MODELING CELL FATE SPECIFICATION MECHANISMS IN HAIR FOLLICLES

By

Breeann Novasel

A capstone project submitted for Graduation with University Honors

May 6, 2022

University Honors

University of California, Riverside

APPROVED

Dr. Qixuan Wang
Department of Mathematics

Dr. Richard Cardullo, Howard H. Hays Chair
University Honors

Abstract

We all have hair, and it is important to many biological features, such as camouflage and thermoregulation. Hairs are produced inside hair follicles (HFs), which are mini organs within the mammalian skin. Due to their richness in stem cells, HFs are able to spontaneously regenerate through life, a process being strictly regulated. It is crucial to understand the cell fate specification mechanisms in HFs, which maintain the functions of the HF and regulate its growth dynamics. In this project, we develop a data-driven mathematical model to study the cell fate specification mechanisms in HFs, where we first apply pseudotime analysis on recently published HF single cell RNA-sequencing data, then apply Boolean model on it, which allow us to identify the key factors in HF cell fate regulation.

Acknowledgements

I cannot thank my faculty mentor, Dr. Qixuan Wang, enough for her endless support in directing me through all aspects of conducting undergraduate research and designing my project during the COVID-19 pandemic. Thank you for all of your guidance in leading me through this project and allowing me to work with you on this project so readily. I was unsure at the beginning of the research regarding how I was going to conduct my research project since it is outside of my field of microbiology, but with your guidance I am now able to confidently present my research in computational biology to others. I would also like to thank my fellow research partner, Christine Pham. Your hard work does not go unnoticed, and working with you on this project has made it more enjoyable than it already is.

I would also like to thank my family for everything they have done for me throughout my undergraduate career. Thank you for motivating me both academically and personally. During the pandemic, I was feeling more anxious and insecure than ever, and I am so grateful for your understanding and for giving your encouragement when I needed it most. Without you all, I would not be the person I am today, and I will forever be grateful for everything you have done for me and for guiding me throughout my time at UC Riverside and in University Honors. Thank you for giving me the strength I needed.

Table of Contents

Introduction

Hair has evolved to become a distinguishing feature of mammals in biology. Hair has come to play a role in various processes, such as in insulation, camouflage, and the facilitation of social interaction (Chen et al 2020). Embedded into our skin are hair follicles. Hair follicles (HFs) are mini organs embedded within the mammalian skin. The mammalian skin is composed of three different layers: the epidermis, dermis, and hypodermis. HFs are present primarily in the epidermal and dermal layers of the skin, and are composed of different structures which all have different functions. One of the main components of HFs are stem cells, which are cells that do not have a designated function at the beginning of their life cycle. This is what makes HFs a unique model system for scientific experiments.

As mentioned previously, HFs contain stem cells in the bulge of the HF. These stem cells will undergo different developmental stages to ultimately gain function. They start as a stem cell and then go into an intermediate stage called the transient fate, and then eventually become terminally differentiated where the stem cells develop their function and will have that function for the rest of their lives.

The decision of a stem cell to choose what function they will perform for the rest of their lives is often controlled by transcription factors, which are proteins that bind to the beginning of a specific sequence of our genetic material, DNA, to either promote or inhibit the conversion of DNA into messenger RNA. The conversion of host DNA into messenger RNA is known as transcription. After undergoing some modifications, the messenger RNA will be translated into protein, and these proteins can contribute to the stem cell's chosen function, and this process is referred to as translation. This is all based on gene expression, in which different genes in our DNA are either expressed (turned on) or suppressed (turned off). If a gene is "turned on," then

the DNA in that gene will be transcribed into messenger RNA (mRNA). Consequently, that mRNA will then be translated into a protein, and the protein will carry out its function to ensure the cell lives.

In addition to containing stem cells, HFs themselves are composed of many structures that all have their own functions. The HF itself has three main components: the infundibulum, isthmus, and the lower follicle (Martel et al 2021). The infundibulum consists of the upper region of the HF (Martel et al 2021). The isthmus is the area of the HF that is between the sebaceous gland and the bulge, which is where stem cells in the HF are located (Martel et al 2021). Lastly, the lower follicle consists of most of the structures important for HF growth, regression, and cell signaling. In the lower follicle, there is the dermal papilla, which is primarily responsible for most of the cell-to-cell communication and promotes division amongst those cells. This region also contains the cellular matrix, which includes many of the divisible cells and cells that ultimately produce the hair itself. The matrix includes three different layers: the inner root sheath (ORS), medulla, and the cortex. The cortex makes up most of the hair (Martel et al 2021).

HFs continuously grow, regress, and sometimes enter a resting state during their development. The growth period of a HF is referred to as anagen; the regression stage is referred to as catagen; and the resting stage is referred to as telogen (Joost et al 2020). The anagen phase is the growth, or proliferation phase. It is in this phase only that the lower follicle, which contains the divisible cells in the matrix, is present (Martel et al 2021). There is also cellular communication that occurs between cells that wish to go from catagen (growth stage) or anagen (regression stage) to telogen, the resting stage of HF development (Oshimori et al 2011).

In order to initiate cell division, the dermal papilla signals to the stem cells present in the bulge. Once these stem cells are stimulated, the lower follicle can grow downwards, forming a

bulb around the dermal papilla. This allows for the dermal papilla to signal matrix cells in the bulb to proliferate, differentiate, and grow upward to form the new hair (Martel et al 2021). Catagen is the regression phase of HF development, and is the shortest stage compared to anagen and telogen (Martel et al 2021).

In order for a HF to initiate catagen, cellular division in the matrix needs to stop happening, and the lower follicle eventually will need to disappear (Martel et al 2021). Once the lower follicle disappears, the bulb around the dermal papilla fades and the dermal papilla moves upward, which consequently leads to the development of a hard white bulb (Martel et al 2021). The stem cells in the matrix can coordinate apoptosis, programmed cell death, with one another and consequently the HF will enter its telogen phase

Lastly, if a HF wishes to enter the telogen or resting phase, club hairs are formed and shed after several months (Martel et al 2021). The shedding of the club hair can act as a signal to initiate anagen so a new hair can be developed.

In this project, the proliferating cells that belong to the matrix will be examined more in depth. These cells are stem cells, and are undifferentiated in the germinative layer of the HF. The matrix itself is composed of three layers (from inside to outside): the medulla, cortex, and inner root sheath (IRS). Somehow, the stem cells in the germinative layer are able to differentiate into these three different structures independent of the other cells. These three layers branch out from the germinative layer and are distinct from one another. In this capstone project, the changes in gene expression in these stem cells will be modeled with a gene regulatory network, apply Boolean logic, and with bioinformatic methods visualize changes in gene expression in the stem cells in the matrix.

Methods

The first technique used is to make a preliminary Boolean network to model the gene regulatory network (GRN) from literature, and the primary data coming from single-cell RNA (scRNA) sequence data (Joost et al 2020). The preliminary gene regulatory, or Boolean, network will be applied to the dividing cells in the matrix of the lower follicle in the HF. As mentioned in the introduction, a gene regulatory network will be constructed and applied. A gene regulatory network, also referred to in this project as a Boolean network, contains different nodes, and these nodes have two different states- ON (1) and OFF (0). Since gene expression changes frequently, the future state of each node in the network depends on the current state of its regulators in the network. Boolean logic is applied to the gene regulatory network and refers to the actions of the regulators in the network. These operators to represent the actions are NOT, AND, and OR.

For example, a Boolean network can have nodes X, Y, and Z. These nodes have 2^n states, where n represents the number of nodes. The Boolean operator 'AND' is represented by X=Y and Y=X ∧ Z, which translates into Y activates X and Y can be activated by X and Z respectively. The Boolean operator 'OR' is written as Y = X ∨ Z, which means Y can be activated by either X or Z. Lastly, the Boolean operator 'NOT' is written as Z = ¬X, which translates to X cannot activate Z.

For the Boolean network designed in this project, there are three states that a node can have: cell division, differentiation, and apoptosis. We looked at cells undergoing anagen, which is the stage in which a HF stops growing. The genes *Krt27, Krt71, Krt35, Krt73, Selenbp1, Rgcc, Aldhla3,* and *Foxq1* were applied to the first method and were generated by code run on the platform MatLab. The expression of each gene was measured randomly at different points in time, and due to this randomness, the code generated different figures every time. A final figure

was developed tracking the overall expression levels of these genes. These genes were then assigned different node positions, and the validity of the genes as nodes was evaluated with a Python code provided by Dr. Fiona Hamey (Hamey et al 2018). This code was utilized for assigning a z3-score to each gene, and the z3-score is from a command line tool that was previously built by Microsoft Windows. The z3-score is a decimal, with decimals closer to the number 1 (i.e. 0.9, 0.8) more preferable than lower z3-scores. This provides insight into genes that could be potential attractors and stable motifs in the proposed GRN. This process was repeated for each gene in the different propagative layers of the matrix: the medulla, inner root sheath (IRS), and cortex. The command used to run the code is as follows: "python booleanRules.py [gene name] [(IRS/MED/CX)_Data_Table_MxGRN.txt] 5 [(IRS/MED/CX)_Ord_MxGRN.txt] 3 3 [Intl_Graph_(IRS/MED/CX).txt] 0.95 0.05." A .sh file was submitted to the UC Riverside HPCC as a non-interactive job with the command "sbatch file_name.sh."

```
#!/bin/bash -l

#SBATCH --nodes=1
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=2
#SBATCH --mem-per-cpu=1G
#SBATCH --time=5-00:00:00      # 5 day and 00 minutes
#SBATCH --output=IRS_Foxq1.log
#SBATCH --mail-user=bnova003@ucr.edu
#SBATCH --mail-type=ALL
#SBATCH --job-name="IRS_Foxq1"
#SBATCH -p batch # This is the default partition, you can use any of the following; intel, batch, highmem, gpu


# Print current date
date

# Load python
module load z3
python3 booleanRules.py Foxq1 IRS_DataTable_MxGRN.txt 5 IRS_Ord_MxGRN.txt 3 3 IntlGraph_IRS_MxGRN.txt 0.95 0.05
exit
```

Figure 1. This .sh (shell script) file was submitted for the gene *Foxq1* in the inner root sheath of the matrix of the HF. Both modules Python and z3 were called to successfully run the command in the second-to-last line of the file. The name of the file is IRS_Foxq1.sh.

The second method used in this project was to refine the initial Boolean network created from the first method. From a literature review, sequenced single-cell RNA (scRNA-seq) data from Mx cells in the HF was used. Genes that were found to not be expressed in these cells were removed from the project. It was at this point that we applied the pseudotime Boolean (PB) network inference method to revise the Boolean model made from the scRNA-seq data. Each node in a Boolean network is scored depending on how often a node agrees with the pairs of input-output cells along the pseudotime trajectory. For this, cells belonging to the three main sections of the matrix, the medulla, cortex, and inner root sheath, were scored with the z3-solver on UCR's High Performance Computer Cluster (HPCC). A diffusion map illustrated that these three layers were able to form distinctly from one another, even as they all stemmed from the main germinative layer in the matrix (Joost et al 2020). Each of these components represents a trajectory of the stem cells in the matrix, so to measure these trajectories, diffusion maps were made.

One single run of the pseudotime Boolean network does not lead to a node reaching a steady state. Some nodes oscillate, in which cells can "commit" multiple fates at the same time, so more than one node can be on or off at the same time. Boolean networks can also have attractors, which describe the long-time behavior of the overall system. There are two categories of attractors: fixed points and complex attractors. Fixed points represent stable states where the state of the system does not change. Complex attractors represent wherein the system oscillates among a set of states. These attractors were analyzed with motifs. Only stable motifs were considered in this project; stable motifs are the smallest and strongly connected component of a network that does not contain both a node and its negation, and if composite nodes are present, it also needs to contain the inputs of the composite nodes. Nodes of a stable motif will have a

steady state in any attractor of the network. With the preliminary GRN that was constructed, attractor and motif analysis was performed via Microsoft Visual Studio Code and the computer package "pystablemotifs" by GitHub user jcrozum. The nodes in this project are associated with the genes: *Rnaset2b, Rgcc, Foxq1, Aldhla3, Krt27, Krt35, Krt71, Krt73,* and *Selenbp1*.

Currently, we are in the process of running the necessary code to perform pseudotime analysis and to apply the preliminary Boolean model made in method two of the project to the data from Simon Joost's paper "The Molecular Anatomy of Mouse Skin During Hair Growth and Rest." Diffusion maps, which use a diffusion distance, are conceptually relevant to how differentiation data is generated biologically, as cells follow noisy diffusion-like dynamics in the course of taking several differentiation lineage paths (Haghverdi et al 2015). Diffusion maps preserve the non-linear structure of data as a continuum and are robust to noise (Haghverdi et al 2015). Diffusion maps have been made for the inner root sheath (IRS), but not yet for the medulla and the cortex. The data from Simon Joost's paper was made with the computer programming language Python, and will be run on UC Riverside's HPCC.

Results

As mentioned in the previous section, a preliminary Boolean network was created to model the GRN of the stem cells in the HF. Though there are many genes present in these stem cells, the only genes that were observed in this project were: *Rnaset2b, Rgcc, Foxq1, Aldhla3, Krt27, Krt35, Krt71, Krt73,* and *Selenbp1*. The reason for doing this was to not overwhelm personal computers that were being used to initially run the code, but jobs were also performed on HPCC to ensure everything on the computer was running correctly without causing the computer to crash. The first step was to determine if all of these genes were actively being

expressed, so one diagram was generated to randomly track the expression of these genes at different points in time.

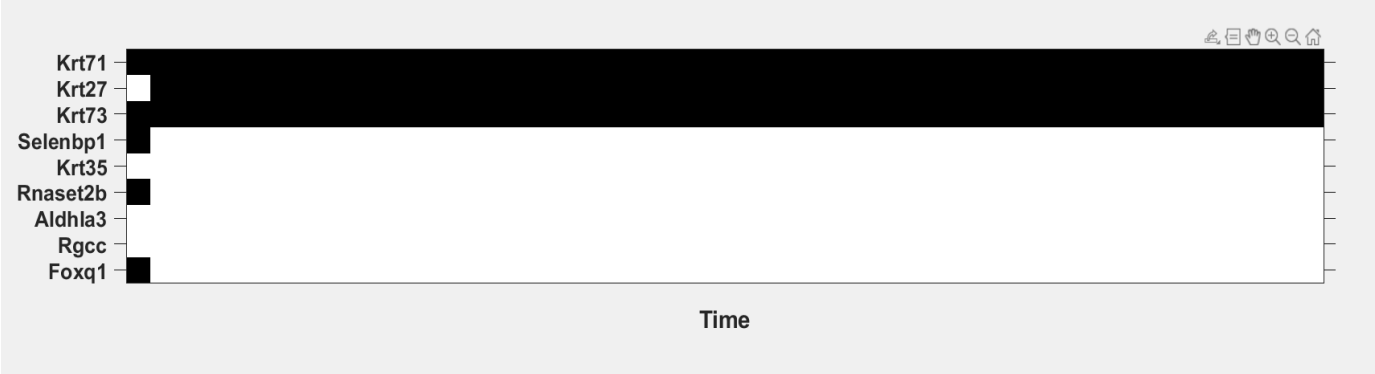


Figure 2. Time graph of measuring gene expression among the genes selected (left axis). Time is on the x-axis. The black represents time the gene is expressed and the white represents the time that the gene is not expressed. Since this process of getting the graph is a random one, the ratio of white and black will change each time the code is run on MatLab.

Figure 3. Another representation of a time graph measuring gene expression. This is the second graph generated from the MatLab code written to produce this graph.

Some of the genes listed are always expressed ("on"), such as *Krt71*, while others are mostly expressed, such as *Krt27*, *Krt73*, and *Selenbp1*. Figures 1 and 2 did not have a dynamic expression of gene expression, as gene expression is something that can change over time and be influenced by different transcription factors. However, though these two figures measure gene

expression at different and random points, the gene *Aldhla3* was not expressed at any point in time. Therefore, it was removed from the genes of interest.

Figure 4. Shows a more dynamic level of gene expression for each gene. The gene *Aldhla3* was removed from the graph since it did not contribute to expression. This graph measures the dynamics of the 8 genes being worked with at random points in time until time reaches 1000.

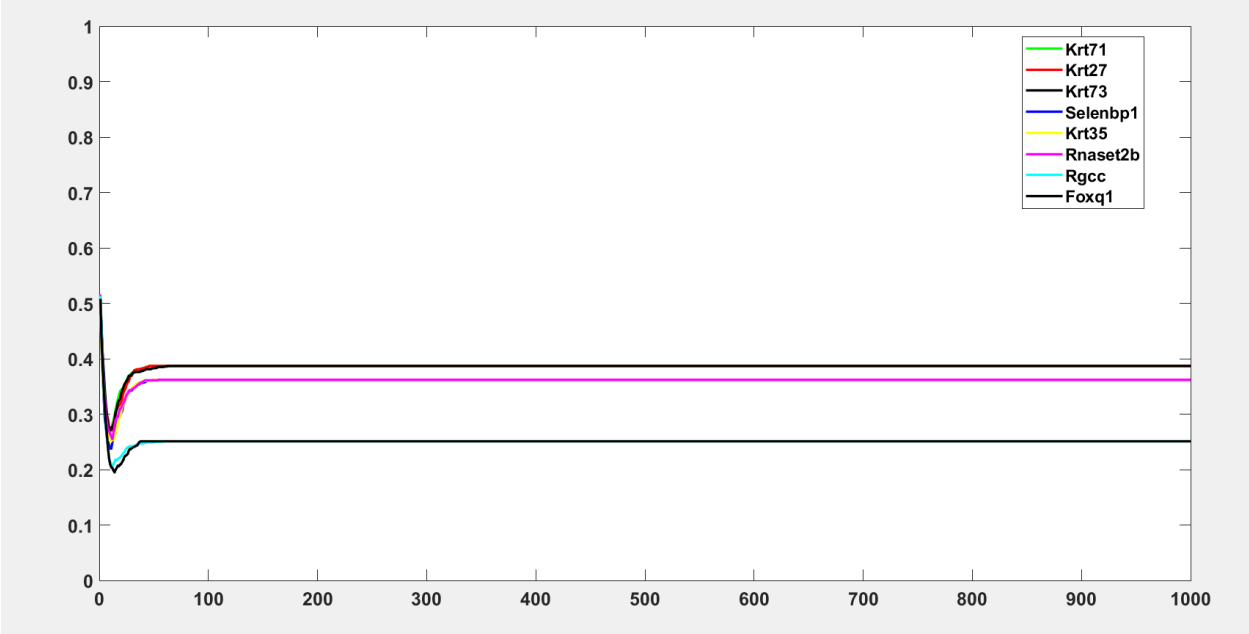


Figure 5. Overall dynamics of the gene expression of the genes selected. This graph was also produced by MatLab, and shows the relative stability of gene expression over longer amounts of

time. The frequency level ranges from anywhere between 0.2 and 0.5. Time is on the x-axis and expression level is on the y-axis.

After the initial dynamics were established, the python script booleanRules.py was run on UC Riverside's HPCC with the command provided by code from the literature as a non-interactive job. Using the .sh file created in the methods section, the Boolean rules were able to be computed. This allowed for the establishment of different boolean rules for each gene in each of the different layers of the matrix. After the rules were generated by the computer, the general rules for each gene were compared with the z3-scores provided for each gene in the cortex, medulla, and inner root sheath. The average z3-score and standard deviation was then calculated.

```
Agreement level = 0.950000
The best rules were:
((281, [('a1', 'Krt35'), ('r1', 'Krt71'), ('a2', 'Rnaset2b'), ('r0', 'or'), ('a0', 'or'), ('r2', 'Krt27')]), 'z3 score 0.959044')
((281, [('r4', 'Krt73'), ('r1', 'or'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'Rnaset2b'), ('r2', 'Krt71')]), 'z3 score 0.959044')
((281, [('r1', 'Krt71'), ('r0', 'or'), ('a0', 'Krt35'), ('r2', 'Krt27')]), 'z3 score 0.959044')
((281, [('r4', 'Rgcc'), ('r1', 'or'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'Krt35'), ('r2', 'Krt71')]), 'z3 score 0.959044')
((281, [('a1', 'Krt35'), ('r1', 'Krt71'), ('a2', 'Rnaset2b'), ('r0', 'or'), ('a0', 'and'), ('r2', 'Krt27')]), 'z3 score 0.959044')
((281, [('r4', 'Rgcc'), ('r1', 'or'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'Rnaset2b'), ('r2', 'Krt71')]), 'z3 score 0.959044')
((281, [('r4', 'Krt73'), ('a1', 'Krt35'), ('r1', 'or'), ('a2', 'Rnaset2b'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'and'), ('r2',
'Krt71')]), 'z3 score 0.959044')
((281, [('r1', 'Krt71'), ('r0', 'or'), ('a0', 'Rnaset2b'), ('r2', 'Krt27')]), 'z3 score 0.959044')
((281, [('r4', 'Rgcc'), ('a1', 'Krt35'), ('r1', 'or'), ('a2', 'Rnaset2b'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'or'), ('r2', 'Krt71')]),
'z3 score 0.959044')
((281, [('r4', 'Foxq1'), ('a1', 'Krt35'), ('r1', 'or'), ('a2', 'Rnaset2b'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'or'), ('r2', 'Krt71')]),
'z3 score 0.959044')
((281, [('r4', 'Foxq1'), ('r1', 'or'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'Krt35'), ('r2', 'Krt71')]), 'z3 score 0.959044')
((281, [('r4', 'Foxq1'), ('r1', 'or'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'Rnaset2b'), ('r2', 'Krt71')]), 'z3 score 0.959044')
((281, [('r4', 'Krt73'), ('r1', 'or'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'Krt35'), ('r2', 'Krt71')]), 'z3 score 0.959044')
((281, [('r4', 'Rgcc'), ('a1', 'Krt35'), ('r1', 'or'), ('a2', 'Rnaset2b'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'and'), ('r2', 'Krt71')]),
'z3 score 0.959044')
((281, [('r4', 'Foxq1'), ('a1', 'Krt35'), ('r1', 'or'), ('a2', 'Rnaset2b'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'and'), ('r2',
'Krt71')]), 'z3 score 0.959044')
((281, [('r4', 'Krt73'), ('a1', 'Krt35'), ('r1', 'or'), ('a2', 'Rnaset2b'), ('r3', 'Krt27'), ('r0', 'or'), ('a0', 'or'), ('r2', 'Krt71')]),
'z3 score 0.959044')
```

Figure 6. Boolean rules produced by UC Riverside's HPCC. The rules computed are for the gene *Selenbp1* in the IRS of the matrix of the HF. The code provided to get these rules is from the Hamey et al paper and a batch .sh file was constructed as mentioned in the methods section. The z-score for this rule is 0.959044, which indicates that the rule has a high agreement level. The higher the z3 score is, the better confidence we have in the Boolean rule computed.

Gene	Rule	Z-Score 1 (IRS)	Z-Score 2 (MED)	Z-Score 3 (CX)	Mean	SD
Krt71	= (Krt71 | Krt27 | Krt73) &! Selenbp1 &! Krt35 &! Rnaset2b &! Rgcc &! Foxq1	0.911263	0.834395	0.769863	0.838507	0.0707896 3
Krt27	= (Krt71 | Krt27 | Krt73) &! Selenbp1 &! Krt35 &! Rnaset2b &! Rgcc &! Foxq1	0.672355	0.81741*	0.617808	0.7025243 3	0.1031643 4
Krt73	= (Krt71 | Krt27 | Krt73) &! Selenbp1 &! Krt35 &! Rnaset2b &! Rgcc &! Foxq1	0.648464	0.957537**	0.943836	0.8499456 7	0.1746226 7
Selenbp1	= (Selenbp1 | Krt35 | Rnaset2b) &! Krt71 &! Krt27 &! Krt73 &! Rgcc &! Foxq1	0.959044	0.944798	0.768493	0.8907783 3	0.1061414 8
Krt35	= (Selenbp1 | Krt35 | Rnaset2b) &! = Krt71 &! Krt27 &!	0.832765	0.774947	0.708219	0.771977	0.0623261

	Krt73 &! Rgcc &! Foxq1					
Rnaset2b	= (Selenbp1 | Krt35 | Rnaset2b) &! Krt71 &! Krt27 &! Krt73 &! Rgcc &! Foxq1	0.539249	0.636943	0.684932	0.6203746 7	0.0742412 7
Rgcc	= (Rgcc | Foxq1) &! Krt71 &! Krt27 &! Krt73 &! Selenbp1 &! Krt35 &! Rnaset2b	0.921502	0.681529***	0.757534	0.786855	0.122644
Foxq1	= (Rgcc | Foxq1) &! Krt71 &! Krt27 &! Krt73 &! Selenbp1 &! Krt35 &! Rnaset2b	0.969283	0.770701	0.90137	0.8804513 3	0.1009301 5

*Krt27 rules short → Foxq1, Rgcc, Selenbp1, Rnaset2b, and Krt35 not shown in Boolean rules
**MED_Krt73 not in best rules set (Krt73 rule)
***MED_Rgcc not in best rules set (Rgcc rule)

Figure 7. Table summarizing the Boolean rules for each gene in all three layers of the matrix of the HF (medulla (MED), inner root sheath (IRS), and cortex (CX)). Next to each gene are the best rules for that gene, its z3-scores in the IRS, MED, and CX respectively, its mean, and standard deviation. The average z3-scores were computed and the standard deviation was calculated, then added to the table.

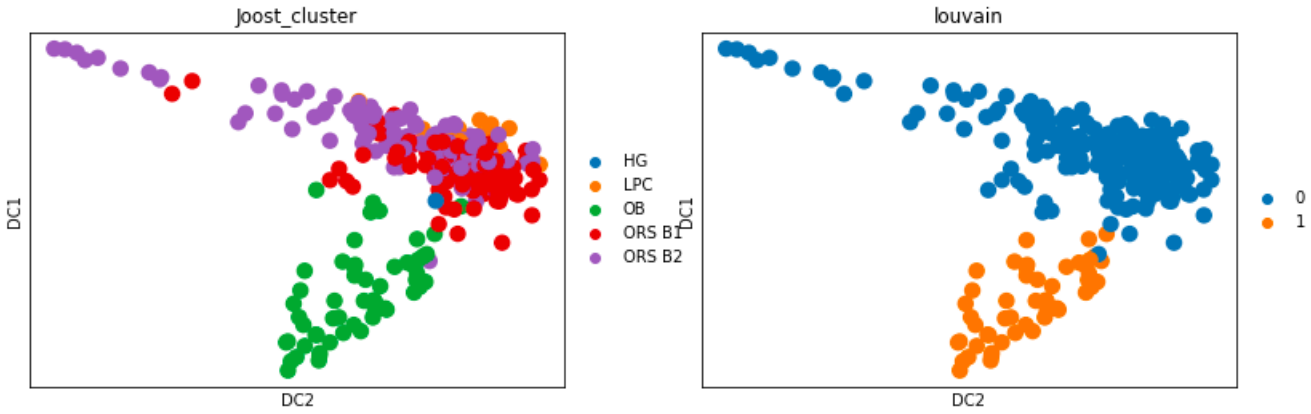
The next step in the procedure was to perform attractor and stable motif analysis. As stated previously, attractor analysis is performed to determine the fixed points of a small Boolean

network. With that, the time-dependency from the Boolean network can be removed and the resulting sets of equations can be solved: $X_i = Y_i$, $Y_i = X_i \wedge Z_i$, and $Z_i = \vee X_i$. Stable motifs are the smallest strongly connected component that does not contain both a node and its negation. Also, if it contains composite nodes, it also needs to contain the inputs of each node. Nodes of a stable motif will have a steady state in any attractor of the network. The stable motif and attractor analysis allows us to identify expanded networks, stable motifs, and reduce the network's complexity using the state of one of multiple stable motifs. Attractors describe the long-time behavior of the network, so attractors are able to describe the behavior of the network when it changes and when it does not change at any time.

```
RULES
Foxq1* = Foxq1 & !Krt27 & !Krt35 & !Krt71 & !Krt73 & !Rnaset2b & !Selenbp1 | Rgcc
Krt27* = !Foxq1 & !Krt35 & Krt73 & !Rgcc & !Rnaset2b & !Selenbp1 | Krt71 | Krt27
Krt35* = !Foxq1 & !Krt27 & !Krt71 & !Krt73 & !Rgcc & Rnaset2b | Selenbp1 | Krt35
Krt71* = !Foxq1 & !Krt35 & Krt73 & !Rgcc & !Rnaset2b & !Selenbp1 | Krt71 | Krt27
Krt73* = !Foxq1 & !Krt35 & !Rgcc & !Rnaset2b & Selenbp1 | Krt73 | Krt71 | Krt27
Rgcc* = Foxq1 & !Krt27 & !Krt35 & !Krt71 & !Krt73 & !Rnaset2b & !Selenbp1 | Rgcc
Rnaset2b* = !Foxq1 & !Krt27 & !Krt71 & !Krt73 & !Rgcc & Rnaset2b | Selenbp1 | Krt35
Selenbp1* = !Foxq1 & !Krt27 & !Krt71 & !Krt73 & !Rgcc & Rnaset2b | Selenbp1 | Krt35

Analyzing network . . .
Analysis complete.

There are 11 attractors.
{'Foxq1': 0, 'Krt27': 1, 'Krt35': 1, 'Krt71': 1, 'Krt73': 1, 'Rgcc': 0, 'Rnaset2b': 1, 'Selenbp1': 1}

{'Foxq1': 0, 'Krt27': 1, 'Krt35': 0, 'Krt71': 1, 'Krt73': 1, 'Rgcc': 0, 'Rnaset2b': 0, 'Selenbp1': 0}

{'Foxq1': 1, 'Krt27': 1, 'Krt35': 1, 'Krt71': 1, 'Krt73': 1, 'Rgcc': 1, 'Rnaset2b': 1, 'Selenbp1': 1}

{'Foxq1': 1, 'Krt27': 1, 'Krt35': 0, 'Krt71': 1, 'Krt73': 1, 'Rgcc': 1, 'Rnaset2b': 0, 'Selenbp1': 0}

{'Foxq1': 1, 'Krt27': 0, 'Krt35': 1, 'Krt71': 0, 'Krt73': 1, 'Rgcc': 1, 'Rnaset2b': 1, 'Selenbp1': 1}

{'Foxq1': 1, 'Krt27': 0, 'Krt35': 1, 'Krt71': 0, 'Krt73': 0, 'Rgcc': 1, 'Rnaset2b': 1, 'Selenbp1': 1}

{'Foxq1': 0, 'Krt27': 0, 'Krt35': 1, 'Krt71': 0, 'Krt73': 1, 'Rgcc': 0, 'Rnaset2b': 1, 'Selenbp1': 1}

{'Foxq1': 0, 'Krt27': 0, 'Krt35': 1, 'Krt71': 0, 'Krt73': 0, 'Rgcc': 0, 'Rnaset2b': 1, 'Selenbp1': 1}

{'Foxq1': 1, 'Krt27': 0, 'Krt35': 0, 'Krt71': 0, 'Krt73': 1, 'Rgcc': 1, 'Rnaset2b': 0, 'Selenbp1': 0}

{'Foxq1': 1, 'Krt27': 0, 'Krt35': 0, 'Krt71': 0, 'Krt73': 0, 'Rgcc': 1, 'Rnaset2b': 0, 'Selenbp1': 0}

{'Foxq1': 0, 'Krt27': 0, 'Krt35': 0, 'Krt71': 0, 'Krt73': 0, 'Rgcc': 0, 'Rnaset2b': 0, 'Selenbp1': 0}

Time reversal:
Foxq1* = !Rgcc & Selenbp1 | !Rgcc & Rnaset2b | Krt73 & !Rgcc | Krt71 & !Rgcc | Krt35 & !Rgcc | Krt27 & !Rgcc | Foxq1 & !Rgcc
Krt27* = Krt27 & !Krt71 & Selenbp1 | Krt27 & !Krt71 & Rnaset2b | Krt27 & !Krt71 & Rgcc | Krt27 & !Krt71 & !Krt73 | Krt27 & Krt35 & !Krt71 | Foxq1 & Krt27 & !Krt71
Krt35* = Krt35 & !Rnaset2b & !Selenbp1 | Krt35 & Rgcc & !Selenbp1 | Krt35 & Krt73 & !Selenbp1 | Krt35 & Krt71 & !Selenbp1 | Krt27 & Krt35 & !Selenbp1 | Foxq1 & Krt35 & !Selenbp1
Krt71* = !Krt27 & Krt71 & Selenbp1 | !Krt27 & Krt71 & Rnaset2b | !Krt27 & Krt71 & Rgcc | !Krt27 & Krt71 & !Krt73 | !Krt27 & Krt35 & Krt71 | Foxq1 & !Krt27 & Krt71
Krt73* = !Krt27 & !Krt71 & Krt73 & !Selenbp1 | !Krt27 & !Krt71 & Krt73 & Rnaset2b | !Krt27 & !Krt71 & Krt73 & Rgcc | !Krt27 & Krt35 & !Krt71 & Krt73 | Foxq1 & !Krt27 & !Krt71 & Krt73
Rgcc* = Rgcc & Selenbp1 | Rgcc & Rnaset2b | Krt73 & Rgcc | Krt71 & Rgcc | Krt35 & Rgcc | Krt27 & Rgcc | !Foxq1 & Rgcc
Rnaset2b* = !Krt35 & Rnaset2b & !Selenbp1 | !Krt35 & Rgcc & !Selenbp1 | !Krt35 & Krt73 & !Selenbp1 | !Krt35 & Krt71 & !Selenbp1 | Krt27 & !Krt35 & !Selenbp1 | Foxq1 & !Krt35 & !Selenbp1
Selenbp1* = !Krt35 & !Rnaset2b & Selenbp1 | !Krt35 & Rgcc & Selenbp1 | !Krt35 & Krt73 & Selenbp1 | !Krt35 & Krt71 & Selenbp1 | Krt27 & !Krt35 & Selenbp1 | Foxq1 & !Krt35 & Selenbp1

TR Stable Notifs:
{'Krt71': 1}
{'Krt27': 1}
{'Krt35': 1}
{'Rgcc': 1}
{'Krt35': 0, 'Rnaset2b': 0, 'Selenbp1': 0}
{'Foxq1': 0, 'Rgcc': 0}
{'Selenbp1': 1}
{'Krt73': 1}
Producing and exporting STG . . .
done.
```

Figure 8. Stable Motif and Attractor analysis for the small network created. Boolean rules were tested for each gene. The "!" in the "RULES" segment of the file represents the logic operator "NOT" and the "&" represents the logic operator "AND". The computer generated eleven attractors and eight stable motifs.

After obtaining the z3 scores from each gene and the stable motif and attractor analysis, the next part of the project was to determine the diffusion maps that will model the pseudotime analysis of our network. This is where we currently are in this project. The diffusion maps are being used to portray the relationships between different nodes on the network, and how each gene can contribute to activating the expression of another or repressing the expression of another gene. Modeling this was done through Google CoLab and based on code from the paper that provided their own code via JupyterNotebook (Joost et al 2020). Diffusion maps were developed with different resolution values to see whether the expression would change in any way in the network. The resolution values tested ranged from 0.5 to 1.7, with the main values of interest being 0.5 and 1.7.



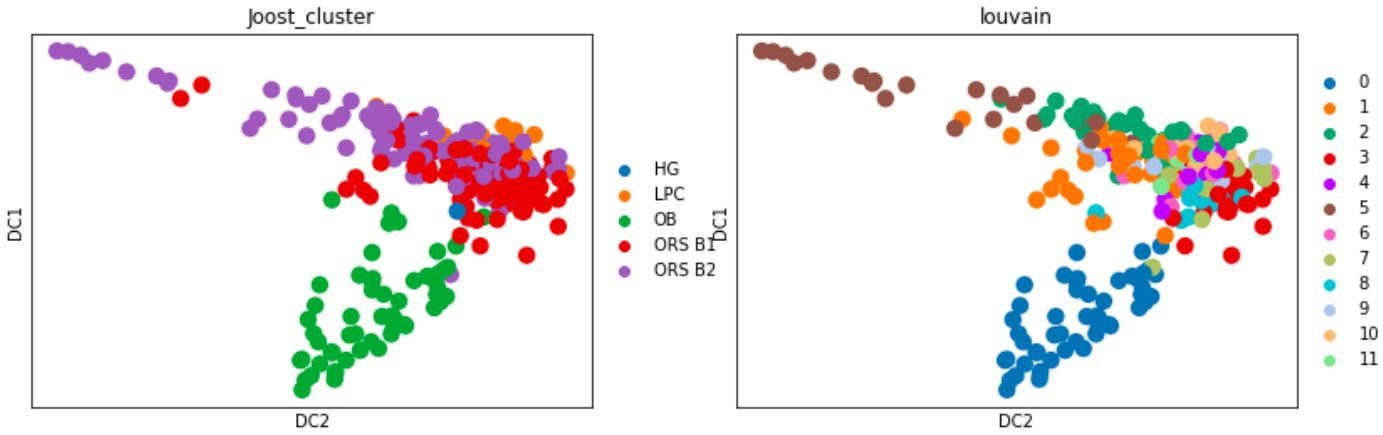


Figure 9. Diffusion maps generated from the google colab code with resolution values 0.5 (top) and 1.7 (bottom). As resolution value increased, complexity of the network also increased. These are diffusion maps from stem cells in the outer root sheath (ORS) of the HF.

Currently in the project, we are attempting to apply our Boolean network to pseudotime analysis. Code from JupyterNotebook will be used to apply the network to the pseudotime, and this will allow for us to determine the patterns of expression, activation or repression, for the initial genes studied in the first part of the methods.

From measuring the dynamics of the overall change in gene expression for each gene to modeling diffusion maps for each one at different values, these results are significant. Though the dynamics of gene expression in figures 1 and 2 change randomly due to the dependency on time, it is still possible to observe the changes of gene expression that occur in the stem cell in the HF. Genes in stem cells can be both turned off and turned on, and eventually in its last stage of development, the genes needed for that cell's function will be expressed and those not required will be inhibited. Measuring the z3 scores for each gene in the small network being studied allows for the measurement of confidence for the gene being an activator or an inhibitor, and this logic was then applied to the stable motif and attractor analysis.

Conclusion and Future Applications

Hair has evolved to serve many different functions in mammals, such as insulation and camouflage. The hair itself is surrounded by many different structures, from which the hair grows outward and upward. Much of the current knowledge about HFs comes from the discovery of HFs containing their own stem cells. HFs contain various structures, and within these structures are stem cells. These stem cells reside in the matrix of the HF and will eventually divide into three different layers of the HF: the IRS, MED, and CX. In order for these stem cells to function and differentiate from one another, the gene expression in these cells must be tightly regulated. During periods of growth, or anagen, the stem cells will proliferate until a signal to stop dividing is received. The same goes for the HF's stages of regression, telogen, and rest, catagen. HFs have recently become a model organ to study gene expression due to this coordination and the natural presence of stem cells in the fields of stem cell and skin biology.

The genes studied in this project are present in the stem cells, and their expression in the matrix and ORS have been studied primarily. In the HF matrix, it was observed that the genes can activate and inactivate each other, and this is ultimately the process to be modeled with the GRN constructed. The eight genes chosen were studied not only due to their expression in stem cells, but also for their contribution to larger GRNs previously studied. This will allow for potential studies in the future regarding gene expression in stem cells, since stem cells are undefined until they mature enough to fully differentiate from other cells.

Studying the stable motif and attractor analysis allows for the understanding of the role each gene plays. In the network studied in this project, there were eleven attractors and eight stable motifs, and the stable motifs indicate that those motifs are common states of the GRN,

even without gene expression randomly changing every time a code is run. This could mean that for a majority of the time, the motifs are the expression patterns being observed.

As mentioned previously, the Boolean network constructed is currently being applied to pseudotime analysis. With this analysis, it will be possible to further understand the patterns of gene expression in the HF and for the visualization of gene expression and inhibition. This will also contribute greatly to the fields of skin biology, for which HFs are embedded in the skin and for seeing how skin can react to certain molecular and physiological stimuli. This will also contribute greatly to stem cell biology as there is still much that is not understood about stem cells, so understanding these mechanisms will provide significant insight into these disciplines.

References

Chen, C.-L., Huang, W.-Y., Wang, E. H. C., Tai, K.-Y., & Lin, S.-J. (2020). Functional complexity of hair follicle stem cell niche and therapeutic targeting of niche dysfunction for hair regeneration. *Journal of Biomedical Science*, *27*(1), 43. https://doi.org/10.1186/s12929-020-0624-8

fionahamey. (2021). *Fionahamey/Pseudotime-network-inference* [Python]. https://github.com/fionahamey/Pseudotime-network-inference (Original work published 2016)

Gao, Z., Chen, X., & Başar, T. (2018). Controllability of Conjunctive Boolean Networks With Application to Gene Regulation. *IEEE Transactions on Control of Network Systems*, *5*(2), 770–781. https://doi.org/10.1109/TCNS.2017.2746345

Haghverdi, L., Buettner, F., & Theis, F. J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*, *31*(18), 2989–2998. https://doi.org/10.1093/bioinformatics/btv325

Hamey, F. K., Nestorowa, S., Kinston, S. J., Kent, D. G., Wilson, N. K., & Göttgens, B. (2017). Reconstructing blood stem cell regulatory network models from single-cell molecular profiles. *Proceedings of the National Academy of Sciences*, *114*(23), 5822–5829. https://doi.org/10.1073/pnas.1610609114

Joost, S., Annusver, K., Jacob, T., Sun, X., Dalessandri, T., Sivan, U., Sequeira, I., Sandberg, R., & Kasper, M. (2020). The Molecular Anatomy of Mouse Skin during Hair Growth and Rest. *Cell Stem Cell*, *26*(3), 441-457.e7. https://doi.org/10.1016/j.stem.2020.01.012

Lab, K. (2021). *Kasperlab/Joost_et_al_2020_Cell_Stem_Cell* [Jupyter Notebook]. https://github.com/kasperlab/Joost_et_al_2020_Cell_Stem_Cell (Original work published 2020)

Martel, J. L., Miao, J. H., & Badri, T. (2022). Anatomy, Hair Follicle. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK470321/

Oshimori, N., & Fuchs, E. (2012). Paracrine TGF-β Signaling Counterbalances BMP-Mediated Repression in Hair Follicle Stem Cell Activation. *Cell Stem Cell*, *10*(1), 63–75. https://doi.org/10.1016/j.stem.2011.11.005

Rozum, J. (2022). *Pystablemotifs* [Jupyter Notebook]. https://github.com/jcrozum/pystablemotifs (Original work published 2020)

Saadatpour, A., & Albert, R. (2013). Boolean modeling of biological regulatory networks: A methodology tutorial. *Methods*, *62*(1), 3–12. https://doi.org/10.1016/j.ymeth.2012.10.012

Saxena, N., Mok, K.-W., & Rendl, M. (2019). An updated classification of hair follicle morphogenesis. *Experimental Dermatology*, *28*(4), 332–344. https://doi.org/10.1111/exd.13913

Wang, Q. (2021, June 24). *Math260-022-Fall2020-Lec15-11252020.pptx*. https://docs.google.com/presentation/d/12W27a0v3o8ewVgcbSZXyZLEuw_cd1ym4

Zañudo, J. G. T., & Albert, R. (2013). An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *23*(2), 025111. https://doi.org/10.1063/1.4809777