The International Young Researchers' Conference
DOI: doi.org/10.34614/iyrc0047

Direct reprogramming of various somatic cell types into fibroblasts cells using transcription factors: a potential cure for COPD and chronic wound healing

Akshara Karthik
International Academy Central Campus, United States
karthik.akshara09@bloomfield.org

Abstract

In the United States, 6 in 10 adults suffer from chronic diseases, like diabetes and chronic obstructive pulmonary disease (COPD), with resultant poor wound healing abilities and irreversible lung damage (*Chronic Diseases*, 2021). Direct cell reprogramming, could potentially reverse the effects of lung damage caused by smoking, as well as quicken wound healing. This novel project aimed to identify transcription factors to directly reprogram various somatic cell types into fibroblasts. Fibroblast cells actively participate in tissue repair and wound healing. Ten transcription factors associated with fibroblasts were identified through the online database Amazonia! and their functions were researched through GeneCards: MMP3, ID1, SNAI2, FGF2, SP1, FAP, FIBP, PRRX1, LBH, and AEBP1. Each of the transcription factors were further explored to confirm that high gene expression was limited to fibroblasts. This will ensure that the use of the chosen transcription factors would lead to fibroblast production rather than an undesired cell type. The National Center for Biotechnology Information (NCBI) and KEGG Pathway Databases were then employed, to study the metabolic pathways (both disease and regulatory) that each of the selected transcription factors are involved in. Metabolic pathways were analyzed to eliminate any transcription factors that are involved in known disease pathways, deeming them not suitable for direct cell reprogramming. Overall, SNAI2 and ID1 were identified as the ideal transcription factors to directly reprogram somatic cells into fibroblasts, because they have limited association with diseases and are not highly expressed in other cell types.

Keywords: Fibroblasts, Transcription Factors, Direct Cell Reprogramming, COPD, Diabetic Wound Healing

Introduction

Clinical Problems

According to the American Lung Association, "The toxins in cigarette smoke weaken your lungs' defense against infections, narrow air passages, cause swelling in air tubes and destroy air sacs—all contributing factors for COPD," (*COPD Causes and Risk Factors*, 2021). The lung damage from COPD, emphysematous, and other cystic lung disorders is often caused by the destruction of fibroblasts in the distal lung and airway. Chronic diseases, like diabetes, can cause weakened wound healing ability, which could lead to wound infection or even the amputation of the limb if not treated (*Non-Healing Wounds*, n.d.). In 2018, over 34.2 million

American citizens were diabetic, making diabetes a relatively common health ailment in the United States (*National Diabetes, 2020*). Alarmingly, diabetes can cause poor blood circulation, which in turn can affect the movement of fibroblasts to the wound site and cause tissue injuries to heal poorly or not at all (Lerman et. al, 2003). Direct cell reprogramming can be a potential solution to reverse lung damage from COPD and heal chronic diabetic wounds faster.

Overview of Direct Cell Reprogramming and Transcription Factors

Direct cell reprogramming, also known as transdifferentiation, is the revolutionary practice of transforming an initial adult somatic cell type into another, without the need to transition through an induced pluripotent state (Gam et. al, 2010).

Transcription factors are a diverse group of small proteins that can regulate both the rate of the transcription of DNA into mRNA, as well as gene expression. These tiny proteins bind to specific DNA promoter regions or even RNA polymerase, which can inhibit or activate gene transcription (Robertson et.al, 2019). Transcription factors also play a pivotal role in creating divergent cell identities, since different transcription factors are activated in varying cell types and different cell types make contrasting proteins.

The principal transdifferentiation method, used in vitro, utilizes transcription factors and takes place in three major steps. Exogenous transgenes, external genes from outside the organism's genome, are first exposed to the initial cell type through viral vectors, which causes the initial cells to start overexpressing crucial transcription factors which catalyze the reprogramming process (Grath & Dai, 2019). Next, the organism's endogenous genes are either upregulated or downregulated by the transcription factors, and finally, pharmacological agents target a variety of transcriptional pathways, which generates a cellular response that ultimately transforms the initial cell type into the product cell type (Grath & Dai, 2019).

Rationale for Reprogramming Somatic Cells to Fibroblasts

Fibroblast cells were chosen as the target cell type of this project because they are widely recognized for their active contribution to the therapeutic resistance of many human pathologies, including cancer, chronic inflammatory diseases, and systemic autoimmune diseases. Fibroblasts are the main connective tissue cells in the human body, and they are responsible for forming the extracellular matrix and generating collagen (Bainbridge, 2013). Fibroblasts also play a pivotal role in wound healing, by relocating to the injured site and releasing collagen to initiate healing (Bainbridge, 2013). When fibroblasts are activated, as a result of tissue harm, they actively differentiate into myofibroblasts, which create sizable contractions and generate extracellular matrix proteins that further facilitate wound healing and closure (Li & Wang, 2012).

Methodology

Investigating Transcription Factors Associated with The Target Cell Type

Once the target cell type was determined as fibroblasts, the Amazonia! microarray database was utilized to identify the relevant transcription factors active or associated with fibroblast cells. The GeneCards database was then employed to research each of the transcription factor's unique functions. NCBI was used to explore whether or not the chosen transcription factors were associated with diseases by researching the transcription factors' summaries and their phenotypes.

Data Analysis of Transcription Expression in Target Cell Type and Other Cells

To efficiently and accurately conduct direct reprogramming of the initial cell type into a fibroblast cell, the transcription factors utilized must be reasonably specific to fibroblasts. If the transcription factor is highly expressed in more than five other cell types, then the direct cell reprogramming experiment greatly risks creating another cell type rather than the target cell trying to be created, which would be undesirable.

In the “Human Body Index” section of Amazonia!, the average microarray signal that is present in the fibroblast samples was calculated. Following that calculation, 90% of that average was also calculated, to identify the other cell types that highly expressed that specific transcription factor. The 90% threshold was chosen since it would most accurately explain which transcription factors are highly expressed compared to the target cell type, which in this case was fibroblasts. If at least one-quarter of the samples of a certain cell type were 90% or greater than that of the average signal for fibroblasts, then the graph exemplified that the transcription factor in question is expressed at a lower rate in fibroblasts, compared to other cells. As a result, the transcription factor may not be good for transdifferentiation purposes, since an undesired cell type could be made. For some transcription factors, in which there were multiple graphs, the same steps were repeated for each graph. The “Embryonic and Adult Normal Tissues” section was ignored for this project since totipotent stem cells were not considered. Additionally, "HESC" and "HIPS" stem cells were disregarded, due to the fact that they are unspecialized and would not be able to carry out the desired functions for this project.

Examining Transcription Factor Pathways

With the KEGG Pathway Database, the signaling pathways for each of the selected transcription factors were investigated through the pathway maps provided, as well as the identification of downstream events. If the transcription factor is involved in many disease pathways, rather than regulatory pathways, the transcription factor, along with other components that could be present in the initial cell, can inadvertently catalyze the pathogenesis of the disease in the new cell, which could potentially be injurious.

Results

Ten transcription factors were identified and investigated in this study: MMP3, FGF2, SP1, FAP, FIBP, PRRX1, LBH, SNAI2, ID1, and AEBP1.

MMP3, PRRX1, AEBP1, LBH, FGF2, SP1, FAP, and FIBP, were not viable transcription factors, given that they were highly expressed in more than five other cell types and/or were involved in disease pathways. Based on the data collected in Table 1, SP1 and FGF2 were eliminated, since they are highly expressed in five or more other cell types and are involved in a large number of metabolic disease pathways, deeming them not specific enough for the transdifferentiation of other cell types into fibroblast cells. Moreover, as seen in Figure 1, they were minimally expressed in the fibroblast cell samples, compared to the other transcription factors. FAP and MMP3 are involved in metabolic disease pathways, including a variety of cancers. Even though the manifestation of these diseases in the new cell is not guaranteed, the transcription factors are directly involved in multiple serious disease pathways which may increase the chances of disease materializing in the product cell. PRRX1, LBH, AEBP1, and FIBP were deemed inconclusive since the KEGG Pathways database did not have any data regarding the metabolic pathways these transcription factors were respectively involved in.

Through this study, ID1 was deduced as an ideal transcription factor, since it is not associated with metabolic disease pathways, and is only highly expressed in fibroblasts and endometrial cells, as seen in Table 2. ID1 is solely involved in regulatory pathways like Rap1 signaling pathway, TGF-beta signaling pathway, Hippo Signaling Pathway, and other various signaling pathways regulating pluripotency, the ability to give rise to any cell type in the human body, of stem cells.

TABLE 1. Analysis on MMP3, FGF2, SP1, FAP, FIBP, PRRX1, LBH, and AEBP1 Transcription Factors Using Data from Amazonia!, NCBI, GeneCards, and KEGG Pathway

Target Cell Type	Important Transcription Factor	Other Cell Types That Express It	Highly expressed in five or more other cell types?	Known Functions (According to NCBI and Genecards)	Associated Diseases (According to NCBI and Genecards)	Metabolic Regulatory Pathways (According to KEGG Pathways Database)	Metabolic Disease Pathways (According to KEGG Pathways Database)
Fibroblasts	MMP3	Synovial Membrane	No	Involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis.	Coronary Heart Disease and Alzheimers Disease	IL-17 Signaling Pathway TNF Signaling Pathway	Coronavirus Disease- COVID-19 Transcriptional Misregulation in Cancer Prostate Cancer Rheumatoid Arthritis
Fibroblasts	PRRX1	**Graph 1:** none **Graph 2:** Deltoid Muscle, Synovial Membrane, Joint Tissue	No	A transcription co-activator, enhancing the DNA-binding activity of serum response factor, a protein required for the induction of genes by growth and differentiation factor	Agnathia-otocephaly complex	N/A	N/A
Fibroblasts	AEBP1	Coronary Artery, Saphenous Vein, Synovial Membrane	No	Encodes a member of carboxypeptidase A protein family. The encoded protein may function as a transcriptional repressor and play a role in adipogenesis and smooth muscle cell differentiation	EHLERS-DANLOS SYNDROME	N/A	N/A
Fibroblasts	LBH	Ovary, Lung, Salivary Gland, Saphenous Vein, Heart, Heart Ventricle, PBMC media resting, CD4+ antiCD3+antiCD28 activated, CD4+ resting, CD8+ antiCD3+ anti-CD28 activated, CD8+ resting, T cells antiCD3, T cells resting, Thymus Gland, Tonsil, Spleen, B cells resting, Uterus	Yes	Transcriptional activator which may act in mitogen-activated protein kinase signaling pathway. Developmental protein with roles in transcription and transcription regulation.	Rheumatoid Arthritis	N/A	N/A
Fibroblasts	FGF2	**Graph 1:** MSC (stem cells *may not be involved in this project*) **Graph 2:** MSC (stem cells)	No	Involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction, and tissue remodeling, as well as in disease processes, such as arthritis and metastasis.	Irritable Bowel Syndrome	EGFR tyrosine kinase inhibitor resistance MAPK signaling pathway Ras signaling pathway Rap1 signaling pathway Calcium signaling pathway P13K-Akt signaling pathway Signaling pathways Regulating the Pluripotency of Stem Cells Regulation of actin cytoskeleton	Melanoma Breast Cancer Gastric Cancer Proteoglycane in cancer Pathways in Cancer Kaposi sarcoma-associated herpesvirus infection
Fibroblasts	SP1	**Graph 1:** (first five) Ovary, Oocyte, Frontal Lobe, Occipital Lobe, Caudate **Graph 2:** (first five) Ovary, Oocyte, Testes, Cerebral Cortex, Frontal Lobe **Graph 3:** (first five) Ovary, Testes, Cerebral Cortex, Frontal Lobe, Cerebellum **Graph 4:** (first five) Ovary, Testes, Cerebral Cortex, Frontal Lobe, Occipital Lobe	Yes	Transcription factor that can activate or repress transcription in response to physiological and pathological stimuli. Binds with high affinity to GC-rich motifs and regulates the expression of a large number of genes involved in a variety of processes such as cell growth, apoptosis, differentiation and immune responses. May have a role in modulating the cellular response to DNA damage.	None	Endocrine resistance Mitophagy- animal TGF-beta signaling pathway Estrogen signaling pathway Cortisol synthesis and secretion Parathyroid hormone synthesis, secretion, and action	Cushing syndrome Endocrine resistance Huntington disease Spinocerebellar ataxia Human cytomegalovirus infection Pathways in cancer Transcriptional misregulation in cancer Breast cancer Choline metabolism in cancer
Fibroblasts	FAP	None	No	FAP expression is high in reactive stromal fibroblasts of epithelial cancers, granulation tissue of healing wounds, and malignant cells of bone and soft tissue sarcomas. FAP is thought to be involved in the control of fibroblast growth or epithelial-mesenchymal interactions during development, tissue repair, and epithelial carcinogenesis.	Colorectal Cancer	N/A	Colorectal Cancer
Fibroblasts	FIBP	Testes, Amygdala, Thalamus, HepG2, HUVEC Cell Line, MSC (stem cells), T cells antiCD3 30h activated	Yes	The FIBP protein is an intracellular protein that binds selectively to acidic fibroblast growth factor (aFGF). It is postulated that FIBP may be involved in the mitogenic action of aFGF.	Acne vulgaris, inflammatory bowel disease, Thauvin-Robinet-Faivre syndrome	N/A	N/A

TABLE 2. Analysis of Applicable Transcription Factors: SNAI2 and ID1 Using Data from Amazonia!, NCBI, GeneCards, and KEGG Pathways

Target Cell Type	Important Transcription Factor	Other Cell Types That Express It	Highly expressed in five or more other cell types?	Known Function(s)	Associated Diseases	Metabolic Regulatory Pathways	Metabolic Disease Pathways
Fibroblasts	SNAI2	Ovary	No	Transcriptional repressor that modulates both activator-dependent and basal transcription. Involved in the generation and migration of neural crest cell. Involved in the regulation of ITGB1 and ITGB4 expression and cell adhesion and proliferation in epidermal keratinocytes	Partial Albanism and Waardenburg Syndrome Type 2D	Hippo Signaling Pathway Adherens Junction	Piebaldism Waardenburg Syndrome
Fibroblasts	ID1	Endometrium	No (but highly expressed in Endometrium cell samples)	Prevent skeletal muscle differentiation by sequestering the E proteins, thus blocking the activity of MyoD and other myogenic bHLH proteins	None	Rap1 Signaling Pathway TGF-beta Signaling Pathway Hippo Signaling Pathway Signaling Pathways Regulating the Pluripotency of Stem Cells	None

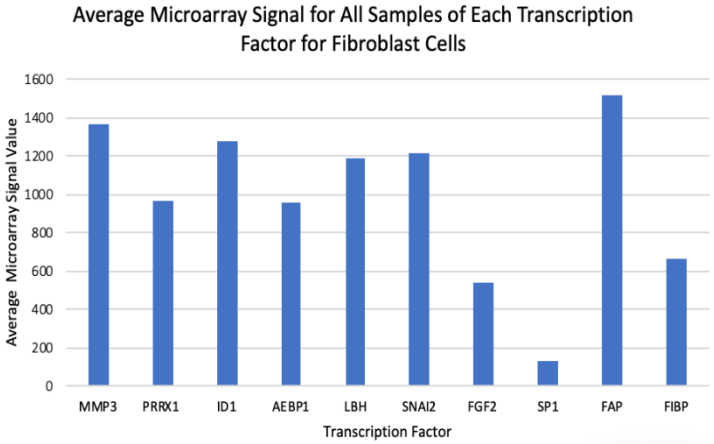

FIGURE 1. Average Microarray Signal (Gene Expression) for All Samples of Each of the Chosen Transcription Factors for Fibroblast Cells.

When the other cell samples that had signal values that were at least 90% of the average fibroblast signal were investigated, it was found that ID1 was also highly expressed in endometrial cell samples, as seen in Figure 2. This could pose a minimal but potential threat to direct cell reprogramming purposes and could unintentionally transform the initial cell type into an endometrial cell rather than that of a fibroblast during direct reprogramming. However, since ID1 is only highly expressed in one other cell type, it is still specific to fibroblast cells and can therefore be utilized for transdifferentiation purposes to create fibroblast cells.

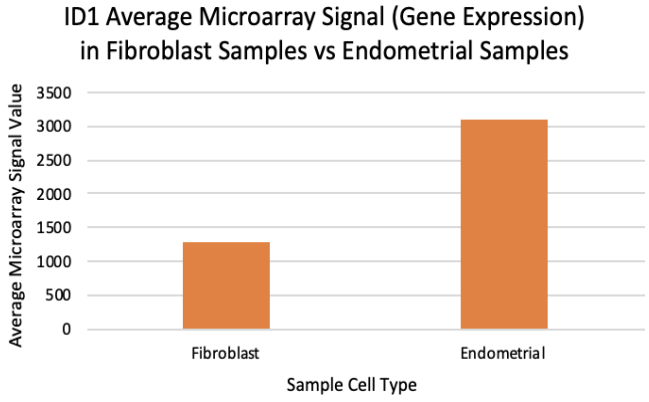

FIGURE 2. ID1 Average Microarray Signal (Gene Expression) in Fibroblast Samples vs Endometrial Samples.

SNAI2 is another ideal transcription factor, since it is associated with few diseases, not as highly expressed in other cell types, and has no role in undesirable pathways, as seen in Table 2. SNAI2 is associated with the Piebaldism disease pathway, a rare autosomal disorder caused by mutations in the KIT and SNAI2 genes, and the Waardenburg Syndrome disease pathway, which is another rare heterogeneous genetic disease caused by mutations in the SNAI2 gene. Furthermore, SNAI2 is also involved in the Hippo Signaling Pathway, which regulates organ size through cell proliferation and apoptosis, as well as Adherens Junctions, which are protein complexes that are found in epithelial tissues at cell-cell junctions. Since SNAI2 is involved in mainly regulatory pathways and rare genetic-based disease pathways, transdifferentiation can be accomplished with minimal concern over the transcription factor causing undesired downstream events.

Since SNAI2 is associated with disease pathways, unlike ID1, the latter would probably be the most ideal transcription factor to use for reprogramming efforts to create fibroblast cells. It is important to note that even if defects in a transcription factor's gene, like SNAI2, may cause disease, it does not mean that it will necessarily translate into disease post-transdifferentiation. Additionally, the only other cell type from the microarray samples that SNAI2 has high gene expression in is the ovary, though the average of the ovary microarray signal values is less than that of SNAI2 (as seen in Figure 3). Since the expression of SNAI2 in the ovary samples is less than the fibroblast samples, creating another cell type other than fibroblasts, during transdifferentiation, is possible but highly unlikely. This credits SNAI2 as still being a viable transcription factor that can be used for direct cell reprogramming efforts with ID1.

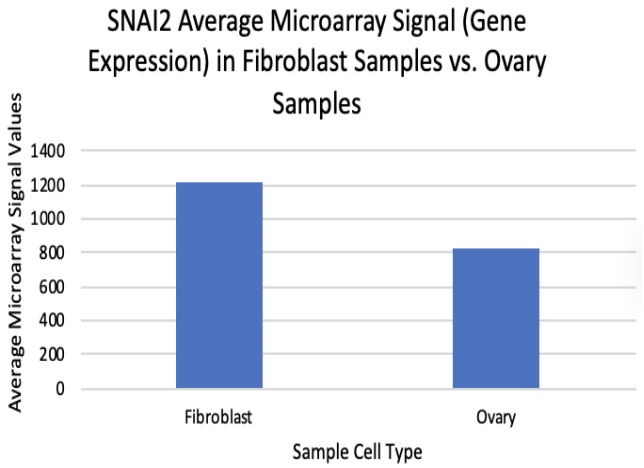

FIGURE 3. SNAI2 Average Microarray Signal (Gene Expression) in Fibroblast Samples vs. Ovary Samples.

Discussion

SNAI2 and ID1 were the transcription factors determined to be viable for the direct reprogramming of other cell types into fibroblast cells.

Previous transdifferentiation experiments have utilized fibroblast cells, along with a variety of other cell types, including hair keratinocytes, blood monocytes, melanocytes, hepatocytes, astrocytes, neural stem cells, T cells, and urine cells, to create other cell types like induced pluripotent stem cells (iPSCs) (Gam et. al, 2010). In addition, researchers were able to utilize transcription factors to directly convert fibroblast cells into functioning neurons (Vierbuchen et. al, 2010). This investigation is novel and deviates from previous transcriptome studies, in that the purpose of this project was to create fibroblasts through direct cell reprogramming, unlike past studies where fibroblast cells have been used as the initial cell type and reprogrammed to create cells of another type.

There are various clinical implications for applying directly reprogrammed fibroblast cells. For example, smoking is a major cause of a variety of lung-related health problems like emphysema and COPD. The findings from this study could potentially be applied to help reverse lung damage caused by smoking and vaping, by replacing damaged fibroblast cells found in the distal lung and airway with new, healthy fibroblast cells made from direct cell reprogramming.

Furthermore, wound healing is significantly prolonged in patients with underlying medical problems, like cardiovascular diseases, diabetes, and immunodeficiency conditions, and consequently are at a higher risk of infection (*Non-Healing Wounds*, n.d.). Fibroblast cells created through transdifferentiation could be applied to faster wound healing. In-vitro transdifferentiation using autologous somatic cells to create fibroblast cells can be transplanted to a patient's wound site. This application could not only improve the patient's wound healing ability but reduce adverse reactions caused by immune rejection since autologous cell types would be utilized.

Due to the COVID-19 pandemic, this study became purely based on computational microarray database values, as well as research on functions and pathways of merely 10 transcription factors associated with fibroblasts. Additionally, due to database limitations, four transcription factors had to be eliminated since there was not enough pathway data available to formulate a conclusion on their effectiveness for the transdifferentiation of other cell types into fibroblast cells. Furthermore, this examination does not account for a variety of external factors a lab would have to offer. First and foremost, in the future, ID1 and SNAI2 should be tested in a lab to determine if they can be used to transdifferentiate other cell types into fibroblast cells and evaluate whether using them individually or together for direct reprogramming would yield better results. Additionally, more transcription factors associated with fibroblasts should be investigated to deduce whether more transcription factors could be used, either individually or in conjunction, to create functioning fibroblast cells. Furthermore, cell types similar to fibroblasts, like osteoblasts and chondroblasts, should also be examined, to uncover any similarities between their respective functions, associated transcription factors, and their microarray gene expressions, and therefore their

efficacy to be created through direct cell reprogramming efforts.

References

Bainbridge, P. (2013). Wound Healing and the Role of Fibroblasts. *Journal of Wound Care, 22*(8). https://doi.org/10.12968/jowc.2013.22.8.407

Carrour, T. L., Assou, S., Tondeur, S., Lherrmitte, L., Lamb, N., Rème, T., Pantesco, V., Hamamah, S., Klein, B., Vos, J. D. (2010). Amazonia!: An Online Resource to Google and Visualize Public Human whole Genome Expression Data. *The Open Bioinformatics Journal*, *4*, 5-10. http://dx.doi.org/10.2174/1875036201004010005

Chronic Diseases in America. (2021, January 12). Centers for Disease Control and Prevention: National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP). Retrieved March 12, 2021, from https://www.cdc.gov/chronicdisease/resources/infographic/chronic-diseases.htm

COPD Causes and Risk Factors. (2021, March 5). American Lung Association. Retrieved March 11, 2021, from https://www.lung.org/lung-health-diseases/lung-disease-lookup/copd/what-causes-copd

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Stein, T. I., Rosen, N., Kohn, A., Twik, M., Safran, M., Lancet, D., Cohen, D. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*, *2017*, bax028. https://doi.org/10.1093/database/bax028

Gam, R., Sung, M., Pandurangan, A. (2019). Experimental and Computational Approaches to Direct Cell Reprogramming: Recent Advancement and Future Challenges. *Cells, 8*(10), 1189. https://doi.org/10.3390/cells8101189

Grath, A., Dai, G. (2019). Direct cell reprogramming for tissue engineering and regenerative medicine. *Journal of Biological Engineering*. https://doi.org/10.1038/4631031a

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genome. *Nucleic Acids Research,* 28(1), 27-30. https://doi.org/10.1093/nar/28.1.27

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Science*, *28*(11), 1947-1951. https://doi.org/10.1002/pro.3715

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Research*, *49*(D1), D545-D551. https://doi.org/10.1093/nar/gkaa970

Lerman, O. Z., Galiano, R. D., Armour, M., Levine, J. P., Gurtner, G. C. (2003). Cellular Dysfunction in the Diabetic Fibroblast. *The American Journal of Pathology, 162*(1), 303-312. https://doi.org/10.1016/S0002-9440(10)63821-7

Li, B., & Wang, J. H. (2012). Fibroblasts and Myofibroblasts in Wound Healing: Force Generation and Measurement. *Journal of Tissue Viability*, *20*(4), 108-120.https://doi.org/10.1016/j.jtv.2009.11.004

National Center for Biotechnology Information (NCBI). (n.d.). National Center for Biotechnology Information. Retrieved February 20, 2021, from https://www.ncbi.nlm.nih.gov/

National Diabetes Statistics Report, 2020. (2020, February 11). Centers for Disease Control and Prevention. Retrieved February 28, 2020, from https://www.cdc.gov/diabetes/library/features/diabetes-stat-report.html

NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(D1), D8–D13. https://doi.org/10.1093/nar/gkx1095

Non-Healing Wounds. (n.d.). https://www.mainehealth.org/Services/Wound-Care-Ostomy-Care/Non-Healing-Wounds

Robertson, S., Barnett, C. (2019, February 26). *Role of Transcription Factors*. News Medical Life Sciences. https://www.news-medical.net/life-sciences/Role-of-Transcription-Factors.aspx

Stelzer, G., Rosen, R., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Iny, T., Nudel, R., Lieder, I., Mazor, Y., Kaplan, S., Dahary, D., Warshawsky, D., Guan-Golan, Y., Kohn, A., Rappaport, N., Safran, M., and Lancet, D. (2016) The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analysis. *Current Protocols in Bioinformatics*, *54*(1), 1.30.1-1.30.33. https://www.genecards.org/

Vierbuchen, T., Ostermeier, A., Pang, Z., Kokubu, Y., Südhof, T., Wernig, M. (2010). Direct Conversion of Fibroblasts to Functional Neurons by Defined Factors. *Nature*, (463)1035–1041. https://doi.org/10.1038/nature08797