THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF VETERINARY AND BIOMEDICAL SCIENCE

A GENE NETWORK ANALYSIS OF T CELL DEVELOPMENT

ALLISON JAIME NEELY
Spring 2010

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in ANIMAL BIOSCIENCE
with honors in ANIMAL BIOSCIENCE

Reviewed and approved* by the following:

Avery August, Ph.D.
Professor of Immunology
Thesis Supervisor

Lester C. Griel, Jr. V.M.D. M.S.
Professor of Veterinary Science
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

Abstract

Many factors are involved in the determination of lineage fates and the activation of T lymphocytes. In this project an analysis of a proposed gene network for T cell development was performed using Boolean modeling. A synthesis of the gene network was collected from the literature and used to determine the various relationships between the genes and factors involved in T lymphocyte activation. Network rules were determined and analyzed using the computer program Booleannet. This experiment specifically attempted to determine what CD4/CD8 lineage choice (if either) a developing T cell would be diverted to if certain genes where turned on or certain genes were knocked out. The hypothesis was that creating network rules from a gene model and running them in a simulator would allow one to determine which genes were necessary for either a $CD4^+$ T cell or a $CD8^+$ T cell. The results from running a simulation of the pre-TCR signal showed that T cells will follow a $CD4^+$ lineage choice only, rather than both CD4 and CD8. However, further analysis of the results suggests that this may be due to errors in the network rules may have led to that result, so more research and experimentation is needed before definitive conclusions could be made. The method of analyzing a gene network using Boolean modeling seems to have merit, and more experimentation in this vein could be help research various other cell lineage pathways beyond just the CD4/CD8 T cell lineage choice.

Acknowledgements

I would like to thank Dr. Griel for his immense assistance as my advisor, Dr. August for his continued efforts to help me get through my thesis, and Dr. Ozment for her help in getting me back on track in the Honors College.

TABLE OF CONTENTS

Section: Page

Abstract .... *i*

Acknowledgements .... *ii*

Chapter 1: T Cell Development *1*
Introduction: *2*
T cell Development in the Thymus: *3*
$CD4^+$ and $CD8^+$ T-cells: *5*
The importance of Gene Regulatory Networks: *8*
CD4/CD8 Choice Mechanism: *9*

Chapter 2: Materials and Methods *11*
Creating the Network Rules: *12*
Use of Network Rules in Boolean analysis: *15*

Chapter 3: Results and Discussion *16*
Regulatory Network Rules: *17*
Analysis of the gene regulatory network using asynchronous Boolean Modeling: *22*
Results: *25*
Discussion: *32*
Conclusions: *33*

Appendices
A: List of Figures
B: Academic Vita

Chapter 1:

T Cell Development

Introduction:

T cells arise from hematopoietic stem cells (HSC) (Goodell, 1996). Hematopoietic stem cells are multipotent, self-renewing, and have a high proliferation and differentiation capacity (Goodell, 1996). These cells reside in the bone marrow and are used to replenish all adult hematopoietic cell lineages in a human or mammal throughout its life (Goodell, 1996). An HSC can become one of two types of progenitor cells, which have lost their self-renewal ability and are now dedicated to a particular cell lineage (Kindt, 2007). A myeloid progenitor cell can differentiate into red blood cells and white blood cells, while a lymphoid progenitor cell can become B cells, T cells, or Natural Killer cells (Spits, 1998). The differentiation of these stem cells is regulated at the genetic level by the expression of lineage specific genes in transcription factors at specific times and specific orders (Rothenburg, 2005). Immunologists can learn how particular genes affect hematopoiesis from studying mice by knocking out a particular gene and following the effects such disruption has on cell differentiation (Rotheburg, 2005). If, for example, a mouse cannot produce red blood cells after a certain gene is knocked out, the immunologist concludes that the particular gene must be necessary for red cell development.

An example of a transcription factor that affects the differentiation of multiple cell lineages is GATA-2, which will be seen later in this paper. GATA-2 is a family member of a group that recognizes GATA, a tetranucleotide sequence motif found in target genes, and is required for the development of lymphoid, erythroid, and myeloid genes (Kindt, 2007). On the other hand, the transcription factor Ikaros is required for cells to differentiate into the myeloid lineage, and this gene encodes a family of early hematopoietic- and lymphocyte-restricted transcription factors (Georgopoulos, 2005).

The Notch family also provides concrete examples of factors that are required for T cell development, and has actually been shown to redirect a potential B cell to the T cell lineage (Rothenburg, 2005). If Notch1 is knocked out in mice models, no T cells can develop so B cells will develop in the thymus instead (Rothenburg, 2005). Notch is also a factor that is required for embryonic viability and has effects on TCR-dependent selection in the thymus (Rothenburg, 2005).

T cell Development in the Thymus:

A lymphoid progenitor cell can differentiate into a progenitor T cell that will migrate to the thymus from the blood marrow and undergo gene rearrangement to differentiate and mature (Spits, 1998). T-cell development begins when $CD34^+$ pluripotent stem cells leave the bone marrow and migrate to the thymus (Spits, 1998). These progenitor T cells stem cell are not yet committed to the T-cell lineage when they arrive in the thymus, and have the capability to develop into T cells, natural killer (NK) cells, and dendritic cells (Spits, 1998). Thymocytes in the thymus undergo two main selection processes: positive and negative selection. Positive selection allows only those thymocytes that can recognize self-MHC molecules to survive, and negative selection eliminates cells that react against self MHC molecules and self peptides (Kindt, 2007).

Progenitor T cells that enter the thymus do not express the T-cell receptor, the CD3 complex, or the co-receptors CD4 and CD8 on their cell surface (Kindt, 2007). Because these cells lack both co-receptors, they are called double negative (DN) cells, and can be divided into four subsets (DN1-4), depending on the presence or absence of outer surface molecules (Kindt, 2007).

Important molecules to remember in this process are the stem cell growth factor c-Kit, the adhesion molecule CD44, and the alpha chain of the IL-2 receptor CD25 (Kindt, 2007).

Figure 1 graphically details the stage of a hematopoietic stem cell as it enters the thymus until it exits into the peripheral tissues as a single positive CD4 or CD8 T cell. The figure shows that the cells that enter the thymus are DN1 cells, and are c-Kit^{+}, CD44 high, and CD25^{-} (Kindt, 2007). Once they begin to proliferate, they become DN2 cells that are c-Kit^{+}, CD44 low, and CD25^{+} (Kindt, 2007). DN2 cells undergo rearrangement of the genes for the TCR gamma, delta, or beta chains (Kindt, 2007). Cells that progress to the DN3 stage are c-Kit^{-}, CD44 low, and CD25^{+} as rearrangement of TCR gamma, delta, or beta progress (Kindt, 2007). Once most cells reach the DN3 stage, they form the pre-T-cell receptor, which activates a signal transduction pathway (Kindt, 2007). This pathway induces developmental progression through the DN4 (c-Kit^{-}, CD44^{-}, CD25^{-}) stage to the CD4^{+}CD8^{+} double positive (DP) stage, where both co-receptors are expressed (Kindt, 2007). Double positive thymocytes that survive the selection processes then develop into immature single-positive CD4^{+} cells or single-positive CD8^{+} cells (Kindt, 2007). These cells will then go undergo even more negative selection before they are can migrate from the cortex of the thymus to the medulla, and from there they can pass out of the thymus into the circulatory system (Kindt, 2007).

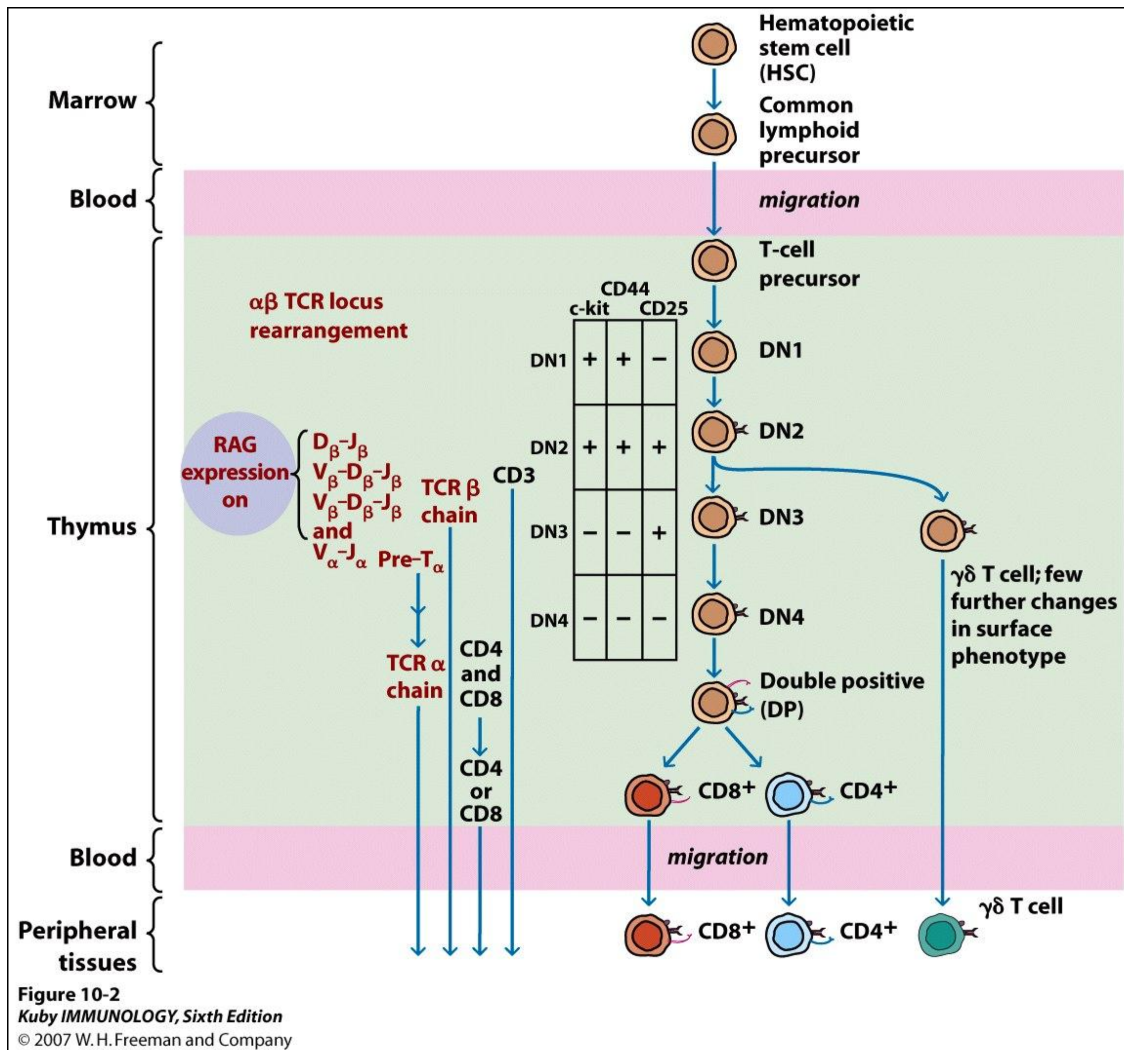

Figure 1: T cell developmental pathway This figure shows the fate of a hematopoietic stem cell beginning in the marrow that develops through the blood and thymus to become a single positive T cell. (taken from (Kindt, 2007))

$CD4^+$ and $CD8^+$ T-cells:

There are two main subpopulations of T cells, known as T helper cells and T cytotoxic cells, and the normal ratio in the human body between T helper cells and T cytotoxic cells is 2:1 (Kindt, 2007). T helper cells display CD4 membrane glycoproteins, and T cytotoxic cells display CD8 membrane glycoproteins (Kindt, 2007). CD4 is a 55-kDa monomeric glycoprotein that contains four extracellular immunoglobulin-like domains, a hydrophobic transmembrane region,

and a long cytoplasmic tail (Kindt, 2007). Figure 2 graphically shows the differences between the domains of the CD4 and CD8 co-receptors. CD8 takes the form of a disulfide-linked alpha-beta heterodimer or alpha-alpha homodimer, and the alpha and beta chains of CD8 are small glycoproteins of 30-38 kDa (Kindt, 2007). Each chain is made up of an extracellular immunoglobulin-like domain, a stalk region, a hydrophobic transmembrane region, and a cytoplasmic tail (Kindt, 2007). The CD4 and CD8 complexes are classified as co-receptors because they recognize the antigen peptide-MHC complex.

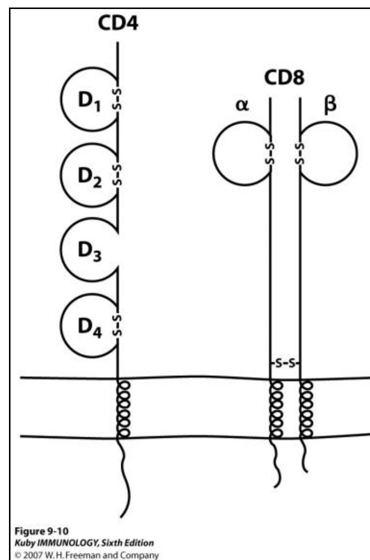

Figure 2: General Structure of the CD4 and CD8 Co-receptors This figure depicts the CD4 and CD8 co-receptors, showing how CD4 has four extracellular domains, a hydrophobic transmembrane region, and a cytoplasmic tail; while CD8 has an alpha-beta heterodimer, a stalk region, a hydrophobic transmembrane region, and a cytoplasmic tail. (taken from (Kindt, 2007))

The extracellular domains of the co-receptors bind to the conserved regions of MHC molecules on antigen-presenting cells or target cells (Kindt, 2007). The signal transduction properties of CD4 and CD8 molecules are mediated by their cytoplasmic domains (Kindt, 2007).

T cell receptors only recognize antigen that is bound to cell membrane proteins called major histocompatibility complex (MHC) molecules (Kindt, 2007). There are two major types of MHC molecules: MHC I and MHC II. Class I MHC molecules are expressed on nearly all nucleated cells, and class II MHC molecules are expressed on only a few types of cells that are specialized to be antigen presenting cells, such as macrophages, dendritic cells, and B cells (Kindt, 2007). Following activation by interaction with MHC II complexes, T helper cells can differentiate into effector cells that present antigen to B cells, macrophages, or cytotoxic T cells, or into memory cells (Kindt, 2007). When T cytotoxic cells are activated by interaction with MHC 1 complexes, T cytotoxic cells differentiate into an effector cytotoxic T cell (CTL) or a memory cell (Kindt,

2007). A CTL effector cell functions by monitoring body cells and eliminating any that display foreign antigen complexed with MHC I molecules (Kindt, 2007).

The importance of Gene Regulatory Networks:

One of the key features explained by gene regulatory networks is the exclusion of alternative fates in development of cell lineages (Constantin, 2008). Gene activation for specific cell types is regulated by three tenants of gene network architecture: positive auto-regulation of specific transcription factors, feed-forward relationships between those factors and their collaborators, and mutual antagonisms between the drivers of alternative cell fates (Constantin, 2008). In systems based on stem cells in adult mammals, multipotency can remain though over several cycles of cells because significant delay can occur before a cell's fate is determined, allowing the differing balances of regulatory factors to affect the differentiation process during intermediate stages (Constantin, 2008). These factors drive different cell fates depending on their presence and ratios.

T lymphocyte development is one example of this mode of specification, where cells can experience a variety of developmental options and can still proliferate extensively even after being committing to a T cell lineage (Constantin, 2008). Lineage exclusion is a slow and discontinuous process for T cell precursors after they reach the thymus (Constantin, 2008). By discontinuous, it is meant that progress between various developmental stages can require repeated stimulation. While B cells follow a more linear feed-forward cascade with specified steps for two transcription factors, T cell specification uses either few or no dedicated factors and is created by the activities of at least eight nonspecific transcription factors that are signaled by the Notch pathway in the thymic microenvironment (Constantin, 2008). PU.1 and GATA-3 are

transcription factors that are required for early T cell development, because if either transcription is not present at the prethymic stage, T cell development is eliminated (Constantin, 2008). These transcription factors can also have other effects, and high doses of them can inhibit certain T cell genes (Constantin, 2008). To understand T cell specification, one must understand that this process is not regulated by very few dedicated transcription factors that work in simple regulatory pathway similar to the B cell specification pathway (Constantin, 2008). Instead, the T cell pathway depends on multiple transcription factors that are also used for other hematopoietic programs (Constantin, 2008).

CD4/CD8 Choice Mechanism:

Differentiation of immature T cells into either CD4+ or CD8+ T cells occurs in the T-cell receptor (TCR)-mediated process of positive selection, where cells with receptors that respond to class II MHC molecules are chosen to become CD4+ T cells, and those that respond to class I MHC molecules are chosen to become CD8+ T cells (Rothenburg, 2009). These separate fates are determined by differences in signals, both their intensity and their duration (Rothenburg, 2009). Gene regulatory networks help us understand the way these decisions are made, revealing that key factors are used for different jobs. For example, one factor can be used to turn a temporary state into a permanent one, another factor can set a threshold necessary for an event, and another factor can make states mutually exclusive to each other (Rothenburg, 2009).

The transcription factor ThPOK has been discovered to be a key player in this mechanism, and is thought to be the “master regulator” necessary for $CD4^+$ cell fates (Rothenburg, 2009). Evidence are the facts that ThPOK is only induced by signals that cause CD4 positive selection; when ThPOK is absent those cells that should be CD4 become CD8; and

when ThPOK is forcibly expressed, cells that are on the CD8 path are redirected to become CD4 cells. GATA-3 is another factor that can promote $CD4^+$ differentiation and silence CD8 expression (Rothenburg, 2009). The factor TOX has been implicated in both $CD4^+$ and $CD8^+$ expression, and it has been seen that when TOX is knocked out in mice, the mice lose all $CD4^+$ cell lineages (Rothenburg, 2009). Runx3 is a transcription factor that has been identified as a player in $CD8^+$ cell development, and can be involved in silencing CD4 expression (Rothenburg, 2009). Overall, studies show that the CD4/CD8 lineage choice mechanism can mostly be explained by the antagonism between ThPOK and Runx3, as both can work to repress the other.

This project set out to test the roles of various genes and transcription factors involved in the T cell development pathway, and specifically to determine what lineage choice (if any) the T cell would be diverted to if certain genes where turned on or certain genes were knocked out. The hypothesis was that creating network rules from a gene model and running them in a simulator would allow one to determine which genes were necessary for either a $CD4^+$ T cell or a $CD8^+$ T cell.

Chapter 2:

Materials and Methods

Creating the Network Rules:

This project began by examining the literature to collect information on gene regulatory network models to determine the relationships between the various genes and transcription factors necessary for T lymphocyte development. (Constantin, 2008; Rothenburg, 2009). Figure 3 shows an example gene network model from the literature.

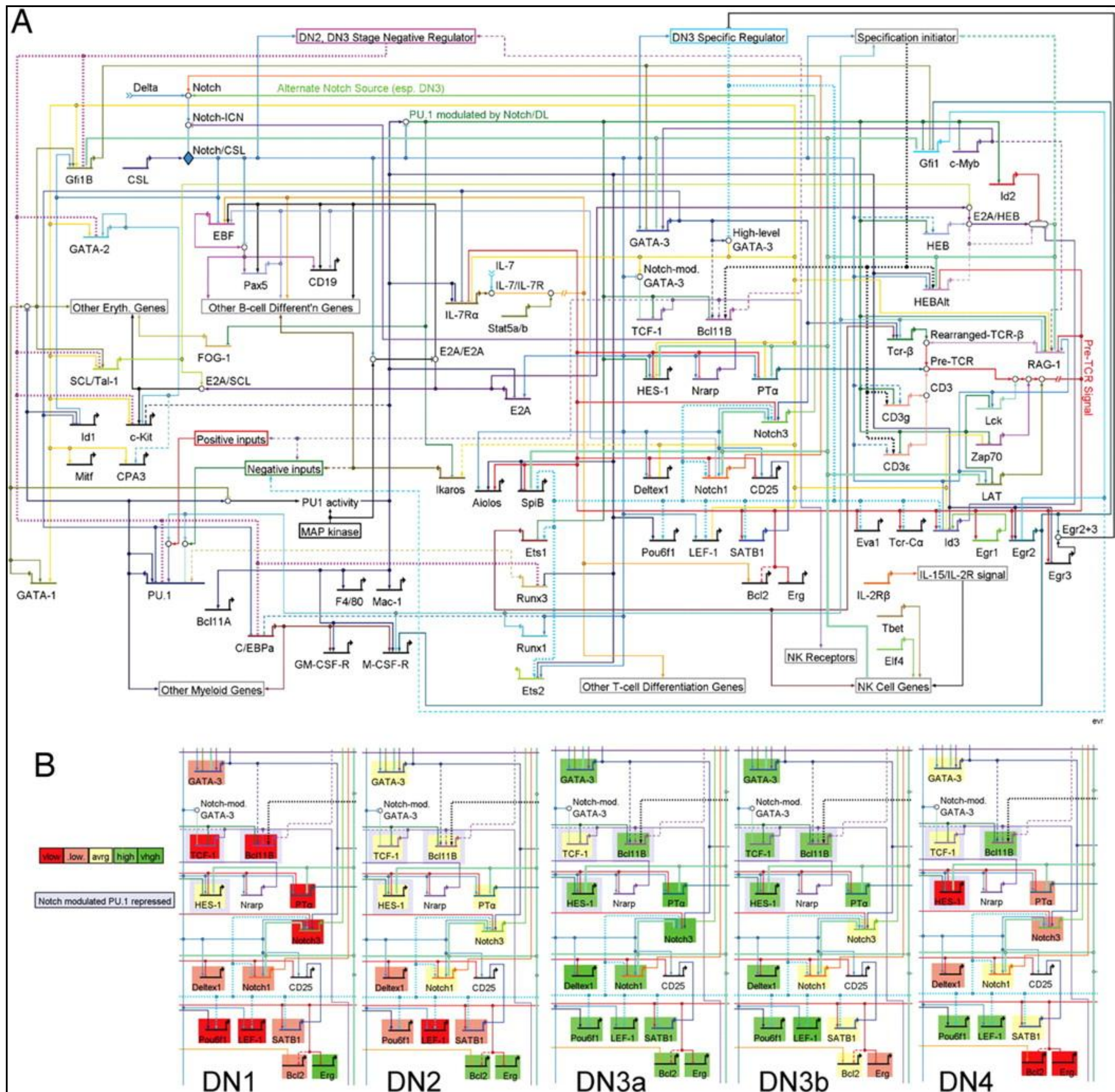

Figure 3: Example Network Model This is an example of a network model for T cell development.This depicts the effects various genes and transcription factors have on each other in the T cell development pathway. (from Georgescu, Constantin; William J. R. Longabaugh, et. al. "A gene regulatory network armature for T lymphocyte specification." *Proceedings of the National Academy of Sciences* 105 (2008): 20100-20105. Print.)

The colored arrows determine what effect a particular gene would have on another. An arrow indicates that a certain factor promoted the development of the factor it pointed too, while a path that ended in a "T" intersection implies that the second factor was inhibited by the

presence of the first. By slowly teasing out each relationship indicated in these models, a set of network rules regarding T lymphocyte activation could be made. Figure 4 shows in a simplified manner how such models can be read.

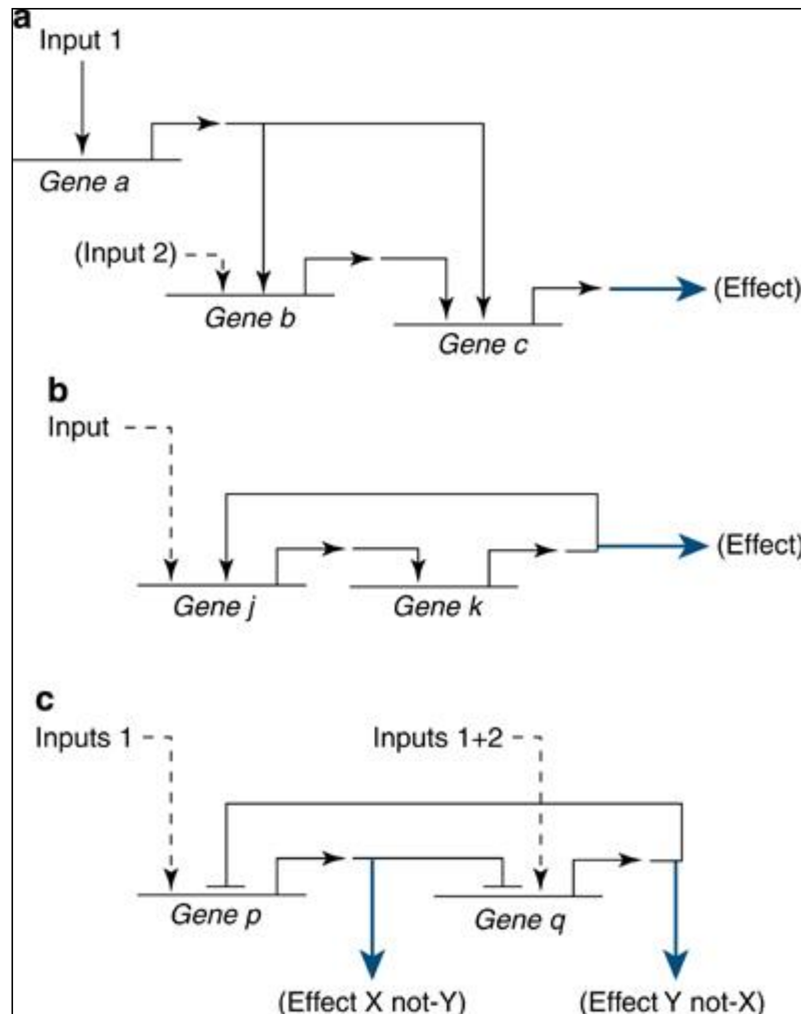

Figure 4: How to Read a Network Model This figure depicts classic gene network motifs. **A**) A feed-forward relationshIp where gene c requires both products from genes a and b. **B**) Mutual regulatory reinforcement where a transient input signal is converted to a sustained transcriptional response. **C**) A mutual exclusion circuit where two regulatory genes can be turned on by related signals but still induce mutually exclusive responses. (from Rothenburg, Ellen V. "Decision by committee: new light on the CD4/CD8-lineage choice." *Immunology and Cell Biology* 87 (2009): 109-112. Print.)

Use of Network Rules in Boolean analysis:

After compiling a list of network rules concerning T lymphocyte activation, the software package Boolean Network Modeling was used for the next phase (Booleannet, 2010). Booleannet provides tools used to simulate biological regulatory networks, enables creation of a biological system in a text language, and promotes research of a variety of dynamic behaviors in that system via a web or application programming (Booleannet, 2010). By inputting my network rules into Boolean, experiments can be run to determine what would happen when certain genes were turned on and off. The data collected from running the network rules in Booleannet was compiled numerically in an Excel spreadsheet, from which graphs could be drawn for the results. The graphs were then used as the basis for interpretation.

Chapter 3:

Results and Discussion

Regulatory Network Rules:

Using data from *A gene regulatory network armature for T lymphocyte specification* (Constantin, 2008), a set of gene regulatory network rules was created to use for the experiments, as detailed in Table 1.

Table 1: Regulatory Network Rules

Target	**Influencing Factors**
AIOLAS* =	PRETCR or NOTCH/CSL
BCL1A* =	PU1
BCL1B* =	GATA3 or TCF1 and not PU1
BC12* =	not PRETCR
CD3* =	CD3G
CD3G* =	NOTCHCSL
CD4* =	THPOK or GATA3 and not RUNX3
CD8* =	RUNX3 or EOMES and not THPOK
CD19* =	E2A or E2AE2A or EBF or PAX5
CD25* =	NOTCHCSL and not (PRETCR or SATB1)
CEBPA* =	not (DN2DN3SNR or GATA3 or NOTCHCSL)
CKIT* =	GATA2 or E2A or GATA3 or E2ASCL and not (DN2DN3SNR or PU1)
CMYB* =	not PU1

Target	**Influencing Factors**
CPA3* =	GATA2
DELTEX1* =	PRETCR or NOTCHCSL
DN2DN3SNR* =	not (NOTCHCSL or BCL1B)
DN3SR* =	NOTCHCSL and not (EGR2 and EGR3)
EBF* =	E2A orE2AE2A or PAX5 or IL7RSTAT and not NOTCHCSL
EGR1* =	PRETCR
EGR2* =	PRETCR or PU1 or (EGR2 and EGR3) and not GFI1B
EGR3* =	PRETCR or EGR3 and not PU1
EOMES* =	RUNX3 and not THPOK
ERG* =	not PRETCR
ETS1* =	DN3SR and not PU1
ETS2* =	NOTCHCSL and DN3SR and not PU1
EVA1* =	DN3SR and not PRETCR
E2A* =	NOTCHCSL and not PU1
E2AE2A* =	E2A and not (NOTCH and MAPK)
E2AHEB* =	(E2A and HEBALT)
E2ASCL* =	E2A
FOG1* =	not PU1
F480* =	PU1
GATA1* =	GATA1 or GATA3
GATA2* =	GATA2 or GATA3 and not DN2DN3SNR

Target	Influencing Factors
GATA3* =	TCR or NOTCHCSL or NETE or CMYB and not DN3SR
GFI1* =	NETE and not (GFI1B or PU1)
GFI1B* =	ID1 or EGR2 or GATA3 and not DN2DN3SNR
GMCSFR* =	PU1 or CEBPA
HEB* =	NOTCHCSL or GATA3 and not PU1
HEBALT* =	NOTCHCSL or NETE and not PU1
HES1* =	NOTCHCSL or GATA3 or NETE and not PRETCR
ID1* =	not GATA3
ID2* =	PU1
ID3* =	PRETCR or EGR1 or DN3SR and not (PU1 or NOTCHCSL or GATA3)
IKAROS* =	GATA3 and not PU1
ILTRALPHA* =	PRETCR
IL15IL2R* =	IL2RBETA
IL7IL7R* =	(IL7 and IL7RALPHA)
IL7RALPHA* =	PU1 and not (PRETCR or GATA3)
IL7RSTAT* =	IL7IL7R or STAT5AB
LAT* =	NOTCHCSL and not (PU1 or NETE)
LCK* =	NOTCHCSL and not PU1
LEF1* =	NOTCHCSL or DN3SR and not GATA3
MAC1* =	PU1

Target	Influencing Factors
MCSFR* =	EGR2 or PU1 or CEBPA or RUNX1
MITF* =	GATA3
NETE* =	HEBALT or E2AHEB and not ID3
NK* =	TBET or IL15IL2R or ELF4 or ETS1 and not NETE
NOTCH* =	DELTA or NOTCH1
NOTCHCSL* =	(CSL and NOTCHICN)
NOTCHICN* =	NOTCH and not NRARP
NOTCH1* =	NOTCHCSL or DN3SR and not (PRETCR or PAX5)
NOTCH3* =	NOTCHCSL or GATA3 or DN3SR and not PRETCR
NOTCHGATA3* =	GATA3 and not NOTCHCSL
NRARP* =	NOTCHCSL and not PRETCR
PAX5* =	E2A or E2AE2A or EBF
POU6F1* =	PU1 or DN3SR
PRETCR* =	RTCRBETA or PTALPHA
PTALPHA* =	NOTCHCSL or NETE and not (PRETCR or GATA3)
PU1* =	MAPK or PU1 or RUNX1 or RUNX3 or and not (GATA3 or DN2DN3SNR or NOTCHCSL)
RAG1* =	NOTCHCSL or CMYB and not (PU1 or GATA3)
RTCRBETA* =	TCRBETA and RAG1
RUNX1* =	NOTCHCSL
RUNX3* =	TBET and not (DN2DN3SNR or PU1 or THPOK or GFI1B)

Target	Influencing Factors
SATB1* =	PRETCR or DN3SR
SCLTAL1* =	GATA3 and not DN2DN3SNR
SPIB* =	DN3SR and not (PRETCR or PU1)
TCF1* =	NOTCHCSL or NK and not (NOTCHGATA3 or PU1)
TCRALPHA* =	PRETCR or DN3SR
TCRBETA* =	ETS1 or GATA3 or RUNX1
THPOK* =	(GATA3 and TOX) and not RUNX3
TOX* =	TCRALPHA
ZAP70* =	not (PU1 or GATA3)

Table 2 lists the functions of some of the genes that are a part of the gene regulatory network.

Table 2: Function of Genes/Transcription Factors

Gene/Transcription Factor	Function
CD3g	encode T cell receptor complex proteins (Kindt, 2007)
c-kit	Required for Notch and IL-7 signaling-induced proliferation and differentiation of cD44^{+}cD25^{-}c-kit^{high} and CD44^{+}CD25^{+}c-kit^{high} thymocytes (Massa, 2006)
CSL	transcription factor for Notch for the Notch-signaling pathway (Kindt, 2007)
Eomes	complement the actions of *T-bet* and act as a key regulatory gene in the development of cell-mediated immunity, invoke attributes of effector CD8^{+} T cells, including interferon-γ (IFN-γ), perforin, and granzyme B (Pearse, 2003)
GATA-1	promotes erythroid lineage (Kindt, 2007)

Gene/Transcription Factor	Function
GATA-2	promotes erythroid, myeloid, lymphoid lineage (Kindt, 2007)
GATA-3	required for early T cell development (Kindt, 2007)
HEB	a.k.a. HeLa E-box binding protein, is a basic helix-loop-helix transcription factor thought to regulate E-box sites present in many T cell-specific gene enhancers, including TCR-α, TCR-ß, and CD4 (Barndt, 1999)
Ikaros	lymphoid lineage (Kindt, 2007)
MAP kinase (MAPK)	This pathway leads to activation of Elk, which is necessary for the expression of Fos, which associates with Jun to form AP-1, which is an essential transcription factor for T –cell activation (Constantin, 2008)
Notch-Delta signaling	required from DN1 stage through commitment and into beta-selection of T lymphocytes, targets Deltex1, Hes1, and Ptcra genes that peak in expression at DN3a stage (Constantin, 2008)
PU.1 activity	promotes erythroid (maturation stages), myeloid (later stages), and lymphoid lineage, required for early T cell development (Kindt, 2007)
Pre-TCR signal	first signal that directs differentiation and proliferation of T cells during the transition from $CD4^-CD8^-TCR^-$ double negative (DN) stage to $CD4^+CD8^+$ double positive (DP) stage (Barndt, 1999)
RAG-1	required for T cell gene rearrangement and modification (Kindt, 2007)
Runx1	critical regulator of hematopoietic development, and required for CD8 T cell development (Kurokawa, 2006)
Runx3	involved in CD8+ development, can work to inhibit CD4 development (Rothenburg, 2009)
ThPOK	induces CD4+ development, can work to inhibit CD8 development (Rothenburg, 2009)

Analysis of the gene regulatory network using asynchronous Boolean Modeling:

Following the formation of the Regulatory Network Rules, tests could be run in Booleannet. The first model ran involved inputting the network rules into the Booleannet

software and specifying that the node PRETCR = true. In the network rules that were put into Booleannet, PRETCR coded for the Pre-TCR signal that occurs during T lymphocyte development, and its function is listed in Table 2. Setting the node PRETCR = true to start the test meant that at the beginning of the experimental run all the nodes were turned off except PRETCR, which was the only node to be turned to "on." The rules were then run using Booleannet to determine what genes would be affected by PRETCR being on. The data was collected and the response of specific genes plotted. This data was compiled into an Excel spreadsheet with the fate of the nodes being reflected by numbers from zero to one. Example data from fourteen important nodes is shown below in Table 3. These nodes were picked to analyze out of rest because many of them had a direct relation to the CD4/CD8 lineage choice, and the rest showed something interesting about how the system worked. Only steps one through eight are shown for space reasons, but none of those nodes showed any change from step eight to step twenty.

Table 3: Excel Data of Nodes

Nodes	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step8
TCRALPHA	0	1	1	1	1	1	1	1
TOX	0	1	1	1	1	1	1	1
CMYB	0	1	1	1	1	1	1	1
PU.1	0	0	0	0	0	0	0	0
GATA3	0	1	1	1	1	1	1	1
NOTCH	0	0	0	0	0	0	0	0
THPOK	1	1	1	1	1	1	1	1
RUNX3	0	0	0	0	0	0	0	0
DN2DN3SNR	0	1	0.47	0	0	0	0	0
BCL1B	0	0	1	1	1	1	1	1
CEBPA	0	1	0	0	0	0	0	0
GMCSFR	0	1	0.48	0	0	0	0	0
CD4	0	0	1	1	1	1	1	1
CD8	0	0	0	0	0	0	0	0

Legend When interpreting the data, zero means that the node is completely turned OFF, while one means that the node is completely turned ON.

From the Table 3, when PRETCR is ON, it was found that TCRALPHA, TOX, CMYB, GATA3 and ThPOK were turned ON in the second step and BCL1B was turned on in the third step; all remained ON for the duration of the simulation. DN2DN3SNR, CEBPA, and GMCSFR were turned ON in the second step but were turned OFF again before the end of the simulation. PU.1, NOTCH, and RUNX3 remained OFF for the entire simulation. The result of the simulation showed that the T cells would become $CD4^+$ rather than $CD8^+$ cells.

Results:

Next graphs were made of the data in Excel from the data collected in Table 3.

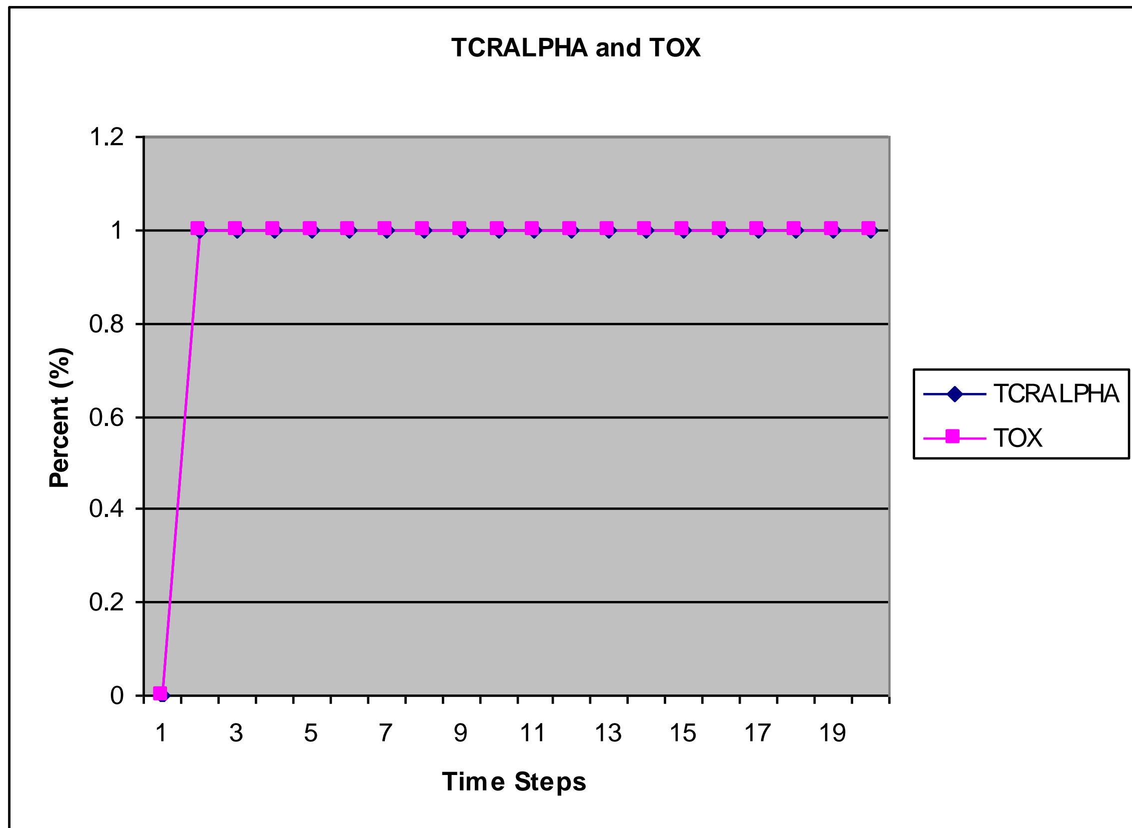

Figure 5: Expression behavior of TCRALPHA and TOX genes. The graph shows that TCRALPHA and TOX were both turned on in step two, and both remained on for the duration of the simulation.

From the network rules, it can be determined that TCRALPHA depends on PRETCR to become activated, and so was turned on quickly because PRETCR was set to ON at the initiation of the experiment. TCRALPHA is required to activate TOX, so TOX was turned on once TCRALPHA was ON (Figure 5).

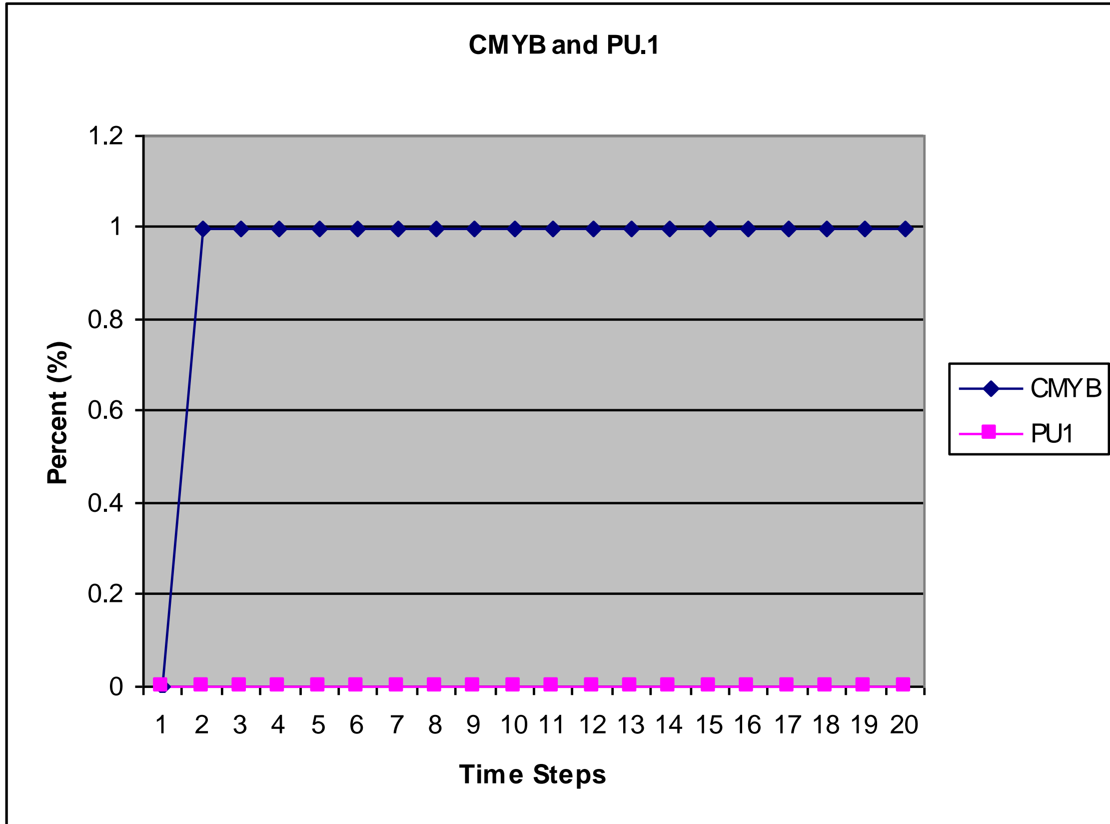

Figure 6: Expression behavior of CMYB and PU.1 genes The graph shows that CMYB is turned on in time step 2, while PU.1 is never turned on.

CMYB is turned on because the rules state that this factor is always on when PU.1 is not on and doesn't list another requirement. PU.1 is never turned on because MAPK, RUNX1, and RUNX3 was never turned on.

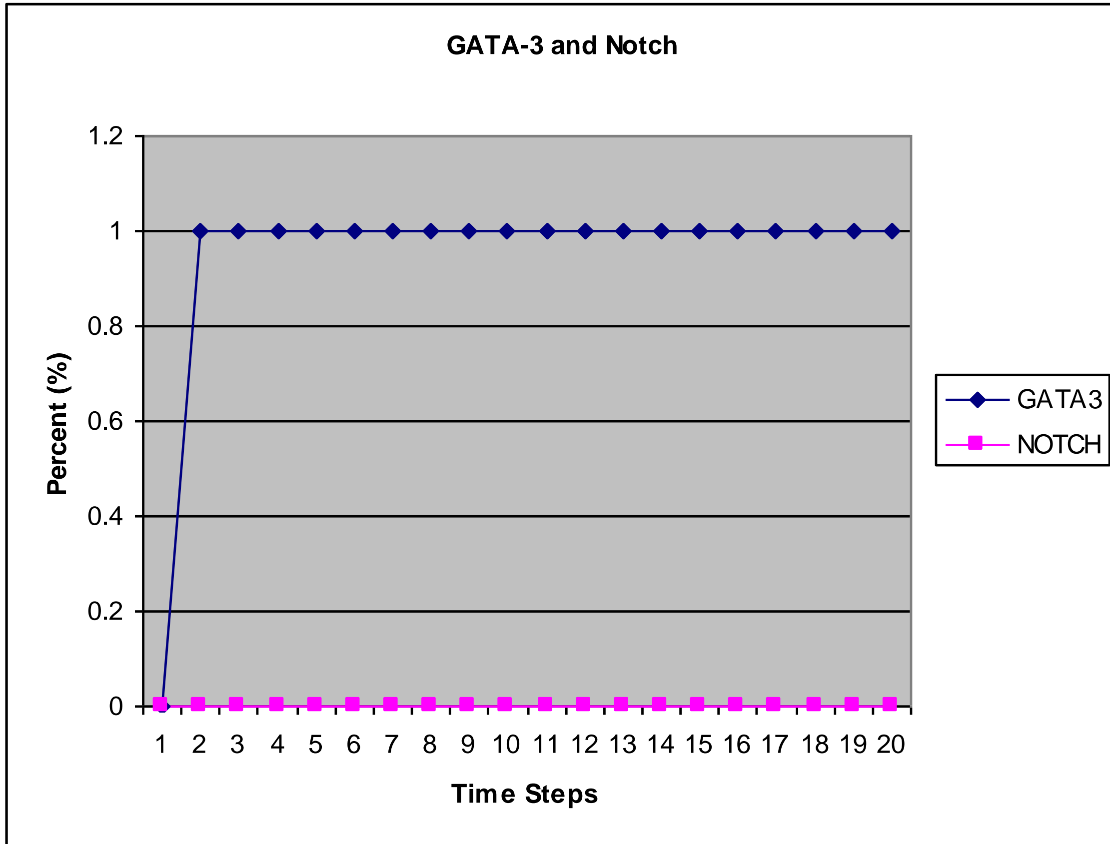

Figure 7: Expression behavior of GATA-3 and NOTCH genes GATA-3 is turned on in step two while Notch is never turned on.

GATA-3 gets turned on when CMYB is turned on (Figure 6), and NOTCH is not turned on because DELTA and NOTCH1 are never turned on.

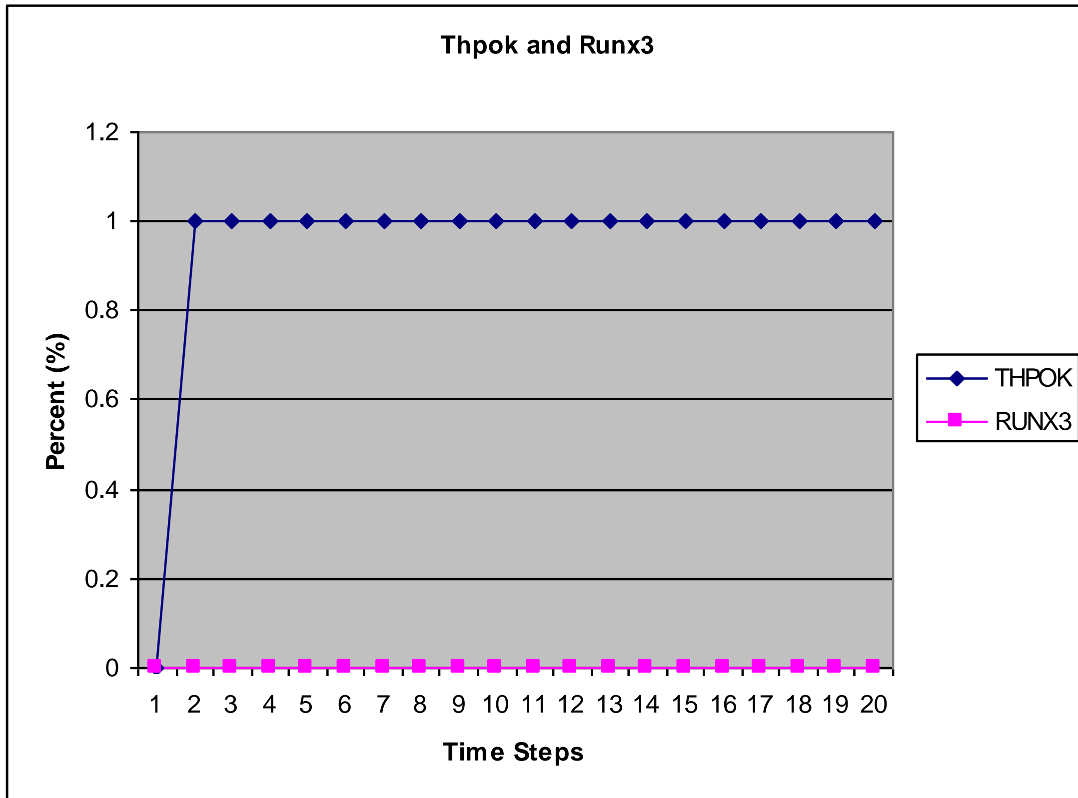

Figure 8: Expression behavior of ThPOK and RUNX3 genes This graph shows that while THPOK is turned on in step two, RUNX3 is never turned on.

ThPOK gets turned on when GATA-3 and TOX are turned on (Figure 7 and Figure 5 respectively), while RUNX3 is never turned on because TBET was never turned on, and because ThPOK actively represses RUNX3.

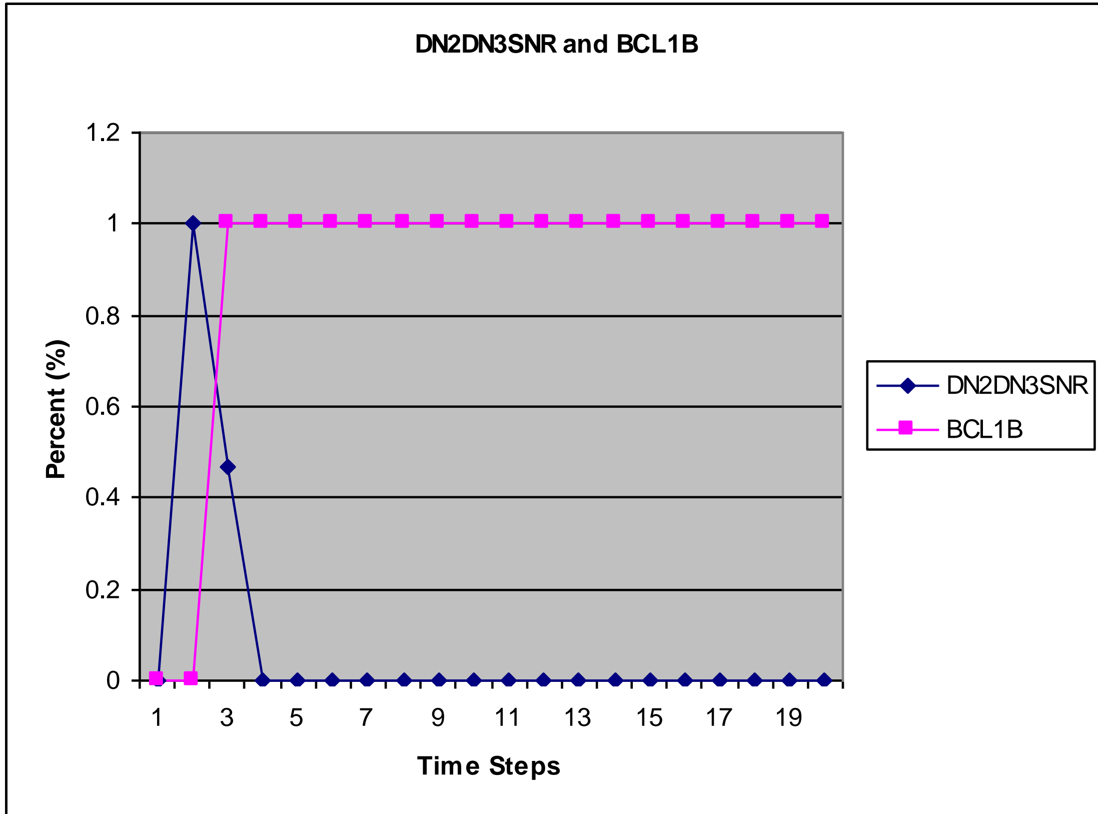

Figure 9: Expression behavior of DN2DN3SNR and BCL1B genes The graph shows that DN2DN3SNR was turned on in step two but was completely turned off by step four, while BCL1B was turned on in step three and remained on for the duration.

DN2DN3SNR was turned on because in the beginning of the experiment when just PRETCR was on, neither NOTCHCSL nor BCL1B was on. When GATA-3 was turned on in step two (Figure 7), it turned OFF DN2DN3SNR.

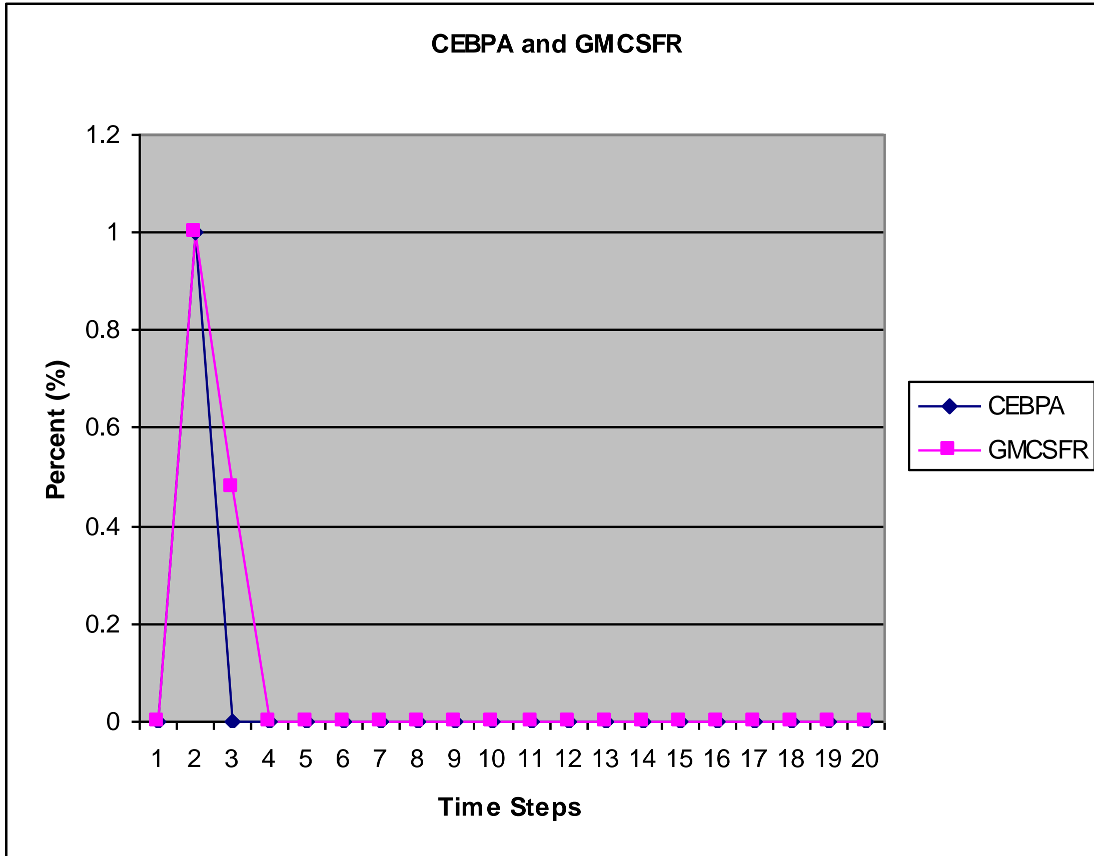

Figure 10: Expression behavior of CEBPA and GMCSFR genes The graph shows that CEBPA was turned on in step two and was turned off by step three, while GMCSFR was turned on in step two and was turned off by step four.

CEBPA was turned on because, when just PRETCR was turned on, BCL1B and GATA3 weren't on yet to turn it off. While CEBPA was ON, it turned on GMCSFR. When BCL1B and GATA3 were turned ON (Figure 9 and Figure 7 respectively), they worked to turn CEBPA OFF. With CEBPA OFF, GMCSFR quickly turned OFF also because it depended on CEBPA.

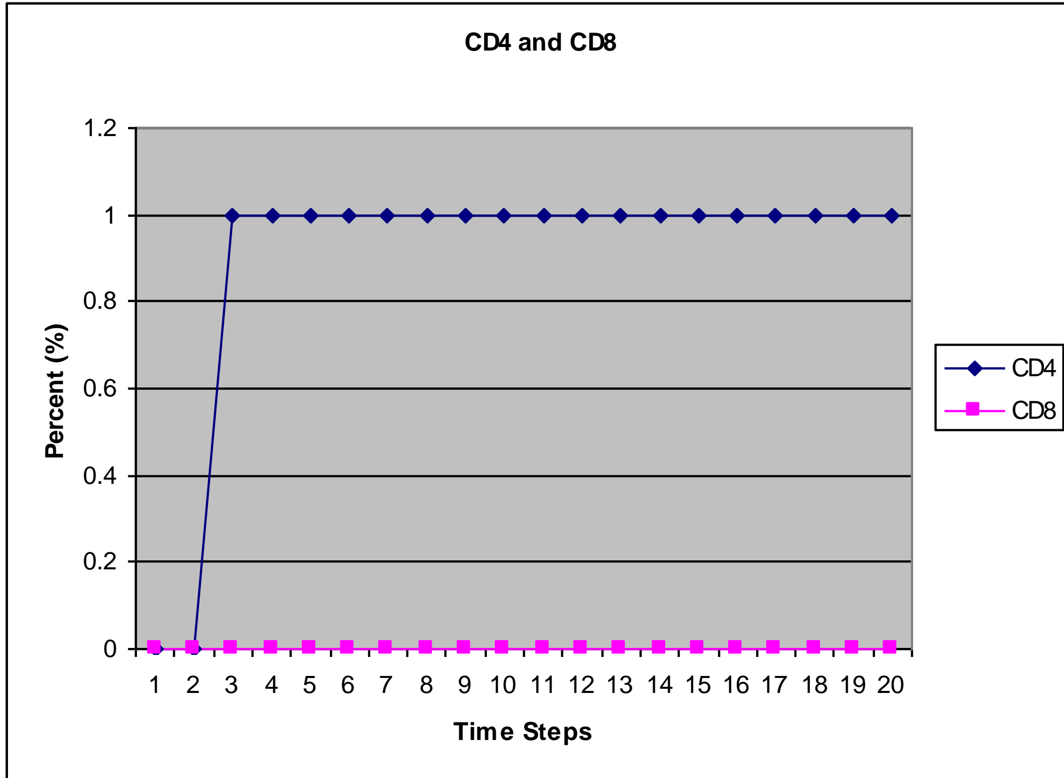

Figure 11: Expression behavior of CD4 and CD8 Results This graph shows that the T cells differentiated into CD4+ cells at the third step and never became CD8+ cells.

CD4 and CD8 expression are mutually exclusive, so it makes sense that if a cell differentiates into a CD4+ cell, it could not express CD8. Figure 6 showed that CMYB was turned on because PU.1 was not present, and CMYB acted to turn on GATA-3. PRETCR also turned on TCRALPHA, which turned on TOX. GATA-3 and TOX working together turned on THPOK. By my system rules, either THPOK or GATA-3 or both can turn on CD4 expression. Also, THPOK directly inhibits RUNX3 expression, which is a factor that turns on CD8 expression.

Discussion:

While this test seems to imply that turning on PRETCR leads to CD4 expression in T cells, it may not be the case. Looking back into the network rules revealed there may be some errors or omissions that affected the results. RUNX3 is turned on by TBET, but TBET is missing from the rules, so in effect it can't be turned on. Because of this, cells in this network could only become CD8+ in the absence of GATA3 and THPOK. While Rothenburg has stated that ThPOK is the master regulator over RUNX3 for the CD4/CD8 lineage choice (Rothenburg, 2009), the omission of TBET still should be corrected in the data and then run the simulation again to see what, if any, difference would occur in the results.

Other possible errors could be with certain transcription factors like BC12, ERG, and ID1. Their rules for being turned on simply state "not PRETCR" for the first two and "not GATA3" for the third. This implies that these genes are always on unless a certain factor is present. This could be true, or there could be a factor that turns on those genes that wasn't listed in the literature used to make the network model. More research to cross-reference the transcription factors could help validate the network rules.

This experiment did seem to be successful in modeling certain factors that depend on the pre-TCR signal to activate, including TCRALPHA, TOX, CMYB, GATA3 and ThPOK, and by graphing the generated results, it was much easier to see what factors were affected by a certain gene rather than scanning a long list of network rules. This type of modeling helped to visually determine not just whether a gene would be turned on or off because of the presence of another factor, but when that gene was affected in relation to other factors, helping one determine a timeline of a lineage pathway.

Conclusions:

Further refinement is needed to correct the above-listed possible errors, and more research would be done to fix and update the rules. Following that, the PRETCR test would be rerun to note any changes from the original. This retesting of the system would help to better determine if the pre-TCR signal exclusively leads to $CD4^+$ T cell lineage fates. From the background research done for this project, the pre-TCR signal most likely does not lead to solely $CD4^+$ T cells. Bardnt states that the pre-TCR signal directs differentiation and proliferation of T cells during the transition from $CD4^-CD8^-TCR^-$ double negative (DN) stage to $CD4^+CD8^+$ double positive (DP) stage during T cell development, so theoretically the pre-TCR signal should be able to lead to both $CD4^+$ and fates $CD8^+$ (Barndt, 1999). Keeping that in mind, some transcription factor(s) following the induction of the pre-TCR signal should decide the CD4/CD8 lineage choice, so more experimentation would need to be done to better identify which factors are responsible for that cell lineage divergence.

This experiment was successful in showing the theory that biological systems can be tested using simulation models, so with more work this method could be a valuable tool in researching various other factors beyond the CD4/CD8 lineage choice. More experiments could be run in the same way to determine the effects of turning on or off different genes, like NOTCH or PU1. The data generated by these tests would be graphed as shown above in the PRETCR test to follow the effects of these factors on other factors in the pathway and on T cell development. The experiments would not need to be limited to even T cell lineage fates; one could research other systems like B cell development or myeloid cell lineages to create gene network models like the ones shown in this paper, and use those rules in the Booleannet software to experiment and learn the importance of various transcription factors and genes in those lineages.

References

1. Barndt, Robert; Mei-Fang Dei and Yuan Zhuang. "A Novel Role for HEB Downstream or Parallel to the Pre-TCR Signaling Pathway During αß Thymopoiesis." *The Journal of Immunology* 163 (1999): 3331-3343. Print.

2. "Boolean Network Modeling." *Booleannet.* Google, 2010. Web. February 11, 2010.

3. Georgescu, Constantin; William J. R. Longabaugh, et. al. "A gene regulatory network armature for T lymphocyte specification." *Proceedings of the National Academy of Sciences* 105 (2008): 20100- 20105. Print.

4. Georgopoulos, Katie; Michael Bigby; et. al. "The ikaros gene is required for the development of all lymphoid lineages." *Cell* 79 (2004): 143-156. Print.

5. Goodell, M.A.; K. Brose; G. Paradis; A.S. Connor and R.C. Mulligan. "Isolational and functional properties of murine hematopoietic stem cells that are replication in vivo." *JEM* 183 (1996): 1797-1806. Print.

6. Kindt, Thomas J.; Richard A. Goldsby and Barbara A. Osborne. *Kuby Immunology: Sixth Edition*. New York: W.H. Freeman and Company, 2007.

7. Kurokawa, Mineo. "AML1/Runx1 as a Versatile Regulator of Hematopoiesis: Regulation of Its Function and a Role in Adult Hematopoiesis." *International Journal of Hematology* 84 (2006): 136-142. Print.

8. Massa, Steffen; Gina Balciunaite; Rhodri Ceredig and Antonius G. Rolink. "Critical role for c-kit (CD117) in T cell lineage commitment and early thymocyte development in vitro." *European Journal of Immunology* 36 (2006): 526-532. Print.

9. Pearse, Erika L.; Allen C. Mullen; et. al. "Control of Effector CD8$^+$ T Cell Function by the Transcription Factor Eomesodermin." *Science* 302 (2003): 1041-1043. Print.

10. Spits, H.; B. Blom; et. al. "Early Stages in the development of human T, natural killer, and thymic dendritic cells." *Immunologic Reviews* 165 (1998): 75-86. Print.

11. Rothenburg, Ellen V. "Decision by committee: new light on the CD4/CD8-lineage choice." *Immunology and Cell Biology* 87 (2009): 109-112. Print.

12. Rothenburg, Ellen V.; Tom Taghon. "Molecular genetics of T cell development." *Annual Reviews of Immunology* 23 (2005): 601-649. Print.

Appendix A: List of Figures

Figures: *Page:*

Figure 1: T cell developmental pathway 5

Figure 2: General Structure of the CD4 and CD8 Co-receptors 7

Figure 3: Example Network Model 13

Figure 4: How to Read a Network Model 14

Table 1: Regulatory Network Rules 17

Table 2: Function of Genes/Transcription Factors 21

Table 3: Excel Data of Nodes 24

Figure 5: Expression behavior of TCRALPHA and TOX genes. 25

Figure 6: Expression behavior of CMYB and PU.1 genes 26

Figure 7: Expression behavior of GATA-3 and NOTCH genes 27

Figure 8: Expression behavior of ThPOK and RUNX3 genes 28

Figure 9: Expression behavior of DN2DN3SNR and BCL1B genes 29

Figure 10: Expression behavior of CEBPA and GMCSFR genes 30

Figure 11: Expression behavior of CD4 and CD8 Results 31

Appendix B: Academic Vita of Allison Jaime Neely

Local Address: 255 East Beaver Ave., Apt. 1004
State College, PA 16801

Home Address: 1795 Hedwig Drive
Allison Park, PA 15101

E-Mail Id: ajn5046@psu.edu

Education

Major: Animal BioScience

Honors Adviser: Dr. Lester Griel

Thesis Title: The Gene Network Concerning T Lymphocyte Activation

Thesis Supervisor: Dr. Avery August

Work Experience

Fecal Parasite Project at the Penn State Horse Barns

Date: June 2009- December 2009

Description: I worked with veterinarian Dr. Jedrzejewski to analyze equine fecal samples to quantify parasite eggs. This project is working off of the theory that some horses in a herd shed the bulk of the total parasites, and so if we can identify which horses are the high shedders and create an effective de-worming program, we can greatly reduce the parasite population in the herd.

Institution/Company: Penn State Horse Barns

Supervisor's Name: Dr. Ed Jedrzejewski

Job at Penn State's Dairy Barns

Date: May 2009 – August 2009

Description: As a dairy barns worker I fed, cleaned up after, handled, and milked the dairy cows. I drove a tractor to push up feed and a Gator to drive cows to pasture. I assisted my boss in pulling calves, then took care of, fed, and vaccinated the newborn calves. I cleaned the facilities, bailed hay, and assisted in recordkeeping. When veterinarians visited our farm, I shadowed them, asked questions about their duties, observed surgeries, and assisted wherever needed.

Institution/Company: Penn State Dairy Barns

Supervisor's name: Nadine Houck

Volunteer Work at Centre Wildlife Care

Date: September 2008- May 2010

Description: Headed by wildlife rehabilitator Robyn Graboski, Centre Wildlife Care takes in sick, orphaned, and injured wildlife to rehabilitate and hopefully release. My duties included feeding the animals, cleaning cages, administering medicine, cleaning dishes and the facility, and general handling and restraining of the animals.

Institution/Company: Centre Wildlife Care

Supervisor's name: Robyn Graboski

Internship at the Animal Rescue League of Pennsylvania Wildlife Center

Date: May 2008- August 2008

Description: At the Animal Rescue League Wildlife Center I interned under wildlife rehabilitator Jill Argall. My duties included feeding the animals; tubing rabbits and pidgeons; cleaning cages, habitats, and the facility; intaking and examining new cases; adminstering medicine and vaccines; performing small surgeries like setting and wrapping a broken wing, and working with the public to solicit donations and give advice about dealing with wildlife.

Institution/Company: Animal Rescue League of Pennsylvania Wildlife Center

Supervisor's name: Jill Argall

Job at Dickie, McCamey & Chilcote, P.C. (Law Firm)

Date: May 2007 – August 2007 and May 2008- August 2008

Description: I worked at this law firm for two summers. The first summer I worked on a data entry project for the computer program ProLaw, and the second summer I worked for secretaries, helping manage case files. While this experience does not relate directly to veterinary medicine, I did learn and practice many valuable office skills like organization, data entry, copying, and work with computer programs like ProLaw, Word, and Excel.

Institution/Company: Dickie, McCamey & Chilcote, P.C.

Job at Penn State's Animal Diagnostic Lab

Date: September 2007 – May 2008

Description: Working for veterinarian Dr. Key at the Penn State Animal Diagnostic Lab, I worked in the serology department preparing and analyzing blood samples with PCR (polymerase chain reaction) to look for BVD (Bovine Virus Diarrhea) I and II. I prepped the samples, helped run the PCR, and wrote and organized the reports.

Institution/Company: Penn State's Animal Diagnostic Center

Supervisor's name: Dr. Douglas Key

Awards: Awarded Dean's List six times, member of Phi Eta Sigma

Penn State Clubs: Pre-Vet Club, Swing Dance club, Circle K, Intramural Football, Intramural Volleyball, Students for the Responsible Use of Animals

Community Service Involvement: Member of Circle K, volunteered at Centre Wildlife Care