Predicting Vitiligo Neoantigens Utilizing Machine Learning

Grant Proposal

Lauren Kim

Massachusetts Academy of Math and Science at WPI

85 Prescott St, Worcester, MA 01605

Executive Summary (Eng)

Vitiligo is a B and T lymphocyte-mediated autoimmune condition resulting in destruction of melanocytes, manifesting as depigmented patches of skin with a lack of melanocytes. Approximately 1% of the global population is affected. Vitiligo is associated with cochlear dysfunction, inner ear diseases, diabetes mellitus, thyroid disease, metabolic syndrome, among others (Wang et al., 2024). Various studies report antibodies directed against melanocyte-specific protein autoantigens, such as tyrosinase, leading to its destruction (Faraj et al., 2021). Currently utilized immunosuppressive treatments, such as corticosteroids, have demonstrated some efficacy. However, due to its non-specific nature, patients often suffer from various side effects and potential harm (Wang et al., 2024).

To address these critical gaps, we propose to identify patient-specific melanocyte antigens by developing Python-based machine learning from an existing database (Gupta et al., 2019). Identified neo-antigens can be targeted to develop personalized therapy. Once trained, the machine learning program will be validated through MHC-I and MHC-II binding predictions.

The current dataset includes gene RPKM values for approximately 9902 genes associated with vitiligo. We plan on using these values, combined with their z-score deviation from normal red blood cells, to predict the gene with the strongest association to vitiligo. Three sample melanocytes in the dataset have p-values less than 0.01, which make them statistically significant. We will feed this data into a series of LSTMs for an output of genes and correlation values. Our findings will build a foundation for disease-specific machine learning tools, aimed at identifying actionable drug targets.

Keywords: vitiligo, T-cells, autoimmune disease, neo-antigens

Predicting Potential Neoantigens in Vitiligo Utilizing Machine Learning

Imagine waking up one day and finding out that patches of your skin have turned white. Researching more into the issue, you find that the condition, vitiligo, has no cure. Even worse, you learn that it will always continue to spread, with treatments only helping to mitigate the process. Today, over 70 million people worldwide are affected by vitiligo, a chronic autoimmune condition where patches of skin lose pigment and turn white (Bergqvist & Ezzedine, 2020). Depigmentation of the skin occurs because the body's immune system, specifically T cells, attacks melanocytes, cells that produce melanin, a pigment that gives skin its color, as shown in figure 1.

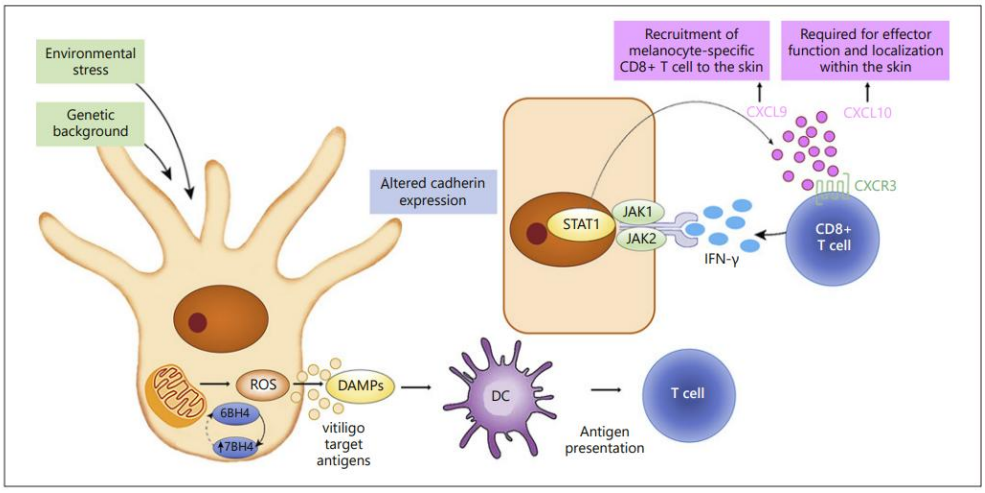

Fig. 1: This figure displays the risk factors and reactions involved with antigens and T cells in the pathogenesis of vitiligo (Bergqvist & Ezzedine, 2020).

There are two main types of vitiligo: segmental and non-segmental. Non-segmental is more common and spreads slowly on both sides of the body. Segmental causes rapid color loss in one side of the body (Bergqvist & Ezzedine, 2020). Although no definitive cure exists, scientists have identified various risk factors and genes related to this immune response. Vitiligo is not life threatening, but it affects the quality of life for individuals who have it by weakening the immune system and increasing the susceptibility to sunburn in patients.

Current treatments include topical treatments such as Corticosteroids, Calcineurin inhibitors, photo treatments like Psoralen and UVA therapy, or surgical treatments like skin grafting (Bergqvist & Ezzedine, 2020). Currently, new and upcoming treatments include advanced treatment medical products (Ghashghaei et al., 2023) and gene therapies like JAK inhibitors shown through ruxolitinib cream (Passeron et al., 2024). Another therapeutic method is Cas9 gene therapy using CRISPR for different autoimmune diseases (Lee et al., 2022). However, gene therapies often completely inhibit immune response by "knocking out" genes, leading to various severe side effects and potential harm (Perez-Bootello et al., 2023). Certain stress correlated signaling proteins – such as HSP70i – can be used to treat vitiligo. However, treatments often risk further depigmentation for certain patients (Mosenson et al., 2013). Targeting patient specific triggers could potentially be more effective and reduce side effects. General disease etiology – potentially in the form of neoantigens – still remains poorly understood.

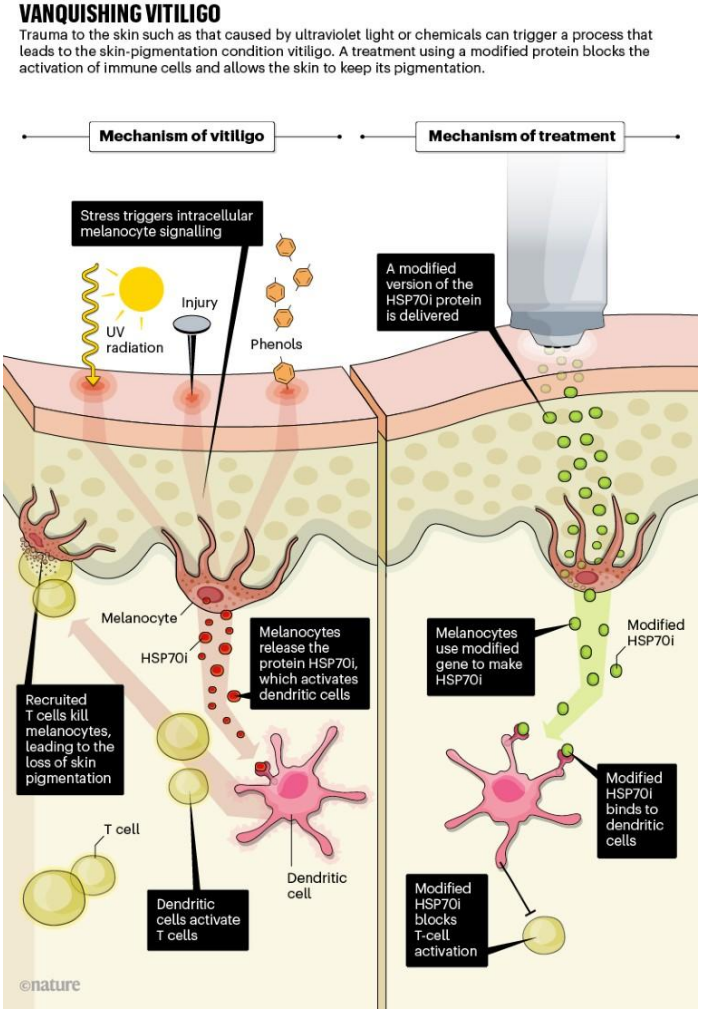

Fig. 2: This figure shows how the alteration of HSP70i can inhibit the immune response associated with vitiligo. Clinical trials are still in progress, so it is unsure what potential side effects this therapy could have (Schmidt, 2020).

Neoantigens are antigens that are produced by cells due to genetic mutation. In oncology, they are produced by tumor cells by tumor-specific mutated genes. Cancer immunotherapies using neoantigens are rising in popularity, as they are an effective and facile way to limit the expression of mutated genes (Lu & Robbins, 2015). In the context of vitiligo, they are antigens produced by cells that trigger the immune response due to genetic mutations in the reaction. These antigens are thought to be produced when melanocytes encounter oxidative stress, altering proteins and causing the immune system to detect the cell as non-self. These cells are particularly vulnerable due to a compound called ROS (reactive oxygen species) produced during melanin production (Faraj, 2021). ~~E~~xposure to environmental factors like UV, certain chemicals, and pollutants can generate ROS, further contributing to oxidative stress. Prediction and identification of neo-antigens have been rising in popularity in the field of cancer and oncology due to their potential impact. In vitiligo, it has substantial potential in identifying the antigens responsible for triggering autoimmune response. These antigens could be predicted with machine learning, another field currently gaining popularity.

Machine Learning is a branch of Artificial Intelligence that develops algorithms to make predictions based on given data. Neoantigen prediction is a rising field as it has a large impact and the potential to help develop a cure for the condition (Cai et al., 2023).

Several antigens tied to vitiligo such as VIT 90, 75, 40, gp100, MART1, and Tyrosinase are known (Cui et al., 1995). Stress-induced proteins like Heat Shock Protein 70 (HSP70i) have also been considered as enhancers of immune response, since they serve as signaling molecules for environmental stress (Schmidt, 2020). However, the “trigger” antigen, or the neo-antigen that initiates autoimmune response, is not yet known. If identified, specific treatment could be manufactured to inhibit the response as a whole instead of at different points along with the reaction. This novel treatment could reduce side effects while maximizing treatment capabilities for patients.

Despite significant advances in understanding vitiligo's pathogenesis, critical knowledge gaps remain regarding these specific trigger antigens. Current studies have identified melanocyte-associated proteins as potential immune targets, but these findings do not explain the variability in patient-specific immune responses to different treatments. Additionally, the role of stress-induced neoantigens, which may arise from oxidative stress and are potentially linked to HSP70i (Mosenson et al., 2013), is poorly characterized, leaving a large portion of the antigen field of vitiligo unexplored (Faraj, 2021). Machine learning (ML) offers a solution by enabling integration of large-scale datasets to identify patterns and correlations (Javaid et al., 2022). Furthermore, machine learning-driven approaches account for patient-specific variability, providing personalized treatment options and identifying antigens driving responses for different individuals, making it a promising avenue.

The ultimate goal of this project is to identify immunogenic neoantigens specific to individual patients through a machine learning algorithm. Python-based coding and bioanalytic programs will be utilized to find and analyze protein sequences from the existing clinical data. We hypothesize that identified neoantigens will be specific for each patient. Moreover, the group of peptides identified in vitiligo patients would be significantly different from healthy individuals. Based on our findings, we may identify effective therapeutic agents that will improve quality of life for patients diagnosed with vitiligo (Bergqvist & Ezzedine, 2020).

Section II: Specific Aims

We aim to identify the neoantigen that initiates autoimmune response in vitiligo patients. Our long-term goal is to discover individual-specific actionable drug targets in vitiligo patients. The central hypothesis is that there is a neoantigen that triggers the immune response for vitiligo. The rationale is that there are currently several antigens known about the reaction that causes immune response, and it is likely that a singular antigen begins the whole reaction, especially in the context of how neo-antigens

are formed by stress factors which commonly trigger immune response. This strong correlation between neo-antigens and casual stress factors makes it likely that vitiligo has a trigger antigen. The work we propose here will help further research in the field of autoimmune disease as well as neo-antigens. We propose to identify and sequence triggering antigen using machine-learning techniques. Our specific aims are:

Specific Aim 1: Identify which genes contribute the most to the expression of vitiligo.

Specific Aim 2: Validate findings using MHC-I and MHC-II binding prediction.

The expected outcome of this work will be the identification of a casual antigen in the autoimmune response. We hypothesize that there exists an antigen that triggers the immune response. This antigen could differ from patient to patient, which is why we plan on utilizing different data sets to train and personalize antigen predicting to individuals using machine learning.

Section III: Project Goals and Methodology

Relevance/Significance

My project is relevant due to the increasing demand for machine learning, as well as the harmfulness of gene therapy. Since machine learning is a fairly new field, the model would be relevant and impactful. For patients, this research will help reduce extensive side effects of current treatment and improve quality of life for diagnosed individuals (Bergqvist & Ezzedine, 2020).

Innovation

As machine learning is a new and innovative field, personalized treatment and data analysis have become much more accessible and easier to find for patients with different diseases. For patients with vitiligo, this innovation would greatly further research and could possibly help cure the condition with personalized treatment.

Methodology

Python and IEDB Analysis Resource readily available online will be utilized to identify immune epitopes. We will experiment with different gene datasets from the internet, then begin prediction and preparation of the gene based on online datasets.

Specific Aim #1:

The objective of this specific aim is to identify the trigger gene and its DNA sequence for immune recognition in vitiligo through a machine learning algorithm.

Our approach (methodology) uses machine learning to analyze data from databases and libraries. We plan on using several LSTM to analyze genome data. Our rationale for this approach is due to the high processing power of the different programming languages and synthesis capabilities. These programs will be useful in generating personalized treatment plans and help to develop a more effective treatment for vitiligo. Since LSTMs are fairly simple to code, we will be able to produce a complex model by utilizing several LSTMs.

Justification and Feasibility.

The methods in this section are all fairly new and popular, which makes this research novel and relevant. The methods (ML) help to address the specific aim by synthesizing data to predict the most correlated gene for patients not in the data set. These methods are important as it makes treatment much more specialized and reliable as the model is trained on a variety of different datasets.

Summary of Preliminary Data.

The Vitivar dataset will be used to predict the correlated genes. This dataset includes the RPKM values for 4 sample melanocytes, as well as the specific z-scores comparing RPKM from vitiligo

melanocytes and normal red blood cells for 22582 genes found in melanocytes. The two figures below show the results of a regression analysis of the dataset.

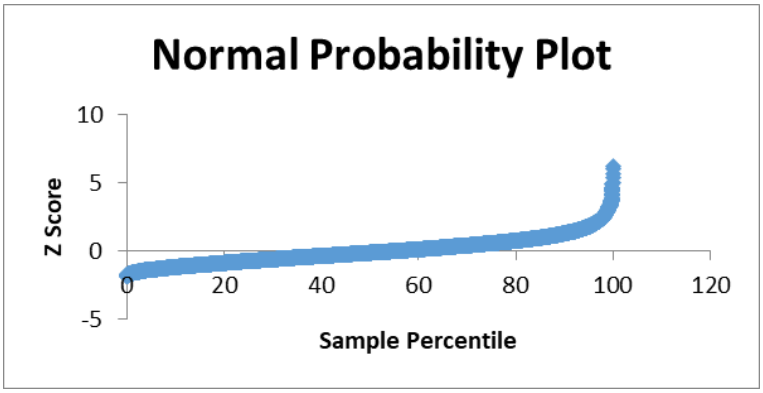

Figure 3: This figure shows the normal probability plot of the dataset. The data is skewed to the left, since many of the values have a high z-score in comparison to the sample percentile.

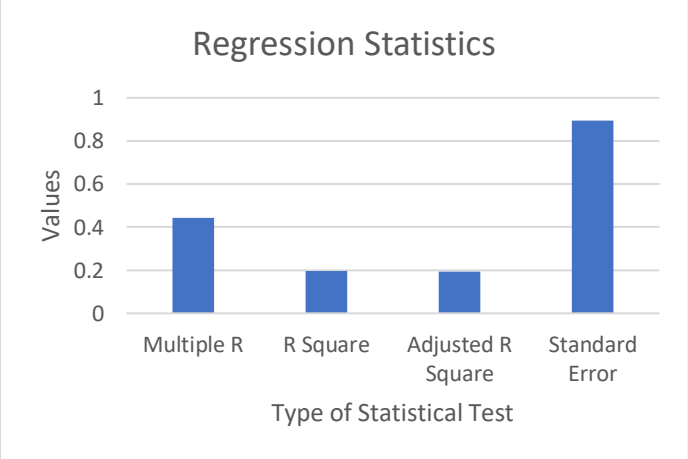

Figure 4: This figure shows the regression statistics of the dataset. The R Square is around 20 percent, meaning the data is representative of around 20% of vitiligo patients.

From these two figures, we can conclude that additional datasets will be needed to produce a fully representative model. Although the data is skewed, it may not need to be adjusted since we are searching for a casual antigen. This data will be used to train the various LSTMs, which make data collection a crucial part of the project.

Expected Outcomes.

The overall outcome of this aim is to identify the gene sequence of the main gene that is most correlated to an autoimmune response in vitiligo. This knowledge will be used to find the trigger neoantigen which can then be used to develop a more personalized treatment with less side effects than gene knockout therapy in the future.

Potential Pitfalls and Alternative Strategies.

We expect data collection may be difficult and potentially not reliable. Thus, the model must be made to be representative and reproducible with different data sets. An alternative to finding data online would be to research and obtain data in a lab setting; however, this can be difficult due to time and space limitations.

Specific Aim #2:

The objective of this specific aim is to validate our findings through the IEDB analysis resources for MHC-I and MHC-II binding prediction.

Our approach (methodology) uses the IEDB analysis resource combined with results from specific aim 1 to rank peptide binding. Our rationale for this approach is due to the high number of antigens associated with genes, finding a dataset with correlated antigens is much more challenging and time consuming in comparison to using gene sequences. Antigens are more easily and effectively found through MHC binding prediction after identification of genes.

Justification and Feasibility

The methods in this section are reliable and facile, since IEDB analysis resource is based on clinical trials from the National Institute of Health. The methods help to address the specific aim by synthesizing data to predict associated antigens for patient specific genes. This is important as it makes treatment more specialized.

Summary of Preliminary Data.

Data will be collected once the first specific aim is complete.

Expected Outcomes.

The overall outcome of this aim is to identify the antigens and protein sequences that cause the autoimmune response in vitiligo causing the depigmentation of skin. This knowledge will be used for the development of more personalized treatment with less side effects than gene knockout therapy in the future.

Potential Pitfalls and Alternative Strategies.

We expect that IEDB Analysis Resource could potentially be unreliable, and findings could vary based on genes. Thus, we will use alternative machine learning algorithms and prediction tools to ensure that our findings are accurate.

Section III: Resources/Equipment

The resources needed are databases to compare healthy, non-vitiligo antigens to vitiligo antigens and neo-antigens. These will most likely be obtained from internet sources and analyzed through a variety of machine learning programs, including Python and IEDB Analysis resource.

Section V: Ethical Considerations

The datasets used will be entirely from public domain, and will only be from non-identifiable, anonymously taken human datasets.

Section VI: Timeline

Within the allotted timeframe, a research plan as well as data will be developed and analyzed. We anticipate that the research portion will take a significant amount of time due to the need to contact professionals or mentors.

2024	August	September	October	November	December	January
Research topics	→	→	→	→	→	
Receive feedback from professionals			→	→	→	
Finalize methodology				→	→	
Accrue datasets				→	→	
Produce code and data from initial datasets						→
Data analysis and poster preparation					→	→

Section VII: Appendix

Section VIII: References

Bergqvist, C., & Ezzedine, K. (2020). Vitiligo: A Review. *Dermatology, 236*(6), 1–22. https://doi.org/10.1159/000506103

Cai, Y., Chen, R., Gao, S., Li, W., Liu, Y., Su, G., Song, M., Jiang, M., Jiang, C., & Zhang, X. (2023). Artificial intelligence applied in neoantigen identification facilitates personalized cancer immunotherapy. *Frontiers in Oncology*, *12*. https://doi.org/10.3389/fonc.2022.1054231

Cui, J., Arita, Y., & Bystrn, J. C. (1995). Characterization of Vitiligo Antigens. *Pigment Cell Research*, *8*(1), 53–59. https://doi.org/10.1111/j.1600-0749.1995.tb00774.x

Faraj, S., Kemp, E. H., & Gawkrodger, D. J. (2021). Patho-immunological mechanisms of vitiligo: the role of the innate and adaptive immunities and environmental stress factors. *Clinical and Experimental Immunology*, *207*(1), 27–43. https://doi.org/10.1093/cei/uxab002

Ghashghaei, S., Abbaszadeh, M., Karimi, S., Ataie-Fashtami, L., Bajouri, A., & Vosough, M. (2023). Advanced Therapy Medicinal Products in Vitiligo; Current Status, Future Prospect, and Approved Treatments. *Cell Journal*, *25*(3), 143–157. https://doi.org/10.22074/cellj.2023.557550.1067

Javaid, M., Haleem, A., Ravi P. S., Suman, R., & Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. *International Journal of Intelligent Networks*, *3*, 58–73. https://doi.org/10.1016/j.ijin.2022.05.002

Lee, M., Shin, J.I., Yang, J.W., Lee, K.H., Cha, D., Hong, J.B., Park, Y., Choi, E., Tizaoui, K., Koyanagi, A., Jacob, L., Park, S., Kim, J., & Smith, L. (2022). Genome Editing Using CRISPR-Cas9 and Autoimmune Diseases: A Comprehensive Review. *International Journal of Molecular Sciences*, *23*(3), 1337. https://doi.org/10.3390/ijms23031337

Lu, Y.-C., & Robbins, P. F. (2015). Cancer immunotherapy targeting neoantigens. *Seminars in Immunology, 28*(1), 22–27. https://doi.org/10.1016/j.smim.2015.11.002

Mosenson, J. A., Zloza, A., Nieland, J. D., Garrett-Mayer, E., Eby, J. M., Huelsmann, E. J., Kumar, P., Denman, C. J., Lacek, A. T., Kohlhapp, F. J., Alamiri, A., Hughes, T., Bines, S. D., Kaufman, H. L., Overbeck, A., Mehrotra, S., Hernandez, C., Nishimura, M. I., Guevara-Patino, J. A., & Le, C. (2013). Mutant HSP70 Reverses Autoimmune Depigmentation in Vitiligo. *Science Translational Medicine*, *5*(174). https://doi.org/10.1126/scitranslmed.3005127

Passeron, T., Harris, J. E., Pandya, A. G., Seneschal, J., Grimes, P., Kornacki, D., Wang, M., Ezzedine, K., & Rosmarin, D. (2024). Repigmentation by body region in patients with vitiligo treated with ruxolitinib cream over 52 weeks. *Journal of the European Academy of Dermatology and Venereology*. https://doi.org/10.1111/jdv.20236

Perez-Bootello, J., Cova-Martin, R., Naharro-Rodriguez, J., & Segurado-Miravalles, G. (2023). Vitiligo: Pathogenesis and New and Emerging Treatments. *International Journal of Molecular Sciences*, *24*(24), 17306–17306. https://doi.org/10.3390/ijms242417306

Schmidt, C. (2020). Temprian Therapeutics: developing a gene-based treatment for vitiligo. Nature. https://doi.org/10.1038/d41586-020-01808-5

Tang, Q., Gross, K. Y., Fakih, H. H., Jackson, S. O., Zain, M., Monopoli, K. R., Blanchard, C., Bouix-Peter, C., Portal, T., Harris, J. E., Khvorova, A., & Alterman, J. F. (2024). Multispecies-targeting siRNAs for the modulation of JAK1 in the skin. *Molecular Therapy — Nucleic Acids*, *35*(1), 102117. https://doi.org/10.1016/j.omtn.2024.102117

Tian, J., Olcott, A., Hanssen, L., Zekzer, D., & Kaufman, D. L. (1999). Antigen-based immunotherapy for autoimmune disease: from animal models to humans? *Immunology Today*, *20*(4), 190–195. https://doi.org/10.1016/s0167-5699(99)01445-0