Final Project Presentation

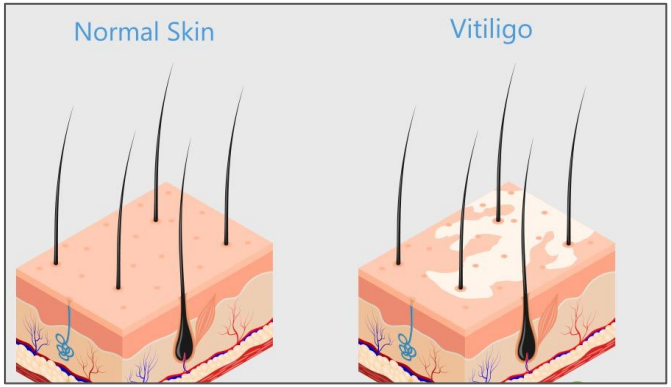
BINF6310: Introduction to Computational Biology

Mitophagy and immune infiltration in vitiligo

Koushik Muthuselvam, Mansi Babu, Rtwick George Moses
Shamitha K V

Vitiligo and Mitophagy

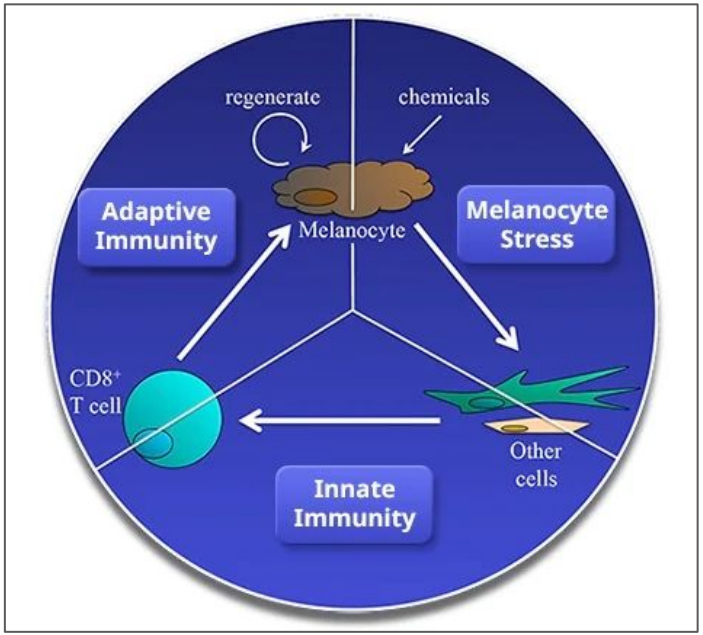
- Vitiligo is autoimmune skin disorder with depigmented patches, affecting 0.5–2% of the population.
- Vitiligo is characterized by a consistent and selective loss of epidermal melanocytes (MCs) as the disease progresses.
- Mitophagy is a natural process that removes damaged or unnecessary mitochondria from a cell to maintain its health
- It is essential for removing damaged mitochondria; its dysfunction exacerbates oxidative stress and immune response in vitiligo.



Overview of vitiligo: autoimmune skin disorder with depigmented patches and loss of epidermal melanocytes.

Vitiligo and Mitophagy

- Vitiligo primarily arises from the interplay of three factors: Adaptive immunity, Innate immunity, Melanocytes stress



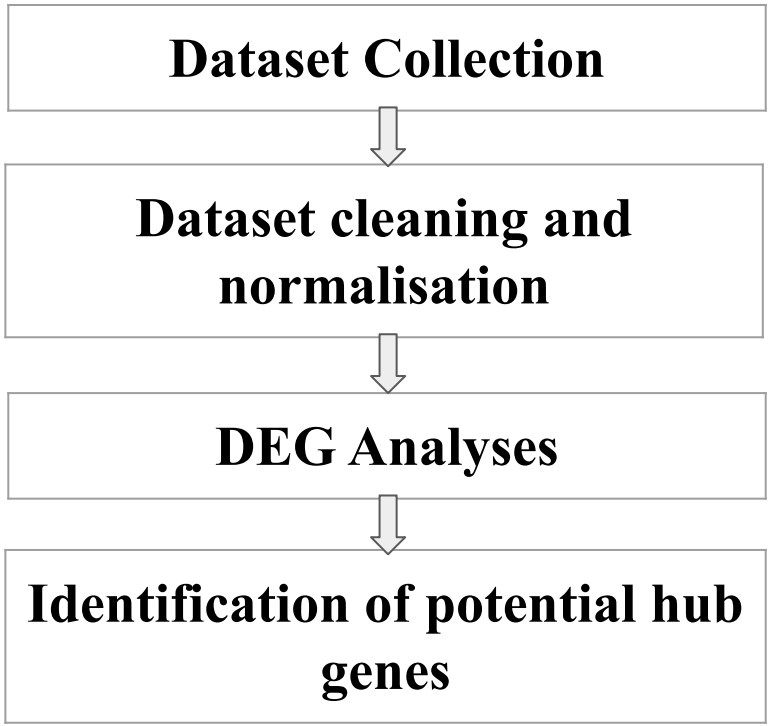
Key factors contributing to vitiligo: adaptive immunity, innate immunity, and melanocyte stress

Problem Statement

- To determine the possible role of mitophagy-associated genes in vitiligo and immune infiltration.

- Vitiligo has unclear pathogenesis and an unsatisfactory response to treatment, it is necessary to explore the mechanism of vitiligo to develop effective target treatments.

Workflow

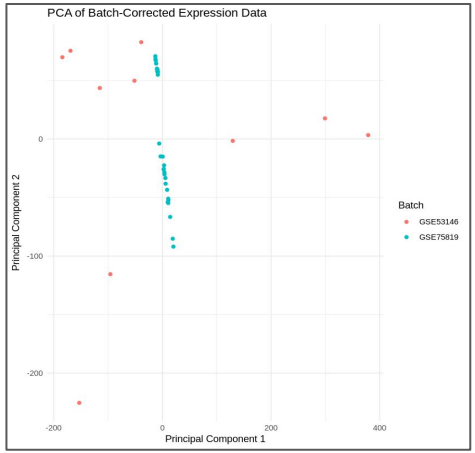

Dataset Collection
Dataset cleaning and normalisation
DEG Analyses
Identification of potential hub genes

Dataset Collection

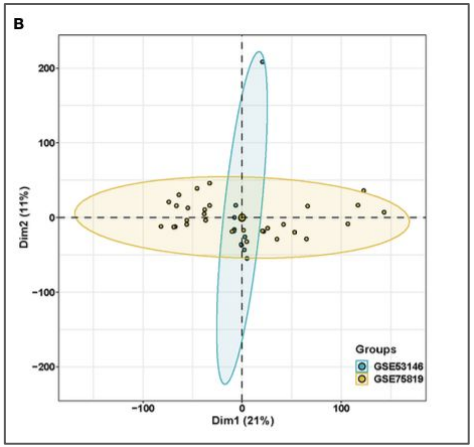
- The paper used raw gene expression data in vitiligo and controls
  - The dataset GSE53146 included 5 samples from vitiligo patients and 5 healthy individuals.
  - The microarray GSE75819, included 30 skin samples from 15 vitiligo patients' lesional and nonlesional skin samples.

Dataset Cleaning and Normalisation

- GSE53146 and GSE75819 were merged based on Gene ID.
- Null values and duplicate values were removed
- Dataset was then normalised using preprocess core and batch effects were removed using limma and sva using standard parameters
- PCA analysis was performed



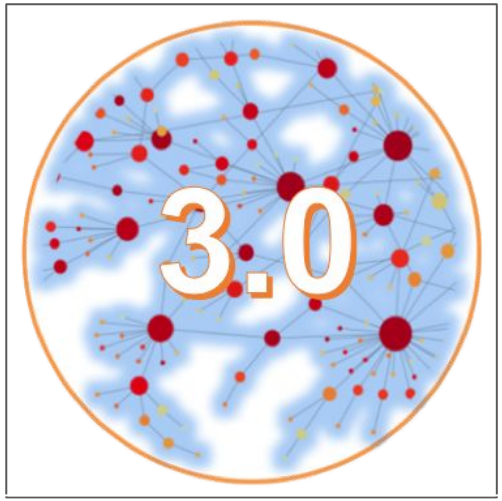
PCA Plot: Combined Dataset



PCA Plot: Combined Dataset From the paper

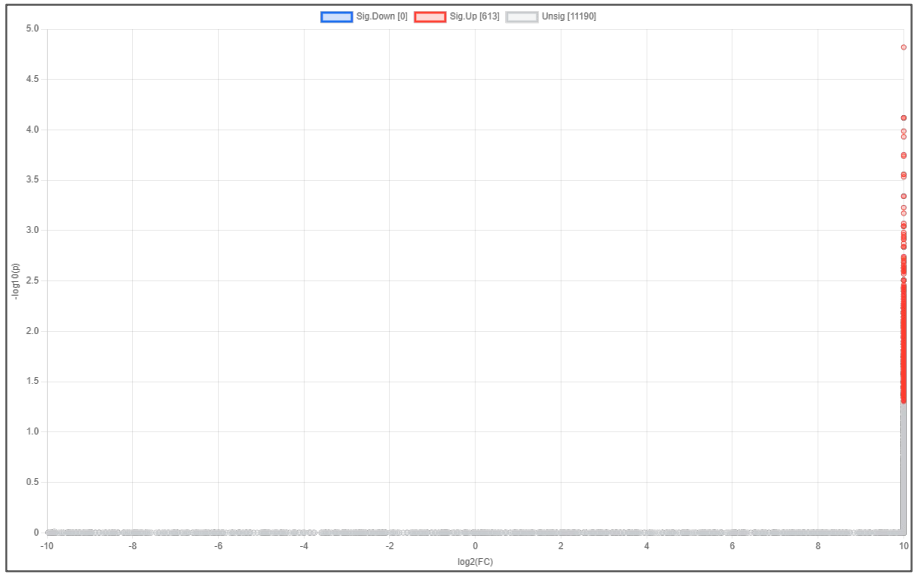
Screening of DEG's in vitiligo

- The combined dataset was analyzed using **NetworkAnalyst**.
- No additional normalization was performed as the data was pre-normalized using the **preprocessCore** package.
- Differential expression analysis was conducted using the **Limma** package.
- Thresholds for the analysis:
  - **log2 fold change (log2FC):** 0.5
  - **p-value cutoff:** 0.05
- The analysis identified **0 significant genes**.



Adjustments Made

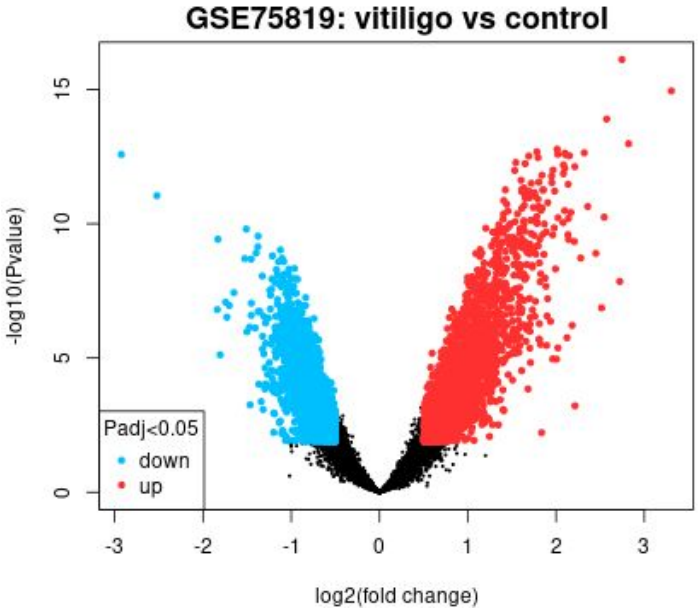
- Parameters were modified, and normalization was performed directly in **NetworkAnalyst.**
- Batch effects were addressed using **Limma**, and **Log2 transformed**.
- The revised analysis yielded **613 significant genes**, all of which were **upregulated**.



Volcano Plot: Combined Dataset

Screening of Individual Datasets' DEG's in vitiligo GSE75819

- The datasets were analyzed individually using **GEO2R**.
- Analysis parameters:
  - **p-value cutoff:** 0.05
  - **log2 fold change (log2FC):** 0.5
- Results for one dataset:
  - **5,006 significant genes** were identified.
  - **2,009 genes** were **downregulated**.
  - The remaining genes (**2,997**) were **upregulated**.
- These findings align more closely with the results of the combined dataset from the paper(3950 DEG's, 2065 upregulated, 1885 downregulated)



Volcano Plot: Individual Dataset (GSE75819)

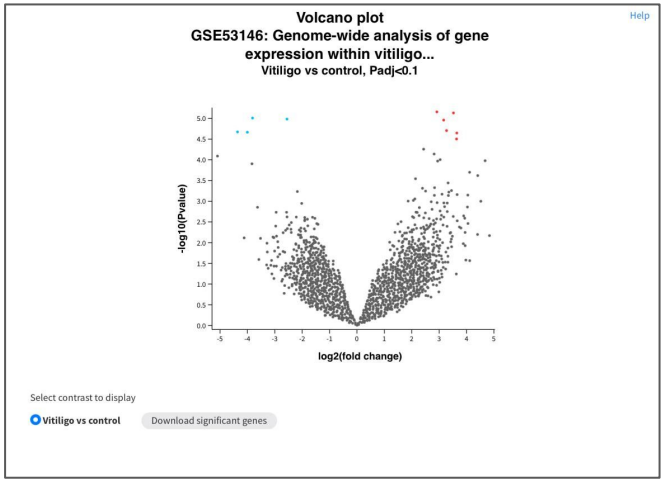
Screening of Individual Datasets' DEG's in vitiligo GSE53146

Initial Analysis:

- A volcano plot was generated for **GSE53146** using thresholds of **p < 0.05** and **LogFC > 0.5**.
- No differentially expressed genes (DEGs) were identified under these stringent criteria.

Adjusted Analysis:

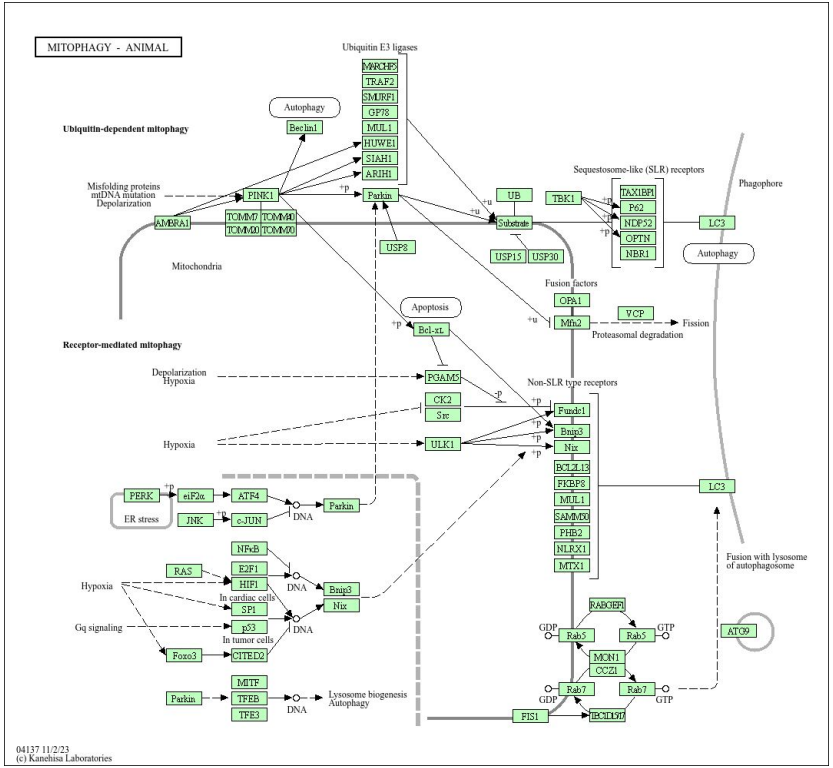
- To explore potential DEGs, the thresholds were relaxed to **p < 0.1** and **LogFC < 0**.
- This adjustment identified **10 significant genes**.



Volcano Plot: Individual Dataset (GSE53146)

Identification of potential hub genes

- Mitophagy related genes were obtained from the KEGG database, hsa04137, which contains 72 genes
- It includes ubiquitin-dependent mitophagy, regulated by key proteins like PINK1 and Parkin, and receptor-mediated mitophagy, involving receptors such as FUNDC1, BNIP3, and NIX.



KEGG pathway, hsa04137

Identification of potential hub genes

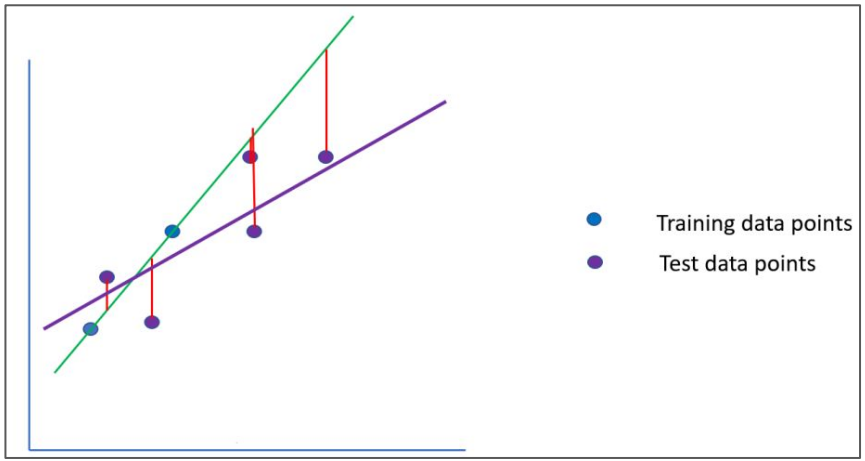
- Common genes between hsa04137 and DEGs from GSE75819 we identified
- **GSE53146** had no common genes

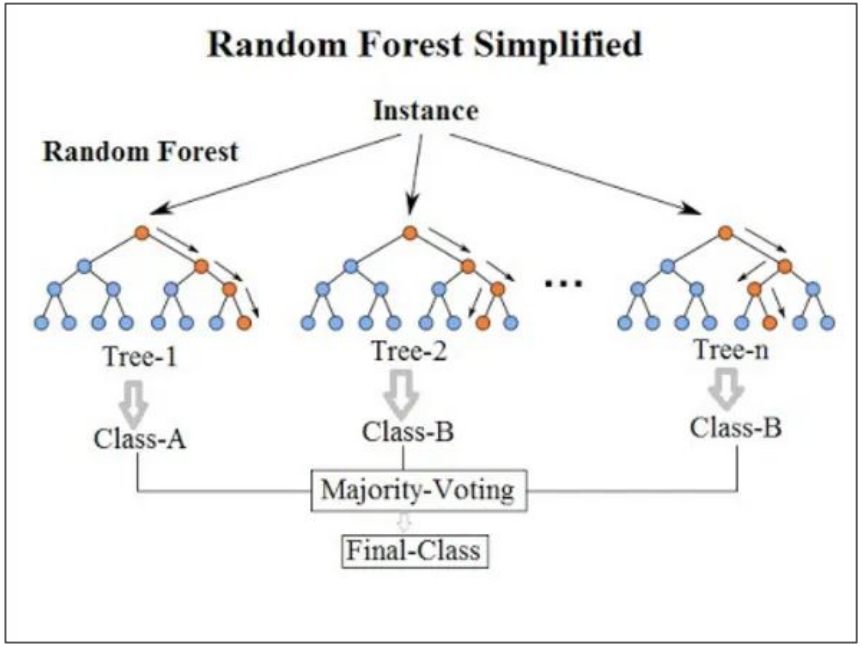
```python
import pandas as pd
from sklearn.linear_model import Lasso
from sklearn.preprocessing import StandardScaler

# Assuming 'df' is your DataFrame
X = df.drop(columns=["target"])
y = df["target"]

# Standardize the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Fit Lasso regression
lasso = Lasso(alpha=0.1)  # Adjust alpha for desired sparsity
lasso.fit(X_scaled, y)

# Get feature coefficients
coefficients = pd.DataFrame({
    "Feature": X.columns,
    "Coefficient": lasso.coef_
})

# Filter features with non-zero coefficients
lasso_selected_features = coefficients[coefficients["Coefficient"] != 0]["Feature"].tolist()

print("Selected Features:")
print(lasso_selected_features)
```

Pipeline for identifying hub genes related to mitophagy using KEGG and DEG analyses.

Identification of potential hub genes

LASSO regression and Random forest were applied on the common genes.



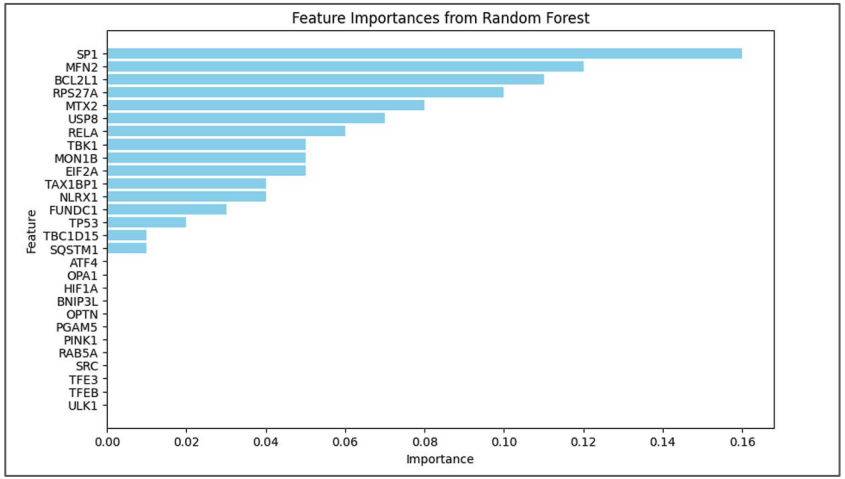


Comparison of hub genes identified through LASSO regression and Random Forest models

Identification of potential hub genes

- For GSE75819 we identified - BCL2L1, MTX2, RPS27A, TAX1BP1, USP8

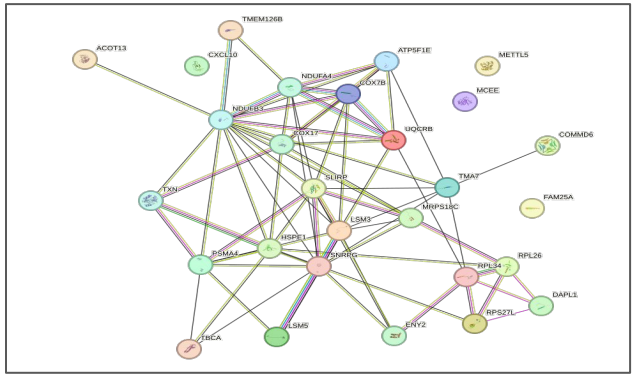
- **LASSO** - BCL2L1, MTX2, PINK1, RPS27A, SQSTM1, TAX1BP1, USP8
- **RFM** - SP1, MFN2, BCL2L1, RPS27A, USP8, RELA, NLRX1, MON1B, MTX2, EIF2A, TBK1, FUNDC1, TAX1BP1, TP53, TBC1D15
- **Overlap** - BCL2L1, MTX2, RPS27A, TAX1BP1, *USP8*
- **Paper** - GABARAPL2, SP1, *USP8*, RELA, TBC1D17

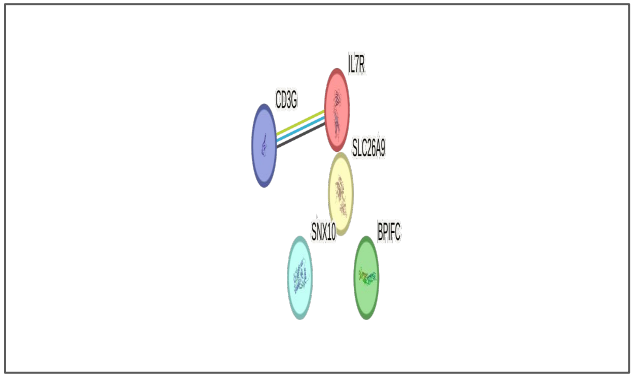


Comparison of hub genes identified through LASSO regression and Random Forest models

Protein Protein Interaction Analysis of DEGs

- Performed **protein-protein interaction (PPI)** analysis using **STRING** on the top 30 upregulated genes in **vitiligo vs. control**.
- Gene ontology (GO) analysis revealed that the most prominent pathways in **GSE75819** were related to **mitochondrial function**.
- GO analysis of **GSE53146** didn't show promising results.





Biological Process (Gene Ontology)

GO-term	description	count in network	strength	signal	false discovery rate
GO:0006119	Oxidative phosphorylation	5 of 122	1.46	0.69	0.0140
GO:1904960	Positive regulation of cytochrome-c oxidase activity	2 of 3	2.67	0.57	0.0434
GO:0042775	Mitochondrial ATP synthesis coupled electron transport	4 of 92	1.49	0.54	0.0402
GO:0019646	Aerobic electron transport chain	4 of 87	1.51	0.54	0.0402
GO:0006091	Generation of precursor metabolites and energy	6 of 411	1.01	0.46	0.0434

Human Phenotype (Monarch)

phenotype	description	count in network	strength	signal	false discovery rate
HP:0005415	Decreased proportion of CD8-positive T cells	2 of 10	2.9	1.22	0.0201
HP:0002721	Immunodeficiency	3 of 192	1.79	1.01	0.0268
HP:0000388	Otitis media	3 of 203	1.77	1.01	0.0268
HP:0002205	Recurrent respiratory infections	4 of 472	1.52	0.98	0.0201
HP:0011947	Respiratory tract infection	4 of 556	1.45	0.95	0.0201

Conclusion

- The paper identified the hub genes - (GABARAPL2, SP1, USP8, RELA, and TBC1D17)
- However we were able to identify only one of them - (BCL2L1, MTX2, RPS27A, TAX1BP1, USP8)
  - Heavy reliance on publicly available datasets.
  - Lack of clarity on how and why the datasets were combined.
  - Parameters used for sva and limma were not provided
  - Hyperparameters for the lasso or ridge models were not provided

Takeaway

The paper sought to address an important issue and yielded valuable results; however, the lack of clarity and accessibility in its methodology made it challenging to conduct a reproducibility study.

References

1. Luo, L., Zhu, J., Guo, Y., & Li, C. (2023). Mitophagy and immune infiltration in vitiligo: Evidence from bioinformatics analysis. Frontiers in Immunology, 14, 1164124. https://doi.org/10.3389/fimmu.2023.1164124
2. Zhou, G., Soufan, O., Ewald, J., Hancock, R. E. W., Basu, N., & Xia, J. (2019). NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic acids research*, *47*(W1), W234–W241. https://doi.org/10.1093/nar/gkz240
3. Rashighi, M., Agarwal, P., Richmond, J. M., Harris, T. H., Dresser, K., Su, M. W., Zhou, Y., Deng, A., Hunter, C. A., Luster, A. D., & Harris, J. E. (2014). CXCL10 is critical for the progression and maintenance of depigmentation in a mouse model of vitiligo. *Science translational medicine*, *6*(223), 223ra23. https://doi.org/10.1126/scitranslmed.3007811
4. Singh, A., Gotherwal, V., Junni, P., Vijayan, V., Tiwari, M., Ganju, P., Kumar, A., Sharma, P., Fatima, T., Gupta, A., Holla, A., Kar, H. K., Khanna, S., Thukral, L., Malik, G., Natarajan, K., Gadgil, C. J., Lahesmaa, R., Natarajan, V. T., Rani, R., … Gokhale, R. S. (2017). Mapping architectural and transcriptional alterations in non-lesional and lesional epidermis in vitiligo. *Scientific reports*, *7*(1), 9860. https://doi.org/10.1038/s41598-017-10253-w

References

1. 5. Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research, 30*(1), 207–210. https://doi.org/10.1093/nar/30.1.207
2. 6. Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research, 28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27
3. 7. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., ... & von Mering, C. (2019). STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research, 47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131
4. 8. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324
5. 9. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological), 58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x