

Supplementary Material

An ensemble feature selection approach for analysis and modeling of transcriptome data in Alzheimer's Disease

Petros Paplomatas *, Marios Krokidis *, Panagiotis Vlamos and Aristidis G. Vrahatis

*Correspondence: p.paplomatas@hotmail.com, mkrokidis@ionio.gr

Table S1. The top 10 enriched biological process for DEGs.

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	oxidative RNA demethylation (GO:0035513)	0.002997	0.07379	454.18	2638.91
2	regulation of core promoter binding (GO:1904796)	0.002997	0.07379	454.18	2638.91
3	mRNA cleavage (GO:0006379)	0.002997	0.07379	454.18	2638.91
4	glandular epithelial cell development (GO:0002068)	0.004193	0.07379	302.76	1657.40
5	negative regulation of keratinocyte proliferation (GO:0010839)	0.004791	0.07379	259.49	1385.98
6	DNA dealkylation (GO:0035510)	0.005985	0.07379	201.81	1032.95
7	DNA dealkylation involved in DNA repair (GO:0006307)	0.005985	0.07379	201.81	1032.95
8	hemidesmosome assembly (GO:0031581)	0.007178	0.07379	165.10	815.05
9	cellular response to thyroid hormone stimulus (GO:0097067)	0.007178	0.07379	165.10	815.05
10	response to thyroid hormone (GO:0097066)	0.007774	0.07379	151.33	735.02

Table S2. The top 10 enriched molecular function for DEGs.

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	endoribonuclease activity, producing 3'-phosphomonoesters (GO:0016892)	0.002997	0.06173	454.18	2638.91
2	guanylate kinase activity (GO:0004385)	0.005388	0.06173	227.05	1185.99
3	cAMP binding (GO:0030552)	0.007178	0.06173	165.10	815.05
4	R-SMAD binding (GO:0070412)	0.01134	0.06173	100.86	451.76
5	nucleoside monophosphate kinase activity (GO:0050145)	0.01194	0.06173	95.55	423.08
6	microtubule plus-end binding (GO:0051010)	0.01194	0.06173	95.55	423.08
7	histone acetyltransferase binding (GO:0035035)	0.01253	0.06173	90.76	397.51
8	cyclic nucleotide binding (GO:0030551)	0.01372	0.06173	82.50	353.87
9	2-oxoglutarate-dependent dioxygenase activity (GO:0016706)	0.01963	0.07851	56.69	222.85

Table S3. The top 10 enriched cellular component for DEGs.

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	endolysosome lumen (GO:0036021)	0.002997	0.02158	454.18	2638.91
2	hemidesmosome (GO:0030056)	0.004193	0.02516	302.76	1657.40
3	platelet dense granule lumen (GO:0031089)	0.008370	0.04305	139.69	668.13
4	basement membrane (GO:0005604)	0.0004304	0.007746	79.75	618.15
5	platelet dense granule (GO:0042827)	0.01253	0.05577	90.76	397.51
6	endolysosome (GO:0036019)	0.01490	0.05577	75.62	318.09
7	endosome lumen (GO:0031904)	0.01549	0.05577	72.59	302.52
8	collagen-containing extracellular matrix (GO:0062023)	0.00005627	0.002026	26.08	255.20
9	ciliary membrane (GO:0060170)	0.01845	0.06037	60.48	241.48
10	intermediate filament (GO:0005882)	0.02960	0.08880	36.99	130.21

Table S4. The top 10 enriched Reactome pathway for DEGs

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	Reversal Of Alkylation Damage by DNA Dioxygenases R-HSA-73943	0.004193	0.04917	302.76	1657.40
2	RUNX3 Regulates CDKN1A Transcription R-HSA-8941855	0.004193	0.04917	302.76	1657.40
3	DNA Damage Reversal R-HSA-73942	0.004791	0.04917	259.49	1385.98
4	Type I Hemidesmosome Assembly R-HSA-446107	0.005388	0.04917	227.05	1185.99
5	Trafficking And Processing of Endosomal TLR R-HSA-1679131	0.007774	0.05675	151.33	735.02
6	Synthesis, Secretion, And Inactivation of Glucose-dependent Insulinotropic Polypeptide (GIP) R-HSA-400511	0.007774	0.05675	151.33	735.02
7	Rap1 Signaling R-HSA-392517	0.009560	0.05725	121.05	562.89
8	Assembly Of Collagen Fibrils and Other Multimeric Structures R-HSA-2022090	0.0005171	0.02221	72.48	548.50
9	Syndecan Interactions R-HSA-3000170	0.01194	0.05725	95.55	423.08
10	Regulation Of Gene Expression in Beta Cells R-HSA-210745	0.01194	0.05725	95.55	423.08

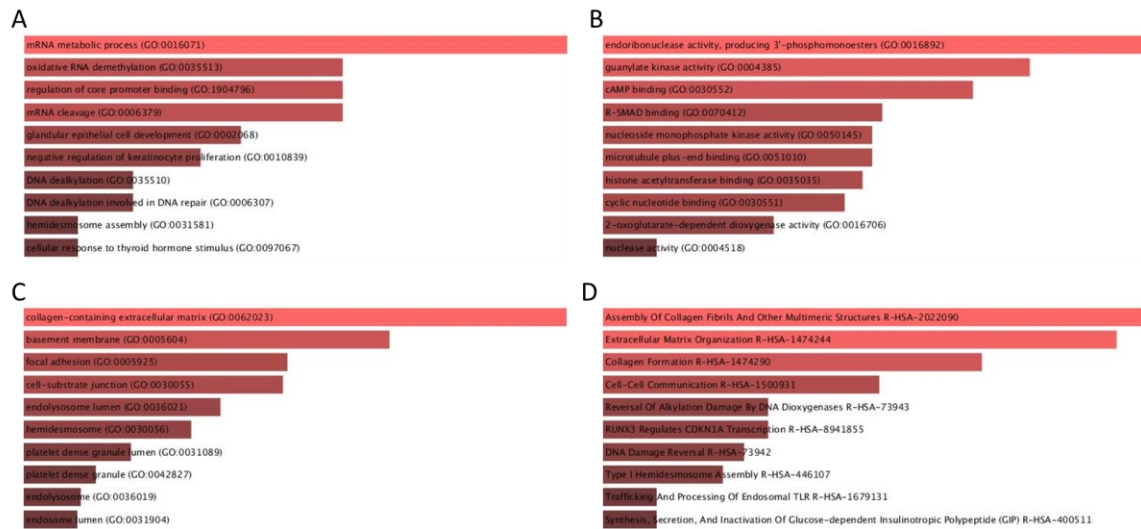


Figure S1. GO term and Reactome pathway enrichment analyses performed using Enrichr on DEGs. (A) The top 10 enriched biological process for DEGs. (B) The top 10 enriched molecular function for DEGs. (C) The top 10 enriched cellular component for DEGs. (D) The top 10 enriched Reactome pathway for DEGs.

Table S5. The top 10 enriched biological process for Dst

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	hemidesmosome assembly (GO:0031581)	0.0006000	0.004800	19988.00	148283.12
2	retrograde axonal transport (GO:0008090)	0.0008000	0.004800	19984.00	142504.37
3	cell-substrate junction assembly (GO:0007044)	0.001700	0.005400	19966.00	127326.11
4	axonal transport (GO:0098930)	0.001800	0.005400	19964.00	126172.23
5	cytoplasmic microtubule organization (GO:0031122)	0.002700	0.006428	19946.00	117971.02
6	wound healing (GO:0042060)	0.003500	0.006428	19930.00	112704.30
7	integrin-mediated signaling pathway (GO:0007229)	0.003750	0.006428	19925.00	111301.34
8	collagen fibril organization (GO:0030199)	0.004450	0.006675	19911.00	107815.38
9	cytoskeleton organization (GO:0007010)	0.006000	0.008000	19880.00	101706.24
10	extracellular matrix organization (GO:0030198)	0.01500	0.01800	19700.00	82734.36

Table S6. The top 3 enriched molecular function for Dst.

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	microtubule plus-end binding (GO:0051010)	0.001000	0.003000	19980.00	138017.41
2	microtubule binding (GO:0008017)	0.01160	0.01535	19768.00	88101.22
3	tubulin binding (GO:0015631)	0.01535	0.01535	19693.00	82250.73

Table S7. The top 10 enriched cellular component for Dst

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	hemidesmosome (GO:0030056)	0.0003500	0.004200	19993.00	159096.45
2	intermediate filament (GO:0005882)	0.002500	0.01040	19950.00	119530.06
3	basement membrane (GO:0005604)	0.002600	0.01040	19948.00	118735.70
4	intermediate filament cytoskeleton (GO:0045111)	0.004200	0.01260	19916.00	108994.00
5	polymeric cytoskeletal fiber (GO:0099513)	0.01280	0.02627	19744.00	86050.65
6	microtubule cytoskeleton (GO:0015630)	0.01655	0.02627	19669.00	80669.99
7	collagen-containing extracellular matrix (GO:0062023)	0.01900	0.02627	19620.00	77760.41
8	focal adhesion (GO:0005925)	0.01935	0.02627	19613.00	77374.66
9	cell-substrate junction (GO:0030055)	0.01970	0.02627	19606.00	76995.59
10	cytoskeleton (GO:0005856)	0.03000	0.03600	19400.00	68027.34

Table S8. The top 10 enriched Reactome pathway for Dst

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	Type I Hemidesmosome Assembly R-HSA-446107	0.0004500	0.005375	19991.00	154056.47
2	RHOV GTPase Cycle R-HSA-9013424	0.001850	0.005375	19963.00	125618.94
3	RHOU GTPase Cycle R-HSA-9013420	0.001950	0.005375	19961.00	124555.53
4	RND3 GTPase Cycle R-HSA-9696264	0.002100	0.005375	19958.00	123057.76
5	RND1 GTPase Cycle R-HSA-9696273	0.002100	0.005375	19958.00	123057.76
6	RND2 GTPase Cycle R-HSA-9696270	0.002150	0.005375	19957.00	122581.99
7	Assembly Of Collagen Fibrils and Other Multimeric Structures R-HSA-2022090	0.002850	0.006107	19943.00	116875.01
8	Cell Junction Organization R-HSA-446728	0.004300	0.007500	19914.00	108514.46
9	Collagen Formation R-HSA-1474290	0.004500	0.007500	19910.00	107587.50
10	Cell-Cell Communication R-HSA-1500931	0.006000	0.009000	19880.00	101706.24

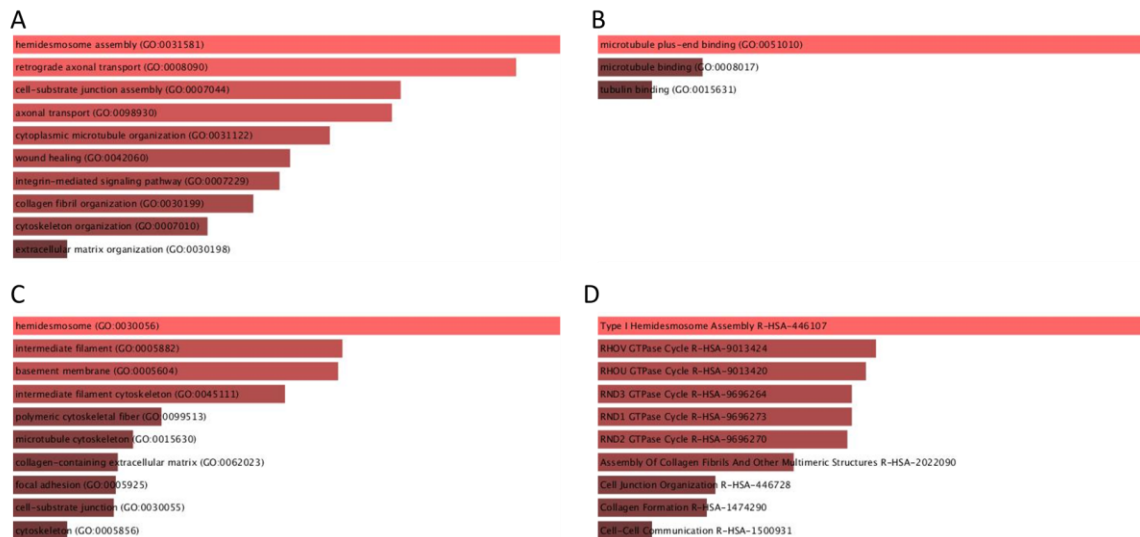


Figure S2. GO term and Reactome pathway enrichment analyses performed using Enrichr on Dst. (A) The top 10 enriched biological process. (B) The top 3 enriched molecular function. (C) The top 10 enriched cellular component. (D) The top 10 enriched Reactome pathway.

Table S9. The top 3 enriched biological process for *Rapgef4*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	regulation of peptide hormone secretion (GO:0090276)	0.003700	0.006250	19926.00	111574.39
2	regulation of insulin secretion (GO:0050796)	0.005200	0.006250	19896.00	104635.25
3	regulation of protein secretion (GO:0050708)	0.006250	0.006250	19875.00	100869.32

Table S10. The top 7 enriched molecular function for *Rapgef4*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	cAMP binding (GO:0030552)	0.0006000	0.004025	19988.00	148283.12
2	cyclic nucleotide binding (GO:0030551)	0.001150	0.004025	19977.00	135204.64
3	guanyl-nucleotide exchange factor activity (GO:0005085)	0.007450	0.01407	19851.00	97261.02
4	small GTPase binding (GO:0031267)	0.008750	0.01407	19825.00	93944.97
5	GTPase binding (GO:0051020)	0.01005	0.01407	19799.00	91079.22
6	adenyl ribonucleotide binding (GO:0032559)	0.01530	0.01680	19694.00	82319.16
7	GTPase activator activity (GO:0005096)	0.01680	0.01680	19664.00	80354.66

Table S11. The top 10 enriched Reactome pathway for *Rapgef4*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	Rap1 Signaling R-HSA-392517	0.0008000	0.006300	19984.00	142504.37
2	Integrin Signaling R-HSA-354192	0.001350	0.006300	19973.00	131975.02
3	Platelet Aggregation (Plug Formation) R-HSA-76009	0.001800	0.006300	19964.00	126172.23
4	Glucagon-like Peptide-1 (GLP1) Regulates Insulin Secretion R-HSA-381676	0.002100	0.006300	19958.00	123057.76
5	Regulation Of Insulin Secretion R-HSA-422356	0.003850	0.009240	19923.00	110765.84
6	Integration Of Energy Metabolism R-HSA-163685	0.005250	0.01050	19895.00	104439.61
7	Platelet Activation, Signaling and Aggregation R-HSA-76002	0.01270	0.02177	19746.00	86214.24
8	Hemostasis R-HSA-109582	0.02880	0.04320	19424.00	68904.42
9	Adaptive Immune System R-HSA-1280218	0.03665	0.04887	19267.00	63703.39
10	Immune System R-HSA-168256	0.09715	0.1118	18057.00	42099.93

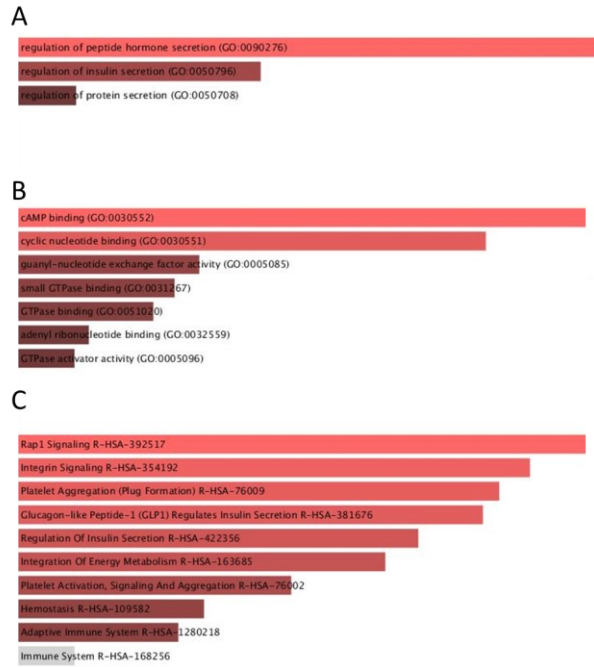


Figure S3. GO term and Reactome pathway enrichment analyses performed using Enrichr on *Rapgef4*. A) The top 3 enriched biological process. (B) The top 7 enriched molecular function. (C) The top 10 enriched Reactome pathway.

Table S12. The top 10 enriched biological process for *Pax6*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	regulation of core promoter binding (GO:1904796)	0.0002500	0.004900	19995.00	165840.18
2	glandular epithelial cell development (GO:0002068)	0.0003500	0.004900	19993.00	159096.45
3	neuron fate commitment (GO:0048663)	0.0007000	0.005880	19986.00	145187.41
4	negative regulation of nervous system development (GO:0051961)	0.0009500	0.005880	19981.00	139049.21
5	positive regulation of transcription regulatory region DNA binding (GO:2000679)	0.001050	0.005880	19979.00	137035.71
6	negative regulation of cell development (GO:0010721)	0.001300	0.006067	19974.00	132735.46
7	negative regulation of neurogenesis (GO:0050768)	0.001750	0.007000	19965.00	126740.99
8	visual system development (GO:0150063)	0.002050	0.007175	19959.00	123544.89
9	sensory organ development (GO:0007423)	0.002800	0.007891	19944.00	117233.87
10	eye development (GO:0001654)	0.002900	0.007891	19942.00	116522.32

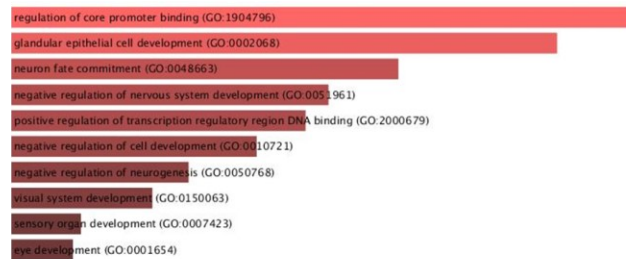
Table S13. The top 10 enriched molecular function for *Pax6*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	R-SMAD binding (GO:0070412)	0.0009500	0.006825	19981.00	139049.21
2	histone acetyltransferase binding (GO:0035035)	0.001050	0.006825	19979.00	137035.71
3	transcription regulatory region nucleic acid binding (GO:0001067)	0.01060	0.04593	19788.00	89974.28
4	kinase binding (GO:0019900)	0.02305	0.05142	19539.00	73663.91
5	protein kinase binding (GO:0019901)	0.02530	0.05142	19494.00	71678.61
6	transcription cis-regulatory region binding (GO:0000976)	0.02745	0.05142	19451.00	69934.03
7	double-stranded DNA binding (GO:0003690)	0.03255	0.05142	19349.00	66270.01
8	sequence-specific DNA binding (GO:0043565)	0.03535	0.05142	19293.00	64486.12
9	sequence-specific double-stranded DNA binding (GO:1990837)	0.03560	0.05142	19288.00	64333.48
10	DNA binding (GO:0003677)	0.04055	0.05271	19189.00	61505.05

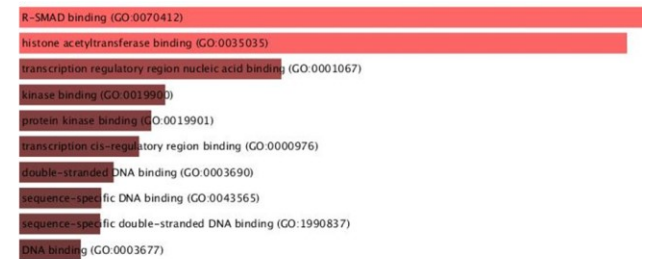
Table S14. The top 10 enriched Reactome pathway for *Pax6*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	Synthesis, Secretion, And Inactivation of Glucose-dependent Insulinotropic Polypeptide (GIP) R-HSA-400511	0.0006500	0.002700	19987.00	146675.88
2	Regulation Of Gene Expression in Beta Cells R-HSA-210745	0.001000	0.002700	19980.00	138017.41
3	Synthesis, Secretion, And Inactivation of Glucagon-like Peptide-1 (GLP-1) R-HSA-381771	0.001050	0.002700	19979.00	137035.71
4	Incretin Synthesis, Secretion, And Inactivation R-HSA-400508	0.001200	0.002700	19976.00	134347.69
5	Regulation Of Beta-Cell Development R-HSA-186712	0.002050	0.003690	19959.00	123544.89
6	Peptide Hormone Metabolism R-HSA-2980736	0.004450	0.005850	19911.00	107815.38
7	Activation Of HOX Genes During Differentiation R-HSA-5619507	0.004550	0.005850	19909.00	107362.11
8	Developmental Biology R-HSA-1266738	0.05365	0.06036	18927.00	55366.73
9	Metabolism Of Proteins R-HSA-392499	0.09450	0.09450	18110.00	42724.35

A



B



C

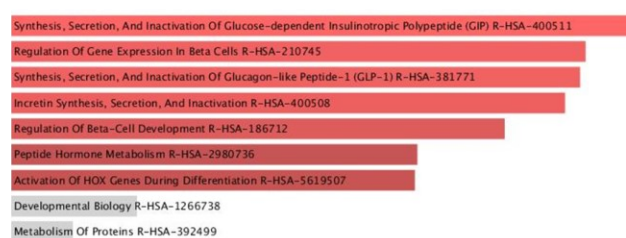


Figure S4. GO term and Reactome pathway enrichment analyses performed using Enrichr on *Pax6*. A) The top 10 enriched biological process. (B) The top 10 enriched molecular function. (C) The top 10 enriched Reactome pathway.

Table S15. The top 10 enriched biological process for *Cask*.

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	negative regulation of keratinocyte proliferation (GO:0010839)	0.0004000	0.004000	19992.00	156418.91
2	regulation of cellular response to growth factor stimulus (GO:0090287)	0.0008500	0.004000	19983.00	141285.77
3	negative regulation of response to wounding (GO:1903035)	0.001050	0.004000	19979.00	137035.71
4	regulation of keratinocyte proliferation (GO:0010837)	0.001200	0.004000	19976.00	134347.69
5	negative regulation of wound healing (GO:0061045)	0.001450	0.004000	19971.00	130534.69
6	negative regulation of cell-matrix adhesion (GO:0001953)	0.001600	0.004000	19968.00	128549.42
7	negative regulation of cell-substrate adhesion (GO:0010812)	0.002000	0.004286	19960.00	124043.95
8	regulation of wound healing (GO:0061041)	0.002350	0.004406	19953.00	120782.64
9	negative regulation of response to stimulus (GO:0048585)	0.003050	0.004562	19939.00	115499.24
10	regulation of cell-matrix adhesion (GO:0001952)	0.003250	0.004562	19935.00	114209.93

Table S16. The top 10 enriched molecular function for *Cask*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	guanylate kinase activity (GO:0004385)	0.0004500	0.001350	19991.00	154056.47
2	nucleoside monophosphate kinase activity (GO:0050145)	0.001000	0.001500	19980.00	138017.41
3	protein serine/threonine kinase activity (GO:0004674)	0.01720	0.01720	19656.00	79859.45

Table S17. The top 10 enriched cellular component for *Cask*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	ciliary membrane (GO:0060170)	0.001550	0.01820	19969.00	129189.85
2	basement membrane (GO:0005604)	0.002600	0.01820	19948.00	118735.70
3	cell projection membrane (GO:0031253)	0.004600	0.02147	19908.00	107139.14
4	cilium (GO:0005929)	0.01185	0.03064	19763.00	87657.54
5	cell-cell junction (GO:0005911)	0.01355	0.03064	19729.00	84861.88
6	actin cytoskeleton (GO:0015629)	0.01580	0.03064	19684.00	81644.38
7	collagen-containing extracellular matrix (GO:0062023)	0.01900	0.03064	19620.00	77760.41
8	focal adhesion (GO:0005925)	0.01935	0.03064	19613.00	77374.66
9	cell-substrate junction (GO:0030055)	0.01970	0.03064	19606.00	76995.59
10	cytoskeleton (GO:0005856)	0.03000	0.04130	19400.00	68027.34

Table S18. The top 10 enriched Reactome pathway for *Cask*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	Syndecan Interactions R-HSA-3000170	0.001000	0.005625	19980.00	138017.41
2	Nephrin Family Interactions R-HSA-373753	0.001100	0.005625	19978.00	136099.47
3	Dopamine Neurotransmitter Release Cycle R-HSA-212676	0.001150	0.005625	19977.00	135204.64
4	Assembly and Cell Surface Presentation of NMDA Receptors R-HSA-9609736	0.001250	0.005625	19975.00	133525.54
5	Non-integrin membrane-ECM Interactions R-HSA-3000171	0.002050	0.006136	19959.00	123544.89
6	Neurotransmitter Release Cycle R-HSA-112310	0.002500	0.006136	19950.00	119530.06
7	Sensory Processing of Sound by Outer Hair Cells of Cochlea R-HSA-9662361	0.002650	0.006136	19947.00	118349.79
8	Neurexins And Neuroligins R-HSA-6794361	0.002850	0.006136	19943.00	116875.01
9	Sensory Processing of Sound by Inner Hair Cells of Cochlea R-HSA-9662360	0.003350	0.006136	19933.00	113594.39
10	Activation Of NMDA Receptors and Postsynaptic Events R-HSA-442755	0.003700	0.006136	19926.00	111574.39

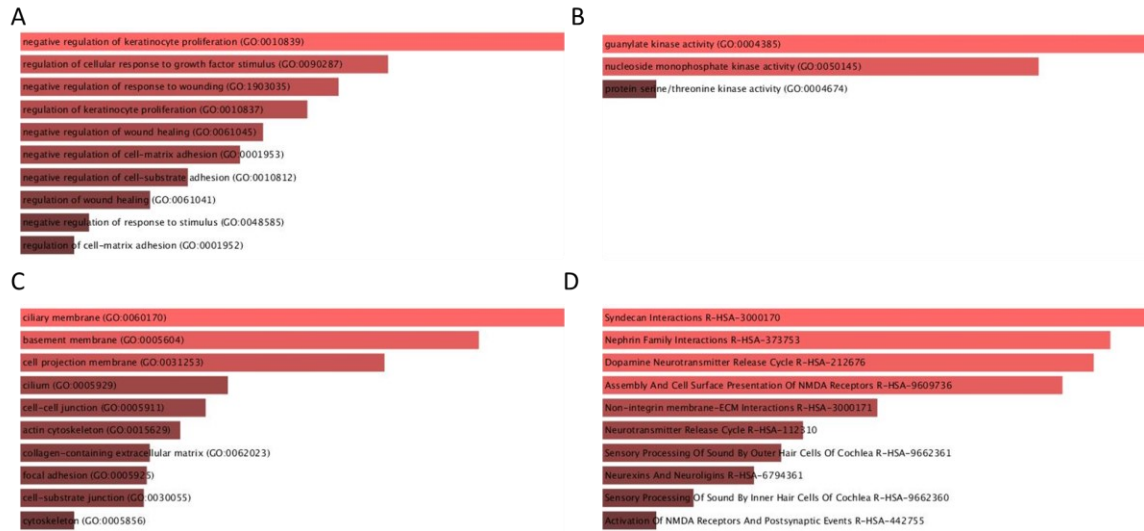


Figure S5. GO term and Reactome pathway enrichment analyses performed using Enrichr on *Cask*. (A) The top 10 enriched biological process. (B) The top 10 enriched molecular function. (C) The top 10 enriched cellular component. (D) The top 10 enriched Reactome pathway.

Table S19. The top 10 enriched biological process for *Ctsb*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	cellular response to thyroid hormone stimulus (GO:0097067)	0.0006000	0.004875	19988.00	148283.12
2	response to thyroid hormone (GO:0097066)	0.0006500	0.004875	19987.00	146675.88
3	cellular protein catabolic process (GO:0044257)	0.003000	0.009469	19940.00	115834.63
4	pattern recognition receptor signaling pathway (GO:0002221)	0.003050	0.009469	19939.00	115499.24
5	proteolysis involved in cellular protein catabolic process (GO:0051603)	0.003700	0.009469	19926.00	111574.39
6	cellular response to hormone stimulus (GO:0032870)	0.003800	0.009469	19924.00	111031.85
7	toll-like receptor signaling pathway (GO:0002224)	0.004550	0.009469	19909.00	107362.11
8	epithelial cell differentiation (GO:0030855)	0.005050	0.009469	19899.00	105233.48
9	epithelium development (GO:0060429)	0.006100	0.01017	19878.00	101367.44
10	regulation of programmed cell death (GO:0043067)	0.009700	0.01455	19806.00	91813.48

Table S20. The top 3 enriched molecular function for *Ctsb*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	cysteine-type peptidase activity (GO:0008234)	0.004900	0.007875	19902.00	105849.45
2	cysteine-type endopeptidase activity (GO:0004197)	0.005250	0.007875	19895.00	104439.61
3	endopeptidase activity (GO:0004175)	0.01575	0.01575	19685.00	81710.92

Table S21. The top 10 enriched cellular component for *Ctsb*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	endolysosome lumen (GO:0036021)	0.0002500	0.002500	19995.00	165840.18
2	endolysosome (GO:0036019)	0.001250	0.004333	19975.00	133525.54
3	endosome lumen (GO:0031904)	0.001300	0.004333	19974.00	132735.46
4	lysosomal lumen (GO:0043202)	0.004300	0.01075	19914.00	108514.46
5	ficolin-1-rich granule lumen (GO:1904813)	0.006150	0.01230	19877.00	101200.08
6	ficolin-1-rich granule (GO:0101002)	0.009200	0.01533	19816.00	92908.55
7	lytic vacuole (GO:0000323)	0.01095	0.01564	19781.00	89299.85
8	collagen-containing extracellular matrix (GO:0062023)	0.01900	0.02375	19620.00	77760.41
9	lysosome (GO:0005764)	0.02385	0.02650	19523.00	72937.49
10	intracellular organelle lumen (GO:0070013)	0.04240	0.04240	19152.00	60532.03

Table S22. The top 10 enriched Reactome pathway for *Ctsb*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	Trafficking And Processing of Endosomal TLR R-HSA-1679131	0.0006500	0.007800	19987.00	146675.88
2	Collagen Degradation R-HSA-1442490	0.002000	0.01090	19960.00	124043.95
3	Assembly Of Collagen Fibrils and Other Multimeric Structures R-HSA-2022090	0.002850	0.01090	19943.00	116875.01
4	Collagen Formation R-HSA-1474290	0.004500	0.01090	19910.00	107587.50
5	MHC Class II Antigen Presentation R-HSA-2132295	0.005200	0.01090	19896.00	104635.25
6	Degradation Of Extracellular Matrix R-HSA-1474228	0.005450	0.01090	19891.00	103674.93
7	Toll-like Receptor Cascades R-HSA-168898	0.008100	0.01389	19838.00	95537.87

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
8	Extracellular Matrix Organization R-HSA-1474244	0.01455	0.02182	19709.00	83372.48
9	Neutrophil Degranulation R-HSA-6798695	0.02340	0.03120	19532.00	73343.17
10	Adaptive Immune System R-HSA-1280218	0.03665	0.04398	19267.00	63703.39

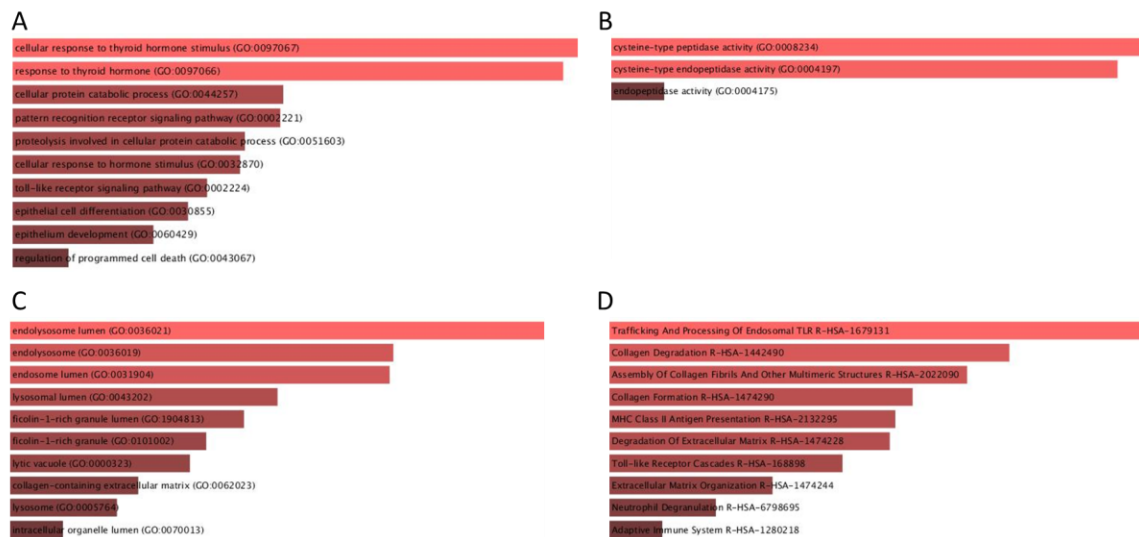


Figure S6. GO term and Reactome pathway enrichment analyses performed using Enrichr on *Ctsb*. (A) The top 10 enriched biological process. (B) The top 10 enriched molecular function. (C) The top 10 enriched cellular component. (D) The top 10 enriched Reactome pathway.

Table S23. The top 10 enriched biological process for *ZFHX3*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	regulation of neuron differentiation (GO:0045664)	0.003950	0.02652	19921.00	110243.90
2	positive regulation of cell adhesion (GO:0045785)	0.004000	0.02652	19920.00	109987.79
3	regulation of cell adhesion (GO:0030155)	0.006650	0.02652	19867.00	99596.26
4	brain development (GO:0007420)	0.007500	0.02652	19850.00	97123.34
5	regulation of cell differentiation (GO:0045595)	0.007800	0.02652	19844.00	96315.69
6	central nervous system development (GO:0007417)	0.01340	0.03797	19732.00	85094.44
7	regulation of nucleic acid-templated transcription (GO:1903506)	0.02150	0.04826	19570.00	75143.11
8	regulation of cellular macromolecule biosynthetic process (GO:2000112)	0.02340	0.04826	19532.00	73343.17
9	positive regulation of nucleic acid-templated transcription (GO:1903508)	0.02555	0.04826	19489.00	71468.59
10	positive regulation of cellular process (GO:0048522)	0.03125	0.05285	19375.00	67148.74

Table S24. The top enriched molecular function for *ZFHX3*.

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	DNA-binding transcription repressor activity, RNA polymerase II-specific (GO:0001227)	0.01280	0.05120	19744.00	86050.65
2	RNA polymerase II cis-regulatory region sequence-specific DNA binding (GO:0000978)	0.05745	0.06795	18851.00	53854.37
3	cis-regulatory region sequence-specific DNA binding (GO:0000987)	0.05745	0.06795	18851.00	53854.37
4	RNA polymerase II transcription regulatory region sequence-specific DNA binding (GO:0000977)	0.06795	0.06795	18641.00	50125.40

Table S25. The top enriched cellular component for *ZFHX3*

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	nucleus (GO:0005634)	0.2242	0.2596	15516.00	23199.80
2	intracellular membrane-bounded organelle (GO:0043231)	0.2596	0.2596	14808.00	19970.28

Table S26. The top enriched Reactome pathway for DEGs

Index	Name	p-value	Adjusted p-value	Odds Ratio	Combined score
1	RUNX3 Regulates CDKN1A Transcription R-HSA-8941855	0.0003500	0.001750	19993.00	159096.45
2	Transcriptional Regulation By RUNX3 R-HSA-8878159	0.004750	0.01187	19905.00	106484.27
3	Generic Transcription Pathway R-HSA-212436	0.05950	0.07245	18810.00	53077.73
4	RNA Polymerase II Transcription R-HSA-73857	0.06560	0.07245	18688.00	50909.53
5	Gene Expression (Transcription) R-HSA-74160	0.07245	0.07245	18551.00	48693.8

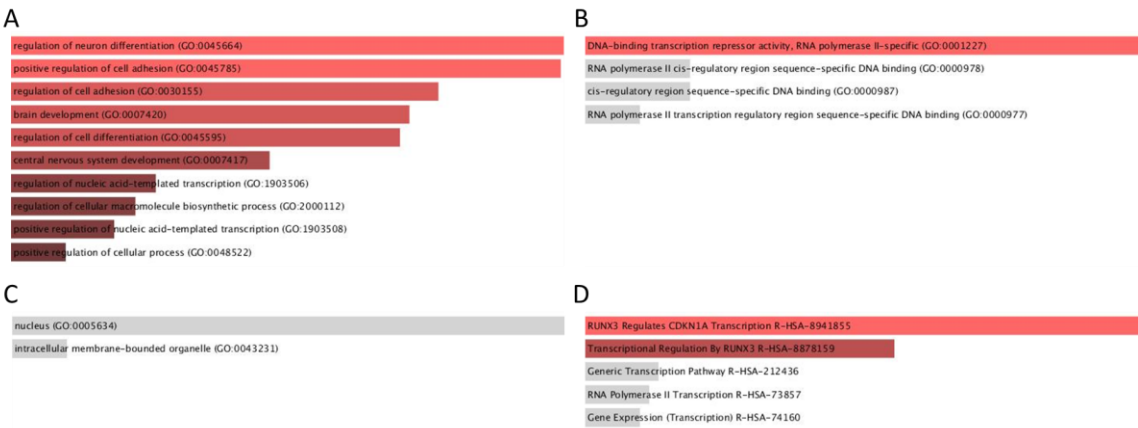


Figure S7. GO term and Reactome pathway enrichment analyses performed using Enrichr on ZFH3. (A) The top 10 enriched biological process. (B) The top enriched molecular function. (C) The top enriched cellular component. (D) The top enriched Reactome pathway.

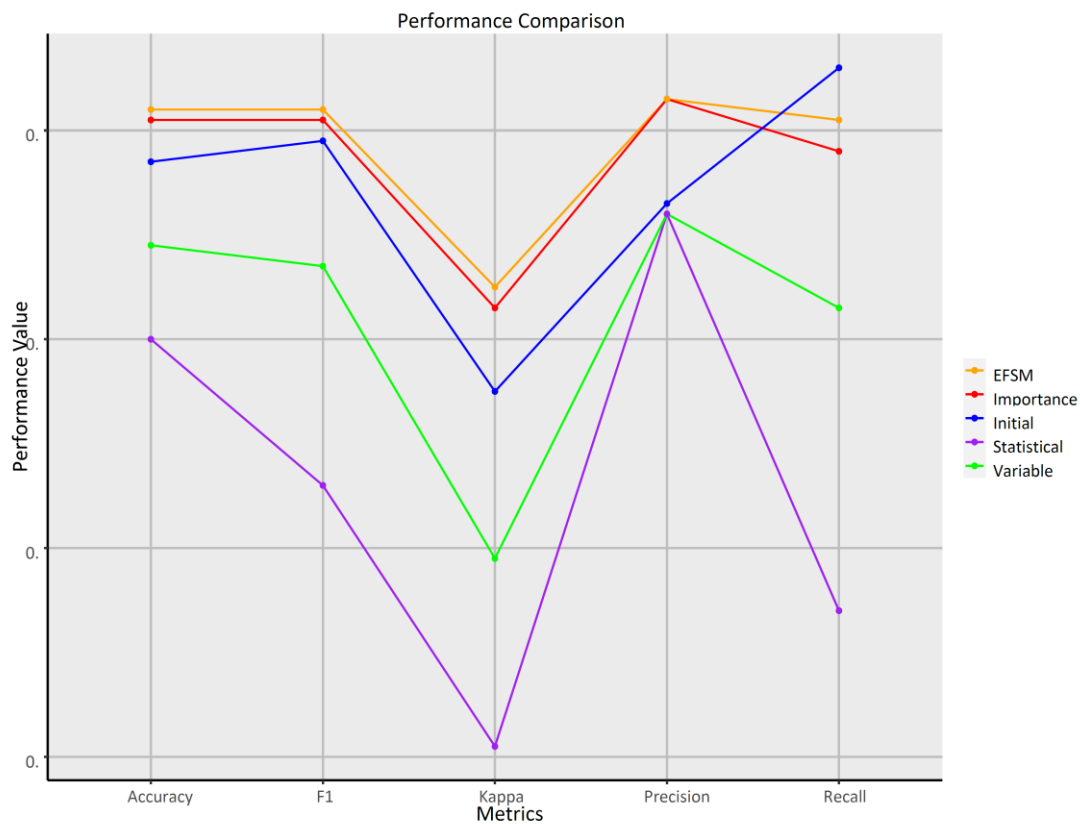


Figure S9. The results of a k-nearest neighbors (kNN) model. The x-axis labeled as "Metrics" and the y-axis labeled as "Performance Value". The different metrics being plotted include accuracy, F1 score, Kappa, Precision, and Recall. The plot utilizes different colors to differentiate between the various methods, with the intention of highlighting the superior performance of the EFSM method.

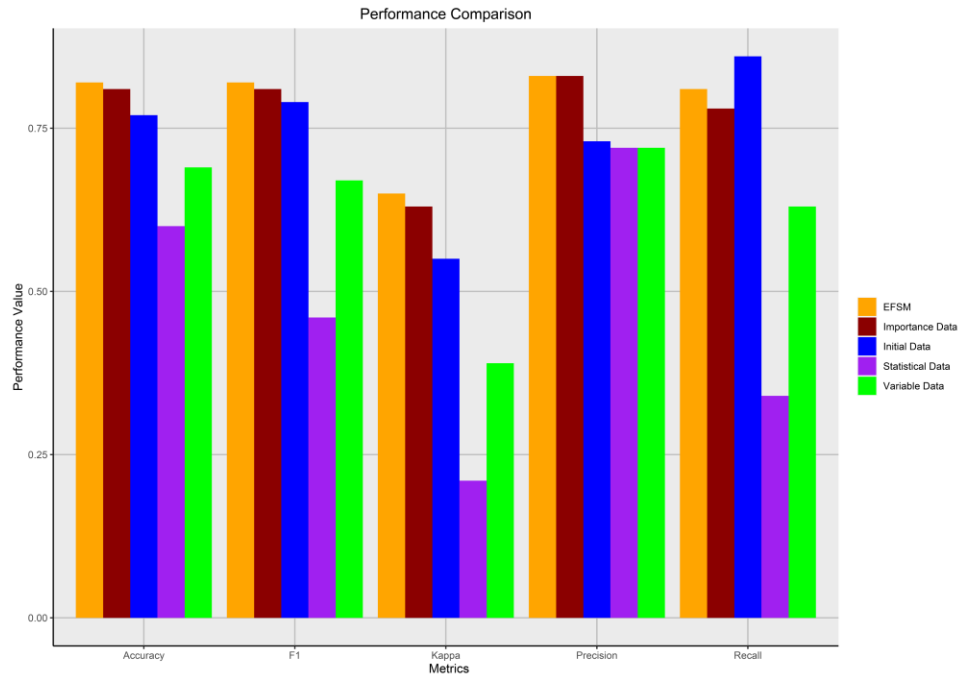


Figure S10. Performance of various methods in terms of accuracy, F1 score, Kappa, Precision, and Recall. The different colors represent different methods, allowing us to compare their performance side by side. The x-axis displays the different metrics, providing a clear visual representation of the effectiveness of each method. The y-axis shows the values of the metrics, with the highest values at the top.