**ICR Summer Fellow:** Ginny Wu, Patricia Beckman Summer Fellow

**School:** University of California, San Diego

**Major:** Biology: **Specialization:** Bioinformatics

**Year in school or (anticipated) graduation date:**

**Future plans (briefly):** Gain experience in the biotech industry through an internship or job for at least a year with the goal to pick up and improve proficiency in more bioinformatic tools and programming languages. Afterwards look for a PhD program in bioinformatics or related field.

**Mentor(s)**: Marisa Korody, Ph.D., Scientist, Conservation Genetics

Iñigo Valiente-Alandi Ph.D., Researcher, Conservation Genetics

**~~~~~~~~~~~~~~~~~~~~ Research Report ~~~~~~~~~~~~~~~~~~~~**

Title of project:

Delving into Northern White Rhino Stem Cells: Metabolic and Transcriptomic Analysis

## Abstract of the project:

The Northern White Rhino (NWR) induced pluripotent stem cells (iPSCs) that the Conservation Genetics team has generated will be utilized for gamete development and assisted reproduction with the aim of maintaining population viability and genetic diversity. To help us optimize our protocols and to improve our understanding of NWR iPSCs, we studied their metabolic and transcriptomic featurescompared them to the features of human iPSCs.The cell lines studied were human and NWR iPSCs, NWR fibroblasts which are a common type of connective tissue, and NWR embryoid bodies (EBs) which are differentiated derivatives of the iPSCs.

## Introduction

In the 1970s, the population of NWRs (*Ceratotherium simum cottoni*) was around 500. Since then, intensive poaching for their coveted horns has reduced them down to 2 individuals: a mother and daughter named Najin and Fatu. Because they are closely related, their genomes do not capture very much genetic diversity. The small gene pool along with lack of male individuals has rendered the NWR functionally extinct. There is no way for them to naturally reproduce or sustain a new genetically viable population.

There is however, hope for revival using the San Diego Frozen Zoo, which has banked and cryopreserved 12 NWR fibroblast cell lines. This pool is projected to have enough genetic diversity to maintain a population in the long term without a bottleneck effect. Thus, the NWR Project was initiated to revive this species using Frozen Zoo samples for stem cell and assisted reproductive technologies.

So far, the NWR Project has reprogrammed stable iPSCs from 9 NWR individuals and generated cardiomyocytes using these stem cells. In theory, pluripotent cells are able to differentiate into cell types from all three embryonic germ layers: the mesoderm, ectoderm, and endoderm. They have also tested the pluripotency of the NWR iPSCs by generating embryoid bodies (EBs). Embryoid bodies form when the iPSCs are allowed to clump together while suspended in culture media. These aggregates are then plated on gel to simulate the implantation of an embryo *in vivo*, where it begins to spontaneously differentiate into specialized cells. We can gather data from the cell types that appear to gauge the differential ability of the iPSCs.

The next steps in the rescue process are to reprogram our iPSCs into gamete predecessors known as primordial germ cells, to further program them so that sperm and egg cells are generated, and to fertilize these gametes into embryos which will be transferred to a Southern White Rhino (SWR) surrogate mother. This process of iPSC to gamete induction was pioneered by Katsuhiko Hayashi and his team at Kyoto University, who reprogrammed mouse iPSCs into gametes and used them to produce fertile offspring.[1] The Conservation Genetics team has completed a round of primordial germ cell reprogramming and are currently analyzing data to optimize the protocol.

There was also another recent landmark that happened during my fellowship in the Reproductive Biology department. A SWR in the San Diego Zoo Rhino Rescue Center gave birth to a healthy baby boy named Edward, showing that the artificial insemination technique was successful and that our captive rhinos could carry a pregnancy to full term.

## Research Plan and Methodology

In order to quantify various metabolic aspects of NWR iPSCs, we conducted several experiments on an iPSC line induced from the NWR individual Najin, who has proven to be the most robust as well as the most reprogrammable. The control samples that we used for comparison were human iPSCs and NWR fibroblasts. We cultured cells in 6-well plates, used part of the 6-well for an experiment, then passaged the remaining cells into a new 6-well for the next experiment.

**Functional Mitochondria**

To determine the energy production pathways of our cells, we first checked our cells for functional mitochondria using the Invitrogen MitoTracker kit. This protocol uses fluorescent proteins that accumulate to active mitochondria. NWR fibroblasts and iPSCs from human and NWR were incubated with the mitochondria probe, then checked for fluorescence using a flow cytometry machine. For the negative control we used NWR iPSCs without the probe and for the positive control we used human iPSCs and NWR fibroblasts.

**ATP Content**

We checked the ATP content in human iPSCs, NWR iPSCs, and NWR fibroblasts using the Sigma-Aldrich ATP bioluminescent somatic cell assay kit. Cells were passaged and cultured for a day in a 96-well plate that was black with clear wells to reduce fluorescent noise. We first released ATP from the cell into the culture media using the kit’s releasing agent, then marked the ATP with a luminescent protein called luciferin. The chemical reaction between ATP luciferin produces an Adneyl-luciferin complex that emits light when oxidized. Luminescence intensity was measured using a luminometer. The negative control was fresh cell culture media, the same type that we used for culturing NWR iPSCs, mTeSR1. The positive control and standard curve were made using a kit reagent containing ATP and disodium salt.

**Oxygen Consumption**

To test if the cells used the oxidative phosphorylation pathway for ATP production, we conducted an oxygen consumption assay using the Cayman Chemical Oxygen Consumption Rate Assay kit. It uses a phosphorescent oxygen probe that emits light until quenched by oxygen, resulting in an inverse relationship between oxygen consumption rate and fluorescence. The cells were also cultured in a black 96-well plate. We had three sets of negative controls, one with fresh culture media, one with media and the oxygen probe, and one with NWR iPSCs and Antimycin A which inhibits the mitochondrial electron transport chain. The positive control was a glucose oxidase solution provided in the kit. Fluorescence was observed using plate reader and excitation wavelength of 480 nm.

**Glycolysis Activity**

The Cayman Chemical Glycolysis Cell-Based Assay Kit was used to detect L-lactate, the end product of glycolysis that is secreted by cells. A kit reagent catalyzes the breakdown of L-lactate, producing a colored formazan that can be quantified by measuring absorbance. A standard curve was made using the kit’s L-lactate standard solution and culture media. The negative control was culture media which was also used as the last serial dilution of the standard curve. Absorbance was measured at 490 nm using a plate reader

**DNA Methylation**

**­** It has been reported that iPSCs are more reprogrammable if their DNA is hypomethylated.[5] This means they have less methyl groups attached to the DNA molecules and therefore more accessible by regulatory proteins. We extracted DNA from our focus cell lines using the Qiagen Blood and Cell Culture Miniprep kit then used the EpiGentek MethylFlashTM Global DNA Methylation (5-mC) ELISA Easy kit (Colorimetric), which detects only 5-mC type methylation using an antibody. There are other types of methylation that have been observed, especially in embryonic stem cells, including 5-hmC and 5-fC, 5-caC; however, these were not measured. I was also given a sample of SWR placenta for DNA extraction and methylation detection due to the birth of Edward a week prior.

After DNA extraction, we coated a 96-well plate with DNA binding protein and added approximately 200 ng of DNA for each sample. The kit provided a positive control solution of 5% 5-mC DNA and a negative control solution of 0% 5-mC DNA. These were also used for the standard curve by mixing them in various ratios. We added the 5-mC detection antibody and signal indicator, then the detection solution which turns blue if there is methylation. After the color was allowed to develop for 2 minutes, we added the stopping solution which stops the reaction and turns the blue coloration to yellow. The absorbance was read using a plate reader at 450 nm to calculate the 5-mC percentage.

**RNAseq Analysis**

The transcriptome of a cell is comprised of all its RNA transcripts. When a cell expresses a gene, it first makes an RNA transcript from the DNA gene, then the transcript is “read” by a ribosome to produce a protein. The RNA is later degraded by enzymes, which means the transcriptome has a level of variability. This variability in the transcriptome can be used as a sort of fingerprint for different cell types as well as cellular conditions. The RNAseq pipeline starts with extracting RNA, sequencing them, counting the similar sequence reads, then aligning them to a known gene name. This captures a snapshot of the cell’s transcriptome which can be analyzed for differential gene expression, in which we look for upregulated (high read count) and downregulated (low read count) genes that may be characteristic of the cell type. One issue is that we do not have a fully annotated NWR genome and neither does the NCBI database. Therefore, RNA sequences were mapped to gene names using what the database did have: the *Equus caballus* or horse.

We started analysis of this data generating pair-wise comparison scatterplots using R. We calculated the logarithmic fold change (logFC) of each gene’s read counts between two cell types. The genes that had a logFC higher than 2 were considered significantly upregulated and those that scored lower than -2 were considered significantly downregulated. Using these differentially expressed genes we identified the biological processes they were associated with using DAVID, a free online webtool and database for mapping gene names to biological process terms developed by the MIT Broad Institute. DAVID also clusters the biological processes by how often they appear associated to the query genes. [2,3] This gave us an overview of all differentially expressed genes, but it was tedious to isolate characterizing genes by comparing only two cell types at a time. Therefore, we developed an analysis that could take into consideration more than two cell types for finding characteristic genes. We modeled the pipeline after methods used by Tomonori Nakamura and other collaborators of the Mitinori Saitou lab at Kyoto University. [6]

The first step is to conduct a principal component analysis (PCA) which calculates the similarity between each sample by giving the genes of each sample a score relative to all the other samples, regardless of cell type. The 3D graph that is generated from this PCA is very easy to read; each sample is a point and is more similar to other points that are closer to it, and more dissimilar to those that are farther away. The PCA graph is able to display all of this data because the 3 axes are scaled by the data from the top 3 samples of the input data that contribute most to overall statistical variability. These 3 samples are given anonymous labels PC1, 2, and 3. In short, this condenses data by leaving out smaller details.

To select for the characteristic genes, we took the gene scores from PC1 and PC2 and filtered out those that had a score within 2 standard deviations of the population mean. The resulting genes were considered significant in characterizing a cell type. We then used an unbiased grouping algorithm called unsupervised hierarchical clustering (UHC) to calculate the similarities in their expression (RNAseq read counts). The UHC draws a dendrogram which we can cut into subclusters using another algorithm. We took the gene names from each cluster and mapped them to associated biological processes using DAVID as we did with the first pair-wise analysis. The output of biological process phrases gives us a window into the workings of our cells and the gene clusters can be used to characterize cell types. We ran the pipeline on RNAseq data extracted from the NWR fibroblast to iPSC reprogramming experiment and from the NWR iPSC to EB experiment where RNA was extracted from cells throughout the differentiation process.

## Results

**Results and Discussion**

Our metabolism data indicates that NWR iPSCs resemble human iPSC and mouse iPSCs in that they all rely on glycolysis and not oxidative phosphorylation to produce ATP for energy. Near 99% of the NWR iPSCs were positive for functional mitochondria and there was ATP produced. However, their oxygen consumption levels were very low when compared with the positive control. On the other hand, they had strong positive results in the glycolysis activity assay. The lack of respiration coupled with evidence of glycolysis products, functional mitochondria, and a standard amount of ATP content means that NWR iPSCs utilize the glycolysis pathway much like other model species’ iPSCs.

The methylation data also showed that the NWR and human iPSC DNA was significantly hypomethylated in comparison the fibroblasts. This aligns with the understanding that more specialized cells have more methylation or are hypermethylated. An interesting result was that the lowest methylation across all cell types came from the NWR Najin. The Najin iPSCs have been the most stable cell line as well as the most susceptible to reprogramming out of all the NWRs. This may be due to the lower levels of methylation.

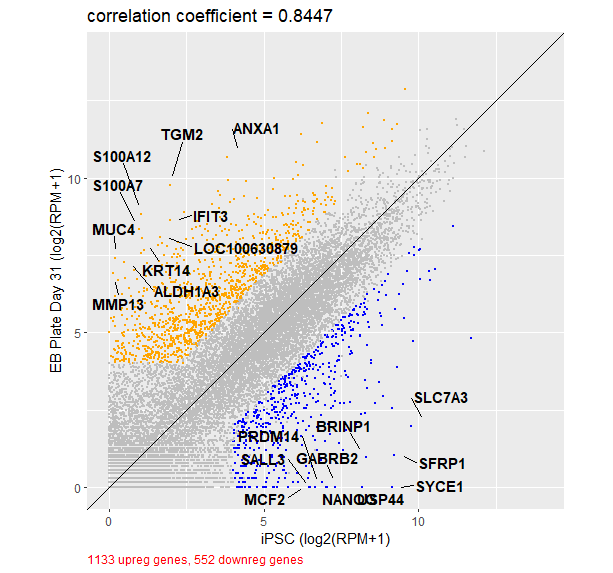
Transcriptome analysis between NWR iPSCs and fibroblasts revealed that the iPSCs display upregulation of neuronal differentiation genes which coincides with results from cross-species stem cell research.[4] The results from the iPSC to EB experiment showed that there was a lot of genes associated with developmental process. A few examples are epithelial, ectoderm, epidermis, ear, and embryonic development. This is provides confirmation that the EBs were beginning to differentiate. Further analysis of the clusters could reveal association with development of all three germ layers.

With more data and larger sample sizes, we could further characterize what makes NWR iPSCs unique and figure out how stem cell techniques could be optimized. Some components of the NWR Project that are still pending include alignment and annotation of the NWR genome. Once we have the full genome, we can more accurately conduct experiments like RNAseq analysis. The RNAseq data that we used was mapped to the NCBI horse genome. A pitfall of using a different species is that putative genes that appear in both the RNAseq data and database genome are marked using the location of the sequence on the database genome. Putative genes in our RNAseq data are labeled with horse genome locations which is not very useful. Nevertheless, the NWR Project can move forward to refining their primordial germ cell production protocol while having the tools to do more in-depth analysis.

A.

NWR Fibroblasts vs. iPSCs NWR iPSCs vs ED Day 31

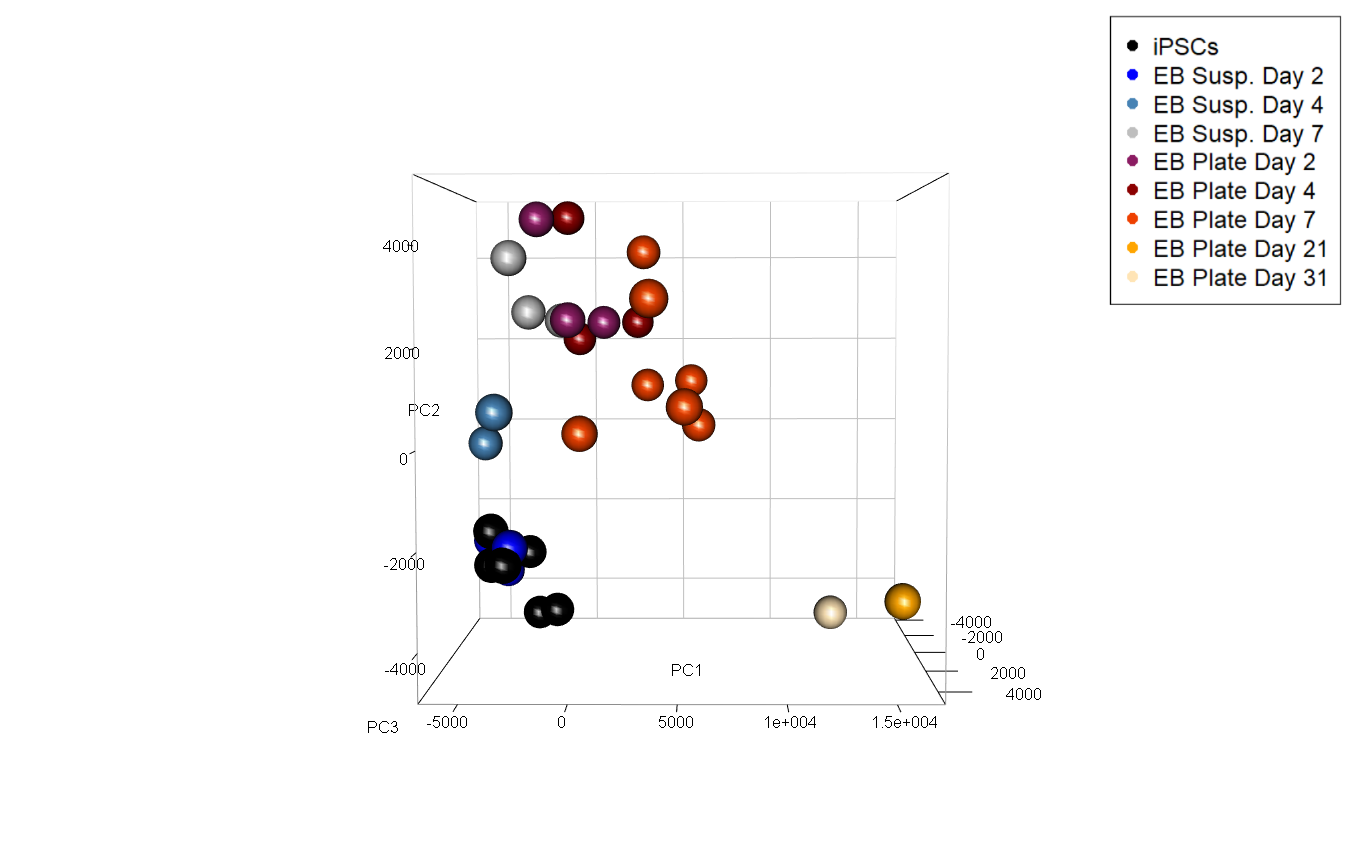
A close up of a map

Description automatically generated

B.

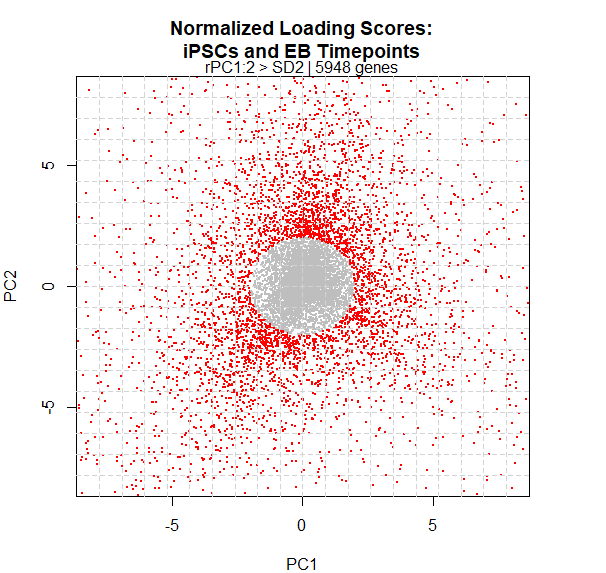
PCA: Fibroblasts and iPSC PCA: iPSCs to EBs

A close up of a map

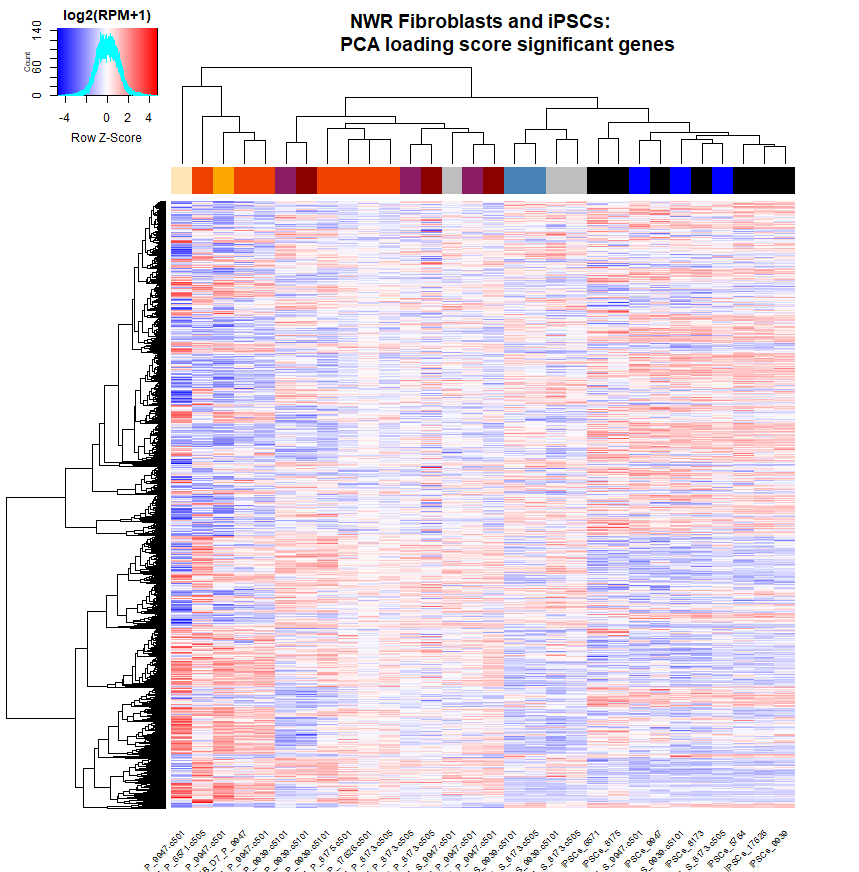
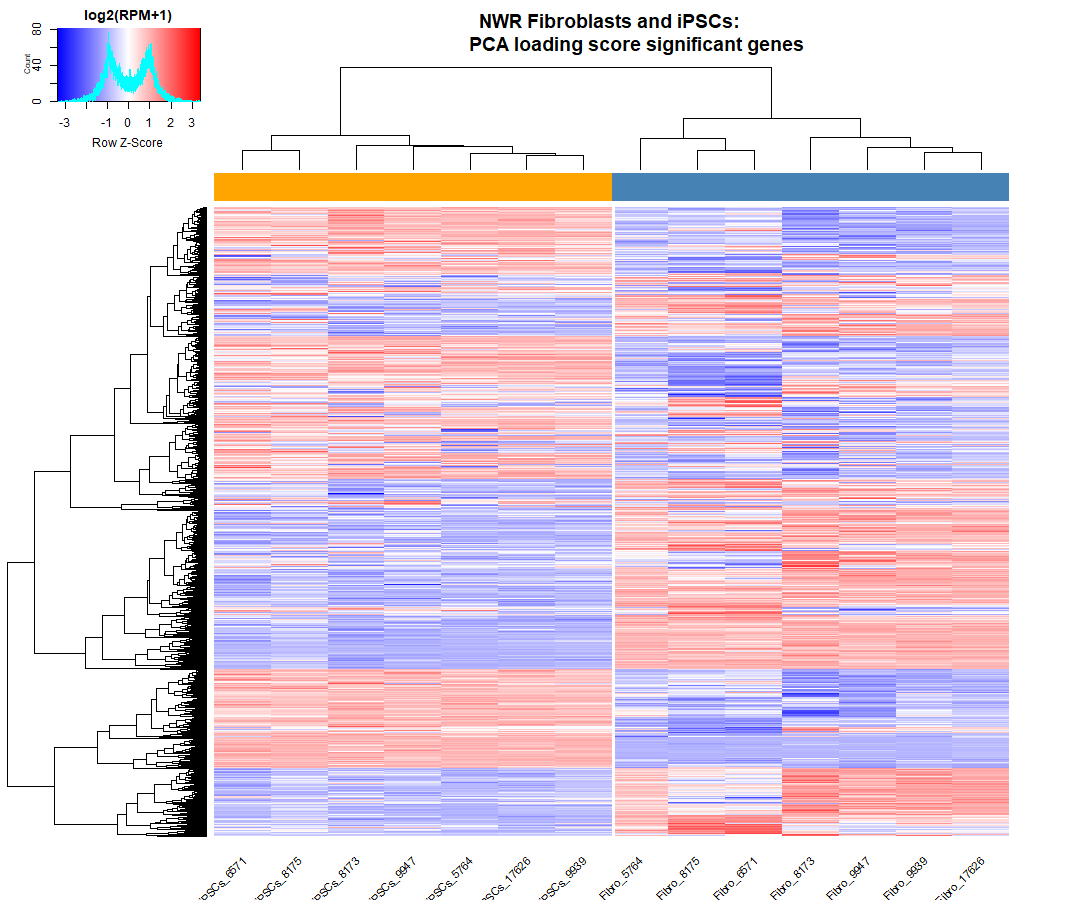
Description automatically generated

C.

A screenshot of a cell phone

Description automatically generated 

D.



1. Preliminary differential expression graphs generated from RNAseq data. The top 10 upregulated and downregulated genes are labeled. The graph on the left shows average expression data from 7 iPSC samples and 7 fibroblast samples collected from different NWR individuals at different times. The graph on the right shows a pair-wise comparison between the average transcriptomic data from iPSCs (n = 7) and one EB cell that was allowed to differentiate for 31 days after plating (n = 1).
2. PCAs of RNAseq data
3. Scatterplots of gene loading scores extracted from PCA. Red points indicate selected genes that scored higher than 2 standard deviations of the mean.
4. Heatmaps of gene expression. Genes shown are selected using PCA scores. UHC clustered by similarity of expression. The colored bars just below the column dendrograms are colored according to their respective PCA graph point colors.

## Discussion/Conclusions, including Broader Impact

The NWR Project embodies the next level of conservation research. With the current state of climate change and the incremental changes to worldwide green practices, it is inevitable that more species will be driven to extinction by human activity. Not only does the Project break ground and push boundaries in by opening stem cell and genetics research to conservation research, but it also bridges these fields with population sustainability and even public awareness campaigns.

## References

## [1] Hayashi, K., et al. 19 August 2011. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* 146, 519-532.

[2] Huang, D.W., Sherman, B.T., Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res. 2009;37(1):1-13.*

[3] Huang, DW., Sherman, B.T., Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. *Nature Protoc. 2009;4(1):44-57*

[4] Nakamura, T., et al. 1 September 2016. A developmental coordinate of pluripotency among mice, monkeys and humans. Nature. Vol. 537.

[5] Panopoulos, A.D., et al. 6 April 2017. Aberrant DNA methylation in human iPSCs associates with MYC binding motifs in a clon-specific manner independent of genetics. *Cell Stem Cell*. Vol. 20(4):505-517.e6.

[6] Yamashiro, C., et al. 19 October 2018. Generation of human oogonia from induced pluripotent stem cells in vitro. Science. Vol. 362.

## Fellow's Perspective

I’ve learned so much from my time here at the San Diego Zoo Institute for Conservation (ICR). It was an incredible opportunity to work with the Conservation genetic team driving the rescue of a nearly extinct species using stem cell technology. The Beckman Institute for Conservation Research has a wide array of conservation projects that I was able to work alongside. Much of my growth came from watching how a lab is funded and run. I had the opportunity to use advanced technologies in my experiments, but I also learned from observing the day-to-day budgeting conversations and practicing green lab habits. I’ve never been in a lab that is mindful of its own waste production or a lab that has collaborative departments that make for near-vertical integration of the research process, from the Rhino Rescue Center to the in-house MiSeq and flow cytometry machine.

There were also many people I got the chance to know and learn from outside of paper headings. It was inspiring to meet the other fellows and learn about the creative ways they were tackling their projects. Working on the NWR Project allowed me to improve my coding project organizational skills as well as my R proficiency in terms of statistical analysis and data visualization. I now have a repository of several tools that the Project can optimize for future experiments.

On a more personal note, while I was data-wrangling results from experiments on samples coming directly from the Frozen Zoo, I came to the realization that in the 7th grade I wrote a – dare I say prophetical – short story. It was about a scientist in a dystopian future and her eleventh-hour attempt to save biodiversity using a “modernized Noah’s Ark,” a DNA bank located in a deep seafloor military base. A plaque that reads ‘20th CENTURY ARK’ is displayed by the Frozen Zoo entrance. Working on an ambitious species rescue project has been such an honor and privilege. My experience here gave me solid direction for my career. My mentors Dr. Inigo Valiente-Alandi and Dr. Marisa Korody showed me the kind of passionate mindset that goes into conservation research. They have inspired me and they have all my admiration and appreciation.