Development of a Frailty Risk Score Model

Shehbeel Arif
Drexel University College of Medicine
Contact: shehbeelarif@gmail.com
Github: @shehbeel

Method

DE Analysis

Volcano plot

EnhancedVolcano
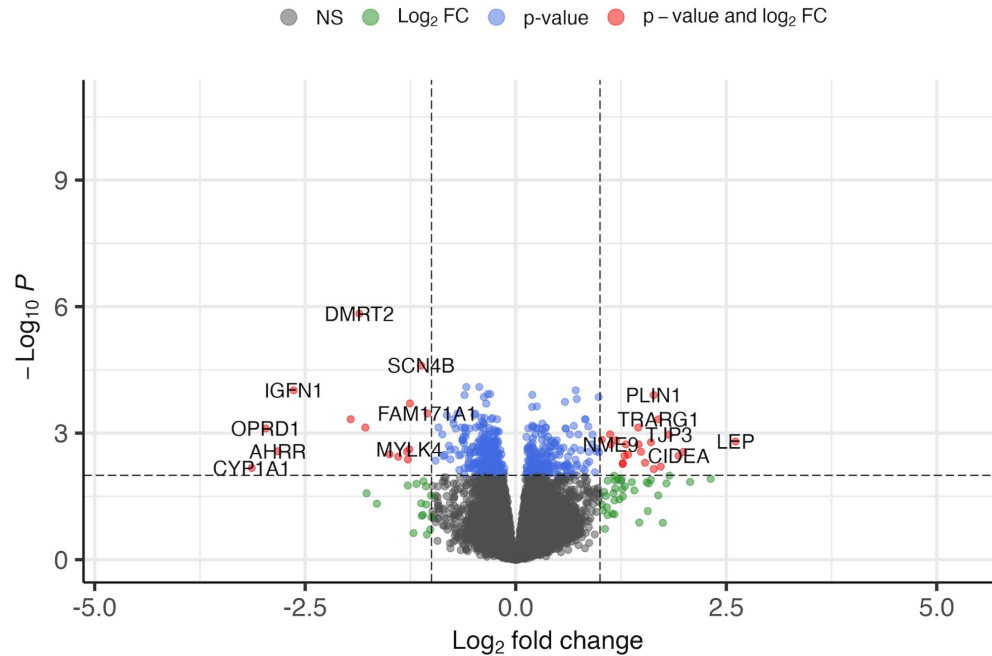
Recursive Feature Elimination using Random Forest
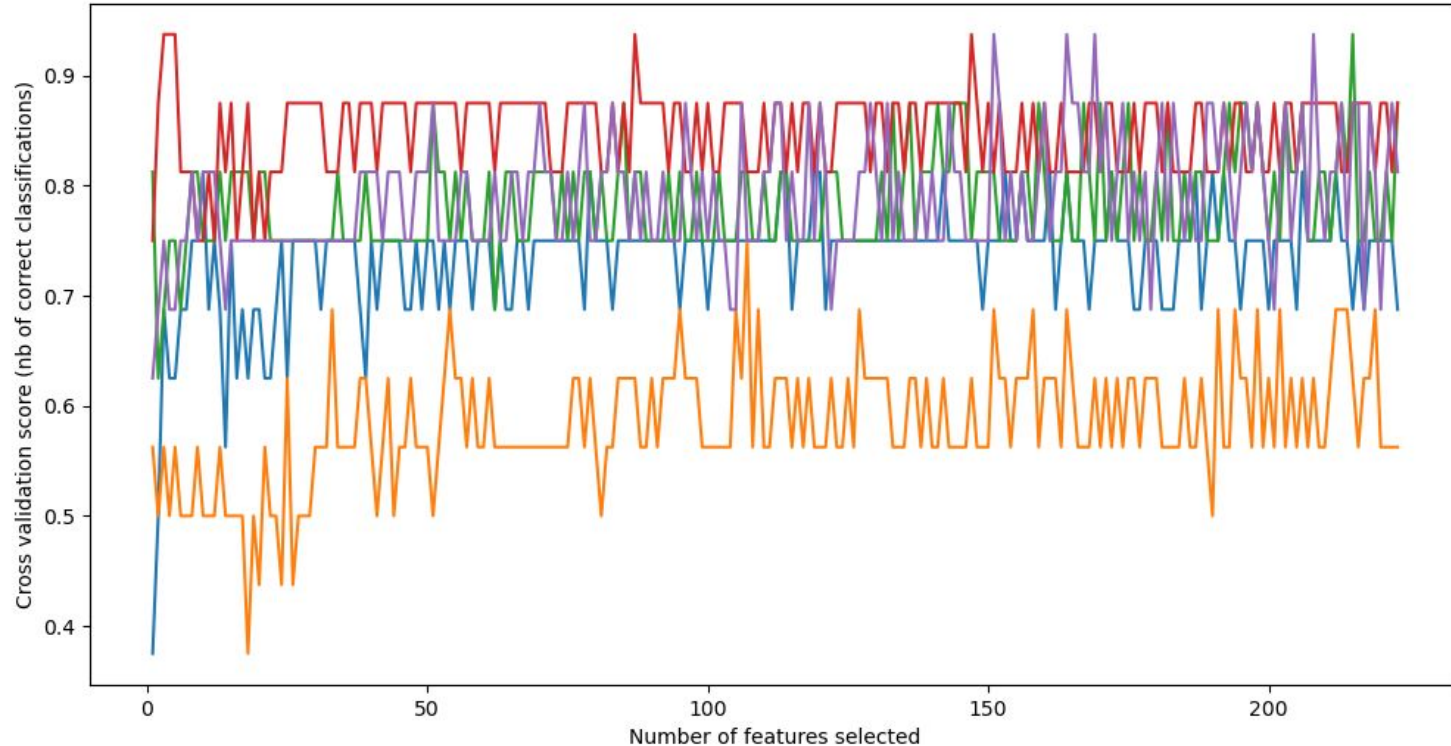
GO Pathways (Molecular Function)

164 Genes

	Homo sapiens (REF)		upload_1 (Hierarchy) NEW! ⑦)					
GO molecular function complete	⚠ #	#	expected	Fold Enrichment	+/-	raw P value	FDR
protein tyrosine kinase activator activity	34	6	.30	20.19	+	1.17E-06	5.92E-03
signaling receptor regulator activity	567	17	4.96	3.43	+	1.38E-05	3.49E-02

Minimum Redundancy Maximum Relevance (MRMR)

Top 5 genes

- Accuracy = 0.9
- AUC = 0.884

	frailty		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.0007	0.0000 – 0.0262	**<0.001**
HABP2	1.3626	1.0735 – 1.7295	**0.012**
MYBPHL	1.2475	0.6636 – 2.3452	0.487
TJP3	1.2777	1.0255 – 1.5918	**0.029**
TEX56P	5.8360	1.8895 – 18.0252	**0.003**
CCL14	1.2661	1.0500 – 1.5268	**0.014**
Observations	80		
R^2 / R^2 adjusted	0.676 / 0.655		

Top 10 genes

- Accuracy = 0.95
- AUC = 0.942

	frailty		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.0000	0.0000 – 0.0119	**0.005**
EFNA5	0.9088	0.5795 – 1.4254	0.673
HABP2	1.5961	1.0245 – 2.4866	**0.039**
MYBPHL	1.2473	0.3812 – 4.0809	0.711
TJP3	1.2752	0.9091 – 1.7886	0.156
TEX56P	11.7245	1.1736 – 117.1343	**0.036**
ACRP1	1.2960	0.1127 – 14.9076	0.833
LRTM1	3.0851	0.6585 – 14.4552	0.150
CCL14	1.5228	1.0408 – 2.2280	**0.031**
SEMA3E	1.2113	0.7475 – 1.9629	0.431
PDF	1.3758	1.0092 – 1.8757	**0.044**
Observations	80		
R^2 / R^2 adjusted	0.815 / 0.795		

Multinomial Logistic Regression Classifier

**Accuracy:** 0.95

AUC: 0.931

Model Equation:

LP = 0.000007422891 + (1.463481209314*$HABP2$) + (1.432561618292*$TJP3$) + (11.787795400887*$TEX56P$) + (1.540721723569*$CCL14$) + (1.324008646151*PDF)

Frailty Risk Score = 1/(1+exp(−**LP**))

Score Range: [0,1.0]

Predictors	Odds Ratios	CI	p
		frailty	
(Intercept)	0.0000	0.0000 – 0.0107	**0.002**
HABP2	1.4635	1.0716 – 1.9986	**0.017**
TJP3	1.4326	1.0968 – 1.8712	**0.009**
TEX56P	11.7878	1.9031 – 73.0128	**0.009**
CCL14	1.5407	1.1132 – 2.1324	**0.010**
PDF	1.3240	1.0459 – 1.6760	**0.020**
Observations	80		
R^2 / R^2 adjusted	0.761 / 0.741		

Gene Expression
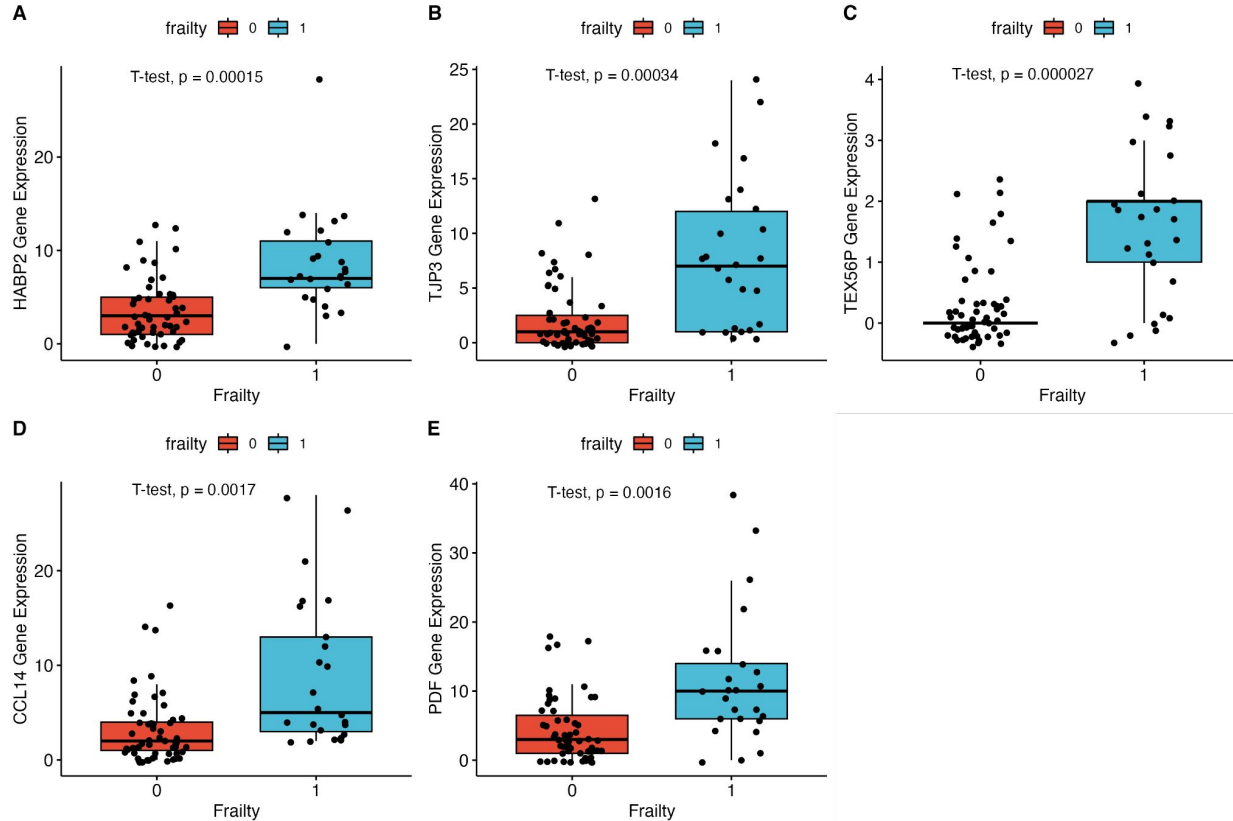
References

- Abbas, M., & El-Manzalawy, Y. (2020). Machine learning based refined differential gene expression analysis of pediatric sepsis. BMC medical genomics, 13(1), 122. https://doi.org/10.1186/s12920-020-00771-4
- Abraham, G., Malik, R., Yonova-Doing, E., Salim, A., Wang, T., Danesh, J., Butterworth, A. S., Howson, J. M. M., Inouye, M., & Dichgans, M. (2019). Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. Nature communications, 10(1), 5819. https://doi.org/10.1038/s41467-019-13848-1
- De Jay, N., Papillon-Cavanagh, S., Olsen, C., El-Hachem, N., Bontempi, G., & Haibe-Kains, B. (2013). mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics (Oxford, England), 29(18), 2365–2368. https://doi.org/10.1093/bioinformatics/btt383
- Grant, S. W., Collins, G. S., & Nashef, S. A. M. (2018). Statistical Primer: developing and validating a risk prediction model. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery, 54(2), 203–208. https://doi.org/10.1093/ejcts/ezy180
- Pavlou, M., Ambler, G., Seaman, S. R., Guttmann, O., Elliott, P., King, M., & Omar, R. Z. (2015). How to develop a more accurate risk prediction model when there are few events. BMJ (Clinical research ed.), 351, h3868. https://doi.org/10.1136/bmj.h3868
- Selby, J. V., & Fireman, B. H. (2021). Building Predictive Models for Clinical Care-Where to Build and What to Predict?. JAMA network open, 4(1), e2032539. https://doi.org/10.1001/jamanetworkopen.2020.32539