PCs and they also express CD27. PCs are considered terminally differentiated and are not able to divide as reviewed by Radbrunch et al (Radbruch et al., 2006).

Once the memory phase has begun, only about 10% of PCs have survived and can continue surviving for 3 weeks, months or even years. These cells are considered long-lived PCs and migrate to take residence in the BM, although a very small portion stays in the spleen. Chemokine (C-X-C motif) receptor CXCR4 expression is considered essential for migration to BM (Kunkel & Butcher, 2003). Long-lived PCs express the surface marker CD32 (an FcγII receptor—for a portion of the IgG and can convey an inhibitory signal to the B cell) which differentiates from known, short-lived PCs. B-cells need certain factors to develop into PCs, such as, presence of interleukin 21 (IL-21) and activation of *PRDM1* (PR domain containing 1) and *IRF4* (interferon regulatory factor 4) genes, which encode for BLIMP1 TF whose expression subsequently enhances XBP-1 (X-box binding protein 1) and suppresses PAX5 (paired box 5) TF. PAX5 is involved in B cell development; therefore it is no longer needed once cells have become PCs. XBP-1 then induces unfolded protein response which is needed for survival in the BM as reviewed by Radbrunch et al. (Radbruch et al., 2006) There are many other genes, IL, TFs, cytokines and chemokines important to cell survival which allows for a complex PC development process—Chemokine ligand 12 (CXCL12), IL-5, IL-6, TGF-β (transforming growth

factor-β) and tumor necrosis factor (TNF) family members to list a few (Shaffer et al., 2004). There are about 1 x 10^9 PCs in the BM (0.1-1% of all cells) under typical circumstances (Terstappen, Z. Hollander, & Loken, 1990). There have been reports suggesting that the population size of PCs seems to be regulated by expression of inhibitory Fc-receptor, FCγRIIB (CD32) on PCs (Xiang et al., 2007).

Memory B cells are a long-lived form of B cells that can differentiate into PCs. After an immune response some B cells differentiate into memory cells. Unlike PCs, these cells do not secrete antibodies but instead express cell-surface antibodies and react quicker to antigen exposure than PCs. Forty percent of B cells in the peripheral blood (PB) in humans are CD27+ memory B cells. Memory B cells have already gone through the GC reaction which means they have had genetic changes, such as SHM and gene rearrangements which result in class switching. For this reason, memory B cells secrete mostly IgG and IgA. Memory B cells can be found in secondary lymphoid tissues after leaving the GC reaction as reviewed by McHeyzer-Williams et al (M. G. McHeyzer-Williams & Ahmed, 1999).

In addition to those mentioned previously, there are two more marker proteins of interest to this project—CD95 and Bcl-2 (B-cell Lymphoma 2).

Expression of CD95, also known as Fas, induces apoptosis and is up-regulated not long after activation. It is a member of the TNF family and is also known to bind TNF-like Fas ligand. Stimulation of Fas results in death-inducing signaling complex formation (Fas-associated protein with Death Domain) FADD, with an adaptor protein (Croker et al). After the Fas receptor trimerizes then FADD is recruited and a cluster of death domains (present on both Fas and FADD) cluster and a caspase-cascade is activated inside the cell. This eventually results in the breakdown of DNA inside the cell nucleus and cell death occurs (Croker et al., 2011; Klein & Dalla-Favera, 2008).

Bcl-2 is an anti-apoptotic mitochondrial protein and belongs to a family of proteins that block the release of cytochrome c, a known trigger of caspase that leads to apoptosis. This family of proteins contains both pro- and anti- apoptotic which help regulate apoptosis. This is important family of proteins because many B cells which go through a stringent affinity selection eventually undergo apoptosis which Bcl-2 is one distinctive pathway regulates this. The translocation between chromosome 14 at the IgH locus and chromosome 18 at the *BCL-2* locus is well known in Follicular lymphoma. This translocation leads to overexpression of Bcl-2 and allows for resistance to apoptosis, which is fundamental for progressing tumors. Loss of expression can bring forth degenerative disorders. Bcl-2 not only is involved in programmed-cell death but also cytotoxic stress-induced apoptosis (Coultas & Strasser, 2003).

1.1.4 Antibody secretion and function

Antibodies, also known as Ig, are the secreted form of the B cell´s antigen receptor. PCs, which mediate the humoral immune response, secrete antibodies. Their structure consists of two H-chains and two L-chains which are composed of a C-region and a V-region. The heavy polypeptide chains are IgG (gamma, γ), IgA (alpha, α), IgM (mu, μ), IgD (delta, δ), and IgE (epsilon, ε). The two types of L-chains are κ and λ. It is this V-region that contains the differences between antibodies due to

hypermutation of genetic segments and the large number of V-genes. Antibodies have three main functions described (IMWG, 2003; Maddaly et al., 2010; Travers et al., 2007).

1. Neutralization: Antibodies bind to foreign molecules rendering them unable to invade and damage host cells.

2. Opsonization: phagocytes then bind to the C-region of antibodies on neutralized pathogens and phagocytosis occurs.

3. Complement activation: coats pathogen with complement fragments then phagocytosis increases and lysis of bacteria occurs.

There are five major forms of an antibody which determines the class and the functional properties. The five classes of Ig are IgM, IgD, IgG, IgE and IgA (Figure 3). While all antibodies will be briefly described only IgM, IgG and IgA will be of interest to this study.

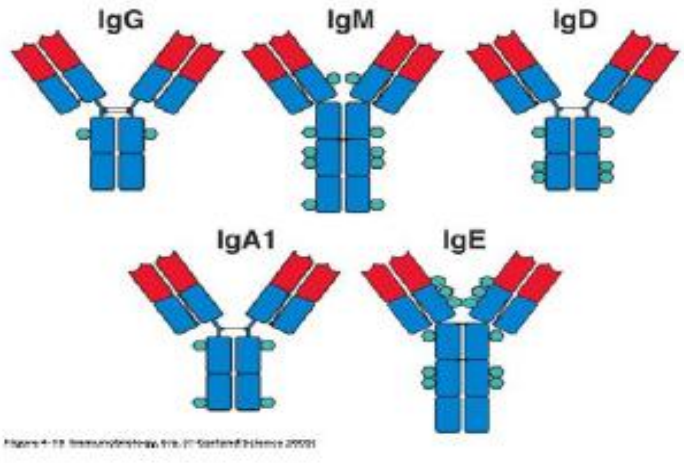


Figure 3: Main immunoglobulin isotypes.

Five main immunoglobulin isotypes include. Black lines indicate disulfide bonds and green hexagons indicate N-linked carbohydrate groups. Distribution differs between isotypes. IgM and IgE do not contain a hinge region (Travers et al., 2007).

IgM is mainly expressed in transmembrane form on the surface of naive B cells before maturation occurs in the GC and is always the initial antibody secreted in immune responses. It accounts for less than 10% of Ig found in plasma and is the heaviest of the Igs because it is found in pentameric form in serum. This gives a greater avidity and compensates for its low affinity. IgD production levels are consistently low and are overshadowed by dominant levels of IgM, with which it is almost always co-expressed, in the early stage of immune responses. Monomeric IgG is the most abundant during an immune response, accounting for approximately 75% of antibody isotopes in humans, and is readily found in the blood and extracellular fluid (Andraud et al., 2012). It is efficient in opsonizing pathogens and operates foremost in the body tissues. IgE can be found in blood or extracellular fluid but at very low levels because its residence is on epithelial tissues. Its major function is to trigger allergic reactions defense against parasites. IgA is mostly found in secretions and is mostly located on respiratory and intestinal epithelial surfaces. It participates in the first function of antibody response—neutralization (Andraud et al., 2012; Travers et al., 2007).

1.1.5 Germinal center

GCs are specialized microenvironments found in lymphoid follicles, within peripheral lymphoid organs, and are comprised of a large number of B cells undergoing proliferation and differentiation into memory or PCs (Figure 4)(Klein & Dalla-Favera, 2008). GC´s are present in the follicles of peripheral lymphoid organs until 3-4 weeks after the initial antigen encounter, peaking around the second week of response (Travers et al., 2007).

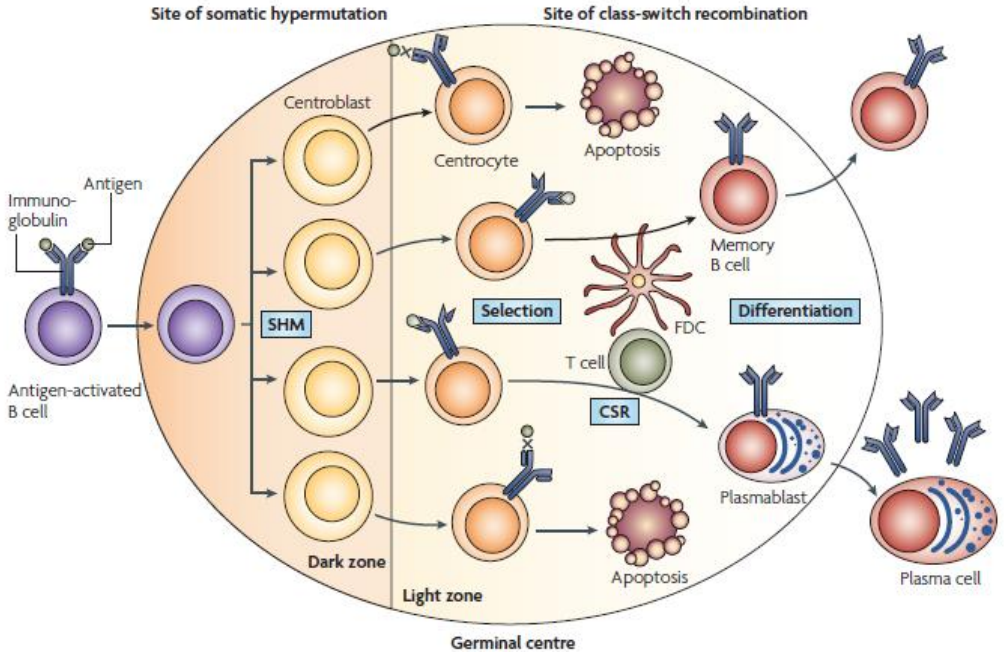


Figure 4: Germinal center reaction.

Antigen-activated B cells differentiate into centroblasts and then go through clonal expansion in the GC dark zone. During proliferation, multiple rounds of SHM introduce base-pair changes into the VDJ sequences on both the H and L-chain V-regions (Klein & Dalla-Favera, 2008)

Once antigen-activated B cells have formed into a GC, they go through SHM which targets the V-region of Ig to become proliferating centroblasts located in the dark zone (Vinuesa, Linterman, Goodnow, & Randall, 2010). After a few days involving fervent proliferation, distinctive light and dark zones become apparent in the GC structure. It is estimated that centroblast complete cell cycle in 6-12 hours (Klein & Dalla-Favera, 2008). Since centroblast cell cycle is short, B cell population can expand greatly. These centroblasts migrate to the light zone, which also consists of a network of macrophages to become centrocytes, T cells, and follicular dendritic cells (FDC), to go through a rigorous selection process to make sure only the B cells with highest affinity survive. Since this process is meticulous and centrocyte cell cycle allows for quick expansion, many cells undergo apoptosis which is accounted for by the lack of anti-apoptotic factors on centrocytes. Therefore the many genes and TFs involved in apoptotic response are important to GC B cell development. The selection process entails expression of *BCL-6 (*B-cell Lymphoma 6), TF PAX5, CD40 ligand, ILs, and adhesion molecules (Vangsted et al., 2012). This process terminates self-reactive or decreases signaling of mutated BcR (Vinuesa et al.,

2010). Surviving centrocytes then undergo CSR. The mechanism behind class switching involves Activation-Induced cytidine Deaminase (AID), which is activated by PAX5, to start a DNA recombination that, along with stimulation from activated T cells, CD40 and its CD154 ligand, and secreted cytokines, leads to class switching (Greta Meyers, 2011; Maddaly et al., 2010). After CSR, cells are considered plasmablasts and will differentiate into either PCs or memory B cells. Development of plasmablasts into PCs, both short- and long-lived depend on many genes (*PRDM1*, *PAX5*, *IRF4*, *BCL-6*), ILs, (IL-21, -5, -6, and -4), TFs (BLIMP1, PAX5, IRF4), cytokines, chemokines and their receptors (CCR7, CXCL12, CXCR4, and CXCR5), to name a few as reviewed by Radbrunch et al (Radbruch et al., 2006). Development into memory B cells requires help from activated T cells and it has been suggested that STAT5 mediates differentiation, as well as the inactivation of PAX5 which is thought to be an important event for differentiation into memory B cells rather than PCs (Klein and Dalla 2008). Although, it is still unclear and controversial as to what specifically causes some GC B cells to differentiate into memory B cells or PCs (Vinuesa et al., 2010). The large number of somatic events can result in aberrant SHM and CSR that lead to damaging DNA, such as DNA strand breaks, which is repaired by DNA repair mechanisms, such as non-homologous end joining (NHEJ) or DNA repair enzymes (Klein & Dalla-Favera, 2008; Vangsted et al., 2012). Repair mechanisms, such as NHEJ, are known to be error-prone, which can introduce mutations and chromosomal translocations to the genome of these GC B cells (Klein & Dalla-Favera, 2008) (Figure 5). Often candidate genes for risk of B cell malignancies and lymphoproliferative disorders are those included in apoptosis, DNA repair, cell cycle, and immune regulation (Liang et al., 2009; Roddam et al., 2010).

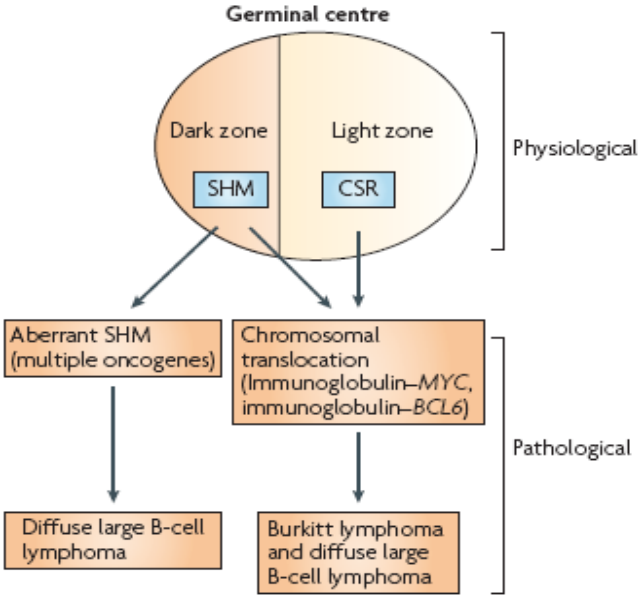


Figure 5: DNA mutations in GC B cells and resulting consequences.

Errors in SHM and CSR processes acting at the *IgH* or *IgL* loci or at the 5´ regulatory regions of various non-Ig genes can result in the creation of free DNA ends that cause chromosomal translocations. Chromosomal translocations can result in dysregulated expression and promoting tumorigenesis (Klein & Dalla-Favera, 2008).

1.2 Immunoglobulin-producing neoplasias

1.2.1 B cell derived tumor development and malignant transformation

Tissue-resident lymphomas can result from a monoclonal neoplastic transformation of B cells. The stage of development which the tumor derives from dictates the characteristics of different lymphoid tumors. Tumors have been found in humans that correspond to all different stages of B cell development (Figure 6) (Travers et al., 2007). Unless derived from early undifferentiated cells, lymphoid tumors feature gene rearrangements that often lead to chromosomal translocations. Mutations that arise during gene rearrangements disturb the controls a normal cell has to prevent growth of these aberrant cells, such as apoptosis. This disruption of normal cellular controls can allow for expansion of tumor cells. For example, a chromosomal translocation to the oncogene *BCL-2*, which is found to prevent apoptosis in B-lineage cells, results in a Bcl-2 protein production increase which is associated with Follicular Lymphoma (FL) development. This aberrant over-expression of Bcl-2, leads to the growth in numbers of B cells that live beyond their normal life-span and promotion of tumorigenesis (Coultas & Strasser, 2003). There are many other examples of mutations and complex disruptions to the normal cell cycle, apoptosis, and DNA repair which contribute to the pathogenesis of different B cell malignancy but the hallmarks remain the same (Hanahan & Weinberg, 2011).

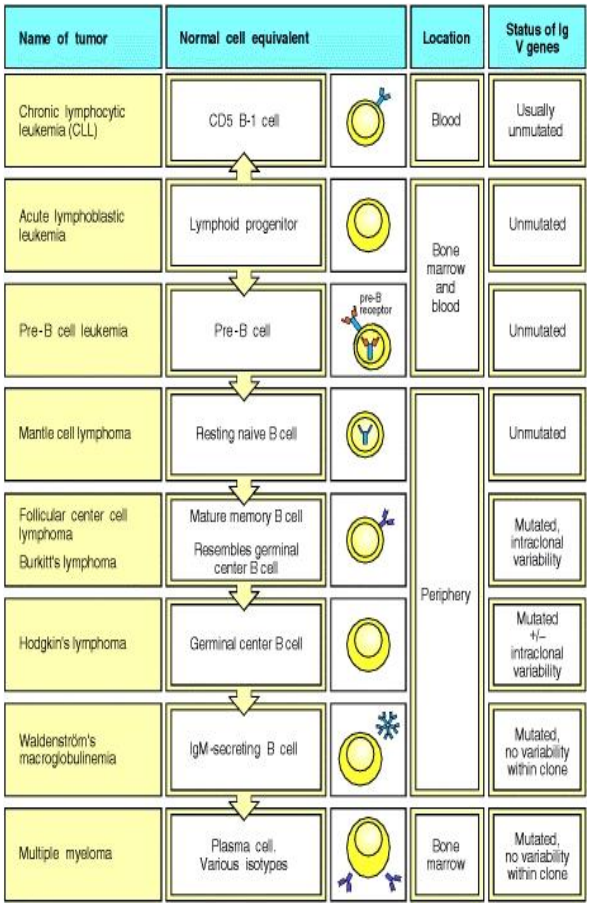


Figure 6: B cell lineage and corresponding tumor development.

Each type of tumor cell has a corresponding normal type B cell counterpart. They behave similarly and reside in similar locations. The type of tumor, normal cell equivalent, location and status of Ig V genes are all indicated (Travers et al., 2007).

1.2.2 Paraproteinemia

Paraproteinemias, also referred to as monoclonal gammopathies (MG), are a group of disorders associated with a benign or malignant expansion of a single clone (monoclonal) of PCs that secrete a monoclonal Ig protein (M-protein, also known as M component or paraprotein) (IMWG, 2003). Most MGs arise via B cell transformation within the GC reaction or post-GC (Roddam et al., 2010). The type of M-protein (IgM, IgA, IgG) depends on the stage of maturation that is affected by the aberrant clone expansion, whether it is an Ig-secreting mature B-cell or a clone of PCs which have already undergone class switching (Helga M. Ögmundsdóttir & Guðríður Ólafsdóttir, 2002; Kyle & Rajkumar, 2010). The most frequent type is IgG, followed by IgM then IgA. MGs prevalence increases with age. Incidence rates of MG are approximately 3.2% in those around 50 years of age, increasing to 7.5% by the age of 70 (IMWG, 2003). There are no major risk factors but chronic antigen stimulation and both genetic and environmental factors have all have been linked to the group of disorders (McMaster & Caporaso, 2007; Ogmundsdottir, Steingrimsdottir, & Haraldsdottir, 2011).

1.2.3 MGUS and SMM

Monoclonal Gammopathy of Undetermined Significance (MGUS) is a premalignant disorder classified by the presence of a serum M-protein at less than 30 g/L, less than 10% proportion of PC infiltration in the BM, and no presence of symptoms typically associated with PC proliferative disorders (bone lesions, hypercalcemia, anemia, and renal failure). Biologically similar to MGUS is smoldering multiple myeloma (SMM), which is asymptomatic like MGUS but the levels of serum M-protein and PC infiltration are larger than the accepted values for MGUS and the risk for malignant transformation is 10-20% per year. MGUS occurs in approximately 3% of the general population of individuals 50 years or older with the rate increasing with age to about 5% in those 70+ years (Kyle & Rajkumar, 2010). MGUS is known to progress to other B cell malignancies, such as Waldenström's Macroglobulinemia (WM) and more commonly MM. A widely accepted prevalence rate of MGUS to MM is around 1% by the age of 50 and increasing to 3% by the age of 70 (IMWG, 2003). However, since MGUS is asymptomatic, it is not always detected before progression to MM has occurred; therefore statistics can be skewed (Ola Landgren, 2009). Possible risk factors for progression from MGUS to MM include a higher level of M-protein (≥15 g/L) and proportion of PC infiltration, family history, age, race, gender and environmental exposures. It is not possible to know if a MGUS patient will progress to malignancy or not (Kyle & Rajkumar, 2010). IgA and IgM type M-protein has also been shown to indicate a higher risk for progression (Helga M. Ögmundsdóttir & Guðríður Ólafsdóttir, 2002; IMWG, 2003; Kyle & Rajkumar, 2010). IgG and IgA MGUS are associated with MM progression, whereas IgM MGUS typically progresses to WM (McMaster & Caporaso, 2007). Cytogenetic data, Ig gene mutational analysis, and chromosomal abnormalities have begun to help create a picture of the progression in several MGs. Fluorescent in situ hybridization (FISH) analysis screens for translocations, deletions and amplification have revealed the presence of translocations at early stages in the disease pathogenesis including MGUS. The most common translocations involve the *IgH* locus on chromosome 14 include: t(4;14), t(11;14)(q13;q32) and t(14;16) and are also common in MM (Chng et al., 2011; Davies et al., 2003; Kyle & Rajkumar, 2010). Evidence of the involvement of genetic factors

has allowed gene expression (GE) and modern cytogenetic tools to be a popular topic of many recent studies concerning MGUS and other related disorders.

1.2.4 Malignancies associated with paraproteinemias: MM and WM

MM is a malignancy characterized by neoplastic proliferation of typically M-protein secreting PCs in the BM. MM accounts for 10% of all hematologic malignancies (John D. Shaughnessy Jr, 2003). MM age-standardized incidence rates for the world are estimated to be around 4.9 per 100,000 for males and 3.2 per 100,000 for females. Iceland´s incidence rates are estimated to be similar to world values (*NORDCAN*, 2011). IgG is found to be the M-protein in around 50% of patients and IgA and FLC each accounts for one-fifth of patients. MM is an incurable malignancy and is characterized by symptoms such as anemia, osteolytic bone lesions, hypercalcemia, and renal failure. Other symptoms relating to profound immunosuppression such as repeated infections can occur as well. Malignant cells survive and expand in the BM and are dependent on the BM microenvironment and the production of different survival factors by the BM stromal cells. The malignant PC can influence the osteoclasts in the bone leading to osteolytic bone lesions that often causes severe bone pain and pathological bone fractures (IMWG, 2003). It is only evidence of symptoms which is necessary for diagnosis of MM because M-protein levels and PC infiltration in BM can vary. MM is associated with IL-6 dependence and alteration in adhesion molecules which makes sense because an increased expression of IL-6 is associated with PC survival within the BM (Davies et al., 2003; Kuehl & Bergsagel, 2002; Richard J. Armitage, 1993).

MM cells can trigger many protective effects. The large variety and quantity of anti-apoptotic signaling mechanisms involved with this disease mediates chemotherapy resistance and eventually leads to a fatal result for most patients (Zhan, 2002). When MM cells bind to BM accessory cells, cytokine secretion is activated and encourages MM cell growth, survival, and migration, as well as chemotherapy resistance (Hideshima, Mitsiades, Tonon, Richardson, & Anderson, 2007). Many MM studies show evidence of complex genetic and cytogenetic changes, which makes the malignancy more difficult to understand (Zhan, 2002). However, modern genetic analysis tools are providing more information about MM. Genetic data combined with well-known clinical data has provided support for a multi-step transformation model from MGUS, a premalignant phase, to MM (Figure 7). For example, a few studies have shown that MGUS has similar gene expression patterns (GEP) to MM, whereas PCs from both phases of MG show a significantly different GEP from normal PCs as seen in Figure 8 (Chng et al., 2011; Davies et al., 2003). Zhan et al. showed in 2007 that within MM PCs there was clustering into subgroups with one group closer to MGUS GEP despite the fairly small differences between MM and MGUS GEPs which further supports a step-wise transformation.

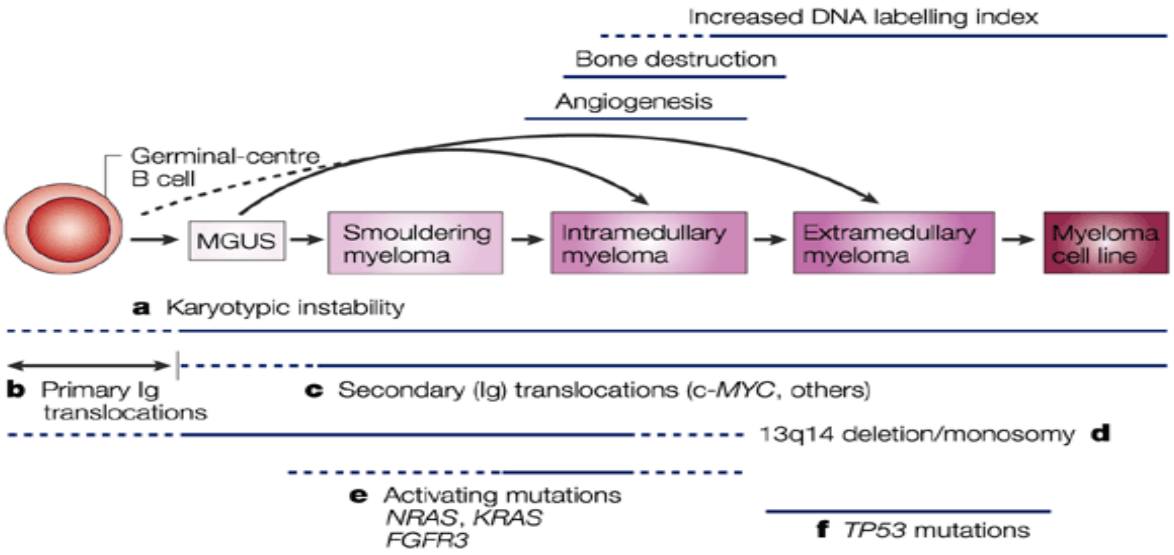


Figure 7: Stages of MM.

MM arises from a normal GC B cell. MM can form from MGUS, the benign PC neoplasm, into SMM and eventually malignancy, but it does not always pass through every phase (Kuehl & Bergsagel, 2002)

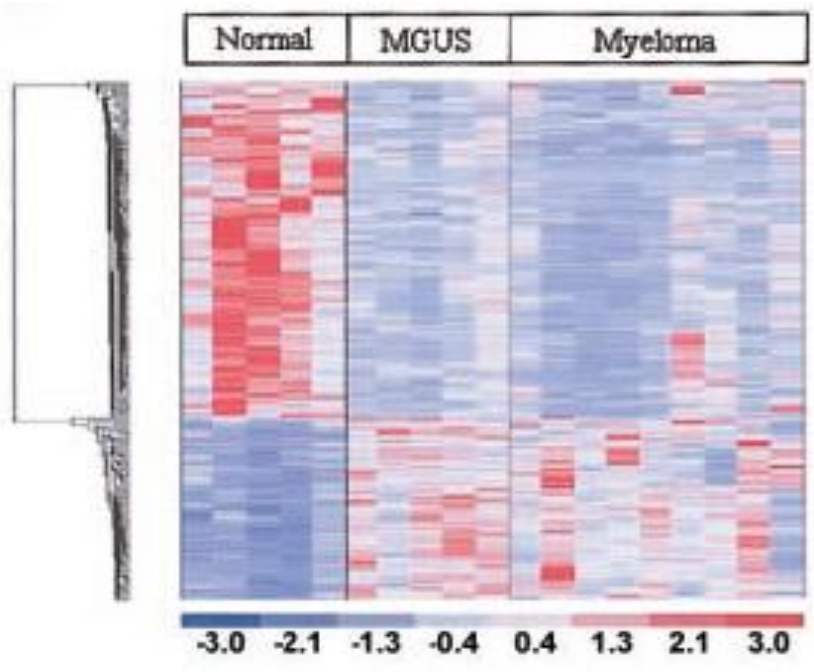


Figure 8: GEP of Normal to MGUS and MM.

A comparison of GEPs between normal PCs and PCs from MGUS and MM patients. Two hundred sixty-three genes were identified and most were down-regulated (Davies et al., 2003).

Within these studies, the most commonly documented significantly different genes involved numerous clusters of gene functions. Ig molecules and MHC genes were down-regulated in malignant PCs rather than normal which would be expected because MG PCs are monoclonal. Other categories included growth-related genes, genes associated with signaling cascades, cell-cycle and proliferation related genes, MYC activation gene sets, DNA synthesis, apoptosis, and translation machinery genes (Chng et al., 2011; Zhan et al., 2007). The most commonly documented chromosomal translocations in MM involve the *IgH* locus on chromosome 14-- t(4;14), t(11;14) and t(14:16) but other chromosomal aberrations have also been associated with the disease. Translocations have been associated with oncogenes that affect cell homeostasis such as—*CCND1* (cyclin D1), *CCND3* (cyclin D3) *FGFR3* (fibroblast growth factor receptor 3), *IRF4*, *MAF* (v-maf *musculoaponeurotic fibrosarcoma oncogene homolog), MAFB*, *MMSET* (Multiple myeloma SET domain-containing protein) *(Htwe et al., 2011; Zhan et al., 2007)*. These translocations are also commonly documented in MGUS and occur at similar frequency in both meaning that the translocations are present before progression to MM (Kuehl & Bergsagel, 2002). This not only supports MGUS as a premalignant phase but also, suggests that class switching events probably take place early in the process of the disease (Davies et al., 2003).

WM is a B-cell malignancy, a type of Non-Hodgkin lymphoma (NHL), which is distinctive by BM infiltration of lymphoplasmacytic cells with an aberrant hypersecretion of monoclonal IgM (McMaster & Caporaso, 2007). WM is asymptomatic in many patients; however, those that do experience symptoms typically develop anemia, aberrant enlargement of the spleen, neuropathy, bleeding in the oronasal areas and other symptoms related to increased viscosity of the blood (Royer et al., 2010). Bone pain is not typically associated.

WM etiology is not clear but there is growing evidence based on genetic and cytogenetic technology. BM masts cells of WM over express CD40 ligand (CD40L), an inducer of B cell expansion, and lack CD23 expression (Vijay & Gertz, 2007). It has been shown that WM is a similar malignancy to MM and chronic lymphocytic leukemia (CLL) in many ways such as B cell origin, development at the later stages of B-cell maturation and BM infiltration, but it has its own distinctive features (McMaster & Caporaso, 2007; Royer et al., 2010). One study showed that the GEP of WM was more similar to CLL than to MM (Chng et al., 2006). WM is different from other related disorders in that it does not share many of the typical translocations on the IgH locus on chromosome 14. A deletion on chromosome 6 (6q21-22.1) is considered its most common chromosomal abnormality although there have been reports of a deletion on the chromosome 20 long arm; however the number of cases were not large (Liu, Miyazawa, Sashida, Kodama, & Ohyashiki, 2006; Mansoor et al., 2001). Vijay et al. also states that *BLIMP-1* is a tumor suppressor gene localized to 6q21 and partial or whole loss of this gene could result in predisposition for WM and other lymphoproliferative disorders (Vijay & Gertz, 2007).

WM is considered rare, with U.S.A incidence rates estimated to be around 3-5 cases per million persons per year, which accounts for about 1% to 2% of hematological malignancies with median survival around 5 years (Ansell et al., 2010). Rates are difficult to estimate because WM was not recorded in the Icelandic cancer registry (ICR) until 1990 (Ögmundsdóttir et al., 2005). Majority of those affected are males and the median age of this malignancy is 65 years, which is similar to that of other lymphoproliferative disorders such as MGUS or MM (Ansell et al., 2010; McMaster & Caporaso,

2007). IgM-MGUS and relation to patients diagnosed with WM are the main risk factors for developing WM (Royer et al., 2010; Vijay & Gertz, 2007)

1.2.5 Detection of paraproteinemia

Currently agarose gel electrophoresis is the most used method for confirming the presence of M-protein in both urine and serum and determining its H-chain and L-chain type. It is sensitive, detecting a serum M-protein of 0.2 g/L and a urine M-protein of 0.04 g/L. When MM, WM and other B cell lymphoproliferative disorders are suspected, it is recommended that both serum and urine be tested. Immunofixation is a more sensitive method detecting even smaller amount of paraprotein and is applied when the paraprotein is not measurable by protein electrophoreses. Serum free light chain (FLC) analysis is also performed to detect FLC of κ or λ type in the serum. That is especially important in cases where the malignant cell only produces the L-chain of the Ig instead of both chains (IMWG, 2003; Kyle & Rajkumar, 2010).

1.3 Familial patterns of Monoclonal Gammopathies--Survey of literature

1.3.1 Family studies

Risk factors associated with MG can vary between disorders; however, an accumulation of studies from many different countries and over decades have reported a familial predisposition in MG which has sparked interest in studying hereditary factors (Kyle & Rajkumar, 2010; Ola Landgren, 2009; Vijay & Gertz, 2007). It is estimated, based on accumulation of published information, that around 130 families contain multiple cases of MM and/or MGUS and WM. Since families most often cohabitate in the same or similar environments and lifestyles, environmental and occupational exposures cannot be ruled out as a reason for these familial observations (Kyle & Rajkumar, 2010; Royer et al., 2010; Steingrimsdottir, Einarsdottir, Haraldsdottir, & Ogmundsdottir, 2011). A Swedish study assessed risk for first-degree relatives and concluded that these relatives of IgA/IgG MGUS patients had a 20-fold elevated risk of developing MGUS, MM and WM. IgM MGUS patient's first-degree relatives had 5.0-fold increased risk for CLL and no significant MM and WM risks. Studying these familial clusters of MG can prove to be a helpful approach for understanding more about the diseases since the pathogenesis of MGs are still weakly understood. Modern technology, such as genetic analysis tools, have allowed a closer look at chromosomal abnormalities and identifying candidate genes to help understand pathogenesis (Ola Landgren, 2009).

1.3.2 Icelandic family studies—previously published

The first Icelandic family containing multiple cases of MG and other lymphoproliferative disorders was described in 1978. At that time, this family (referred to as Family 8) included three cases of MGUS, and one case each of MM and WM. In this study, 45 descendants and 19 spouses were screened and revealed seven individuals with elevated IgM levels (Bjornsson et al., 1978). Since then, several studies have been published in Iceland. In an article published in 1994 by H.M Ögmundsdóttir et al, 35 family members of Family 8 were screened and nine disease-free family members found with

elevated Ig production after *in vitro* Poke-Weed Mitogen (PWM) stimulation assay. Those individuals that showed Ig concentration values ≥2 standard deviations (SD) above normal controls at the peak of immune response (days 7-8) in the *in vitro* PWM were termed Hyper-Responders (HR) due to their B-cell hyperactivity (Ogmundsdottir et al., 1994). Later in 2004, H.M. Ögmundsdóttir published a population-based Iceland Cancer Registry (ICR) study which identified 218 cases of MM between 1955 and 1989, and determined that female relatives of patients were at an elevated risk of developing MM. Eight families were identified within Iceland that contained ≥ 2 cases of MM and/or MM and ≥ 1 of another hematologic malignancy (Ögmundsdóttir et al., 2005). Those eight families were traced back seven generations to ensure they were eight separate pedigrees. Since then a second screening has identified seven new HR (Figure 9) (Steingrimsdottir et al., 2011). Recently, Helga Einarsdóttir and Sóley Valgeirsdóttir used these HRs and normal controls to test a new *in vitro* GC model designed to mimic *in vivo* observations more closely than the PWM model as part of their research. In the *in vitro* GC model, isolated B cells were seeded with CD40L expressed on transfected Chinese-Ovarian hamster (CHO) cells and IL-4, to induce B cell expansion and stimulation. The cells were cultured in RPMI 1640+10%-ultra low IgG FCS medium as described further within materials and methods. IL-4 was the only IL used because it is not produced endogenously. B cells were transferred to fresh CHO cells every 7 days for 21 days. Results of a surface markers expression analysis supported *in vivo* observations within this model. This study uses samples from before stimulation and after 14 days of stimulation within this *in vitro* GC model.

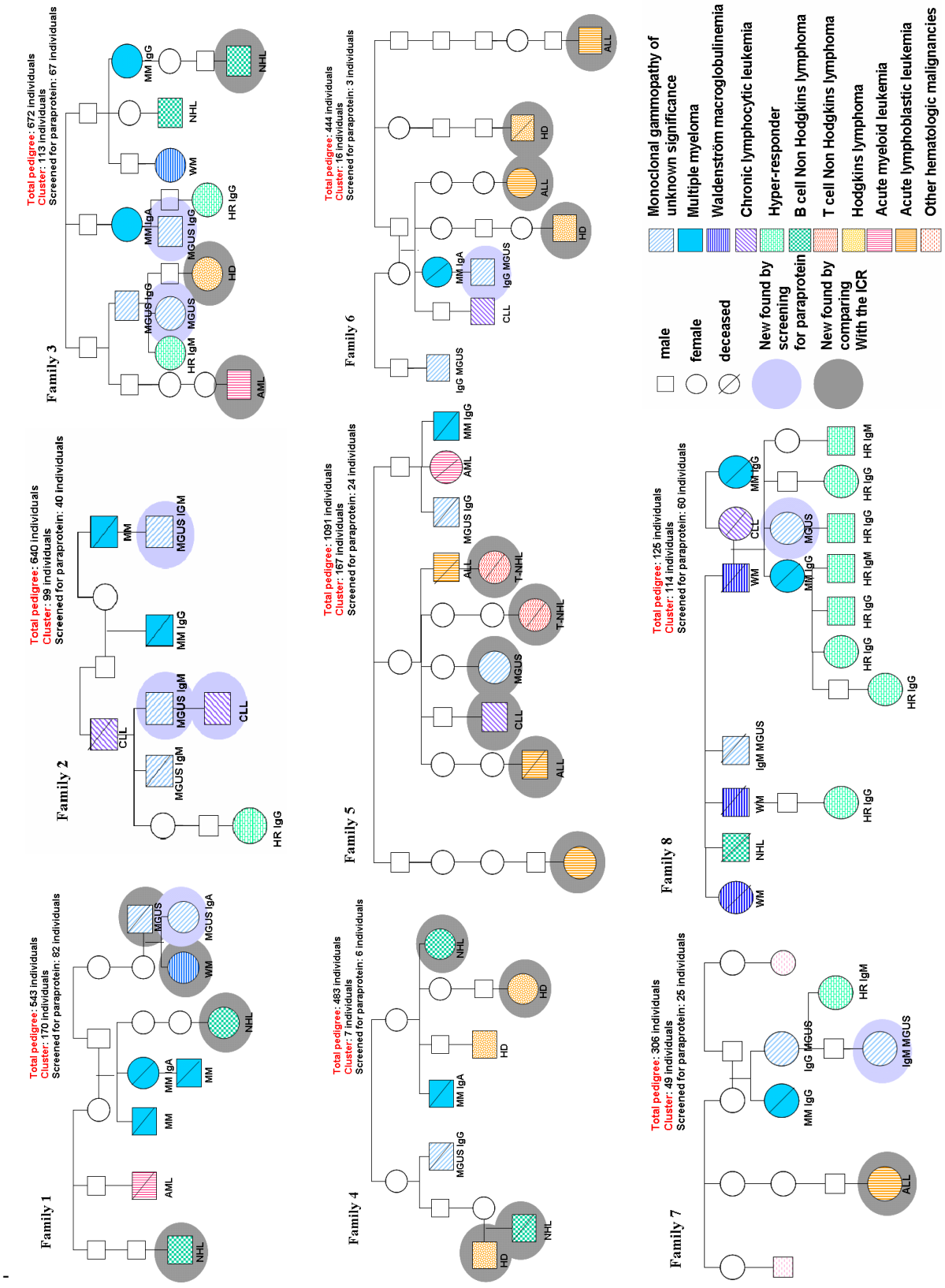


Figure 9: Icelandic MG families.

Eight Icelandic MG family pedigrees with MGUS, MM, or WM and another case of MM, MGUS or > 1 case of an additional hematologic malignancy. ICR, Icelandic Cancer Registry (Steingrimsdottir et al., 2011).

1.3.3 Hyper-responder as an endophenotype

Endophenotype can be defined as an inheritable condition within a population that is linked with illness manifesting in a person whether the illness is active or not. It is typically found in unaffected family members at a higher rate than the general population (Hasler, Drevets, Gould, Gottesman, & Manji, 2006). A well-documented example is the abnormal glucose tolerance test prior to manifestation of clinical diabetes (Kendler & Neale, 2010). Typically, endophenotypes, which sometimes need to be brought forth by a challenge, have been established as a helpful approach to studying complex diseases (Ogmundsdottir et al., 2011). The previously defined HRs seem to fulfill the definition of an endophenotype. Further investigation of HRs compared to controls would be needed to support that type of statement, which is done within this study.

1.4 Gene Expression & Microarrays

Microarray technology has become a popular and powerful tool for investigating molecular biology, and behavior of diseases, ever since it became evident that abnormal GE is a key characteristic of most cancers, if not all (John D. Shaughnessy Jr, 2003). In 1995, Schena et al., developed microarray plates which were processed by chemical and heat treatments to attach cloned DNA sequence fragments to glass surface and denature them. Then fluorescent dye was used to label the experiment *Arabidopsis* cDNA, Cy-3, and a different fluorescent dye was used to label normal *Arabidopsis* cDNA, Cy-5. Samples were then scanned with a laser and signals that saturated the scan detector were recorded and the fluorescence ratio was used as a quantitative measurement of expression levels. At this time, this version of microarrays could examine the relative expression levels for thousands of *Arabidopsis* genes at the same time (John D. Shaughnessy Jr, 2003; Schena, 1995). In sixteen years, the microarray has advanced greatly. Many organism transcriptomes, a complete set of transcripts which covers GE on a genome-wide scale, have become well documented. The length of oligonucleotides has increased which has reduced the need for high-redundancy probe design to ensure capture of target sequences (Mo et al., 2006). Different types of microarrays slides have formed which contain multiple arrays per slide and can use either one-color or two-color labeling and hybridization.

Companies have patented their own techniques for making microarray slides which differ slightly but the concept is the same. The slides used in this project were made using a Maskless Array Synthesizer (MAS) technology to place oligonucleotide arrays on glass slides. This method is possible using computer-controlled aluminum micro-mirrors (Digital Micromirror Device) to focus a laser beam on a specific location (spot) on the array. This laser beam of UV-light cleaves the protecting group at the location where the next nucleotide is to be coupled. A solution containing the nucleotide is added and the specific nucleotide adds on the deprotected chain to build long oligonucleotide sequences at high density that will target a specific portion of a gene transcript within the labeled cDNA samples hybridized to the slide. Any unattached nucleotides from the solution are washed off before the next cycle (Gibson & Muse, 2009; Roche NimbleGen, 2011). cDNA is synthesized from mRNA or tRNA and labeled with either one-color, Cy3, or two-color fluorescent dyes. Hybridization of these samples onto

the slide is done within special humidified chambers under strict conditions in order to minimize cross-hybridization between analogous genes. Slides are then washed and scanned with fluorescence imaging in order to visualize expression levels of targets on the array. This project used one-color labeled cDNA onto high-density oligonucleotide microarray slides that contain twelve arrays per slide(Roche NimbleGen, Iceland Llc.) and use approximately 135,000 probes that cover 45,000 gene transcripts which can record expression levels for comparison between treatment groups (Gibson & Muse, 2009; John D. Shaughnessy Jr, 2003; Tokuzo Arao, 2011). Image processing, data normalization and mining, and statistical testing and analysis are then needed in order to form information on expression levels and significance of differently expressed genes between treatment groups. There are many methods depending on the design of the experiment, therefore for a more detailed description of the methods used in this project, please refer to materials and methods within this report.

There are many useful applications for microarray technology with the most common being detection of candidate genes, annotation of gene function and definition of genetic pathways. This projects main goal was detection of candidate genes and is based on the deduction that genes which are transcribed in one group of samples but not another, are vital or contribute to any biological differences between groups. This type of application is helpful for studies which want to associate genes in diseases or aberrant immune responses(Gibson & Muse, 2009). cDNA libraries/databases have grown to include information on biological and molecular functions relevant to thousands of genes used in high-density oligonucleotide microarrays (Zhan, 2002). These databases, such as the Gene Ontology (GO) website ("Gene Ontology Database," 2012), are essential to make biological sense of all the resulting data at a much faster and efficient pace. GE profiling in studies of diseases has become a useful tool in understanding diseases on a deeper molecular level (Gibson & Muse, 2009). The aim of the present study is to utilize the useful applications of high density oligonucleotide microarray technology to learn more about the GE of defined HRs and to validate our *in vitro* GC model.

There are limitations to microarray data that should be acknowledged. For instance, protein levels are not always parallel to mRNA levels and additional studies and tests will be needed for confirmation of candidate genes (John D. Shaughnessy Jr, 2003). Also, statistical analysis can be complicated depending on the experimental design, number of samples, the number of genes represented on an array and more. All these factors have to be considered and corrected for throughout data processing (Tarca, Romero, & Draghici, 2006). New software programs are being developed to translate genes into biological functional information. There are limitations with each; however, these are the available tools we have to work with data. Therefore, it is important to consider the best program for the data (Tarca et al., 2006; van den Berg, Thanthiriwatte, Manda, & Bridges, 2009).

2 AIMS OF THE STUDY

This study is part of a larger study which uses both gene expression (GE) and cytogenetic tools to investigate possible genetic basis of the hyper-responder (HR) phenotype and paraproteinemias. Prior to this study, individuals within families with multiple cases of monoclonal gammopathies (MG), and other lymphoproliferative disorders, were screened and the previously described HRs were defined. Twelve HR individuals agreed to participate in this study. One had to be removed because of false paternity report. B cells from the participating HRs, related controls (RC) and unrelated controls (UC) were isolated and stimulated in an *in vitro* germinal center (GC) developed for this study. The possible fulfillment of HRs as an endophenotype, along with the GE studies reporting a step-wise transformation from normal plasma cells (PC) to monoclonal gammopathy of undetermined significance (MGUS) to malignancy (Figure 7 and 8), has led to this study's hypothesizes that HRs might play a role in this step-wise transformation as an endophenotype (Davies et al., 2003; Zhan et al., 2007).

In this project the specific aims were as follows:

1. To analyze gene expression data obtained using microarray technology for samples from all participants regardless of classification (HR, RC, or UC) before and after 14 days of stimulation within an *in vitro* GC and compare observed differences with published data from *in vivo* GC reaction for candidate genes-of-interest.

2. To compare the differences between the suggested endophenotype, HR, and controls before and after 14 days of stimulation within an *in vitro* GC for candidate genes-of-interest.

3. To compare previous gene expression patterns (GEP) reports of significantly different genes between normal PCs and MGUS to those found between normal controls and HRs before and after 14 days of stimulation in order to investigate if HR's GE data are more similar to MGUS GE or to normal PC GE.

3 MATERIALS AND METHODS

3.1 Samples

The hyper-responders (HR) were selected from the eight Icelandic families, as described in previous Icelandic studies, and related controls (RC) were selected from the same families and unrelated controls (UC) both matched for age and gender (Table 2). Twelve HRs agree to participate in this study. Each HR and its corresponding RC and UC samples were taken the same date were cultured on the same days. GC ID corresponds to samples which were cultured at the same time within the *in vitro* germinal center (GC). Therefore all GC-1 samples were isolated and cultured at the same time. GC-9 and its samples were removed from this study due to uncertainties in reported paternity. Permits were obtained from the National Bioethics Committee and the Data Protection Authority, and informed consent was obtained from the subjects.

Table 2: Sample information

Sample information for each subject used in this study. F and M denote female and male respectively. Family numbers are from Figure 9. NA is Not Applicable because these samples are not within the families in Figure 9 before of unrelated status.

*HRs 2 and 3 shared a RC from GC 4. GC-4 UC was taken to ensure a control for this GC.

GC ID	Gender	Age	Family #	Group ID
1	F	43	2	HR
	F	46	2	RC
	F	43	NA	UC
2*	F	45	3	HR
	F	45	NA	UC
3*	F	60	3	HR
	F	60	NA	UC
4*	F	55	3	RC
	F	55	NA	UC
5	M	31	8	HR
	M	36	8	RC
	M	31	NA	UC
6	M	56	8	HR
	M	48	8	RC
	M	56	NA	UC
7	M	60	8	HR
	M	52	8	RC
	M	60	NA	UC
8	F	51	8	HR
	F	47	8	RC
	F	51	NA	UC
10	M	49	8	HR
	M	56	8	RC
	M	49	NA	UC
11	F	41	8	HR
	F	36	8	RC
	F	41	NA	UC
12	F	60	8	HR
	F	60	8	RC
	F	60	NA	UC
13	F	24	8	HR
	F	29	8	RC
	F	24	NA	UC

Prior to my involvement, peripheral blood mononuclear cells (PBMNs) were isolated from 70 ml peripheral blood (PB) samples collected into EDTA-tubes by the Ficoll-Hypaque (Sigma, St. Louis, MO, USA) density gradient method. PBMNs were isolated using the standard Sigma-Aldrich Histopaque-1077 procedure.

3.2 Cell isolation and culture

3.2.1 B cell isolation

B cells were isolated from PBMNs by another student, Sóley Valgiersdóttir, M.Sc. project using the CD19 MACS B-cell purification kit from Miltenyi Biotec (Bergisch Gladbach, Germany) and following the manufactures instructions. Isolated B cells were then divided into three equal portions for three different students working with these samples. I received one of those three portions which was added to 400 µl RNA later® (Qiagen, Germany) + 100 µl medium (RPMI-1640 + ultra-low Fetal Calf Serum (FCS)) per 10^7 cells and frozen at -80°C at day 0 without any stimulation for RNA isolation.

3.2.2 Stimulation of B cells in the *in vitro* germinal center model

Stimulation of B cells within the *in vitro* GC model was developed by Helga Einarsdóttir and these samples were stimulated by Sóley Valgiersdóttir as part of her M.Sc. project. In this model isolated B cells were stimulated with CD40L on Chinese-Hamster Ovarian (CHO) cells and Interleukin-4 (IL-4).

The following description is based on Helga Einarsdóttir´s description of the *in vitro* GC model process. After B cell isolation, $3x10^5$ B-cells were seeded in the presence of $50x10^3$ CD40L transfected CHO cells, gamma-irridated with 40Gy. The CHO-cells were seeded one day before in a volume of 250 µl. The assay was performed in 24-well plates (Nunc) in 1 ml per well of RPMI-1640 containing 2mM glutamine, penicillin/streptomycin (50 units/ml pen., 50 µg/ml strept), 10 ng rhIL-4 (R&D Systems, Minneapolis) and 10% ultra-low FCS. After 7 days of stimulation within this model, B-cells were transferred to a fresh CHO-cell coated 24-well plate. Supernatants (150 µl per well) were collected every other day and fresh medium was added. Cells were harvested, counted and their viability evaluated with Trypan Blue dye on days 7 and 14. Day 14 cells were placed into 400 µl RNA later® (Qiagen, Germany) + 100 µl medium (RPMI-1640 + ultra-low FCS) per 10^7 cells for RNA isolation.

3.3 Determination of IgG and IgM by ELISA

Immunoglobulin (Ig), IgG or IgM, concentrations of samples collected from the GC cultures at days 7, 14, and 21 were measured. On few occasions we used day 5 to supplement day 7 and day 19 to supplement day 21. To allow for the lines of the graph to be connected day 19 was marked as day 21 in graphs and this is noted beside those graphs. Using a tailor-made ELISA (Enzyme-Linked ImmunoSorbent Assay) protocol, 96-well MaxiSorp plates (Nunc) were coated with anti-IgG or anti-IgM antibodies (DakoCytomation, Dako, Denmark) at a 1/1000 dilution in a carbonate-bicarbonate coating buffer and left to incubate overnight at 4°C then washed with a Phosphate Buffered Saline

(PBS) washing buffer the next day. Blanks and IgM / IgG controls, pooled from human samples of known concentration, were used at appropriate dilutions, in PBS-washing buffer, of 1/30,000 for IgM and 1/160,000 for IgG assays. Each plate contained a triplicate of the standard, pooled human serum samples, which at highest concentration was diluted to 1/30,000 for IgG and 1/5000 for IgM. The plates were incubated at room temperature for 60 min and then coated with an alkaline phosphatase-conjugated (AP) anti- IgM or IgG antibodies (DakoCytomation, Dako Denmark). Anti-IgM AP was diluted in PBS-washing buffer to 1/3000 and anti-IgG AP was diluted to 1/1000. For each plate a solution of two 5 mg phosphatase substrate tablets (Sigma-Aldrich, Missouri, USA) dissolved into 10 ml of 10% diethanolamine-HCl substrate buffer was added at 100 µl per well to develop an enzyme reaction. The optical density was read after approximately 20-50 minutes (depended on when the highest absorption measurement reached a value of more than 1.0 for IgG plates and close to, but not reaching, 2.0 for IgM plates). The optical densities were measured at 405 nm using a SpectraMax™ 250 Microplate Spectrophotometer (Molecular Devices, California, USA). The antibody concentration of each sample concentration was calculated from the standard curve using the average of the duplicated samples for all of the different concentrations in µg/ml. A two sample t-test was used to compare HR samples to each type of control (related or unrelated) and to compare each type of control to the other type samples collected on days 7, 14, and 21.

3.4 mRNA isolation

mRNA was isolated from all cell samples on day 0 and day 14 using the RNeasy® kit (Qiagen, California, USA). All tubes, pipette tips and water were RNase free (Ambion, California, USA). All laboratory work areas were cleaned using RNase ZAP (Ambion) for all procedures to ensure no RNase contamination. All samples and solutions were kept on ice unless instructed otherwise. Samples were spun down in a centrifuge (5 min at 2000 rpm) and supernatant removed in order to have complete removal of cell-culture medium. For each sample, cellular contents were disrupted by adding 350 µl of a 9.9 % mercaptoethanol (ICN Biomedicals Inc., Ohio, USA) and Buffer RLT (Qiagen) solution and then homogenized by adding 350 µl of 70% ethanol (Sigma-Aldrich, Missouri, USA) all within a 1.5 ml RNase free tube. This 700 µl solution was then placed into the provided RNeasy spin column (Qiagen) to collect precipitated RNA. DNase digestion was necessary for samples to be intact and free of DNA contamination because the samples yielded very small amounts of RNA. Any possible contaminating DNA was digested by first adding 350 µl RW1 buffer (Qiagen) to wash RNA (centrifuged 30 sec. at 10,000 rpm) and then 80 µl of DNase 1 Incubation Mix (Appendix 1) was added to treat the RNA bound to the RNeasy membrane of the Qiagen® spin column and incubated at room temperature for 15 min. DNase 1 Incubation Mix was then removed with a second wash of 350 µl RW1 buffer then centrifuged (30 sec at 10,000 rpm). Then RNA was washed with 500 µl Buffer RPE (Qiagen, Appendix 1) and centrifuged (30 sec at 10,000 rpm). RNA was washed again using 500 µl Buffer RPE (Qiagen) which was added to the Qiagen® spin column. The spin column was centrifuged once (2 min at 10,000 rpm) and then again (1 min at 12,000 rpm). The extra centrifuge time was to ensure that no ethanol was transferred in the RNA elution. RNA was then eluted into an RNase free

1.5 ml tube (Ambion) by adding 50 µl of RNase free water (Ambion), centrifuged (1 min at 10,000 rpm) and spin column was discarded. RNA samples were then frozen at -80°C and RNA aliquots were placed aside to be measured for quantity (1.5 µl) and quality (1.0 µl) which is described in the next section. RNase Zap and DNA-OFF (Ambion) were used as well for steps performed from cDNA synthesis and future procedures.

3.5 Quantity and quality measurements

Using a Nanodrop® Spectrometer (ND 1000, Thermo) RNA yield was quantified (ng/µl) for each sample. An aliquot of 1.5 µl of all samples were measured against 1.5 µl RNase-free water (Ambion) as a control. Samples which measured yields too low for cDNA synthesis were repeated. Purity measurement (Abs_{260}:Abs_{280}) ratio should be > 1.8. Some samples measured very low amounts repeatedly, therefore; amplification of all RNA samples was necessary.

Quality of RNA isolated from samples was measured using the Agilent RNA 6000 Nano Kit on the Agilent 2100 Bioanalyzer (Agilent Technologies, Germany). To start electrodes which read samples were decontaminated by a wash with RNase Zap (Ambion) and RNase free water (Ambion) using separate electrode Nano Labchips (Agilent) included in the kit. The kit included individually sealed Labchips with 16 wells. Four of the 16 wells were designated for ladder and gel loading. To prepare the gel, which samples will be placed on, 1 µl of RNA 6000 Nano dye concentrate (Agilent) was added to a gel aliquot. The gel was then gently agitated until dye had cleared in color and then centrifuged (10 min at 14,000 rpm). Samples were kept on ice unless instructed otherwise. Into each sample, 5 µl of RNA 6000 Nano marker (Agilent) was added to the 1 µl aliquots of sample in order to be read by the Bioanalyzer (Agilent). Those samples were then incubated on a 70°C heat block for 2 min to promote binding and then placed back on ice. The gel-dye mixture was then loaded, at a volume of 9 µl, onto the Labchip (Agilent). Using a plunger connected to the priming station and 30 sec. of air pressure, gel was forced through the Labchip evenly. The following steps had to be done within 5 min or else the chip was compromised. An additional 9 µl of gel-dye mixture was added to a well as a blank. Then the ladder (1 µl) and 5 µl of Nano 6000 marker (Agilent) were added at the ladder indicated well. Then 6 µl of the sample + Nano 6000 marker was added to the 12 remaining wells. All sample positions were noted to enter into the Agilent Bioanalzyer program. The Labchip was then vortexed for 1 min and placed into the Bioanalyzer to be read. Once the program had been started sample identifications were entered into the program so that the print-out indicated which samples linked to their appropriate trace results. Data were then printed and kept. Quality can be determined by RNA integrity number (RIN) measurements. All samples with RIN ≤ 6.5 were repeated. Any samples exhibiting degradation (Figure 10) were not carried on through future steps and were repeated until meeting quality standards. Repeated samples which gave peaks in the correct location but at a lower magnitude and gave RIN results of “N/A” were acceptable for further use to be amplified due to the small volume of RNA.

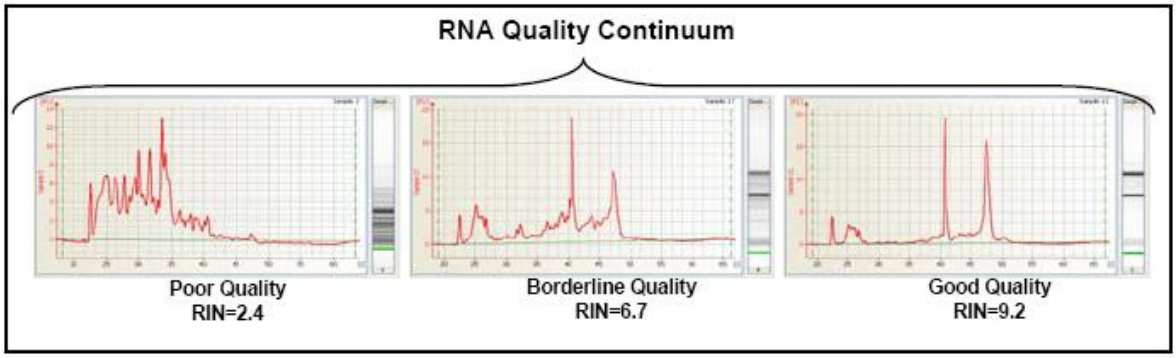


Figure 10: RNA quality standards.

Example quality standard traces for RNA samples using an Agilent 2100 Bioanalyzer (Nugen-Technologies, 2011)

3.6 RNA amplification and cDNA synthesis

Since some samples did not provide enough RNA for cDNA synthesis and labeling, it was necessary to amplify all samples. RNA amplification and cDNA synthesis was performed using the WT-Ovation Pico RNA Amplification System (NuGen, California, USA) (Figure 11). Amplification was initiated at the 3´end and at random throughout the whole transcriptome of each sample. The amplified product of the WT-Ovation System was optimized for the detection of low-, medium- and high-abundance gene transcripts using real-time quantitative PCR (NuGen). After measuring the quantity of all samples and based on the sample with the lowest concentration, the starting volume in the amplification process was chosen to be 5 ng for each sample. This is within the acceptable limits (500 pg-50 ng) of starting volume (NuGen). No fewer than eight samples were to be processed at one time because fewer may affect reagent recovery volumes in the kit. All reagents were thawed as close to use as possible to ensure quality and all samples were on ice throughout the protocol unless instructed otherwise.

The first part of the RNA amplification protocol is First Strand cDNA synthesis. First strand cDNA is prepared using a distinctive DNA/RNA chimeric primer mix and reverse transcriptase (RT). The primers have a DNA portion that hybridizes either to the 5´ portion of the poly(A) sequence or randomly across the transcript. RT extends the 3´ DNA end of each primer generating first strand cDNA (NuGen). The resulting cDNA/mRNA hybrid molecule contains a unique RNA sequence at the 5´ end of the cDNA strand. To start, RNase free water was added to 5 ng samples to a volume of 5 µl. Samples (5µl) were then added to 2 µl of First Strand Primer Mix (NuGen) in a 0.2 ml RNase free PCR tube (Ambion). For primer annealing, samples were then placed in a pre-warmed Veriti™ Thermal Cycler (Applied Biosystems, California, USA) to incubate at 65°C for 2 min then removed, pulse centrifuged and placed on ice. Then, 3 µl of the First Strand Master Mix (Appendix 1) was added to each tube and mixed well by pipette. For first strand synthesis samples were then placed into the thermal cycler again to incubate at the following programmed settings: 4°C for 1 min, 25°C for 10 min, 42°C for 10 min, 70°C for 15 min then removed, pulse centrifuged and placed on ice. During this time reagents for second strand synthesis were thawed on ice. Agencourt® RNAClean® purification beads

were put out at room temperature. While on ice, 10 µl of Second Strand Master Mix (Appendix 1) was added to each first strand reaction tube and mixed by pipette.

Fragmentation of the mRNA within the cDNA/mRNA complex creates priming sites for DNA polymerase to synthesize a second strand, which includes DNA complementary to the 5´ unique sequence from the first strand chimeric primers. The result is a double-stranded cDNA with a unique DNA/RNA heteroduplex at one end (NuGen). For second strand synthesis samples were then placed in the thermal cycle to incubate at the following programmed settings: 4°C for 1 min., 25°C for 10 min., 50°C for 30 min, 70°C for 5 min then removed, pulse centrifuged and placed on ice. The next step is cDNA purification using the Agencourt® RNAClean® beads. To each reaction, 32 µl of bead suspension was added and mixed by pipetting. This step attaches cDNA strands to the beads. Each reaction was then transferred into individual wells of a 96-well MaxiSorp plate and incubated at room temperature for 10 min. After room temperature incubation, samples were transferred to a Agencourt SPRIplate®96R Ring magnet plate (Agencourt Bioscience Corporation) and beads migrated to form a ring around the inside of each well. After 5 min and while still on the magnetic plate, 45 µl of clear supernatant (extra binding buffer) was removed and discarded. Then three washes of 200 µl fresh 70% ethanol followed by 30 sec wait were performed. Following complete removal of 70% ethanol RNAClean beads were air dried for 20 min at room temperature. During this time, reagents for the SPIA amplification were thawed on ice.

The final step is SPIA isothermal linear amplification which is a process developed by NuGEN and is described by the company's user guide as follows. It uses a SPIA DNA/RNA chimeric primer, DNA polymerase and RNase H in a homogeneous isothermal reaction that provides extremely efficient amplification of DNA sequences. RNase H is used to degrade RNA in the DNA/RNA heteroduplex at the 5´ end of the first cDNA strand. This results in the exposure of a DNA sequence that is available for binding a second SPIA DNA/RNA chimeric primer. DNA polymerase then begins replication at the 3´ end of the primer, displaces the existing forward strand. The RNA portion at the 5´ end of the newly synthesized strand is again removed by RNase H, exposing part of the unique priming site for initiation of the next round of cDNA synthesis. The process of SPIA DNA/ RNA primer binding, DNA replication, strand displacement and RNA cleavage is repeated, resulting in swift accumulation of cDNA with sequence complementary to the original mRNA (See Figure 11). An average amplification of 1,500-fold is observed when starting with 5 ng total RNA (Nugen-Technologies, 2011). To start SPIA amplification, the 96-well plate (Maxisorp) was removed from the Agencourt SPRIplate®96R Ring magnetic plate (Agencourt Bioscience Corporation) and 160 µl of SPIA Master Mix (Appendix 1) was added to each well and mixed thoroughly by pipette. For each sample volume of 160 µl, 80 µl (half-reaction) were placed into two separate fresh RNase free 0.2 ml PCR tubes to ensure no spillover of the whole volume in one tube. These tubes were then placed into the thermal cycler and incubated at the following settings: 4°C for 1 min, 47°C for 60 min, 95° for 5 min and then removed, pulse centrifuged and placed on ice. The post-amplification workspace was separate and disinfected with DNA-OFF™ (Ambion). Half-reactions were then recombined into a fresh post-amplification 96-well plate (Maxisorp) and the plate transferred onto the magnetic plate. This step will allow for the magnetic RNAClean® beads to be separated from the supernatant which is eluted cDNA. After 5 min, the eluted

cDNA was removed from the wells and placed into individual fresh RNase-free 1.5 ml tubes and proceed to the amplified cDNA purification protocol or frozen at -20°C until purified later.

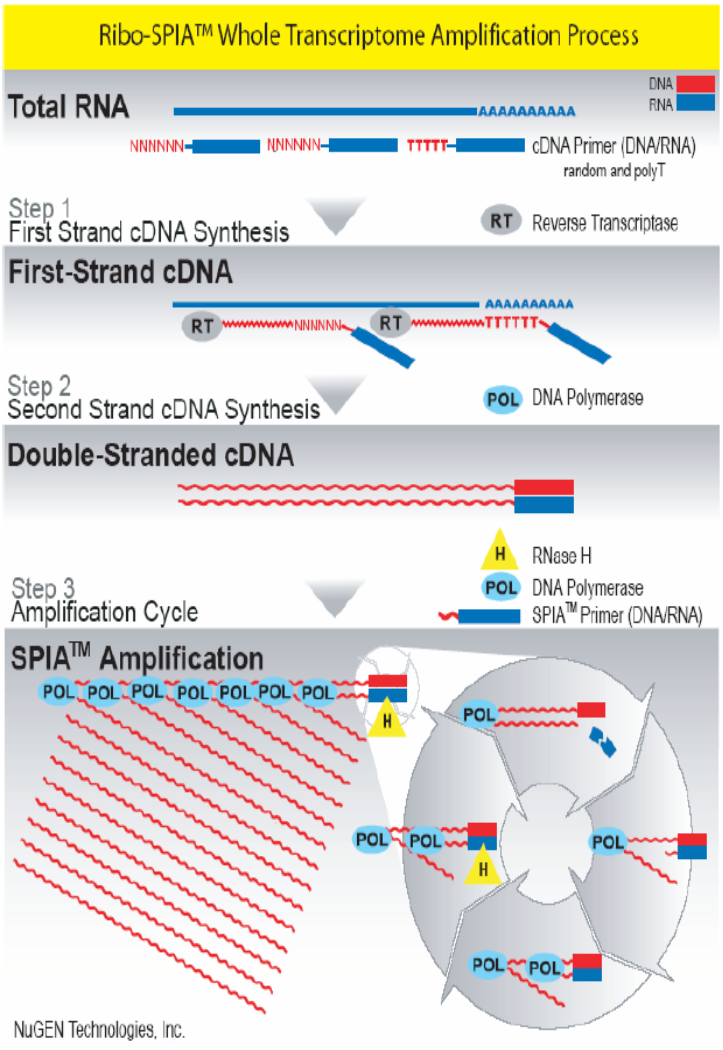


Figure 11: RNA amplification and cDNA synthesis.

Ribo-SPIA RNA amplification process used in the WT-Ovation system (Nugen-Technologies, 2011)

Amplified cDNA was purified using the Qiagen QIAquick® PCR Purification Kit. This amplified cDNA purification was a suggested step by the staff of Roche NimbleGen Iceland LLc. and is necessary to ensure no contamination for cDNA labeling and future hybridization. PB buffer (Qiagen) (800 µl) was added to samples (160 µl) then vortexed and pulse centrifuged. Half-volume (480 µl) of samples were then placed onto one QIAquick spin column (Qiagen) and centrifuged (1 min at 13,000 rpm) then repeated for the remaining half. Flow through was collected and discarded. Once cDNA was isolated to the spin column membrane, it was washed by adding 700 µl of fresh 80% ethanol and centrifuged (1 min at 13,000 rpm)—repeated twice. Before elution of cDNA, the column tip was blotted with filter paper. Then 30 µl of room temperature nuclease-free water was added to the spin column membrane and incubated at room temperature for 5 min to elute purified cDNA. The spin column was centrifuged in a clean nuclease-free 1.5 ml tube (Ambion) to collect eluted cDNA. Quantity and quality were then measured using the same steps as previously described for RNA quantity and quality measurements.

Purity measurement (Abs_{260}:Abs_{280}) ratio should be > 1.8 then placed into 1 µg aliquots for cDNA labeling. Figure 10 does not apply to quality standards now because purified cDNA standards for quality show a different type of peak as seen in Figure 12.

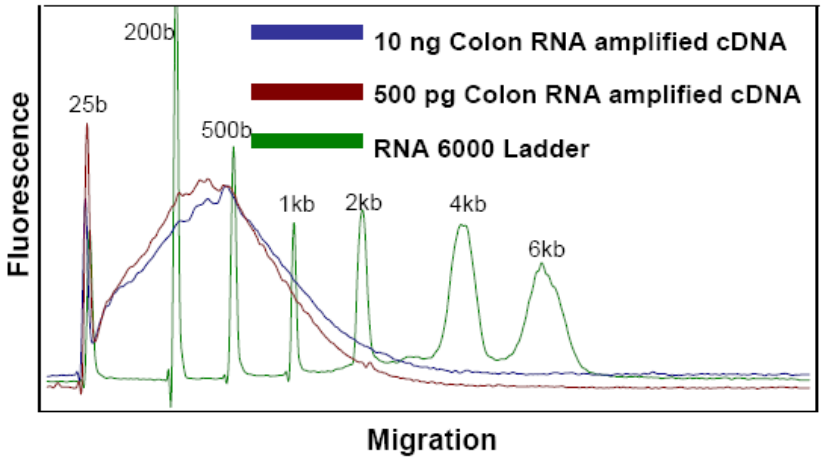


Figure 12: cDNA quality standards.

Standard Bioanalyzer trace of good quality amplified cDNA after cDNA purification (Nugen-Technologies, 2011)

3.7 cDNA labeling

cDNA was labeled using the NimbleGen One-Color DNA Labeling Kit. 40 µl aliquots of Cy-3 primer buffer solutions (Appendix 1) stored at -20°C, which were close to expiration yet quality research has shown still viable, were provided by Roche NimbleGen. Aliquots of 1 µg cDNA samples and Nuclease-free water (Ambion) to volume of 80 µl were added to 40 µl primer aliquots in Nuclease-free 0.2 ml PCR tubes (Ambion) then heat-denatured for 10 min at 98°C in a thermocycler followed by a 10 min ice-water bath. These tubes were carefully protected from light throughout the experiment. 50% dNTP + 10% Klenow master mix and Nuclease-free water (Ambion) to volume of 20 µl per sample was then added and mixed well with samples. Samples were then left for 2 hours at 37°C in a thermocycler protected from light. The reaction was then stopped using 10 µl of 0.5M EDTA followed by 11.5 µl 5M NaCl to each sample. Samples were then washed in 110 µl isopropanol to clean unbound agents. After 10 min room-temperature incubation and 10 min spin in a centrifuge (12,000 rpm), samples were rinsed with ice-cold 80% ethanol (Sigma-Aldrich) then removed and dried in a SpeedVac on low heat. Samples were considered dried when pellet was clear and barely visible which took 5-15 min depending on the size of the pellet. Once all ethanol had been dried off, samples were suspended in Nuclease-free water (Ambion) and quantity was measured in order to aliquot 4 µg samples for hybridization on Roche NimbleGen 12-plex microarrays then dried once again to form a pellet.

3.8 Microarray hybridization and processing

Microarray hybridization, washing and processing was performed at the Roche NimbleGen, Iceland LLc. facility using their NimbleGen Hybridiztion System at 42°C. Unique sample tracking controls were

used for each sample and added to the 4 µg cDNA labeled pellets. Each sample (3.3 µl) was then added to 12µl of hybridization solution master mix (Appendix 1). Tubes were then placed for 5 min in 95°C incubation, followed by 5 min in 42°C incubation. Hx12 mixers were used to apply to the 12x135K array NimbleGen Microarray (Figure 13). Application was performed with a Precision Mixer Alignment Tool and a Mixer Brayer supplied by the NimbleGen facility. Firm application of mixer to array slide is important to prevent air bubbles and leaks. Each sample was carefully loaded, at a volume of 6 µl, onto an individual array. The 59 samples were loaded onto five slides with 12 arrays per slide and covered with port seals for each array. The microarrays were then placed into the NimbleGen Hybridization System at 42°C for 16 - 20 hours. After this point, the washing and scanning of arrays was performed by Roche NimbleGen, Iceland LLc staff.

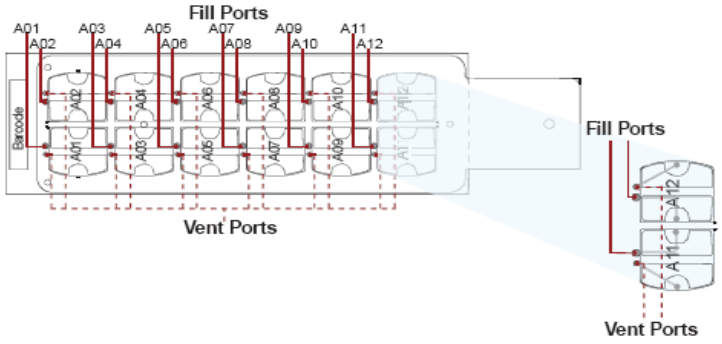


Figure 13: Roche NimbleGen 12x135k microarray slide.
HX12 Mixer and for Roche NimbleGen 12x135k microarray slide (NimbleGen Userguide, version 5.1)

3.9 Data Processing and Statistical Analysis

Data were normalized using a slight variation of Robust Multichip Average (RMA) normalization, which is often used with oligo-nucleotide normalization, with the use of the R-program (Gibson & Muse, 2009; Irizarry, Hobbs, Collin, & Speed, 2003). In typical RMA normalization there are three steps: background correction, quantile normalization, and summarization via median-polish (Irizarry et al., 2003). The final step of summarization via median-polish was not done for this study because keeping the three probe values separate, rather than averaging into a single number, might increase statistical power. Another purpose for no summarizing the three probes was to allow for normalization for slide effect because the arrays are nested within the slides. Once background corrections were made and data were normalized then data was log-transformed on the base 2 scale in order to produce symmetry of relative increases and decreases in fluorescence intensity. Frequencies of the log ratios closely resemble a normal distribution more than the raw ratios which will facilitate statistical analysis (Gibson & Muse, 2009; Tarca et al., 2006).

Unsupervised hierarchical clustering was performed using Euclidean distances and a smoothed histogram plotted (Figure 15) using the R-program to observe any samples that might need repeating due to highly aberrant results or distribution ("R-program," 2012). Any samples distributing abnormally were investigated to decide if repeated sample processing was necessary. Samples which clustered and/or distributed abnormally were repeated to ensure no mistakes were performed during process

and probe intensities were investigated. When all utilizable data were collected unsupervised hierarchical clustering was repeated and heat-maps of calculated Z-scores were produced using the R-program to determine clustering. Z-score is a measure that quantifies the distance a data point is from the mean of a data set and was calculated by (intensity-probe mean)/probe SD). Table 3 indicates which samples were used in data normalization and statistical analysis.

Table 3: Hybridized samples and outliers.

Samples included in hybridization onto microarray slides are indicated with an "X". Samples which were no included because of low B cell volumes are indicated with a "—". *outliers excluded from data.

GC ID	Sample day	HR	RC	UC
1	0	--	X	X
	14	X	X	X
2	0	X	--	X
	14	X	--	X
3	0	X	--	X
	14	X	--	X
4	0	--	X	X
	14	--	X	X
5	1	X	--	X
	14	X	--	X
6	0	--	X	X
	14	X	X*	X*
7	0	--	X	X
	14	X*	X	X
8	0	X	X	X
	14	X	X*	X
10	0	X	--	X
	14	X	X	X
11	0	--	X	X
	14	X	X	X
12	0	X	X	X
	14	X	X*	X
13	0	X	X	X
	14	X	X	X

Models and testing were developed with the help of Jóhannes Guðbrandsson and Arnar Palsson. Random effects models were fitted using the lme4-package in R (Vazquez, Bates, Rosa, Gianola, & Weigel, 2010). It was decided to use a linear mixed-effect model which accounts for confounding factors in order to estimate variances brought forth by each variable (confounding factor) in order to identify interesting variances in measurements between tested factors, which in our study was experiment group compared to controls and day 0 compared day 14 stimulation (Draghici, 2002). Confounding factors within our experimental design include: probe, slide, array, GC culture, and conventional error. When testing for differences between the HRs and all controls (controls were grouped together since they are all considered controls) the day effect was included in confounding factors, as well as the others mentioned about. We also tested for significant differences between individual in vitro GC cultures (gender and day were included in confounding factors) and differences between gender (day was included in confounding factors). The basic model can be seen in Figure 14.

$$y_{ijklm} = P_i + D_j + GC_k + S_l + A_m + \varepsilon_{ijklm}$$
$$GC_k \sim N(0, \sigma^2_{GC}), \quad S_l \sim N(0, \sigma^2_S), \quad A_m \sim N(0, \sigma^2_A) \quad \varepsilon_{ijklm} \sim N(0, \sigma^2)$$
$$i = 1,2,3 \quad j = 1,2 \quad k = 1,\ldots,12 \quad l = 1,\ldots,6 \quad m = 1,\ldots,54$$

Figure 14: Linear-mixed effects formula model.

Basic model for testing interactions within our experiments design. y is normalized intensity, P is the probe effect, D is the day effect, GC is the culture effect, S is the slide effect, A is the array effect and ε is conventional error (Vazquez et al., 2010).

The same basic procedure was used to estimate all effects. That procedure is as follows: First a full model with the effect of interest and a reduced model excluding the effect of interest were fitted using maximum likelihood for each transcript separately. Then, they were compared with likelihood ratio test to produce a large table that included t-statistic, chi-square (χ^2) values, fold-change (FC) values and p-values for all gene transcripts. Finally, resulting p-values were corrected for multiple testing using Benjamini and Hochberg correction (Benjamini & Hochberg, 1995). Both FC and t-statistic methods of gene ranking were considered. Some studies argue against using FC and describe it as an arbitrary unless combined with other reliable statistical methods (Draghici, 2002; Tarca et al., 2006). Typically the FC method of gene ranking produces a similar list; however, some genes, such as transcription factors (TF), might have key biological effects despite their expression change being less than two-fold (Tarca et al., 2006). The t-statistic method of ranking differential expressed genes was ultimately used because it has been shown in previous publications to be a reliable method (Kadota & Shimizu, 2011; Tarca et al., 2006). Previous published literature has also shown that it is likely for p-values to rank the genes in a similar order as the t-statistic. Also, the t-statistic offers the advantage of a lower number of false discovery rates, than FC, and increases the sensitivity of gene ranking (Smyth, 2004).

3.10 Gene Ontology and Bioinformatics

In order to process the list of sometimes thousands of significantly differentially expressed genes, the Onto-Express package from the bioinformatics software Onto-Tools was used ("Onto-Tools," 2012). This was one of many suggested through the Gene Ontology (GO) website--a database of function profiles for genes ("Gene Ontology Database," 2012). A few programs were tested but this specific program was chosen because of its wide range of tools available, such as an annotation browser, search engine and visualization component, statistical analysis component and access to the GO database (van den Berg et al., 2009)). The Onto-Express is an efficient tool available to automatically transfer a list of differentially expressed genes into functional profiles which reduces the time put into researching genes individually. Text files with significantly differentially expressed genes listed were attached as input files and tested against all the genes in the array specifically used within this study, rather than all known genes, which was obtained through design files given by Roche NimbleGen, Iceland Llc. after microarray processing. Testing against the genes on our specific array allows for calculated p-values on whether the biological function is significantly represented by the amount of genes by chance (Khatri et al., 2007).

4 RESULTS

This study began with peripheral blood (PB) samples from 12 hyper-responders (HR) characterized individuals, 9 related controls (RC) and 12 unrelated controls (UC). RNA was extracted from B cells isolated from these samples before and after 14 days of stimulation in an *in vitro* germinal center (GC) culture for a total of 64 samples. Some had to be excluded from RNA isolation due to low B cell counts and 59 samples were left to hybridize onto microarrays. During hybridization of microarray, in 3 different arrays leakage occurred and results were void. Those three samples were repeated—day 0: GC-2 UC, GC-3 UC, and GC-4 RC. A smoothed histogram was produced from all samples in order to view the sample distribution (Figure 15). There were five samples which notably distributed differently from the rest. The original heatmap produced from unsupervised hierarchical clustering also showed five samples including a mixture of HR, RC and UC from day 14 which clustered differently to the left (Figure A1. Appendix 2). When investigating into these abnormal occurrences within samples, it was found that in both figures the five distinctly different samples were the same—day 14: GC-6 UC and RC, GC-7 HR, GC-8 RC, and GC-12 RC. Also, abnormal in the original heatmap was one day 0 sample clustering with day 14 samples and one day 14 sample clustering with day 0. We believed there could have been a labeling error. Therefore, the five differently distributed samples and the two potentially switched day samples were repeated. To fill the 12-plex array, two more samples were chosen--day14: GC-2 UC and GC-7 RC. These 12 samples were repeated from the step of RNA isolation in the Materials and Methods section through microarray hybridization and statistical analysis.

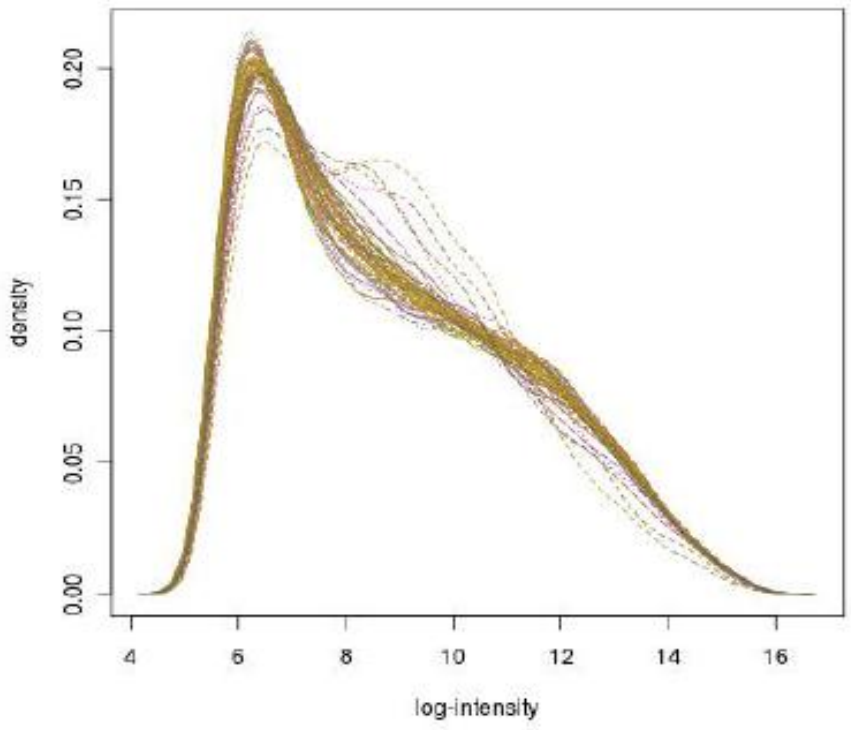


Figure 15: Sample distribution.

Smoothed histogram showing the distribution of RMA-levels for each sample on log2-scale. Arrays with leaked samples were not included.

Once again, repeated samples were plotted in a histogram and unsupervised hierarchical clustering was performed and expressed in a heatmap figure. The results show the same five samples as before distributed and clustered in the same manner as before. The two samples which clustered with opposite days now cluster with their stimulation day. This suggests that labeling between the days was switched. One sample, GC-2 UC day 14, did not pass quality standards possibly due to a handling error in the washing and scanning steps done at the Roche NimbleGen Iceland, Llc. facilities. This sample was one of the two samples added to fill the 12-plex array; therefore, we kept the original result. The other of the two added samples clustered in similar pattern as the original and was thrown out. The mean and median probe intensities of the five distinctly different samples were then compared to the mean and median of all the probes of all the other arrays by creating a box plot (Figure 16). It would be expected that perfect match (pm) probe intensity is higher than background probe (bg) intensity for arrays hybridized with human-RNA. Figure 16 shows that for these five outlier samples the pm/bg intensity ratio was lower than normal which could suggest that very little and/or low quality RNA is found in these samples. Closer examinations showed that RNA concentrations were low but not too low for an adequate volume necessary for microarray hybridization. Bioanalyzer traces showed peaks within quality standards. Another possibility could be a weak or aberrant stimulation within the *in vitro* GC cultures. These five samples were excluded from the study.

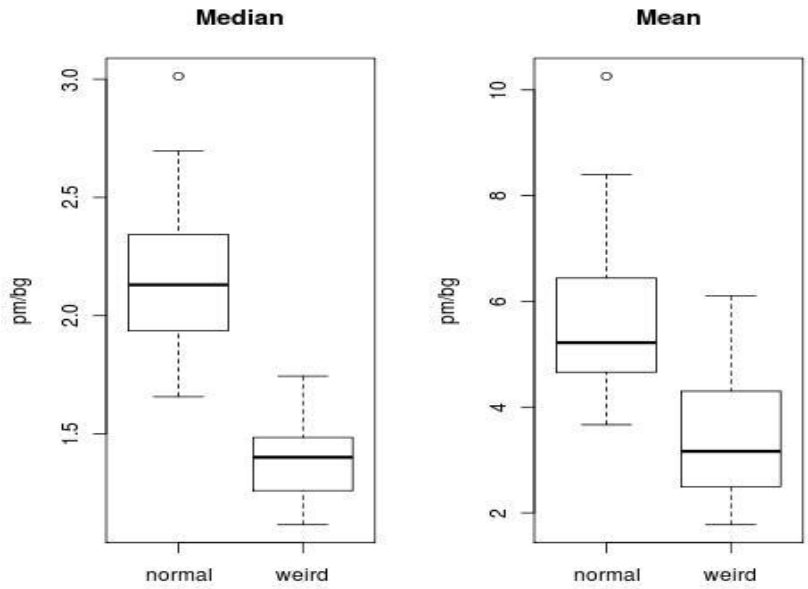


Figure 16: Boxplot of outliers probe intensities.

Box plot of mean and median perfect-match relative to background probe intensities for five outlier samples that distributed differently compared to all other samples and normal samples from all other arrays.

4.1 Gene expression during *in vitro* GC reaction

4.1.1 Significance and clustering analysis

After exclusions, 54 out of the original 59 samples were used to perform unsupervised hierarchical clustering and transformed into a heatmap calculated Z-scores (Figure 17). The Z-score is simply a representation in order to visualize the differences of intensities. The most distinctive feature was the differences between before, day 0, and after 14 days of *in vitro* GC stimulation, day 14. The blue coloring indicates a higher expression and the green color indicates a lower expression. The top and bottom portions of this figure will represent gene transcripts with the most difference between samples and this holds true for all figures. P-values calculated as a result of maximum likelihood tests, were then used to determine which of these gene transcripts were significantly differentially expressed between day 0 and day 14. Table 4 includes those results after adjusting for multiple testing using the Benjamini and Hochberg correction (Benjamini & Hochberg, 1995). Out of the 45,000 gene transcripts on each array, almost one-third (14,277) of them were significantly differentially expressed between day 0 and day 14 after adjusting for confounding factors and multiple testing.

Table 4: Significantly differentially expressed genes for GC stimulation

The number of significantly differentially expressed genes, after adjusting for multiple testing, between day 0 and day 14 at different p-values.

Adjusted p-value	Day
0.05	21558
0.01	17711
0.001	14277

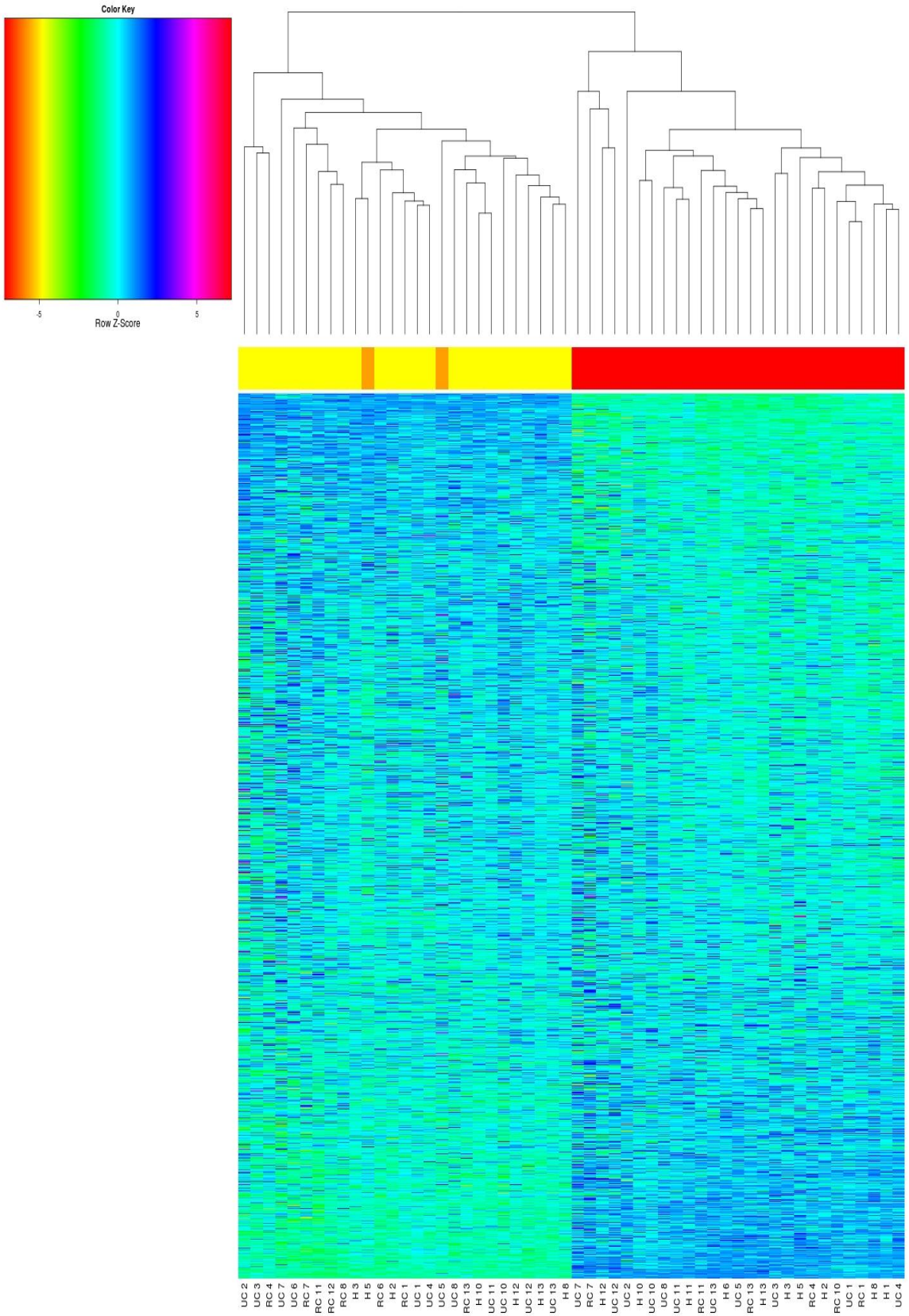


Figure 17: Unsupervised hierarchical cluster analysis heatmap for all samples.

Heatmap for unsupervised hierarchical clustering of all probes for all accepted samples. Samples missing include: GC-1 HR d0, GC-6 HR d.0, GC-6 RC d14, GC-6 UC d14, GC-7 HR d0, GC-7 HR d14, GC-8 RC d14, GC-10 RC d0, GC-11 HR d0, GC-12 RC d14. Yellow is day 0, orange is day 1, and red is day 14.

4.1.2 Gene categories expressed differently

Statistical analysis of differences in gene expression (GE) data between day 0 and day 14 resulted in 14,277 significantly differentially expressed gene transcripts representing genes. All genes were sorted into up- and down- regulated genes based on the t-statistic method.

For analysis of interesting genes the top 50 most up- and down-regulated genes were collected into two tables—Table 5 and 6. Fold change (FC) and t-statistic are included.

Table 5: 50 most up-regulated genes in day 14 compared to day 0 samples.

The 50 most up-regulated genes in day 14 compared to day 0 samples using the t-statistic method to rank genes.

Gene symbol	Description	t-statistic	Adjusted p-value
AICDA	activation-induced cytidine deaminase	85.32	7.80E-54
CCL22	chemokine (C-C motif) ligand 22	74.44	7.80E-54
C11orf41	chromosome 11 open reading frame 41	50.42	4.51E-41
C11orf69	chromosome 11 open reading frame 69	46.14	3.67E-39
WDR65	WD repeat domain 65	42.81	3.02E-40
SNFT	Jun dimerization protein p21SNFT	42.54	1.57E-40
CDK1	cyclin-dependent kinase 1	42.51	4.06E-39
KCNMA1	potassium large conductance calcium-activated channel, subfamily M, alpha member 1	42.36	2.19E-39
IL17RB	interleukin 17 receptor B	39.32	1.40E-40
MELK	maternal embryonic leucine zipper kinase	38.30	3.81E-38
NTRK2	neurotrophic tyrosine kinase, receptor, type 2	37.74	1.75E-34
FXYD2	FXYD domain containing ion transport regulator 2	37.30	3.29E-35
IGSF3	immunoglobulin superfamily, member 3	36.92	2.04E-34
SPBC25	spindle pole body component 25 homolog (S. cerevisiae)	36.56	7.56E-37
MYB	v-myb myeloblastosis viral oncogene homolog (avian)	36.25	3.51E-40
P4HA2	procollagen-proline,2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha polypeptide II	35.85	1.27E-41
DKFZp762E1312	hypothetical protein DKFZp762E1312	35.75	5.79E-35
NEK2	NIMA (never in mitosis gene a)-related kinase 2	35.19	9.72E-35
UBE2C	ubiquitin-conjugating enzyme E2C	34.58	4.31E-36
MCM10	MCM10 minichromosome maintenance deficient 10 (S. cerevisiae)	34.19	2.30E-33
KIF20A	kinesin family member 20A	33.99	9.21E-36
CKAP2L	cytoskeleton associated protein 2-like	33.62	1.65E-37
TRAF1	TNF receptor-associated factor 1	33.31	5.33E-34

CCL1	chemokine (C-C motif) ligand 1	33.20	2.94E-33
BUB1B	BUB1 budding uninhibited by benzimidazoles 1 homolog beta (yeast)	33.01	1.57E-34
C1orf24	chromosome 1 open reading frame 24	32.71	6.09E-35
SCD	stearoyl-CoA desaturase (delta-9-desaturase)	32.39	9.72E-35
PIK3R3	phosphoinositide-3-kinase, regulatory subunit 3 (p55, gamma)	32.27	4.08E-34
ADCY1	adenylate cyclase 1 (brain)	31.72	1.61E-32
TPX2	TPX2, microtubule-associated, homolog (Xenopus laevis)	31.71	3.97E-34
HMMR	hyaluronan-mediated motility receptor (RHAMM)	30.95	7.66E-34
CLIC2	chloride intracellular channel 2	30.80	1.38E-31
LOC643331	similar to Kinase suppressor of ras-1 (Kinase suppressor of ras) (mKSR1) (Hb protein)	30.70	3.43E-30
CD40LG	CD40 ligand (TNF superfamily, member 5, hyper-IgM syndrome)	30.32	3.47E-34
KSR2	kinase suppressor of ras 2	30.17	1.41E-25
BCAT1	branched chain aminotransferase 1, cytosolic	29.85	1.39E-32
HOMER2	homer homolog 2 (Drosophila)	29.19	2.42E-32
GPR55	G protein-coupled receptor 55	29.15	4.57E-31
AURKB	aurora kinase B	28.86	2.48E-32
CCNA2	cyclin A2	28.84	2.44E-31
ESCO2	establishment of cohesion 1 homolog 2 (S. cerevisiae)	28.70	1.57E-34
HIST1H3B	histone 1, H3b	28.69	7.63E-30
DHCR24	24-dehydrocholesterol reductase	28.68	2.93E-31
BIRC5	baculoviral IAP repeat-containing 5 (survivin)	28.62	1.19E-31
TMOD1	tropomodulin 1	28.61	6.72E-30
EXO1	exonuclease 1	28.59	4.02E-30
ASPM	asp (abnormal spindle)-like, microcephaly associated (Drosophila)	28.58	3.36E-31
CDC6	CDC6 cell division cycle 6 homolog (S. cerevisiae)	28.61	5.98E-30
CENPF	centromere protein F, 350/400ka (mitosin)	28.69	1.02E-30
CENPE	centromere protein E, 312kDa	30.17	2.19E-33

Table 6: 50 most down-regulated genes in day 14 compared to day 0.

The 50 most down-regulated genes in day 14 compared to day 0 using the t-statistic method to rank genes.

Gene symbol	Description	t-statistic	Adjusted p-value
LOC653193	similar to Amphiregulin precursor (AR) (Colorectum cell-derived growth factor) (CRDGF)	-63.23	2.52E-48
DUSP1	dual specificity phosphatase 1	-49.26	1.92E-42
KLF4	Kruppel-like factor 4 (gut)	-45.04	3.08E-40
AREG	amphiregulin (schwannoma-derived growth factor	-44.35	1.37E-40
GRASP	GRP1 (general receptor for phosphoinositides 1)-associated scaffold protein	-43.86	1.18E-40
MARCH1	membrane-associated ring finger (C3HC4) 1	-42.52	7.25E-38
ChGn	chondroitin beta1,4 N-acetylgalactosaminyltransferase	-38.54	6.95E-38
S100A8	S100 calcium binding protein A8 (calgranulin A)	-37.62	6.05E-38
ARL4A	ADP-ribosylation factor-like 4A	-37.06	2.04E-36
TBC1D9	TBC1 domain family, member 9	-36.38	1.76E-36
CKAP4	cytoskeleton-associated protein 4	-34.75	2.51E-33
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog	-34.46	6.08E-42
KCNH8	potassium voltage-gated channel, subfamily H (eag-related), member 8	-34.26	4.33E-29
LOC54103	hypothetical protein LOC54103	-32.64	4.57E-34
FOSB	FBJ murine osteosarcoma viral oncogene homolog B	-31.59	6.41E-31
SNF1LK	SNF1-like kinase	-31.54	1.46E-32
LOC643366	similar to ADP-ribosylation factor-like protein 4A	-30.38	3.32E-32
NR4A2	nuclear receptor subfamily 4, group A, member 2	-30.35	1.38E-32
COL19A1	collagen, type XIX, alpha 1	-29.50	1.42E-32
CD55	CD55 molecule, decay accelerating factor for complement (Cromer blood group)	-29.44	4.29E-29
RASA3	RAS p21 protein activator 3	-28.73	1.20E-30
EPC1	enhancer of polycomb homolog 1 (Drosophila)	-28.14	1.39E-31
LILRB1	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 1	-28.13	7.18E-32
CCNL1	cyclin L1	-28.12	2.22E-29
MOP-1	MOP-1	-27.80	6.06E-30
RBMS1	RNA binding motif, single stranded interacting protein 1	-27.29	1.85E-30
RGS2	regulator of G-protein signalling 2, 24kDa	-27.10	5.36E-28
EIF5	eukaryotic translation initiation factor 5	-26.69	1.37E-28
C1orf56	chromosome 1 open reading frame 56	-26.67	5.32E-28
FCRL1	Fc receptor-like 1	-26.59	2.16E-28
LAIR1	leukocyte-associated immunoglobulin-like receptor 1	-26.25	2.24E-29
SAT	spermidine/spermine N1-acetyltransferase	-25.94	3.56E-28
ING3	inhibitor of growth family, member 3	-25.67	1.48E-27

ABCB4	ATP-binding cassette, sub-family B (MDR/TAP), member 4	-25.33	4.40E-26
GPRASP1	G protein-coupled receptor associated sorting protein 1	-25.22	8.52E-28
RGC32	response gene to complement 32	-25.13	1.57E-29
PRKCB1	protein kinase C, beta 1	-24.77	5.69E-27
C6orf192	chromosome 6 open reading frame 192	-24.72	5.89E-27
FCRL2	Fc receptor-like 2	-24.69	5.98E-27
FCGR2B	Fc fragment of IgG, low affinity IIb, receptor (CD32)	-24.42	6.69E-27
PNRC1	proline-rich nuclear receptor coactivator 1	-24.25	2.75E-27
KLF6	Kruppel-like factor 6	-24.22	1.22E-27
LOC153222	adult retina protein	-24.20	5.75E-26
BACH1	BTB and CNC homology 1, basic leucine zipper transcription factor 1	-24.20	1.31E-25
TAGAP	T-cell activation GTPase activating protein	-23.87	9.64E-27
H3F3B	H3 histone, family 3B (H3.3B)	-23.84	2.78E-26
FAM46C	family with sequence similarity 46, member C	-23.81	1.87E-25
PHC3	polyhomeotic like 3 (Drosophila)	-23.66	1.33E-26
KLF9	Kruppel-like factor 9	-23.61	1.69E-27
CD69	CD69 molecule	-23.60	4.47E-26

Due to the large number of significantly differentially expressed genes, only 1,000 were used for analysis using bioinformatic software. Using the Onto-Express bioinformatic software package, see materials and methods, the 500 genes with the highest t-statistic values and 500 lowest t-statistic values separately were categorized into functional profiles from the gene ontology (GO) database as can be seen in Table 7 ("Onto-Tools," 2012). The corrected p-values represent the probability that the specific biological function category has this many genes represented by chance when compared to our reference file of the 45,000 gene transcripts on each array. From Table 7, DNA packaging, response to DNA damage stimulus, cell proliferation, and regulation of cell death were all investigated because of the large number of genes different in this specific function or their previously published role in the GC reaction (Liang et al., 2009; Shen et al., 2004).

Table 7: GO function categories for 1,000 most regulated genes.

The 500 most down-regulated genes and the 500 most-up regulated genes categorized by Onto-Express into function categories. Corrected p-values are the significance values over-representation in that category.

Ontology category	Down-regulated genes		Up-regulated genes	
	Corrected p-value	Total # of genes	Corrected p-value	Total # of genes
Cell cycle				
Cell cycle			0.0	45
Cell cycle checkpoint			0.0	13
Cell cycle phase			0.0	56
Cell division			0.0	35
Cell proliferation			0.0	18
Centrosome cycle			1.8E-02	3
M phase of meiotic cell cycle			1.4E-02	6
Negative regulation of cell cycle process			0.0	5
Regulation of cell cycle process			0.0	12
Cell chemotaxis	1.9E-02	4		
Cell cycle arrest	8.3E-03	7		
Cell differentiation	9.9E-03	35		
Mitosis				
Chromosome movement towards spindle pole			6.1E-03	2
Chromosome segregation			5.6E-04	5
Cytoskeletal protein binding			2.0E-02	16
Establishment of chromosome localization			0.0	5
Establishment of mitotic spindle localization			6.1E-03	2
Establishment of spindle localization			1.7E-02	2
Exit from mitosis			9.8E-03	3
G2/M transition of mitotic cell cycle			1.8E-02	3
Interphase of mitotic cell cycle			0.0	11
M phase of mitotic cell cycle			0.0	45
Mitotic cell cycle			1.3E-02	2
Mitotic chromosome condensation			9.0E-05	5
Mitotic chromosome movement towards spindle pole			2.6E-03	2

Mitotic metaphase plate congression			6.1E-03	2
Mitotic metaphase/anaphase transition			7.0E-05	5
Mitotic spindle organization			4.7E-04	4
Microtubule-based movement			6.0E-05	10
Microtubule cytoskeleton organization			0.0	16
Regulation of mitotic cell cycle			0.0	17
Regulation of ubiquitin-protein ligase activity(mitosis)			1.4E-04	8
Cell death and proliferation				
Programmed cell death	1.0E-04	32	1.5E-02	25
Negative regulation of programmed cell death	5.9E-03	13	3.0E-02	11
Regulation of cell death	7.5E-03	21	2.6E-03	22
Response to DNA damage stimulus			4.5E-05	18
Tumor necrosis factor production	3.3E-02	3		
Regulation of cell proliferation	9.4E-03	20	1.6E-02	18
Mononuclear cell proliferation			1.0E-02	6
Mesenchymal cell proliferation	4.8E-02	2		
Negative regulation of cell proliferation	5.7E-04	15		
SMAD binding	3.9E-03	5		
Immune response				
Lymphocyte activation	3.9E-02	9	1.4E-02	10
Cytokinesis			4.8E-04	6
Lymphocyte homeostasis			2.3E-02	3
Regulation of cyclin-dependent protein kinase activity			1.7E-03	6
Adaptive immune response	9.5E-03	7		
Cytokine biosynthetic process	4.2E-02	5		
Defense response	7.6E-04	23		
ER overload response	4.8E-02	2		
Hemopoietic or lymphoid organ development	8.3E-03	12		
Innate immune response	1.5E-02	8		
Interleukin-6 production	8.3E-04	5		
Leukocyte chemotaxis	1.9E-02	4		
Leukocyte mediated immunity	1.1E-02	7		

Myeloid cell homeostasis	1.9E-02	2		
Myeloid leukocyte activation	3.8E-02	4		
Positive regulation of cytokine production	3.3E-02	4		
Regulation of immune response	1.9E-02	8		
Response to bacterium	3.5E-04	10		
Response to fungus	2.0E-02	3		
Response to molecule bacterial origin	7.5E-03	4		
Response to wounding	9.0E-05	22		
Replication				
DNA packaging			1.2E-04	10
Nucleotide binding			1.4E-02	44
Purine nucleotide binding			5.2E-04	49
Ribonucleotide binding			5.8E-04	47
DNA binding	2.1E-03	57		
Gene expression	3.0E-05	86		
Nucleobase, nucleoside, nucleotide, and nucleic acid metabolic process	4.0E-05	91		
Regulation of binding	3.8E-02	6		
Other				
Cellular response to stress	4.1E-02	14	5.0E-05	18
Phosphorus metabolic process	2.2E-02	27	2.0E-02	27
Carbohydrate catabolic process			7.5E-03	7
Cellular catabolic process			3.4E-03	30
Cellular chemical homeostasis			3.6E-02	11
Chromatin binding			4.7E-02	6
Sister chromatid segregation			0.0	11
Electron carrier activity			4.1E-02	9
Generation of precursor metabolites and energy			3.7E-03	14
Heme binding			2.2E-02	7
Hydrolase activity, acting on acid anhydrides			9.5E-03	23
Macromolecule catabolic process			3.0E-03	25
Negative regulation of glucose transport			4.8E-02	2

Negative regulation of molecular function			3.0E-02	10
Peroxiredoxin activity			2.2E-03	3
Protein binding			0.0	124
Protein C-terminus binding			4.8E-02	6
Organelle localization			7.0E-05	8
Organelle organization			0.0	69
Oxidation reduction			4.7E-04	21
Oxidoreductase activity			7.6E-04	20
Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen			2.3E-02	7
Oxidoreductase activity (on CH-NH2 group of donors)			2.3E-02	3
Regulation of cellular component organization			8.8E-04	16
Response to stimulus			1.4E-02	1
Regulation of catalytic activity			4.1E-03	23
Transferase activity			2.4E-02	31
Transferase activity (phosphorus-containing groups)			2.3E-02	25
Biopolymer metabolic process	0.0	133		
Cation binding	5.2E-03	80		
Cellular biosynthetic process	4.0E-05	95		
Cellular macromolecule metabolic process	0.0	137		
Cellular response to extracellular stimulus	2.5E-02	4		
Macromolecule biosynthetic process	0.0	90		
Negative regulation of intracellular transport	2.0E-02	3		
Neurological system process	1.1E-02	6		
Metal ion binding	5.2E-03	85		
Positive regulation of cytokine production	3.3E-02	4		
Protein complex binding	2.5E-02	7		
Protein dimerization activity	2.0E-02	15		
Signal transduction	3.8E-04	83		
Regulation of establishment of protein localization	1.6E-02	6		
Regulation of growth	5.0E-02	4		
Regulation of macromolecule metabolic process	0.0	91		
Regulation of primary metabolic process	0.0	92		
Regulation of protein localization	6.6E-03	7		

Response to drug	4.3E-02	5
Response to extracellular stimulus	4.9E-03	8
Response to oxygen levels	4.1E-02	5
Response to oxidative stress	1.5E-02	7
Regulation of cellular localization	8.8E-03	8
Regulation of cellular metabolic process	0.0	95
Regulation of biosynthetic process	0.0	85
Regulation of cell motion	3.3E-02	6
Transcription corepressor activity	2.8E-03	9
Transcription factor activity	4.7E-03	29
Transcription factor binding	1.4E-02	16

4.2 Gene expression differences between study groups

The 54 samples used in GE analysis were collected from 3 groups of individuals previously described in materials and methods—HR, RC and UC. GE differences between these groups were then compared using significance and cluster analysis. A tailor-made Enzyme-linked immunosorbent assay (ELISA) method was also used in order to estimate Ig concentrations in each sample.

4.2.1 Significance and cluster analysis

From observing our overall heatmap result of unsupervised hierarchical clustering (Figure 17) there seems to be no specific cluster of groups together. To take a closer look, days were separated and heatmaps were created (Figure A2 and A3, Appendix 2). No clear clustering was observed. To investigate further, p-values were used to determine the number of significantly differentially expressed genes between groups. At this point, since no cluster was observed, RC and UC were grouped together for statistical analysis. There were no significantly differentially expressed genes between controls and HR when adjusting for day effect or with p-values calculated for differences between groups for each day separately. Then p-values which were not adjusted for multiple testing were considered (Table 8). There were 198 genes expressed differently on day 0 and 44 genes on day 14 between HRs and controls at a p-value cut-off of 0.001. These were considered cautiously for interesting genes.

Table 8: Differently expressed genes between HRs and controls.

The numbers of genes expressed differently at specific p-values without adjusting for multiple testing between HRs and controls adjusted for day effect and on each day separately.

Significance value	HR vs Control	HR vs Control d0	HR vs Control d14
0.05	2957	4358	2214
0.01	561	1272	386
0.001	42	198	45

Unsupervised hierarchical clustering was performed and heatmaps were created for all samples and included only the genes different without adjust for multiple testing on each stimulation day (Figure 18 and 19). Within these figures groups placed together show visually noticeable differences; however only a small number of genes are being observed which can skew the perspective of these genes compared to the entire 45,000 and the number of these genes do fall into the number of genes which could be different by chance with the chosen p-value. For day 14 samples, only two genes had a FC greater than 1.0 and no genes with a FC higher or lower than ±2.0. Those two genes were *TGFB3* (transforming growth factor, beta 3) and *LOC387895* (hypothetical LOC387895*)*. Figure 20 shows the average of the 3 probe intensity for each HR and its corresponding controls for *TGFB3*. Figure 21 shows the average of the 3 probe intensity for each HR and its corresponding controls for *LOC387895*.

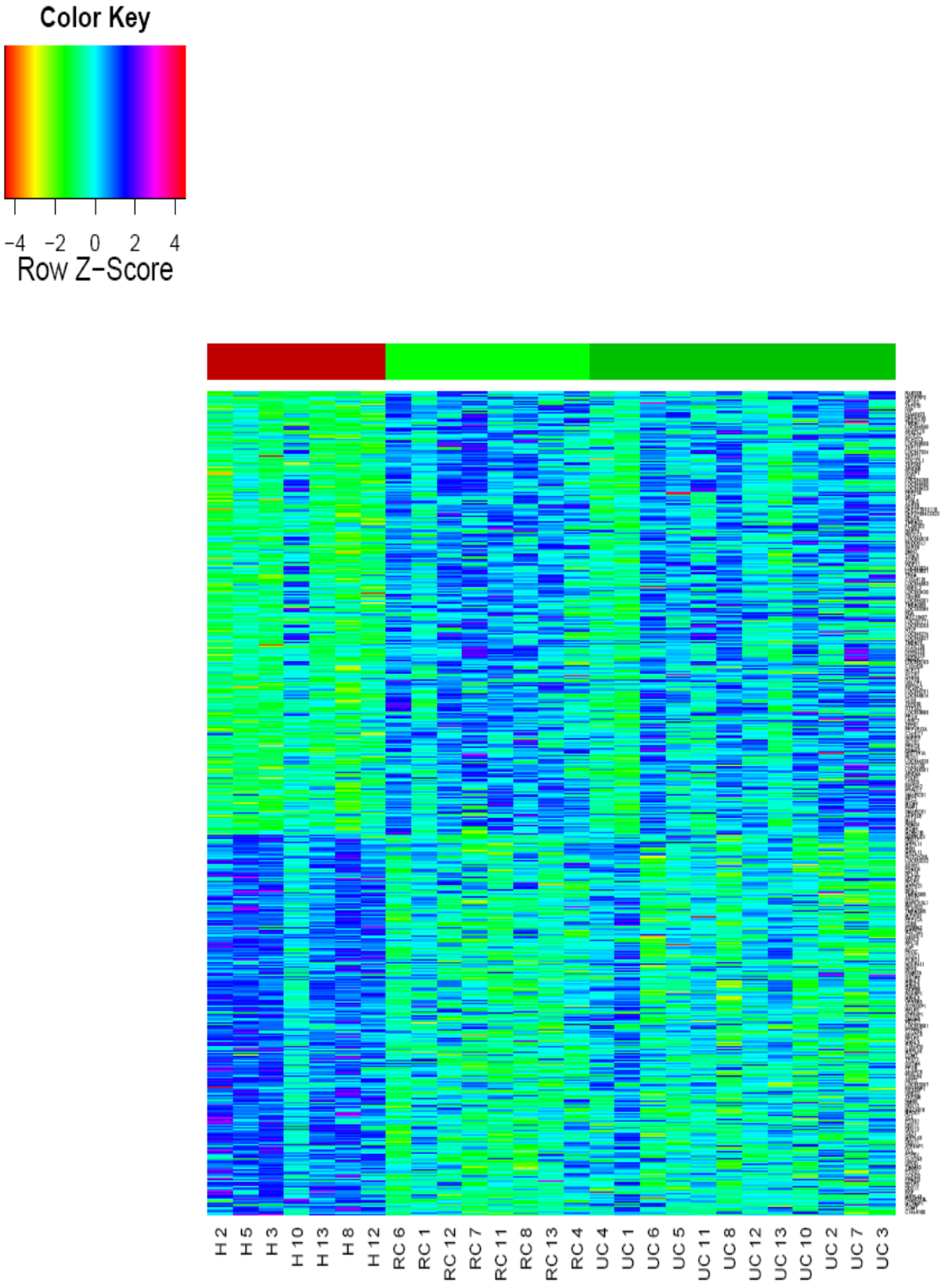


Figure 18: Heatmap for differently expressed genes between HRs and Controls, day 0*.

Heatmap of unsupervised hierarchical clustering of gene probes significantly differentially expressed without adjusting for multiple testing for day 0 samples. Red is H (HR), light green is RC and dark green is UC.

*without adjusting for multiple testing

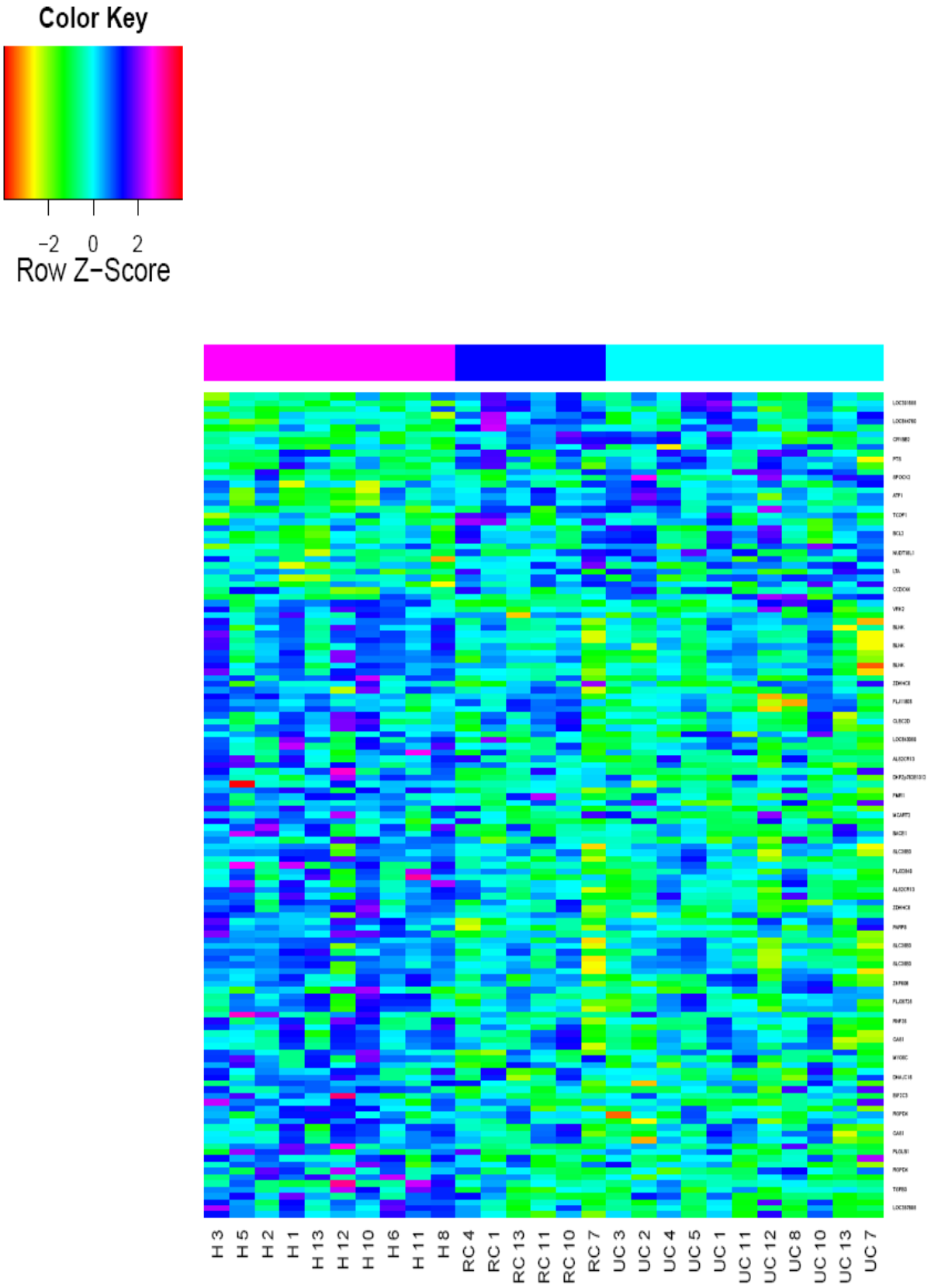


Figure 19: Heatmap for differentially expressed genes between HRs and Controls, day 14*.

Heatmap of unsupervised hierarchical clustering of 44 gene probes significantly differentially expressed without adjusting for multiple testing for day 14 samples. Pink is H (HR), dark blue is RC, and light blue is UC.

*without adjusting for multiple testing

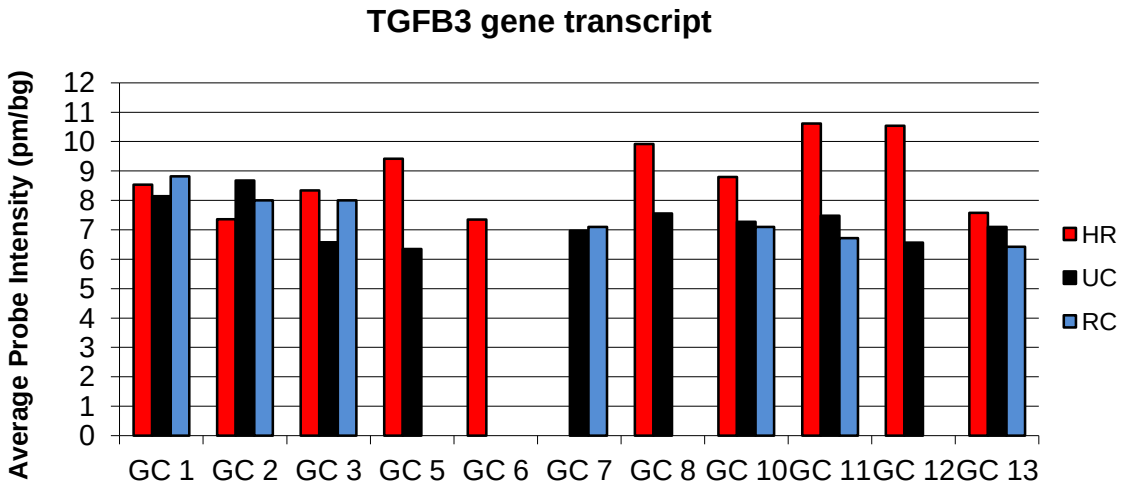


Figure 20: *TGFB3* gene transcript average probe intensities.

Average probe intensities for the *TGFB3* gene transcript for all available samples excluding outliers. HRs are in red, UCs in black, and RC in blue for each GC ID.

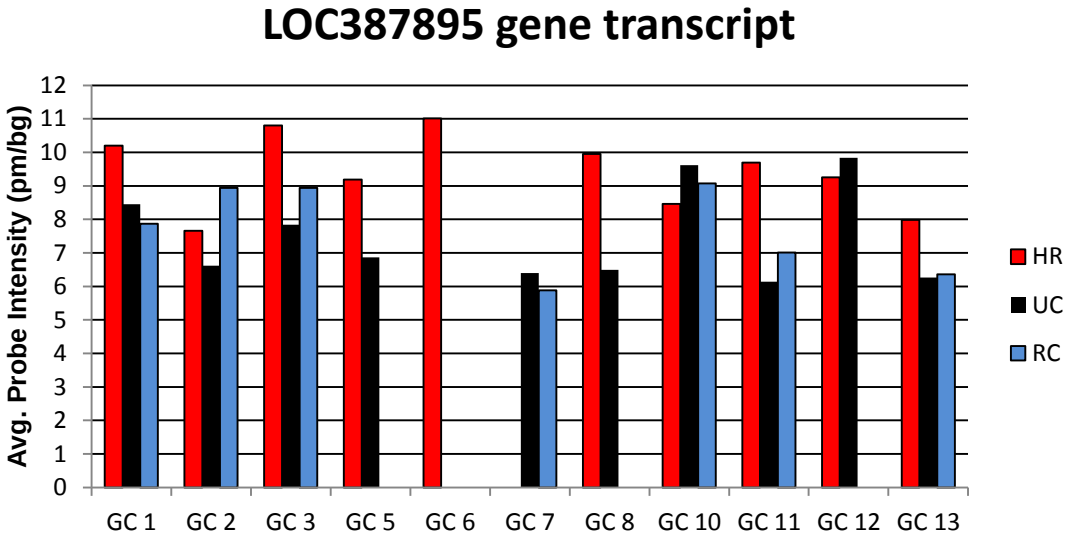


Figure 21: *LOC387895* gene transcript average probe intensities.

Average probe intensities for the *LOC387895* gene transcript for all available samples excluding outliers.

4.2.2 Immunoglobulin production in GC cultures

IgG and IgM concentration levels were measured by using ELISA methods for all samples collected from the *in vitro* GC cultures on days 7, 14, and 21 (see materials and methods). P-values resulting from the two-sample t-tests were all greater than a p = 0.05 significance cut-off value. Controls and HRs did not have significantly different Ig concentrations for IgM or IgG in this *in vitro* GC model. Concentration values were plotted on a line graph to show value changes over the three sample collection days (Figures A4-A9, Appendix 2).

5 DISCUSSION

5.1 Summary of main findings

In this study, B cells were isolated from 32 individuals before and after 14 days of stimulation within an *in vitro* germinal center (GC) culture. Of these 32 individuals, 11 were hyper-responders (HR), 9 were related controls (RC), and 12 were unrelated controls (UC). mRNA was extracted from 59 samples. From these, only 54 samples were used after excluding outliers. Maximum likelihood tests and unsupervised hierarchical cluster analysis was carried and revealed a large number of significantly differentially expressed genes and visualization of groups which cluster into similar gene expression (GE). The initial results showed a distinctive difference between unstimulated and 14 day stimulated samples. Calculations showed almost one-third (14,277) of all gene transcripts were significantly different. Those genes were then grouped into functional categories, using the Gene Ontology (GO) database, to investigate if the *in vitro* GC model mimics *in vivo* observations previously reported by other studies which is to be discussed below (Table 9). The top 50 up- and down-regulated genes (n=100) were considered for any interesting genes (Tables 5 and 6). Some of those genes were included among genes previously described with a GC role. When investigating the difference between HRs and all controls (both related and unrelated), there were no significantly differently expressed genes after adjusting for multiple testing. Genes which were expressed differently at a low p-value ($p < 0.001$) but not adjusted for multiple testing were investigated. These were cautiously considered for comparison to genes which have been previously described as being differently expressed in monoclonal gammopathies (MG) and no similar genes were noted. Two-sample t-tests of Ig concentration levels indicated that there is no significant difference in IgM or IgG concentration levels between HR and controls as a group in this *in vitro* GC model. This is at odds with previous results since the HRs were defined. This discrepancy could be explained by the differences between the poke-weed mitogen (PWM) stimulation and the *in vitro* GC stimulation.

5.2 Gene expression differences of *in vitro* GC model stimulation

There were a large number (14,277) of genes expressed differently between samples before and after 14 days of *in vitro* GC stimulation. This large number of genes was reduced based on the t-statistic in order to obtain a more meaningful analysis. Tables 5 and 6 shows the 50 most up- and down-regulated genes and their descriptive names. Not unexpectedly, we see CD40L included among the most up-regulated genes. Also included are many previously described GC related genes and expectedly involved with cell-cycle related events. Further investigation of these genes and supplementary tests will be needed (Tarca et al., 2006).

For this project, the 500 most up- and down-regulated genes (n=1,000) as selected by the t-statistic method were investigated using bioinformatics software, Onto-Express, that sorted them into functional categories using the GO database (Table 7) ("Onto-Tools," 2012). Function categories that often show up-regulation in GC gene expression patterns (GEP) include: DNA packaging, response to DNA damage stimulus, cell proliferation, and regulators of cell-cycle related events such as centrosome separation/segregation and cytokinesis (Klein & Dalla-Favera, 2008; Radbruch et al.,

2006; Shen et al., 2004). As can be seen from Table 7, data showed similar up-regulation in these categories. These categories are often cited because of their role in events occurring in the GC. Increased expression of DNA damage stimulus, DNA repair, DNA packaging are all categories expected in GC B cells because of double-stranded DNA breaks linked to SHM and CSR. Increased expression of cell proliferation associated genes and cell-cycle related genes is anticipated because of the dramatic expansion of proliferating B cells that occurs in the dark zone of a GC reaction (Figure 4) (Klein & Dalla-Favera, 2008; Shen et al., 2004). Regulation of cell death is another commonly cited function involved in the GC reaction because of the large amount of SHM mutations occurring and DNA repair pathways, such as the NHEJ mechanism are error-prone and apoptosis is needed for those which cannot be repaired correctly (Klein & Dalla-Favera, 2008; Roddam et al., 2010).

When comparing to other published studies, we only considered any genes with a p-value < $1.0e^{-8}$. That cut-off lowered our total number of genes to around 6,000 to make for slightly easier and more significant comparison of data. This pruning of data also allowed most FC values greater than -1.0 and less than 1.0 to be excluded. Any categorization of up- or down-regulated genes is based on these 6,000 genes. There are many examples of genes known to be associated with GC events within our list of differentially expressed genes and there are also many that do not have a known association with GC. Genes encoding for well-known GC associated TFs were often mentioned in review articles were not greatly up- or down regulated in the data including *BCL-6* (B-cell Lymphoma 6), *PAX5* (paired box 5), *IRF4* (interferon regulatory factor 4), *STAT3* (signal transducer and activator of transcription), *STAT5*, *CXCR5* (Chemokine (C-X-C motif) receptor 5), *NFKB1* (nuclear factor of kappa light polypeptide gene enhancer in B-cells 1) (Klein & Dalla-Favera, 2008; L. J. McHeyzer-Williams & McHeyzer-Williams, 2007).

A study performed by Shen et al, used cDNA microarrays to report genes which seem to have a specific expression for the human GC when compared to other major human B cell compartments. That study focused on functional categories which were predominately characterized by increased expression including: cell proliferation and quiescence; DNA repair, replication and protein synthesis; apoptosis and cell survival; and cytokines and chemokines and their receptors. The samples used in that study were *in vivo* micro-dissected mantle zone (MNZ) and GC compartments from the tonsils and FACS-sorted B cells. Samples for the marginal zone (MGZ) compartment were isolated the same way from human spleen tissue. Genes that were significantly differentially expressed and have a FC greater than two when compared to the other two compartments were considered specific to that compartment (Shen et al., 2004). Although the Shen et al. study does not reflect the same methods or materials as those used within our present study, it was used as a good general comparison for genes appearing to be specific to the human *in vivo* GC compartment that might also be found in our data. Table 9 shows the gene symbol and FC values within our analysis which shared an up- or down-regulated expression to *in vivo* observations. As can be seen there are many genes within multiple categories that are similar in expression.

Table 9: Similar GE in previous publication.

Genes that share similar expression with previously published *in vivo* human GC specific genes. Positive fold changes indicate genes that are expressed higher after 14 day stimulation with *in vitro* GC and correlates to up-regulation. Negative fold changes indicate genes that are expressed higher on day 0 and correlates to down-regulation (Shen et al., 2004).

Ontology Category	***Gene symbol***	***Fold change (FC)***
Cell proliferation		
	BUB1	3.43
	CCNA2	5.14
	CCNB1	4.90
	Cyclin F	1.80
	CDK1	3.86
	CDKN2c	1.93
	CENPA	2.02
	CENPE	3.60
	CENPF	4.61
	HMGB2	1.26
	HMMR	4.62
	MAD2L1	2.87
	MAP1B	3.06
	MAP2K6	2.33
	PCNA	3.61
	PPIL5	2.68
	TTK	2.23
	CbLB	-2.57
DNA repair, replication, and protein synthesis		
	CHEK1	2.61
	DDB1	1.46
	DNA2L	2.41
	EXO1	4.53
	LMNB1	2.16
	MCM2	2.09
	MSH2	2.27
	POLE2	2.21
	RAD51	4.10
	RFC4	1.97
	RRM1	2.13
	TCF19	2.17
	TOP2A	5.70
	UBE2C	5.0
	UHRF1	3.70
	UMPS	1.78
Apoptosis and cell survival		
	FAIM	1.32
	PDCD8	1.98
	TNFRSF17	2.88
	BCL2	-1.19
	PDCD4	-1.32

Cytokines and chemokines and their receptors		
	CXCL10	2.61
	SOCS1	2.16
	AKAP13	-2.26
	CCL3	-1.80
	IL7R	-1.73
	MX2	-3.27
Other unique GC markers		
	AICDA	7.12
	CD59	2.76
	GCET2	1.75
	XBP1	1.48
	CD36	-2.70
	TXNIP	-2.94

Notable genes within the apoptosis function category were observed in both studies. Four of the 15 genes listed as significant increase or decrease in expression for the *in vivo* GC compartment were similarly expressed in this study's data. *PDCD8* (Programmed cell death 8) has a role in regulation of apoptosis and showed increased expression as (Jourdan et al., 2009). *PDCD4* (Programmed cell death 4 PDCD4) was within the top 500 down-regulated genes and with a FC values less than -2.7. The role of *PDCD4* is to be an inhibitor of translation by inhibiting eIF4A helicase activity (Yang et al., 2003). This repression of *PDCD4* could therefore facilitate proliferation of centroblasts (Shen et al., 2004). *TNFRSF17* (Tumor-necrosis factor (TNF) receptor superfamily, member 17), a known promoter of B cell survival when up-regulated and preferentially expressed in mature B cells, was within the top 2,000 up-regulated genes with a FC value of 3.03 (Hatzoglou et al., 2000). This imitates *in vivo* observations by Shen et al. and might promote cell expansion that is seen within normal GC reactions (Shen et al., 2004). Table 9 also reveals a similar slightly decreased expression of *BCL-2* (B-cell Lymphoma 2) in samples as does the *in vivo* expression. Over expression of the *BCL-2* gene can results in resistance to apoptosis (Coultas & Strasser, 2003). Apoptosis is complex and includes a variety of apoptotic and survival signals to create a homeostasis; therefore we would expect regulation of apoptosis to include both types of signals (Jourdan et al., 2009).

Cell proliferation was well represented within both *in vivo* and *in vitro* results. Almost 43% of the genes listed specific to the GC compartment showed a similar increased or decreased expression in this study's results (Table 9). Cyclins (CCN) including: *CCNA2*, *CCNB1*, *CCNF*, and *CDK1* (cyclin-dependent kinase 1) were expressed higher in samples from the *in vitro* GC than day 0. These cyclins seem to work together forming a complex of cyclin-cyclin interactions. CCNA2 expression is usually low in G_0 phase and begins to increase in early G_1. In order for the cell to enter G2/M phase, CDK1 is required. CCNB1 is needed to form a complex with CDK1 for the transition and ability to relocate to the nucleus. Nuclear localization is assisted by CCNF which is within our top 800 most up-regulated genes (Kong, Barnes, Ollendorff, & Donoghue, 2000). All of these genes encoding for these cyclins were highly expressed in our data with *CDK1* in the top 10 most up-regulated genes and *CCNA2* and *CCNB1* included in the top 150 most up-regulated genes. Shen et al., found a lower expression of these cyclin genes in the MNZ and MGZ compared to the GC. This is probably due to the resting state

of MNZ and MGZ. Genes that are also associated with cells in a resting state are *CCNG2* and CD72. *CCNG2* was highly expressed in MNZ within the *in vivo* study (Shen et al., 2004). CCNG2 is associated with cell cycle arrest and functions independently of p53 (Horne et al., 1997). We did not find a high or low expression of *CCNG2* which might supports the highly active state of the GC reaction. Higher expression of CD72 may be involved in the quiescent status in MNZ and is found to have decreased expression in the GC (Shen et al., 2004). *CD72* was also shown to be down regulated in this study's *in vitro* results. This is good support for our culture model because we had high expression of these genes involved with cell proliferation, as well as decreased expression of *CD72* and no large change of *CCNG2* that are involved in maintaining a resting state. This implies that our *in vitro* expression patterns mimic the highly proliferating GC expectation.

Seven of the 39 genes associated with cytokines and chemokines and their receptors found to be expressed differently in the *in vivo* GC compartment, also appeared as up- or down-regulated in this project's data (Shen et al., 2004). Many of these genes are related to relevant biological functions. It is important to note that the smaller number of genes listed as similar in this or any category does not imply the other genes expressed opposite intensities from *in vivo* observations. However, our data revealed a few interesting genes within this category that were not highlighted in the Shen et al. article. Human naïve and memory B cells express CCR6 (Chemokine (C-C motif) receptor 6) and expression is mostly absent in GC B cells (Schutyser, Struyf, & Van Damme, 2003). A significant down-regulation in *CCR6* was observed in our top 1,000 most down-regulated genes (FC = -3.0). Another interesting chemokine receptor was also significantly down-regulated. CXCR4 (Chemokine receptor 4) is present on variety of cells, including expression on centroblast and later on PCs migrating to the BM as reviewed by Radbrunch et al (Radbruch et al., 2006). *CXCR4* was within the top 100 down-regulated genes indicating that expression decreases within the 14 days of *in vitro* GC culture stimulation. CXCR4 is thought to decrease in expression as differentiation of plasmablasts occurs. It has been shown that centroblasts in the dark zone of the GC express CXCR4 whereas centrocytes have a loss of CXCR4 expression. It is proposed that this is a better indicator of movement into the light zone than the previously described $CD77^+$ (Caron, Le Gallou, Lamy, Tarte, & Fest, 2009). In a study by Xiang et al., *FCGR2B* (Fc fragment of IgG, low affinity IIb, receptor) was found to be expressed higher in PCs than on mature B cells or less-differentiated PCs. This gene, much like *CXCR4*, is within our 100 most down-regulated genes (FC = -4.5). As seen from our Ig concentration measurements, it seems as though class switching might not have occurred in most cells. Perhaps our *in vitro* model has not reached a large number of cells having undergone CSR or differentiation into plasmablasts by day 14 which is discussed later with Ig concentrations resulting from ELISA methods. Isolation of RNA samples at more days of stimulation as well as qRT-PCR of stage specific factors might give some insight into the GC model.

Almost 45% of genes associated with DNA repair, replication and protein synthesis that were significantly increased or decreased in expression for the *in vivo* GC compartment were similarly expressed in this study's data (Table 9). This is anticipated due the DNA damage that occurs during GC events (Klein & Dalla-Favera, 2008). A few interesting and well known genes from these categories include, *UBE2C* (ubiquitin-conjugating enzyme E2C), *TOP2A* (topoisomerase (DNA) II

alpha), *RAD51* (*RAD51* homolog (*S. cerevisiae*)) and *BRCA1* (breast cancer 1, early onset). *UBE2C* belongs to proliferative genes known to comprise the majority of genes included in prognostic GE signatures in cancer (Desmedt et al., 2008; Wirapati et al., 2008). It is often involved in cellular events through protein regulation, fate and unfolding. *TOP2A*, *BRCA1*, and *RAD51* are all among the 250 most up-regulated genes in data. TOP2A is a part of a family of ATP-dependent enzymes that catalyze topological changes in DNA via DSBs. TOP2A expression is greatly increased in proliferating cells and decreases as cell differentiate (Austin & Marsh, 1998). In DNA homologous recombination to repair DSBs, RAD51 participates by assembling with the single-stranded DNA with the help of mediators for example BRCA1. BRCA1 is a tumor suppressor that functions in DNA homologous recombination and is often reported mutated in breast and ovarian cancers. BRCA1 has been linked to a number of cellular functions including DNA damage repair, cell cycle checkpoint, transcriptional regulation, DNA replication, centrosome function, among others as reviewed by Pageau et al (Pageau, Hall, Ganesan, Livingston, & Lawrence, 2007). The similar up-regulation of these genes which interact together and are involved in a relevant functional category in GC events further supports this model as functional and could suggest along with other GE evidence that by day 14, CSR has not yet occurred on a large scale in samples.

Shen et al. reported 41 other genes unique to the GC compartment. Of those six were similar in increased or decreased expression (Table 9). Two of the most interesting genes included *GCET2* and *AICDA*. *GCET2* (germinal center expressed transcript 2) showed similar up-regulated expression. *GCET2* is highly expressed in normal GC B cells and is one of the most representative genes in the GC B cell signature (Lossos, Alizadeh, Rajapaksa, Tibshirani, & Levy, 2003; Zenggang Pan, 2007). *AICDA* (activation-induced cytidine deaminase) was the most up-regulated gene in our data. The protein for which this gene encodes (AID) is known to be involved in SHM, gene conversion, and CSR of Ig genes. Defects in this gene are the cause of recessive hyper-IgM immunodeficiency syndrome (Di Noia & Neuberger, 2007; Kim, Lim, Kang, Hillsamer, & Kim, 2005; Masayuki Kuraoka, 2011). It has also been reported to remove developing autoreactive B cells in the GC (Greta Meyers, 2011). Both of these are supportive for our GC culture model functioning as GC which follows many expression patterns of *in vivo* observations.

These are just a few examples of previously published genes related to protein functions that are highly involved in GC reaction events. There are some limitations that are important to keep in mind. Protein expression and GE are not always parallel (John D. Shaughnessy Jr, 2003). The comparison we are making is between B cells isolated from peripheral blood mononuclear cells (PBMN) and those stimulated after 14 days in an *in vitro* GC. Also, we are only looking closely a small amount of data and there is much that is left to consider. It is also important to remember that this designed GC could need more work. The next step is to take a closer look at genes which we found to be unique in our data to the GC. This will take more time and supplementary studies and quantification in order to achieve. Overall, it can be seen that the similarities in GE data collected compared to previously published results of *in vivo* observations in the GC reaction and biological knowledge of gene functions supports that our designed *in vitro* GC culture model closely mimics the *in vivo* GC.

5.3 Differences between HR and controls

Previously, individuals classified with elevated levels of Ig were termed HRs within eight Icelandic families with multiple cases of Monoclonal Gammopathy of Undetermined Significance (MGUS), Multiple Myeloma (MM), Waldenström's Macroglobulinemia (WM) and/or another lymphoproliferative disorder. MGs contain step-wise transformation from benign MGs to malignant MM, such as WM or MM. The risk for development increases among relatives of patients. Individuals classified with elevated levels of Ig production in PWM stimulation were termed HRs. These were identified in four of the eight Icelandic families with multiple cases of MGUS, MM, WM and/or another lymphoproliferative disorders. GE results from these 11 HR-classified individuals were statistically compared to controls. Results showed no significant difference between the groups. There were a small number of genes significantly differentially expressed without adjusting for multiple testing on separate days which were investigated further. Ig production did not support differences between HRs and controls either.

5.3.1 Differently expressed genes between HR and controls

There were no significant differences between HR and controls. There were some differences when multiple testing was not taken into account at p-value < 0.001. These differences were considered for day 0 (198 genes) and day 14 (45 genes) separately to investigate if any were note worthy for future studies done with these samples. Day 14 differences gave two genes with a FC greater than 1.0 and no genes with a FC higher or lower than ±2.0. Figures 20 and 21 show probe intensities for all available samples for these two genes. It can be clearly seen that for all these genes HRs intensities are above those of controls. This is not significant but worth noting for future studies. Those genes were *TGFB3* (transforming growth factor, beta 3) located on chromosome 14 and a hypothetical gene of unknown function (*LOC387895*) located on chromosome 12. *TGFB3* is a gene which has been documented with a role in growth, differentiation, and maturation of B lymphocytes. The effects of TGF-β depend on concentration, stage of differentiation in B cells and costimuli. These effects include stimulation of programmed cell death, hindrance of proliferation, and isotype switching (Lebman & Edmiston, 1999). TGF-β can inhibit proliferation, through cell cycle arrest in the G1 phase, of human B cells activated by both mitogen stimulation and CD40L. B cells produce TGF-β themselves which might imply that *in vivo* results are different from those seen *in vitro* (Lebman & Edmiston, 1999). Seemingly, we might expect to see down-regulation in HRs compared with controls because of increased Ig levels observed for HRs in previous studies and the inhibition of proliferation TGF-β has on B cells (Steingrimsdottir et al., 2011; Ögmundsdóttir et al., 2005). However, since we do not see a difference in Ig production in HRs compared to control then perhaps this gene is functioning as a balance to the previously established elevated Ig concentrations. Similar observations of a TGF-β suppressive function was noted a study done by Kim et al (Kim et al., 2005). There is currently no published information on unknown hypothetical gene *LOC387895*. However, it did have a FC of 1.78 indicating a higher expression in HRs and could possibly serve as a candidate gene with repeated or further research.

Day 0 differences (198 genes) between HRs and controls without adjusting for multiple testing were also investigated and no genes were reported with a FC higher or lower than ±2.0. One gene did

have a FC value different from most others (FC = -1.83), *RAB39B*. It is a member of the RAS oncogene family and has a possible role in vesicular trafficking. This could be of interest because Ras-encoded proteins are a part of signal transduction pathways that are involved with cell proliferation and differentiation in response to external signals (Cheng, 2002). There is no published literature relating *RAB39B* to Ig production or B cells.

There were no previously published genes found to be differently expressed between MGUS and normal PCs among our differently expressed genes without adjusted for multiple testing (Chng et al., 2011; Davies et al., 2003; Korac et al., 2009; Leleu et al., 2009; Zhan et al., 2007). This *in vitro* GC model is different from the PWM model. This could account for why we do not see more marked differences in HRs. A previous graduate students, Sóley Valgeirsdóttir, work with the model showed that it does work as seen from the surface markers and transcription factors. However, transcription factors can exhibit fluctuations over weeks as also seen in Sóley‘s work with this GC model. This projects data is only taking a snapshot of one moment in the models reaction. Therefore collecting samples as a later time point might give further insight into differences between HRs and controls, as well as before and after stimulation within the model.

5.3.2 Immunoglobulin production in the GC model

IgG and IgM concentrations were measured for all samples collected on days 7, 14, and 21 (see materials and methods). As can be seen from Figures A4-A9, the IgG and IgM concentration values did not change greatly, over time, for most samples. In a normal GC reaction we would expect to see IgM concentration levels eventually decreasing and IgG concentration levels increasing as CSR takes place and differentiation into PCs or memory B cells occurs. There could be a few reasons for why this does not occur in the GC model. It might be that our model takes longer than expected for class switching to occur as well as the fact that the main site for Ig production is not in the GC but in the BM. Compared to the PWM stimulation originally used to identify and confirm the HR’s phenotype, the *in vitro* GC culture may provide better survival conditions, allowing the cells to survive longer (Ogmundsdottir et al., 2011). This slower culture could account for why we did not see an increase in IgG production by day 21 and levels were similar between time points. Another possibility could be that some aspect of the GC reaction is missing within the cultures. In the original PWM stimulation, the cultured cells were from the whole PBMN population providing cells, such as T cells which provide other cytokines that may be important to the Ig production and are absent from the *in vitro* GC model. A previous graduate students, Sóley Valgeirsdóttir, work with the model showed that it does work as seen from the surface markers and transcription factors. GE results reported here also indicate that the *in vitro* GC does mimic *in vivo* observations. It is also possible that the HR phenotype is unstable and that individuals previously defined as HR have lost the phenotype.

Using a two-sample t-test in the R-program, HRs did not significantly differ from controls in Ig production levels in this *in vitro* GC model. TGF-β expression, which has been reported to suppress IgG production and secretion of IgM, was elevated in HR when not adjusting for multiple testing, implying active feedback control (Richard J. Armitage, 1993). Although it is recognized that this is not significant according to statistical adjustments, it is worth nothing that this might have played a role in

Ig production. If Ig production was slightly inhibited in HR but not in controls then perhaps this could account for similar values between HR and controls. Finally, it is important to consider that the *in vitro* GC might not be a good method of testing this HR phenotype and the PWM simulation model is more reliable form of identification.

5.4 Conclusions

From the GE data and Ig concentrations collected in this project, we have gained further insight into our *in vitro* GC model and the possible HR phenotype.

The *in vitro* GC model seems to mimic the *in vivo* GC reaction as one-third of the genes have significant differences in expression between day 0 and surviving B cells at 14 days of stimulation and many of those genes have been reported within *in vivo* GCs. Further investigation of these genes allows for comparisons to previously published data which reveals some similarities and differences with the *in vivo* GC. Surface markers measured in previously collected data from this GC model also indicated that the model functions similar to *in vivo (Steingrimsdottir et al., 2011)*. When the 500 most up- and down- regulated genes (n = 100) were classified into relevant ontology categories such as DNA repair, replication, cell proliferation, apoptosis, and cell survival, the GC model closely mimicked the previously published *in vivo* GC data. The *in vitro* GC designed and used for this study could be improved to fit an even more realistic version. It is also important to note that GEPs are not always parallel with protein levels which are the functional artifact of the mRNA blueprint. Complimentary studies are necessary to further confirm GE data (John D. Shaughnessy Jr, 2003; Tarca et al., 2006).

HRs do not show overall significant differences in gene expression compared to controls; however, data unadjusted for multiple testing revealed two genes expressed differently between the two groups for day 14 (*TGFB3* and *LOC387895*) and one for day 0 (*RAB39B*) The lack of significantly differentially expressed genes could be due to the differences between classification with the PWM and the GC model used in this study. This GC model allows for a longer survival of cells than the PWM. It could be that the PWM is a better test for the HR phenotype than the GC. It seems through the investigation of particular genes encoding for transcription factors, that taking samples at a later date will give more insight into the *in vitro* GC model. Some GE data and the Ig measurements indicate that class switching might not have occurred yet by day 14 in the *in vitro* GC model.

REFERENCES

Andraud, M., Lejeune, O., Musoro, J. Z., Ogunjimi, B., Beutels, P., & Hens, N. (2012). Living on three time scales: the dynamics of plasma cell and antibody populations illustrated for hepatitis a virus. *PLoS Comput Biol, 8*(3), e1002418. doi: 10.1371/journal.pcbi.1002418

Andrea Cerutti, H. Z., Andras Schaffer, Leif Bergsagel, Nagaradona Harindranath, Edward E. Max and Paolo Casali. (1998). CD40 Ligand and Appropriate Cytokines Induce Switching to IgG, IgA, and IgE and Coordinated Germinal Center and Plasmacytoid Phenotypic Differentiation in a Human Monoclonal IgM +IgD+ B Cell Line. *Journal of Immunology, 160*, 2145-2157.

Ansell, S. M., Kyle, R. A., Reeder, C. B., Fonseca, R., Mikhael, J. R., Morice, W. G., . . . Gertz, M. A. (2010). Diagnosis and management of Waldenstrom macroglobulinemia: Mayo stratification of macroglobulinemia and risk-adapted therapy (mSMART) guidelines. [Review]. *Mayo Clin Proc, 85*(9), 824-833. doi: 10.4065/mcp.2010.0304

Austin, C. A., & Marsh, K. L. (1998). Eukaryotic DNA topoisomerase II beta. *Bioessays, 20*(3), 215-226. doi: 10.1002/(sici)1521-1878(199803)20:3<215::aid-bies5>3.0.co;2-q

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society, 57*(1), 289-300.

Bjornsson, O. G., Arnason, A., Gudmunosson, S., Jensson, O., Olafsson, S., & Valimarsson, H. (1978). Macroglobulinaemia in an Icelandic family. *Acta Med Scand, 203*(4), 283-288.

Caron, G., Le Gallou, S., Lamy, T., Tarte, K., & Fest, T. (2009). CXCR4 expression functionally discriminates centroblasts versus centrocytes within human germinal center B cells. *J Immunol, 182*(12), 7595-7602. doi: 10.4049/jimmunol.0804272

Cheng, H. Y. M. X. N. J., M; et al. (2002). Isolation and characterization of a human novel RAB (RAB39B) gene. *Cytogenetic and Genome Research, 97*, 72-75.

Chng, W. J., Huang, G. F., Chung, T. H., Ng, S. B., Gonzalez-Paz, N., Troska-Price, T., . . . Fonseca, R. (2011). Clinical and biological implications of MYC activation: a common difference between MGUS and newly diagnosed multiple myeloma. *Leukemia, 25*(6), 1026-1035. doi: 10.1038/leu.2011.53

Chng, W. J., Schop, R. F., Price-Troska, T., Ghobrial, I., Kay, N., Jelinek, D. F., . . . Bergsagel, P. L. (2006). Gene-expression profiling of Waldenstrom macroglobulinemia reveals a phenotype more similar to chronic lymphocytic leukemia than multiple myeloma. [Comparative Study Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Blood, 108*(8), 2755-2763. doi: 10.1182/blood-2006-02-005488

Coultas, L., & Strasser, A. (2003). The role of the Bcl-2 protein family in cancer. *Seminars in Cancer Biology, 13*(2), 115-123. doi: 10.1016/s1044-579x(02)00129-3

Croker, B. A., O'Donnell, J. A., Nowell, C. J., Metcalf, D., Dewson, G., Campbell, K. J., . . . Roberts, A. W. (2011). Fas-mediated neutrophil apoptosis is accelerated by Bid, Bak, and Bax and inhibited by Bcl-2 and Mcl-1. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *Proc Natl Acad Sci U S A, 108*(32), 13135-13140. doi: 10.1073/pnas.1110358108

Davies, F. E., Dring, A. M., Li, C., Rawstron, A. C., Shammas, M. A., O'Connor, S. M., . . . Anderson, K. C. (2003). Insights into the multistep transformation of MGUS to myeloma using microarray expression analysis. [Comparative Study Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Research Support, U.S. Gov't, P.H.S.]. *Blood, 102*(13), 4504-4511. doi: 10.1182/blood-2003-01-0016

Desmedt, C., Haibe-Kains, B., Wirapati, P., Buyse, M., Larsimont, D., Bontempi, G., . . . Sotiriou, C. (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. *Clin Cancer Res, 14*(16), 5158-5165. doi: 10.1158/1078-0432.ccr-07-4756

Di Noia, J. M., & Neuberger, M. S. (2007). Molecular mechanisms of antibody somatic hypermutation. [Review]. *Annu Rev Biochem, 76*, 1-22.

Draghici, S. (2002). Statistical intelligence: effective analysis of high-density microarray data. *Drug Discovery Today, 7*(11), 55-63.

Gene Ontology Database. (2012). Retrieved 2011-2012 http://www.geneontology.org/GO.downloads.database.shtml

Gibson, G., & Muse, S. V. (2009). *A Primer of Genome Science* (3rd ed.). Massachusetts, USA: Sinauer Associates, Inc. Publishers.

Greta Meyers, Y.-S. N., Jason M. Bannock, Aubert Lavoie, Jolan E. Walter, Luigi D. Notarangelo, Sara S. Kilic, Guzide Aksu, Marianne Debré, Frédéric Rieux-Laucat, Mary Ellen Conley, Charlotte

Cunningham-Rundles, Anne Durandy, and Eric Meffre,. (2011). Activation-induced cytidine deaminase (AID) is required for B-cell tolerance in humans. *Proceedings of the National Academy of Sciences, 108*(28), 11554-11559.

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

Hasler, G., Drevets, W. C., Gould, T. D., Gottesman, II, & Manji, H. K. (2006). Toward constructing an endophenotype strategy for bipolar disorders. *Biol Psychiatry, 60*(2), 93-105. doi: 10.1016/j.biopsych.2005.11.006

Hatzoglou, A., Roussel, J., Bourgeade, M. F., Rogier, E., Madry, C., Inoue, J., . . . Tsapis, A. (2000). TNF receptor family member BCMA (B cell maturation) associates with TNF receptor-associated factor (TRAF) 1, TRAF2, and TRAF3 and activates NF-kappa B, elk-1, c-Jun N-terminal kinase, and p38 mitogen-activated protein kinase. *J Immunol, 165*(3), 1322-1330.

Helga M. Ögmundsdóttir, V. H., Guðmundur M. Jóhannesson,, & Guðríður Ólafsdóttir, K. B., Helgi Sigvaldason and Hrafn Tulinius. (2002). Monoclonal gammopathy in Iceland: a population-based registry and follow-up. *British Journal of Haematology, 118*, 166-173.

Hideshima, T., Mitsiades, C., Tonon, G., Richardson, P. G., & Anderson, K. C. (2007). Understanding multiple myeloma pathogenesis in the bone marrow to identify new therapeutic targets. [Review]. *Nat Rev Cancer, 7*(8), 585-598. doi: 10.1038/nrc2189

Horne, M. C., Donaldson, K. L., Goolsby, G. L., Tran, D., Mulheisen, M., Hell, J. W., & Wahl, A. F. (1997). Cyclin G2 is up-regulated during growth inhibition and B cell antigen receptor-mediated cell cycle arrest. *J Biol Chem, 272*(19), 12650-12661.

Htwe, S. S., Maeda, M., Matsumoto, R., Sakamoto, N., Murakami, S., Yamamoto, S., . . . Otsuki, T. (2011). Quick detection of overexpressed genes caused by myeloma-specific chromosomal translocations using multiplex RT-PCR. *Int J Mol Med, 27*(6), 789-794. doi: 10.3892/ijmm.2011.648

IMWG. (2003). Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. *Br J Haematol, 121*(5), 749-757.

Irizarry, R. A., Hobbs, B., Collin, F., & Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics, 4*(2), 249-264.

John D. Shaughnessy Jr, B. B. (2003). Interpreting the molecular biology and clinical behavior of multiple myeloma in the context of global gene expression profiling. *Immunological Reviews, 194*, 140-163.

Jourdan, M., Reme, T., Goldschmidt, H., Fiol, G., Pantesco, V., De Vos, J., . . . Klein, B. (2009). Gene expression of anti- and pro-apoptotic proteins in malignant and normal plasma cells. *Br J Haematol, 145*(1), 45-58. doi: 10.1111/j.1365-2141.2008.07562.x

Kadota, K., & Shimizu, K. (2011). Evaluating methods for ranking differentially expressed genes applied to microArray quality control data. [Evaluation Studies Research Support, Non-U.S. Gov't]. *BMC Bioinformatics, 12*, 227. doi: 10.1186/1471-2105-12-227

Kendler, K. S., & Neale, M. C. (2010). Endophenotype: a conceptual analysis. [Research Support, N.I.H., Extramural Review]. *Mol Psychiatry, 15*(8), 789-797. doi: 10.1038/mp.2010.8

Khatri, P., Voichita, C., Kattan, K., Ansari, N., Khatri, A., Georgescu, C., . . . Draghici, S. (2007). Onto-Tools: new additions and improvements in 2006. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *Nucleic Acids Res, 35*(Web Server issue), W206-211. doi: 10.1093/nar/gkm327

Kim, J. R., Lim, H. W., Kang, S. G., Hillsamer, P., & Kim, C. H. (2005). Human CD57+ germinal center-T cells are the major helpers for GC-B cells and induce class switch recombination. [Research Support, Non-U.S. Gov't]. *BMC Immunol, 6*, 3. doi: 10.1186/1471-2172-6-3

Klein, U., & Dalla-Favera, R. (2008). Germinal centres: role in B-cell physiology and malignancy. [Review]. *Nat Rev Immunol, 8*(1), 22-33. doi: 10.1038/nri2217

Kong, M., Barnes, E. A., Ollendorff, V., & Donoghue, D. J. (2000). Cyclin F regulates the nuclear localization of cyclin B1 through a cyclin-cyclin interaction. *EMBO J, 19*(6), 1378-1388. doi: 10.1093/emboj/19.6.1378

Korac, P., Peran, I., Skrtic, A., Ajdukovic, R., Kristo, D. R., & Dominis, M. (2009). FOXP1 expression in monoclonal gammopathy of undetermined significance and multiple myeloma. [Research Support, Non-U.S. Gov't]. *Pathol Int, 59*(5), 354-358. doi: 10.1111/j.1440-1827.2009.02377.x

Kuehl, W. M., & Bergsagel, P. L. (2002). Multiple myeloma: evolving genetic events and host interactions. [Review]. *Nat Rev Cancer, 2*(3), 175-187. doi: 10.1038/nrc746

Kunkel, E. J., & Butcher, E. C. (2003). Plasma-cell homing. *Nat Rev Immunol, 3*(10), 822-829. doi: 10.1038/nri1203

Kyle, R. A., & Rajkumar, S. V. (2010). Monoclonal gammopathy of undetermined significance and smoldering multiple myeloma. [Research Support, N.I.H., Extramural Review]. *Curr Hematol Malig Rep, 5*(2), 62-69. doi: 10.1007/s11899-010-0047-9

Lebman, D. A., & Edmiston, J. S. (1999). The role of TGF-beta in growth, differentiation, and maturation of B lymphocytes. *Microbes Infect, 1*(15), 1297-1304.

Leleu, X., Hunter, Z. R., Xu, L., Roccaro, A. M., Moreau, A. S., Santos, D. D., . . . Treon, S. P. (2009). Expression of regulatory genes for lymphoplasmacytic cell differentiation in Waldenstrom Macroglobulinemia. [Research Support, Non-U.S. Gov't]. *Br J Haematol, 145*(1), 59-63. doi: 10.1111/j.1365-2141.2009.07592.x

Liang, X. S., Caporaso, N., McMaster, M. L., Ng, D., Landgren, O., Yeager, M., . . . Goldin, L. R. (2009). Common genetic variants in candidate genes and risk of familial lymphoid malignancies. [Research Support, N.I.H., Intramural]. *Br J Haematol, 146*(4), 418-423. doi: 10.1111/j.1365-2141.2009.07790.x

Liu, Y. C., Miyazawa, K., Sashida, G., Kodama, A., & Ohyashiki, K. (2006). Deletion (20q) as the sole abnormality in Waldenstrom macroglobulinemia suggests distinct pathogenesis of 20q11 anomaly. *Cancer Genet Cytogenet, 169*(1), 69-72. doi: 10.1016/j.cancergencyto.2006.03.013

Lossos, I. S., Alizadeh, A. A., Rajapaksa, R., Tibshirani, R., & Levy, R. (2003). HGAL is a novel interleukin-4-inducible gene that strongly predicts survival in diffuse large B-cell lymphoma. *Blood, 101*(2), 433-440. doi: 10.1182/blood-2002-06-1931

MacLennan, I. C. M., Toellner, K.-M., Cunningham, A. F., Serre, K., Sze, D. M.-Y., iga, E. Z. n., . . . Vinuesa, C. G. (2003). <Maclennan2003.pdf>. *Immunological Reviews, 194*, 8-18.

Maddaly, R., Pai, G., Balaji, S., Sivaramakrishnan, P., Srinivasan, L., Sunder, S. S., & Paul, S. F. (2010). Receptors and signaling mechanisms for B-lymphocyte activation, proliferation and differentiation--insights from both in vivo and in vitro approaches. *FEBS Lett, 584*(24), 4883-4894. doi: 10.1016/j.febslet.2010.08.022

Mansoor, A., Medeiros, L. J., Weber, D. M., Alexanian, R., Hayes, K., Jones, D., . . . Bueso-Ramos, C. E. (2001). Cytogenetic findings in lymphoplasmacytic lymphoma/Waldenstrom macroglobulinemia. Chromosomal abnormalities are associated with the polymorphous subtype and an aggressive clinical course. *Am J Clin Pathol, 116*(4), 543-549. doi: 10.1309/6u88-357u-ukj5-ypt3

Masayuki Kuraoka, T. M. H., Dongmei Liao, Mandy Womble, Derek W. Cain, Alexander E. Reynolds, and Garnett Kelsoe. (2011). Activation-induced cytidine deaminase mediates central tolerance in B cells. *Proceedings of the National Academy of Sciences, 108*(28), 11560-11565.

McHeyzer-Williams, L. J., & McHeyzer-Williams, M. G. (2007). Memory B cell evolution: B cell biology. *Adv Exp Med Biol, 596*, 31-45. doi: 10.1007/0-387-46530-8_3

McHeyzer-Williams, M. G., & Ahmed, R. (1999). B cell memory and the long-lived plasma cell. *Curr Opin Immunol, 11*(2), 172-179.

McMaster, M. L., & Caporaso, N. (2007). Waldenstrom macroglobulinaemia and IgM monoclonal gammopathy of undetermined significance: emerging understanding of a potential precursor condition. [Review]. *Br J Haematol, 139*(5), 663-671. doi: 10.1111/j.1365-2141.2007.06845.x

Mo, X. Y., Ma, W. L., Li, L., Xu, Q. L., Zhang, Y. L., & Zheng, W. L. (2006). The effects of different sample labelling methods on signal intensities of a 60-mer diagnostic microarray. [Comparative Study Research Support, Non-U.S. Gov't]. *J Virol Methods, 134*(1-2), 36-40. doi: 10.1016/j.jviromet.2005.11.017

NORDCAN, (2011).

Nugen-Technologies. (2011). WT-Ovation Pico RNA Amplification System Userguide Version 1.0, 2010-2011

Ogmundsdottir, H. M., Johannesson, G. M., Sveinsdottir, S., Einarsdottir, S., Hegeman, A., Jensson, O., & Ogsmundsdottir, H. M. (1994). Familial macroglobulinaemia: hyperactive B-cells but normal natural killer function. *Scand J Immunol, 40*(2), 195-200.

Ogmundsdottir, H. M., Steingrimsdottir, H., & Haraldsdottir, V. (2011). Familial paraproteinemia: hyper-responsive B-cells as endophenotype. [Research Support, Non-U.S. Gov't]. *Clin Lymphoma Myeloma Leuk, 11*(1), 82-84. doi: 10.3816/CLML.2011.n.014

Ola Landgren, S. Y. K., Lynn R. Goldin, Neil E. Caporaso, Cecilie Blimark, Ulf-Henrik Mellqvist, Anders Wahlin, Magnus Bjorkholm and Ingemar Turesson. (2009). Risk of plasma cell and lymphoproliferative disorders among 14621 first-degree relatives of 4458 patients with monoclonal gammopathy of undetermined significance in Sweden. *Blood, 114*, 791-795. doi: 10.1182/blood-2008-12-19167610.1182/blood-2008-12191676

. Onto-Tools. (2012), 2011-2012, from http://vortex.cs.wayne.edu/projects.htm

Pageau, G. J., Hall, L. L., Ganesan, S., Livingston, D. M., & Lawrence, J. B. (2007). The disappearing Barr body in breast and ovarian cancers. *Nat Rev Cancer, 7*(8), 628-633. doi: 10.1038/nrc2172

Perez-Andres, M., Paiva, B., Nieto, W. G., Caraux, A., Schmitz, A., Almeida, J., . . . Orfao, A. (2010). Human peripheral blood B-cell compartments: a crossroad in B-cell traffic. [Research Support, Non-U.S. Gov't Review]. *Cytometry B Clin Cytom, 78 Suppl 1*, S47-60. doi: 10.1002/cyto.b.20547

Podojil, J. R., & Sanders, V. M. (2005). CD86 and beta2-adrenergic receptor stimulation regulate B-cell activity cooperatively. [Research Support, N.I.H., Extramural Research Support, U.S. Gov't, P.H.S. Review]. *Trends Immunol, 26*(4), 180-185. doi: 10.1016/j.it.2005.02.005

. R-program. (2012), 2010-2012, from http://www.r-project.org/

Radbruch, A., Muehlinghaus, G., Luger, E. O., Inamine, A., Smith, K. G. C., Dorner, T., & Hiepe, F. (2006). Competence and competition: the challenge of becoming a long-lived plasma cell. [10.1038/nri1886]. *Nat Rev Immunol, 6*(10), 741-750.

Richard J. Armitage, B. M. M., Melanie K. Spriggs, and William C. Fanslow. (1993). Human B Cell Proliferation and Ig Secretion Induced by Recombinant CD40 Ligand Are Modulated by Soluble Cytokines. *The Journal of Immunology, 150*(9), 3671-3680.

Roche NimbleGen, I. (2011). NimbleGen, 2010-2011, from http://www.nimblegen.com/index.html

Roddam, P. L., Allan, J. M., Dring, A. M., Worrillow, L. J., Davies, F. E., & Morgan, G. J. (2010). Non-homologous end-joining gene profiling reveals distinct expression patterns associated with lymphoma and multiple myeloma. [Research Support, Non-U.S. Gov't]. *Br J Haematol, 149*(2), 258-262. doi: 10.1111/j.1365-2141.2010.08088.x

Royer, R. H., Koshiol, J., Giambarresi, T. R., Vasquez, L. G., Pfeiffer, R. M., & McMaster, M. L. (2010). Differential characteristics of Waldenstrom macroglobulinemia according to patterns of familial aggregation. [Research Support, N.I.H., Intramural]. *Blood, 115*(22), 4464-4471. doi: 10.1182/blood-2009-10-247973

Schena, M. S., Dari; Davis, Ronald W;Brown, Patrick O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science, 270*(5235), 467-470.

Schutyser, E., Struyf, S., & Van Damme, J. (2003). The CC chemokine CCL20 and its receptor CCR6. *Cytokine Growth Factor Rev, 14*(5), 409-426.

Shaffer, A. L., Shapiro-Shelef, M., Iwakoshi, N. N., Lee, A. H., Qian, S. B., Zhao, H., . . . Staudt, L. M. (2004). XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. *Immunity, 21*(1), 81-93. doi: 10.1016/j.immuni.2004.06.010

Shen, Y., Iqbal, J., Xiao, L., Lynch, R. C., Rosenwald, A., Staudt, L. M., . . . Chan, W. C. (2004). Distinct gene expression profiles in different B-cell compartments in human peripheral lymphoid organs. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S.]. *BMC Immunol, 5*, 20. doi: 10.1186/1471-2172-5-20

Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol, 3*, Article3. doi: 10.2202/1544-6115.1027

Steingrimsdottir, H., Einarsdottir, H. K., Haraldsdottir, V., & Ogmundsdottir, H. M. (2011). Familial monoclonal gammopathy: hyper-responsive B cells in unaffected family members. [Research Support, Non-U.S. Gov't]. *Eur J Haematol, 86*(5), 396-404. doi: 10.1111/j.1600-0609.2011.01593.x

Tarca, A. L., Romero, R., & Draghici, S. (2006). Analysis of microarray experiments of gene expression profiling. [Research Support, N.I.H., Extramural Research Support, N.I.H., Intramural Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *Am J Obstet Gynecol, 195*(2), 373-388. doi: 10.1016/j.ajog.2006.07.001

Terstappen, L. W. M. M., Z. Hollander, H. M., & Loken, M. R. (1990). <TerstappenEtal_1990.pdf>. *Journal of Leukocyte Biology, 48*, 138ö148.

Tokuzo Arao, K. M., Mari Maegawa, and Kazuto Nishio. (2011). Target Therapy for Cancer: Anti-cancer Drugs Targeting Growth-Factor Signaling Molecules. What Can and Cannot Be Done Using a Microarray Analysis? Treatment Stratification and Clinical Applications in Oncology. *Biol. Pharm. Bull, 34*(12), 1789-1793.

Travers, P., Murphy, K. M., & Walport, M. (2007). *Janeway´s Immunobiology* (7th ed.): Garland science.

van den Berg, B. H., Thanthiriwatte, C., Manda, P., & Bridges, S. M. (2009). Comparing gene annotation enrichment tools for functional modeling of agricultural microarray data.

[Comparative Study Research Support, U.S. Gov't, Non-P.H.S.]. *BMC Bioinformatics, 10 Suppl 11*, S9. doi: 10.1186/1471-2105-10-S11-S9

Vangsted, A., Klausen, T. W., & Vogel, U. (2012). Genetic variations in multiple myeloma I: effect on risk of multiple myeloma. [Review]. *Eur J Haematol, 88*(1), 8-30. doi: 10.1111/j.1600-0609.2011.01700.x

Vazquez, A. I., Bates, D. M., Rosa, G. J., Gianola, D., & Weigel, K. A. (2010). Technical note: an R package for fitting generalized linear mixed models in animal breeding. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S.]. *J Anim Sci, 88*(2), 497-504. doi: 10.2527/jas.2009-1952

Vijay, A., & Gertz, M. A. (2007). Waldenstrom macroglobulinemia. [Review]. *Blood, 109*(12), 5096-5103. doi: 10.1182/blood-2006-11-055012

Vinuesa, C. G., Linterman, M. A., Goodnow, C. C., & Randall, K. L. (2010). T cells and follicular dendritic cells in germinal center B-cell formation and selection. *Immunological Reviews, 237*, 72-89.

Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., . . . Delorenzi, M. (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. *Breast Cancer Res, 10*(4), R65. doi: 10.1186/bcr2124

Xiang, Z., Cutler, A. J., Brownlie, R. J., Fairfax, K., Lawlor, K. E., Severinson, E., . . . Smith, K. G. (2007). FcgammaRIIb controls bone marrow plasma cell persistence and apoptosis. *Nat Immunol, 8*(4), 419-429. doi: 10.1038/ni1440

Yang, H. S., Jansen, A. P., Komar, A. A., Zheng, X., Merrick, W. C., Costes, S., . . . Colburn, N. H. (2003). The transformation suppressor Pdcd4 is a novel eukaryotic translation initiation factor 4A binding protein that inhibits translation. *Mol Cell Biol, 23*(1), 26-37.

Zenggang Pan, Y. S., Baosheng Ge, Cheng Du, Timothy McKeithan, and Wing C Chan. (2007). Studies of a germinal centre B-cell expressed gene, GCET2, suggest its role as a membrane associated adapter protein. *Br J Haematol, 137*(6), 578-590.

Zhan, F. (2002). Global gene expression profiling of multiple myeloma, monoclonal gammopathy of undetermined significance, and normal bone marrow plasma cells. *Blood, 99*(5), 1745-1757. doi: 10.1182/blood.V99.5.1745

Zhan, F., Barlogie, B., Arzoumanian, V., Huang, Y., Williams, D. R., Hollmig, K., . . . Shaughnessy, J. D., Jr. (2007). Gene-expression signature of benign monoclonal gammopathy evident in multiple myeloma is linked to good prognosis. [Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't]. *Blood, 109*(4), 1692-1700. doi: 10.1182/blood-2006-07-037077

Ögmundsdóttir, H. M., Haraldsdóttir, V., Jóhannesson, G. M., Ólafsdóttir, G., Bjarnadóttir, K., Sigvaldason, H., & Tulinius, H. (2005). Familiality of benign and malignant paraproteinemias. A population-based cancer-registry study of multiple myeloma families. *haematologica, 90*, 66-71.

APPENDIX 1

Buffer for B-cell isolation:

200 mL PBS

0.15 g EDTA

10 mL 10% BSA

Sterile-filtered, pH adjusted to 7.2

RNA Isolation Solutions

DNase 1 Incubation Mix (volumes for 1 sample):

10 µl DNase 1 stock solution (Qiagen)

70 µl Buffer RDD (Qiagen)

RPE Buffer (volumes for 50 samples):

11 ml Buffer RPE

44 ml 99.6% ethanol

cDNA Synthesis Solutions (volumes for 1 sample):

First Strand Master Mix:

2.5 µl First Strand Buffer Mix

0.5 µl First Strand Enzyme Mix

Second Strand Master Mix:

9.75 µl Second Strand Buffer Mix

0.25 µl Second Strand Enzyme Mix

SPIA Master Mix:

80 µl SPIA Buffer Mix

40 µl SPIA Primer Mix

40 µl SPIA Enzyme Mix

Sample Labeling Solutions:

10X TE (for use in dNTP Mix)

1.5 ml 1 M Tris HCl, pH 7.4

0.3 ml 0.5M EDTA

13.2 ml VWR water

50x dNTP Mix (volumes for 50 samples):

250 µl VWR water

50 µl 10X TE

50 µl 100 mM dATP

50 µl 100 mM dGTP

50 µl 100 mM dTTP

50 µl 100 mM dCTP

Random Primer Buffer(42 µl/Cy3-labeled primers):

860.75 µl VWR water

125 µl 1 M Tris HCL, pH 7.4

12.5 µl 1 M $MgCl_2$

1.75 µl ß-Merceptoethanol

Hybridization Solution Master Mix:

88.5 µl 2X Hybridization Buffer

35.4 µl Hybridization Component A

3.6 µl Alignment Oligo

Buffers for ELISA:

Coating Buffer:

15 mM Na2CO3

35 mM NaHCO3

0.2 g NaCN3

Diluted to 1L with dH20, pH adjusted to 9.6

Substrate Buffer:

114 mg MgCLx6 H2O

100 ml diethanolamine (Merck, NJ, USA)

Stirred overnight

890 ml dH2O

10 ml conc HCL

Extra HCL titrated until pH reaches 9.8

Washing Buffer:

8 g NaCl

0.2 g KH2PO4

1.15 g Na2HPO4x2H2O

0.2 g KCL

0.2 g NaN3

0.5 ml Tween 20 (Sigma)

Diluted to 1 L with dH2O and pH adjusted to 7.4

APPENDIX 2

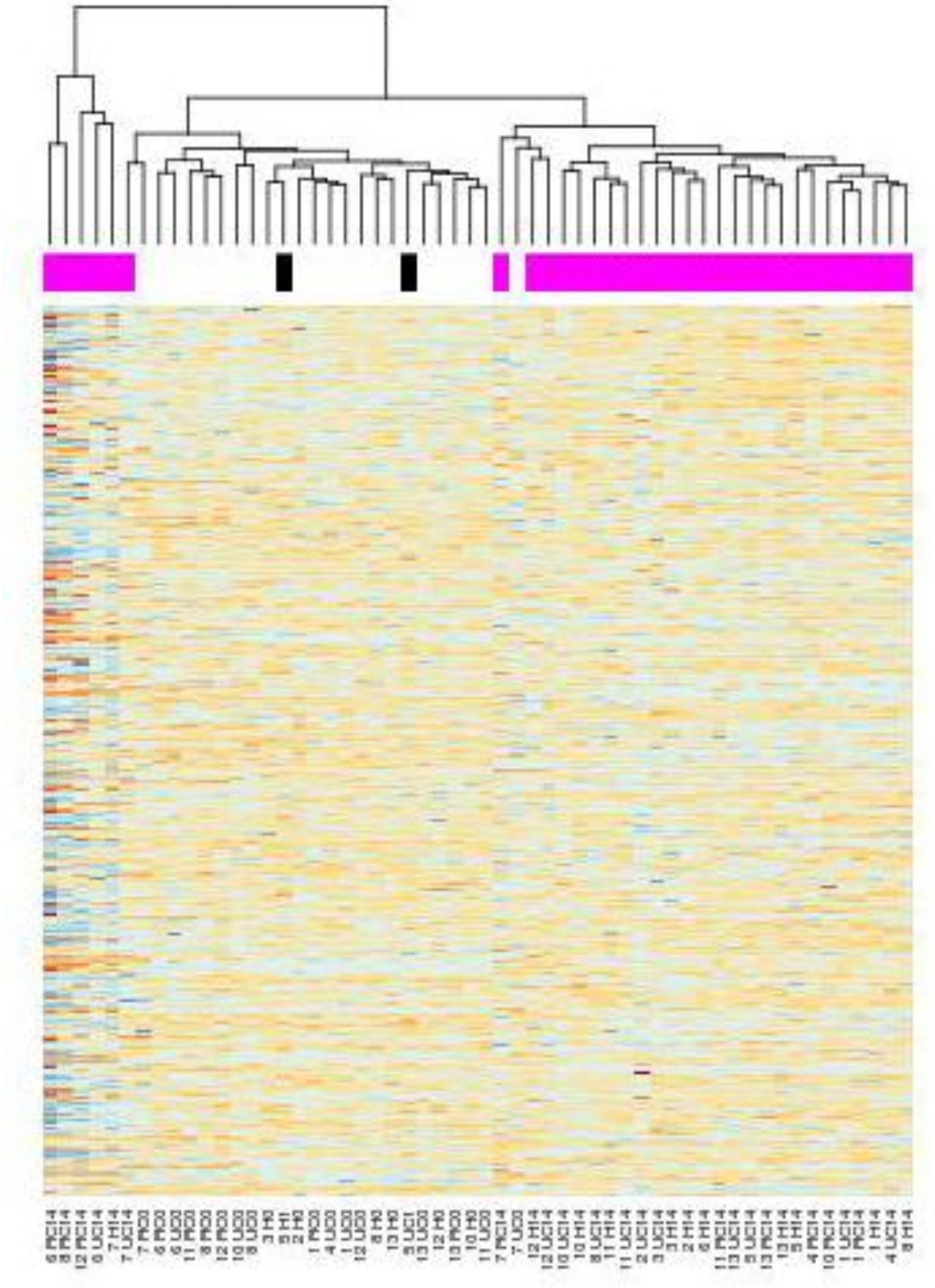


Figure A1: Heat map of unsupervised hierarchical cluster analysis for all samples.

Original heatmap result of unsupervised hierarchical cluster analysis including outliers. In the tree at the top, White indicates day 0 samples, black indicates day 1 samples and pink indicated day 14 samples.

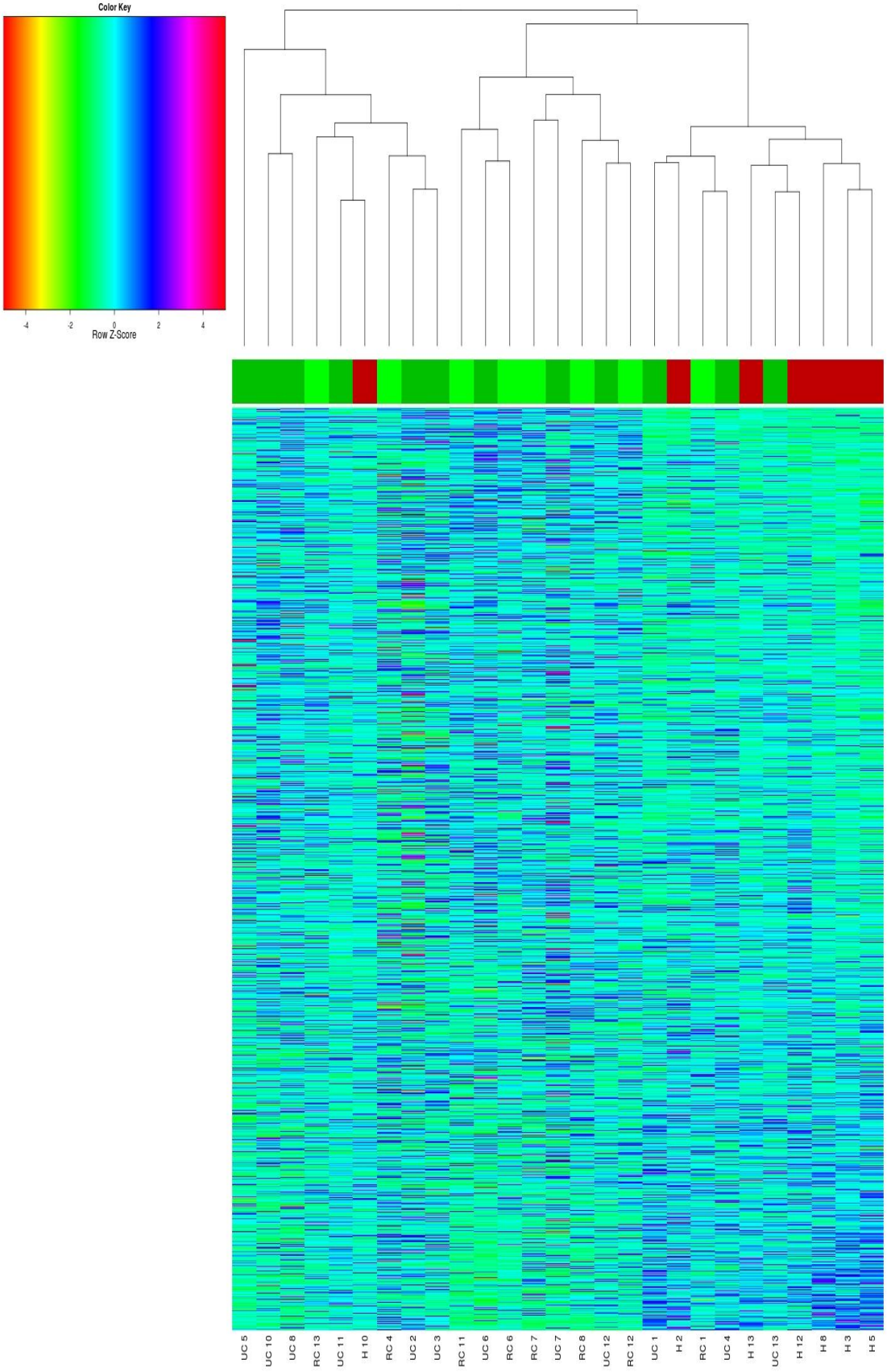


Figure. A2: Heatmap for unsupervised hierarchical clustering of all probes for day 0 samples. Missing samples include: GC-1 HR, GC-6 HR, GC-7 HR, GC-10 RC, and GC-11 HR due to low volume in B cell isolation. In the tree at the top, red is HR (Hyper-responder), light green is RC (related control), and dark green is UC (unrelated control).

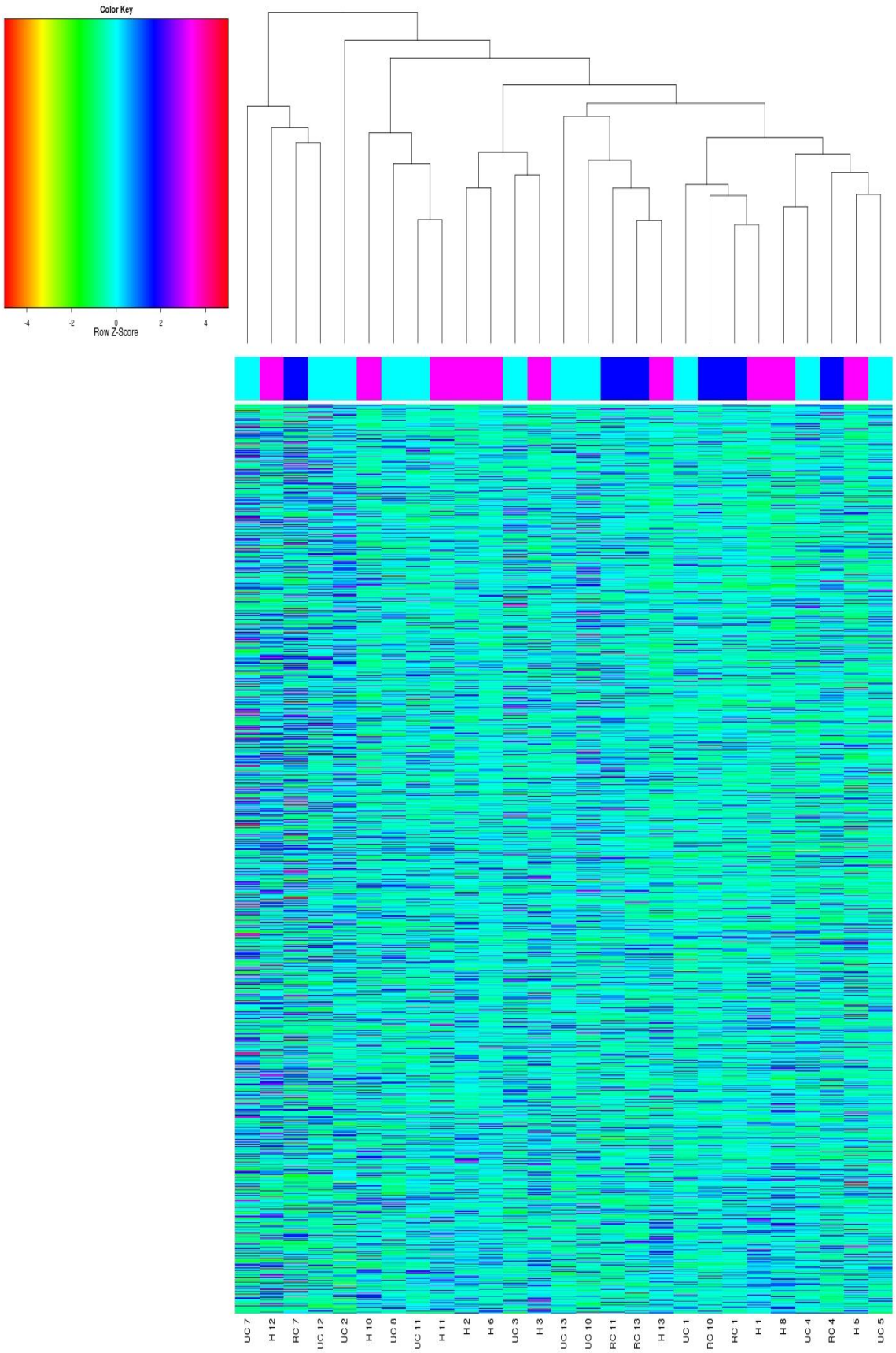


Figure A3: Heatmap of unsupervised hierarchical clustering of all probes for day 14 samples. Missing samples include: GC-6 RC, GC-6 UC, GC-7 HR, GC-8 RC, GC-12 RC due to low B cell volume or outlier status. In the tree at the top, pink is HR, dark blue is RC, and light blue is UC.

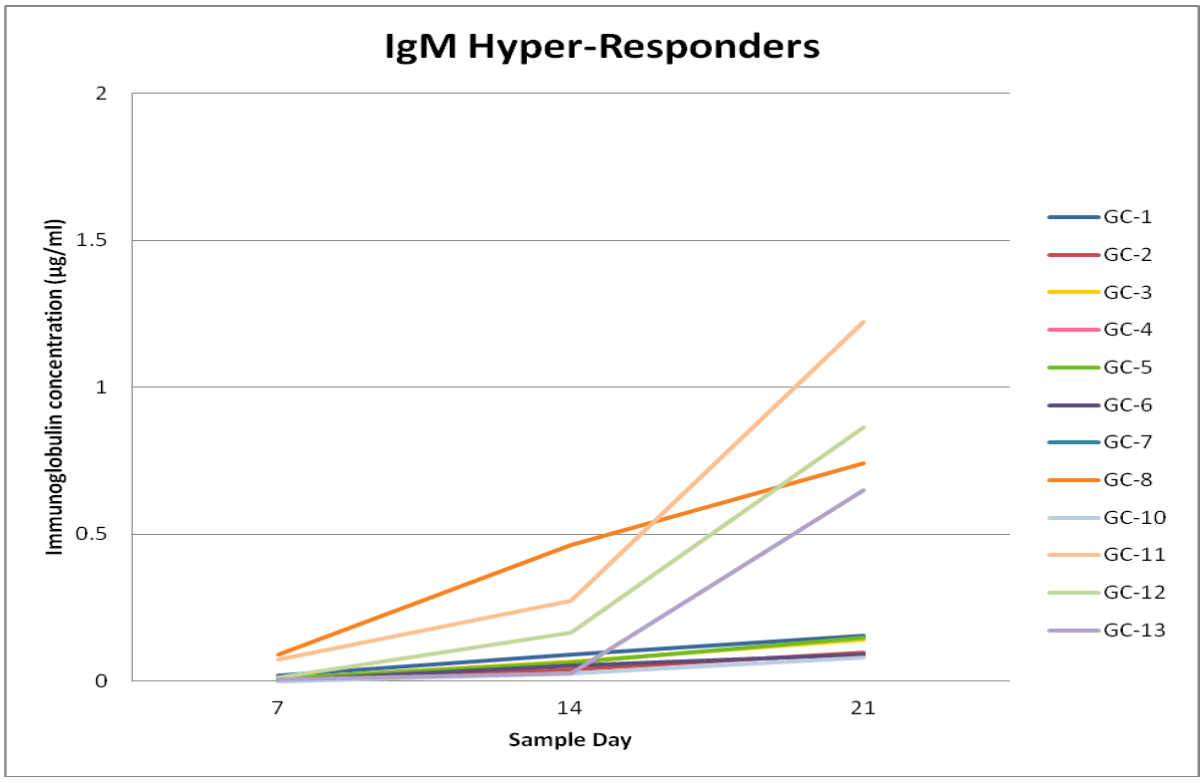


Figure A4: IgM concentrations for HRs.

Measurements were obtained, using an ELISA method, for samples collected on days 7, 14, and 21 for all HR samples.

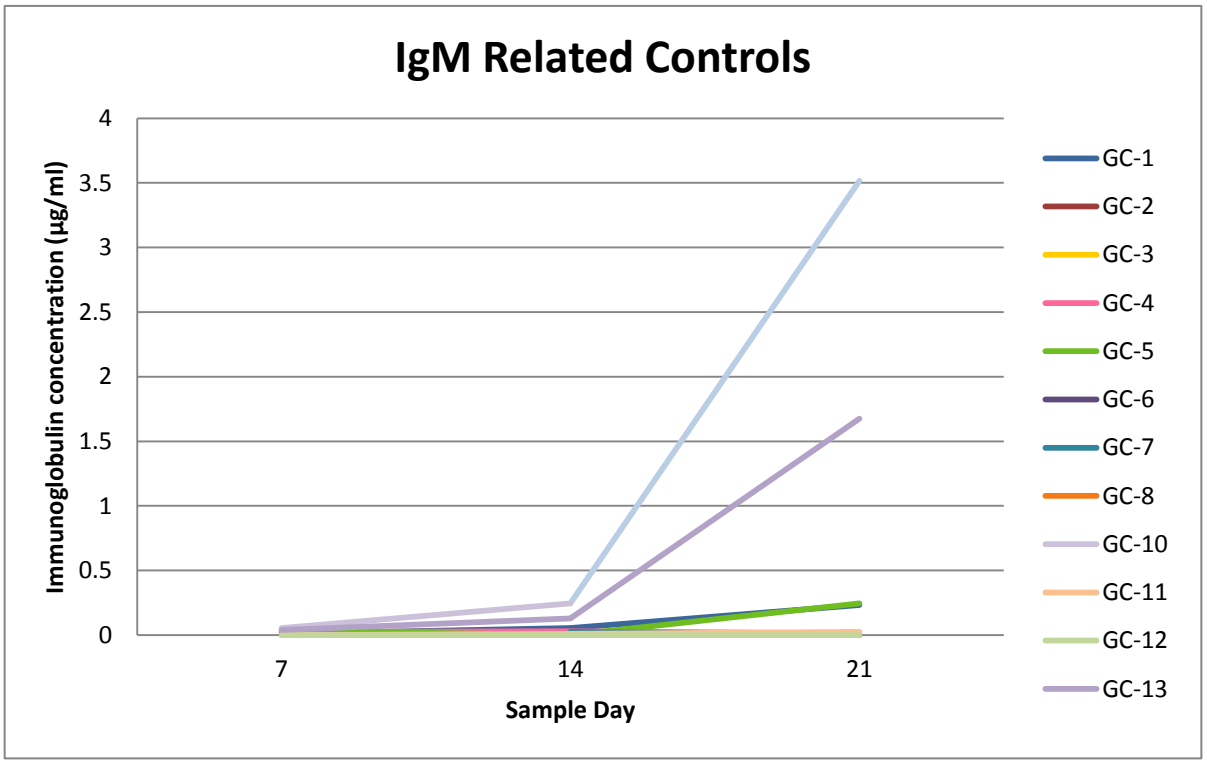


Figure A5: IgM concentrations for RCs.

Measurements were obtained, using an ELISA method, for samples collected on days 7, 14, and 21 for all RCs.

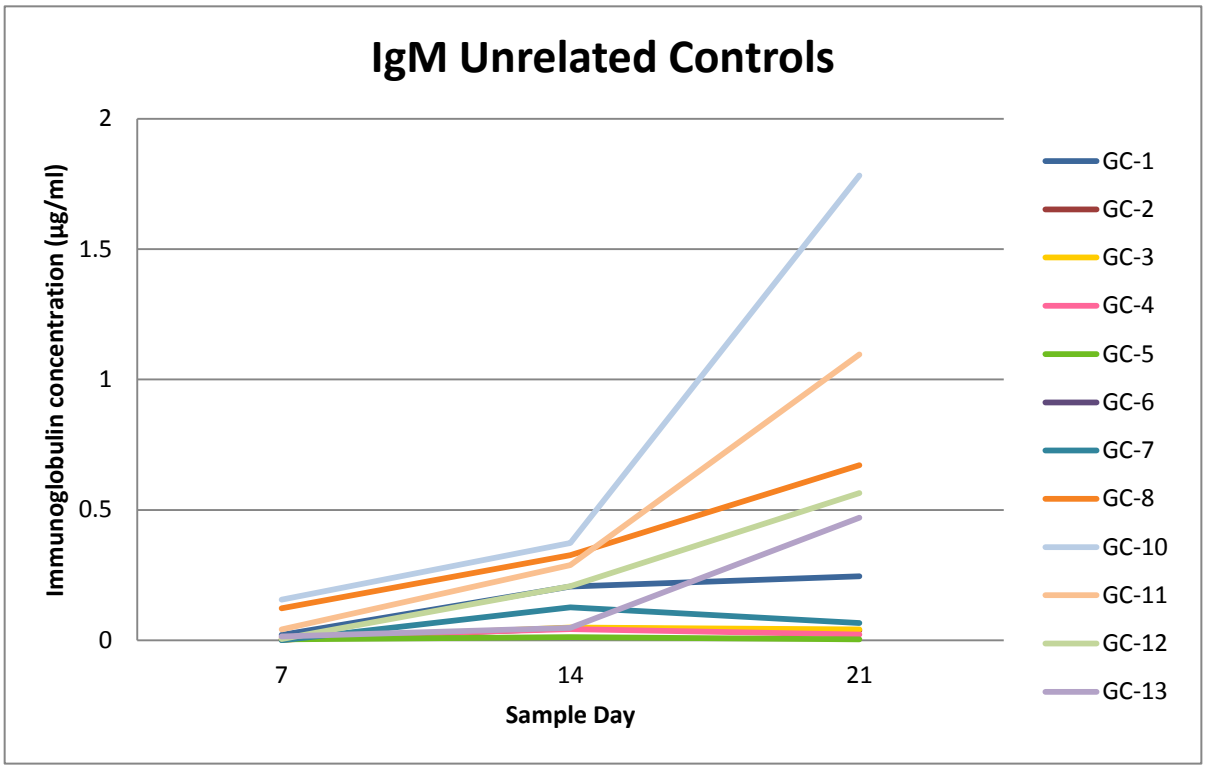


Figure A6: IgM concentrations for UCs.

Measurements were obtained, using an ELISA method, for samples collected on days 7, 14, and 21 for all UCs.

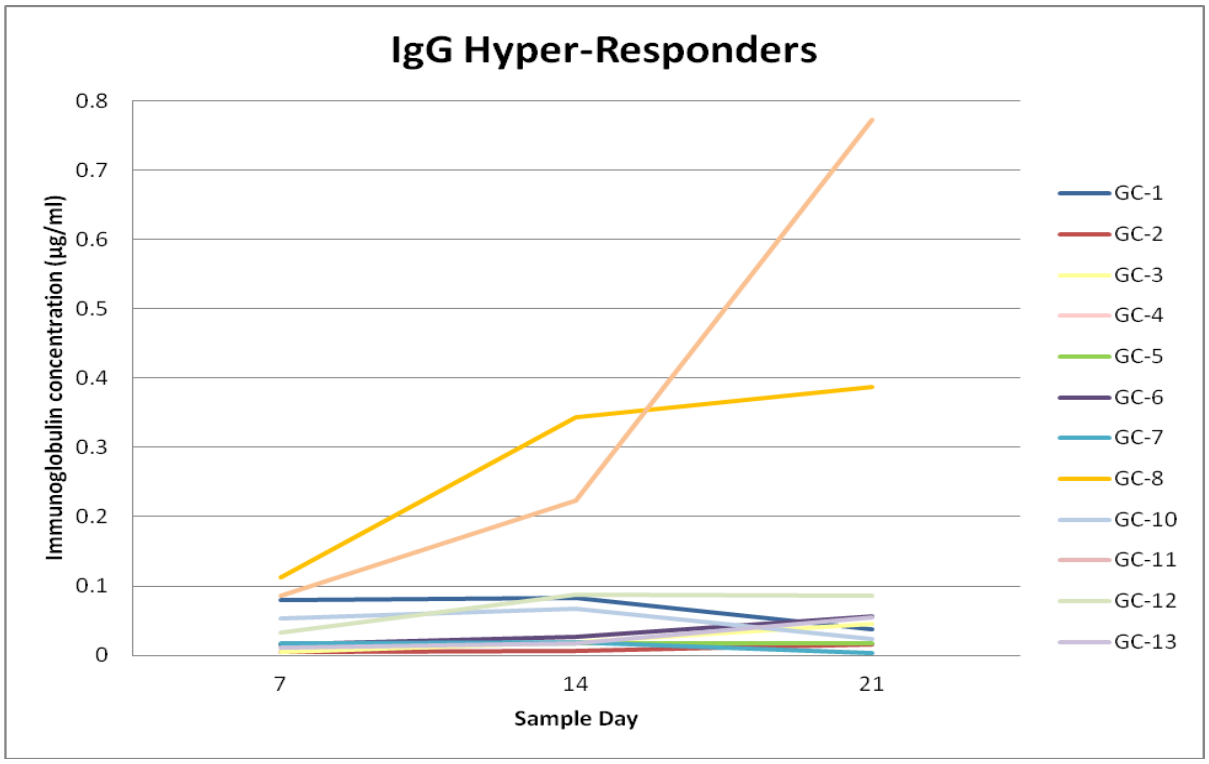


Figure A7: IgG concentrations for HRs.

Measurements were obtained, using an ELISA method, for samples collected on days 7, 14, and 21 for all HRs.

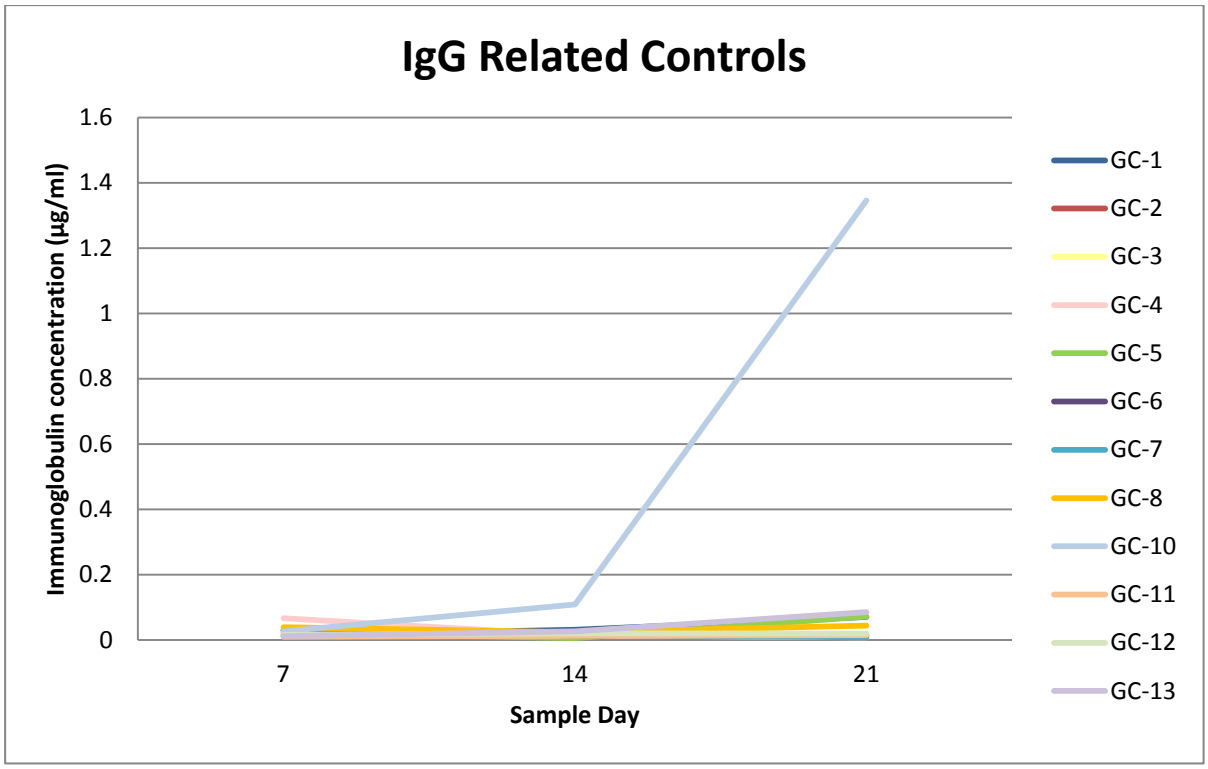


Figure A8: IgG concentrations for RCs.

Measurements were obtained, using an ELISA method, for samples collected on days 7, 14, and 21 for all RCs.

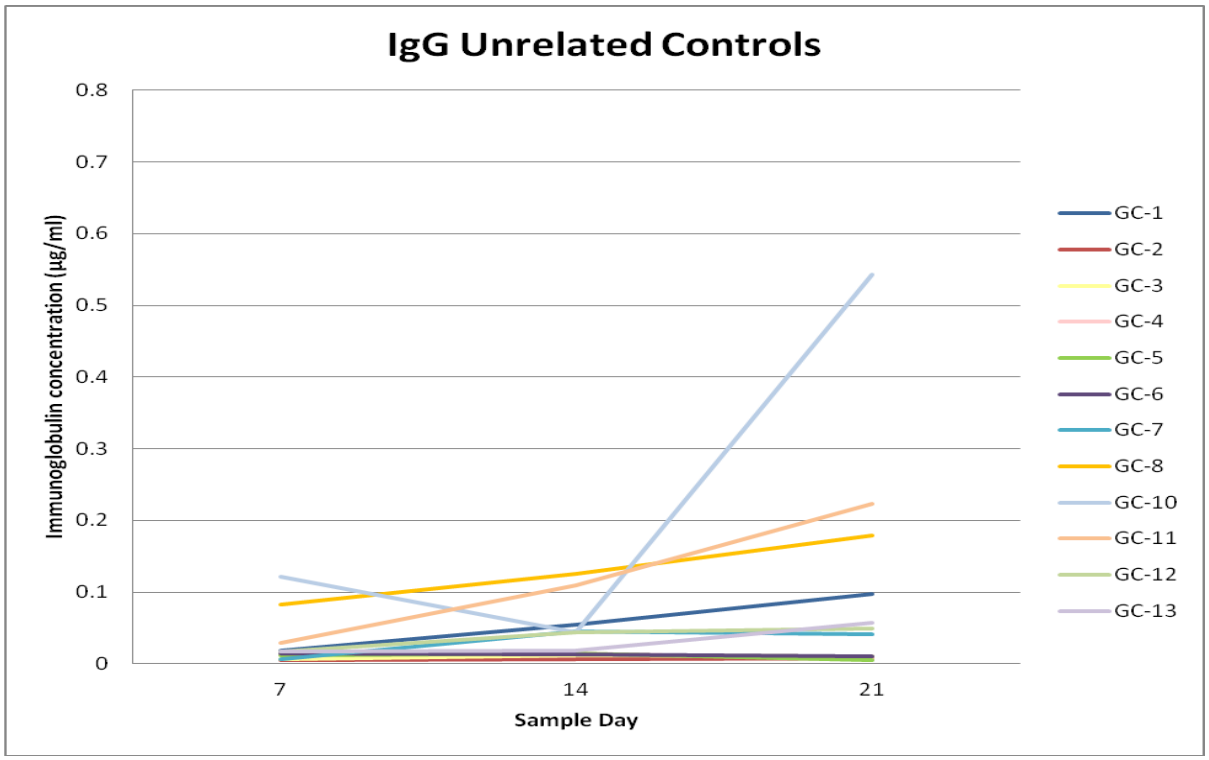


Figure A9: IgG concentrations for UCs.

Measurements were obtained, using an ELISA method, for samples collected on days 7, 14, and 21 for all UCs.