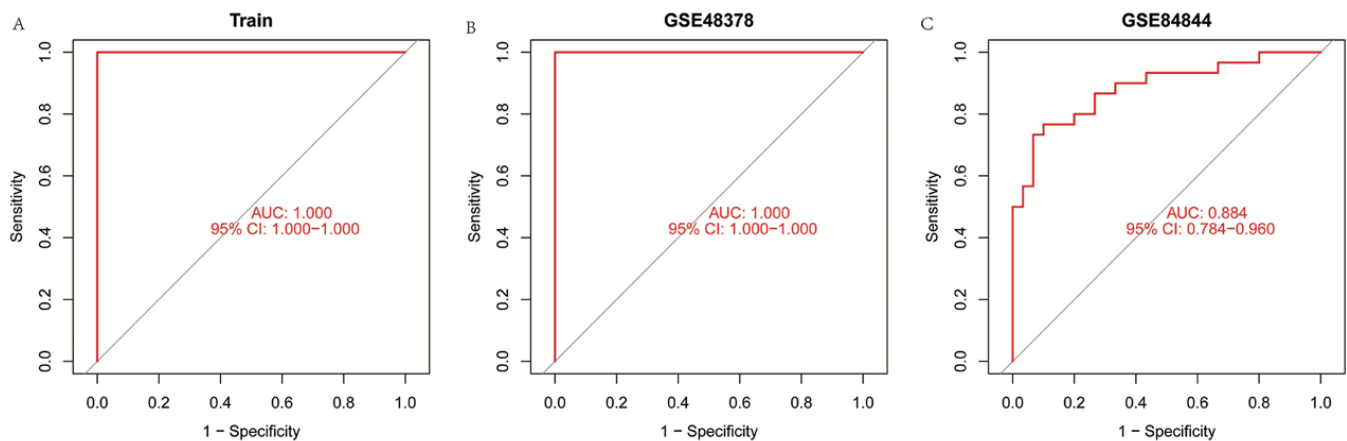
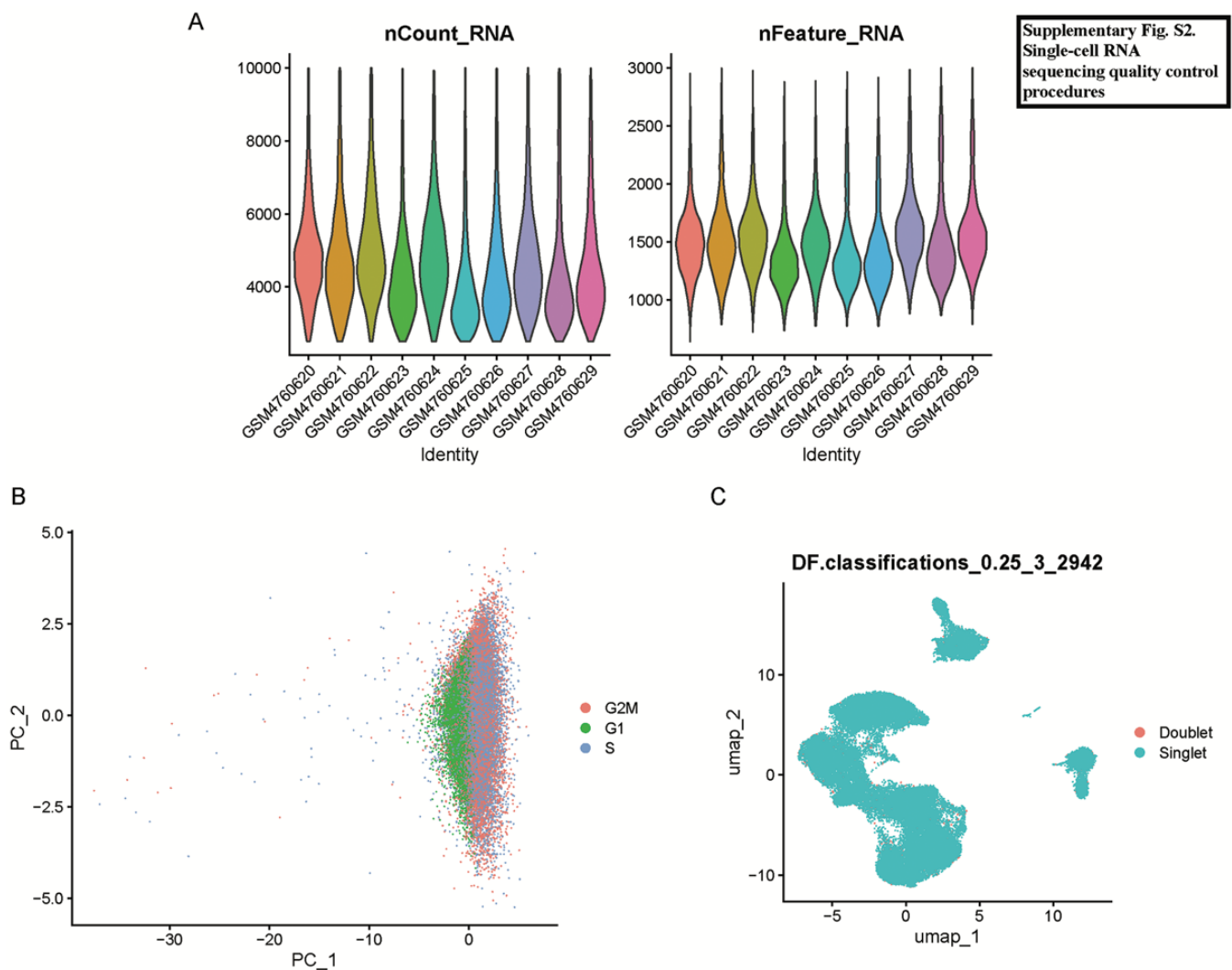




Supplementary Fig. S1. Random Forest model performance validation.



Supplementary Fig. S2. (A) The gene descriptions and total gene numbers for the profiles. (B) Cell cycle correction. (C) Removal of doublets in UMAP.

Supplementary Table S1. Dataset information and sample characteristics.

A summary of GEO datasets					
Dataset Type	GEO Number	Platform	Sjögren's disease	Healthy controls	Tissue-type
Train sets	GSE51092	GPL6884	190	32	Blood
	GSE66795	GPL10558	131	29	Blood
Validation sets	GSE48378	GPL5175	11	16	Blood
	GSE84844	GPL570	30	30	Blood

Supplementary Table S2. Differentially expressed genes in SjD.

id	logFC	AveExpr	t	P.Value	adj.P.Val	B
EPSTI1	1.760017	10.97143	11.03121	9.06E-25	9.07E-21	45.39839
IFI44	2.027869	10.03123	10.33886	2.85E-22	9.93E-19	39.78938
XAF1	1.417631	9.775846	10.3012	3.88E-22	9.93E-19	39.49001
PARP9	0.796362	9.45405	10.29844	3.97E-22	9.93E-19	39.46809
IFI44L	2.585577	10.4601	10.23001	6.92E-22	1.39E-18	38.92564
RSAD2	2.430908	9.426359	10.16302	1.19E-21	1.99E-18	38.39665
STAT2	0.692944	11.44229	10.01745	3.85E-21	5.50E-18	37.25396
GBP1	1.006587	9.759889	9.980779	5.16E-21	6.46E-18	36.96764
STAT1	0.638375	11.46842	9.796151	2.24E-20	2.50E-17	35.53526
EIF2AK2	1.018073	10.73713	9.766084	2.85E-20	2.85E-17	35.30349
PARP14	0.824768	9.760872	9.741064	3.47E-20	3.16E-17	35.11095
ISG15	1.880634	10.53759	9.664683	6.33E-20	5.28E-17	34.52499
LY6E	1.65092	11.21047	9.589979	1.14E-19	8.76E-17	33.95458
PARP12	0.836944	9.713414	9.449876	3.39E-19	2.42E-16	32.8921
IFIH1	0.896326	8.90688	9.363894	6.58E-19	4.39E-16	32.24482
IFIT1	1.639313	9.820379	9.346222	7.54E-19	4.72E-16	32.11224
LOC400759	1.098267	7.28095	9.186252	2.57E-18	1.43E-15	30.91925
IFITM3	1.374084	12.67301	9.16273	3.07E-18	1.54E-15	30.74493
IFI27	3.319316	9.815726	9.145319	3.51E-18	1.67E-15	30.61607
OAS1	1.332424	8.63655	9.066994	6.35E-18	2.89E-15	30.03837
IFI35	0.79358	10.0725	8.989938	1.14E-17	4.79E-15	29.47314
DDX58	0.757413	7.827583	8.988478	1.15E-17	4.79E-15	29.46246
IFIT2	1.049203	11.86813	8.975327	1.27E-17	5.07E-15	29.36631
LAP3	0.888488	9.368405	8.931748	1.76E-17	6.77E-15	29.04836
IFIT3	1.5035	9.698886	8.912903	2.02E-17	7.50E-15	28.91118
SERPING1	1.334161	7.20609	8.856312	3.08E-17	1.09E-14	28.50037
GBP	0.847418	10.15268	8.853766	3.14E-17	1.09E-14	28.48193
MX1	1.347592	12.54399	8.766868	5.99E-17	2.00E-14	27.85457
HERC6	0.94582	8.721721	8.742348	7.18E-17	2.32E-14	27.67829
PRIC285	0.738248	10.61041	8.723514	8.25E-17	2.58E-14	27.54311
BATF2	1.274256	6.726104	8.71334	8.90E-17	2.70E-14	27.47017
LAMP3	1.456391	6.113771	8.677776	1.16E-16	3.41E-14	27.21563
IRF7	0.796727	9.397474	8.653786	1.38E-16	3.94E-14	27.04433
CEACAM1	0.627646	8.348458	8.599947	2.04E-16	5.69E-14	26.66103
OAS2	1.192864	8.020956	8.587942	2.23E-16	6.04E-14	26.57579
RTP4	1.030196	7.39906	8.563654	2.66E-16	7.02E-14	26.40356
SCO2	0.699049	9.79822	8.45621	5.81E-16	1.42E-13	25.64564
HES4	1.355457	8.509939	8.43479	6.78E-16	1.62E-13	25.49532
FBXO6	0.66648	8.444892	8.300468	1.78E-15	3.96E-13	24.55864
OASL	1.27383	8.228607	8.023651	1.26E-14	2.67E-12	22.66134
USP18	1.187404	5.057521	7.998366	1.50E-14	3.12E-12	22.4903
MT1A	0.665701	10.38262	7.979515	1.71E-14	3.49E-12	22.36303
TTC21A	0.797179	5.566892	7.748002	8.39E-14	1.62E-11	20.81755
MS4A4A	0.722674	5.932627	7.705965	1.12E-13	2.07E-11	20.54047
IFI6	1.250298	10.64295	7.664056	1.48E-13	2.70E-11	20.26531
MT2A	0.748107	10.79601	7.632276	1.84E-13	3.23E-11	20.05739
FEZ1	-0.68273	4.783738	-7.59369	2.38E-13	4.11E-11	19.80581
CXCL10	1.160179	6.241615	7.59027	2.44E-13	4.14E-11	19.78354
ATF3	0.809456	5.239162	7.313875	1.52E-12	2.21E-10	18.0096
MMP28	-0.75465	4.391199	-7.17646	3.71E-12	4.89E-10	17.14604

Continued

id	logFC	AveExpr	t	P.Value	adj.P.Val	B
LOC389816	-0.75611	5.218743	-7.13751	4.77E-12	6.05E-10	16.90353
CCR1	0.586799	8.859799	7.134414	4.86E-12	6.09E-10	16.8843
ETV7	0.930812	4.88188	7.127316	5.09E-12	6.30E-10	16.84023
CCL8	0.891288	4.52859	7.062223	7.72E-12	8.99E-10	16.43764
PLSCR1	0.784539	7.37419	6.958851	1.49E-11	1.64E-09	15.80413
TNFAIP6	0.62393	9.30215	6.923989	1.85E-11	1.95E-09	15.5921
OTOF	1.392087	4.714488	6.87978	2.44E-11	2.49E-09	15.32442
TIMM10	0.904019	8.569732	6.836737	3.19E-11	3.14E-09	15.06508
LOC644532	-0.68744	4.301911	-6.78843	4.31E-11	4.11E-09	14.77555
SLC16A10	-0.63452	5.578325	-6.61953	1.21E-10	1.08E-08	13.77588
ZCCHC2	0.831017	4.674956	6.524508	2.15E-10	1.80E-08	13.22222
ANKRD22	0.859468	6.939332	6.365998	5.53E-10	4.13E-08	12.31293
CCL2	0.976336	4.579503	6.354033	5.94E-10	4.40E-08	12.24502
ZNF684	0.66131	5.515881	6.346003	6.22E-10	4.52E-08	12.19951
TOP2A	0.67713	4.533242	6.226839	1.25E-09	8.54E-08	11.52953
USP41	0.894563	4.379862	6.172048	1.71E-09	1.14E-07	11.22495
LGALS3BP	0.69874	6.928774	5.928981	6.79E-09	3.74E-07	9.900443
DSC1	-0.71468	4.718011	-5.91633	7.28E-09	3.94E-07	9.832687
HLA-DRB6	0.690713	10.18741	5.80619	1.34E-08	6.84E-07	9.24812
APOBEC3A	0.75286	8.48507	5.778303	1.56E-08	7.81E-07	9.101561
C9orf109	0.610244	4.716669	5.684909	2.59E-08	1.22E-06	8.615043
CD248	-0.86972	5.280291	-5.61497	3.77E-08	1.70E-06	8.255062
LOC650761	-0.68959	5.30006	-5.55141	5.29E-08	2.26E-06	7.931227
AGRN	0.605073	4.61668	5.516242	6.37E-08	2.70E-06	7.753388
FCGR1A	0.635817	7.215048	5.505643	6.74E-08	2.82E-06	7.699975
PDGFRB	0.668138	5.406526	5.489298	7.34E-08	3.04E-06	7.617783
RBPM52	-0.63185	6.426872	-5.17428	3.69E-0	1.24E-05	6.074712
PTGDS	-0.59786	7.195321	-5.07702	5.99E-07	1.90E-05	5.614213
MYOM2	-1.07938	6.89682	-4.96205	1.05E-06	3.12E-05	5.079756
KIAA1324	-0.73072	7.972861	-4.85796	1.73E-06	4.68E-05	4.605168
LRRN3	-0.77865	6.72177	-4.753	2.84E-06	7.07E-05	4.135584
MGC52282	-0.65634	5.316922	-4.73519	3.08E-06	7.58E-05	4.056819
IND	0.627564	6.242278	4.366697	1.62E-05	0.000289	2.486385
GSTT1	-0.63188	5.89635	-3.11901	0.001951	0.013052	-1.95764
HDC	-0.62215	6.789636	-2.90587	0.003874	0.021908	-2.57699

Supplementary Table S3. RF machine learning identification of 26 hub genes.

features	algorithm
CD248	RF
MMP28	RF
STAT2	RF
RBPM52	RF
PARP9	RF
TOP2A	RF
XAF1	RF
EPSTI1	RF
KIAA1324	RF
FEZ1	RF
LRRN3	RF
DSC1	RF
IFI44	RF
ZNF684	RF
SLC16A10	RF
EIF2AK2	RF
GBP1	RF
LY6E	RF
MYOM2	RF
HES4	RF
DDX58	RF
STAT1	RF
GBP5	RF
MS4A4A	RF
IFIH1	RF
PARP14	RF

Supplementary Table S4. Gene Ontology enrichment results.

ONTOLOGY	ID	Description	Gene	Bg Ratio	Rich Ratio	Fold Factor	Score Enrichment	pvalue	p.adjust	qvalue	geneID	Count
BP	GO:0009615	response to virus	8/22	425/18805	0.01882333	16.0898396	10.7686266	1.5442E-08	9.1724E-06	6.388E-06	STAT2/PARP9/IFI44/EIF2AK2/GBP1/STAT1/GBP5/IFIH1	8
BP	GO:0051607	defense response to virus	7/22	321/18805	0.02180685	18.6390337	10.909497	5.4144E-08	1.6081E-05	1.1199E-05	STAT2/PARP9/EIF2AK2/GBP1/STAT1/GBP5/IFIH1	7
BP	GO:0071346	cellular response to type II interferon	5/22	98/18805	0.05102041	43.6108534	14.4738871	8.5118E-08	1.6853E-05	1.1737	E-05/PARP9/GBP1/STAT1/GBP5/PARP14	5
BP	GO:0140888	interferon-mediated signaling pathway	5/22	112/18805	0.04464286	38.1594968	13.4987426	1.6638E-07	2.4707E-05	1.7207E-05	STAT2/PARP9/STAT1/IFIH1/PARP14	5
BP	GO:0034341	response to type II interferon	5/22	121/18805	0.04132231	35.3211871	12.9620603	2.4487E-07	2.909E-05	2.026E-05	PARP9/GBP1/STAT1/GBP5/PARP14.5	5
BP	GO:0060333	type II interferon-mediated signaling pathway	3/22	27/18805	0.11111111	94.9747475	16.7233148	3.9916E-06	0.00039517	0.00027521	PARP9/STAT1/PARP143	5
BP	GO:0045088	regulation of innate immune response	6/22	482/18805	0.02244813	10.6403244	7.33788429	1.4463E-05	0.00110835	0.0007719	STAT2/PARP9/EIF2AK2/GBP5/IFIH1/PARP14.6	6
BP	GO:0050792	regulation of viral process	4/22	142/18805	0.028169012	4.078105	9.44729706	2.0509E-05	0.00135361	0.00094271	TOP2A/EIF2AK2/STAT1/IFIH1.4	4
BP	GO:0048525	negative regulation of viral process	3/22	76/18805	0.03947368	33.7410287	9.78804155	9.2437E-05	0.00392198	0.00273141	EIF2AK2/STAT1/IFIH1.3	3
BP	GO:0016032	viral process	5/22	421/18805	0.011870481	0.1516951	6.4994494	0.00010558	0.000278474	0.00278474	PARP9/TOP2A/EIF2AK2/STAT1/IFIH1.5	5
CC	GO:0032982	myosin filament	1/23	24/19880	0.041666673	6.0144928	5.84133273	0.02741594	0.2621071	0.2323387	MYOM21	1
CC	GO:0031430	M band	1/23	28/19880	0.035714293	0.8695652	5.38282363	0.03191479	0.2621071	0.2323387	MYOM21	1
CC	GO:0005921	gap junction	1/23	30/19880	0.033333332	8.8115942	5.18812729	0.03415674	0.2621071	0.2323387	DSC11	1
CC	GO:0030057	desmosome	1/23	30/19880	0.033333332	8.8115942	5.18812729	0.03415674	0.2621071	0.2323387	DSC11	1
CC	GO:0090575	RNA polymerase II transcription regulator complex	2/23	254/19880	0.00787402	6.80588839	3.1693778	0.03448778	0.2621071	0.2323387	STAT2/STAT12	2
CC	GO:0031672	A band	1/23	41/19880	0.024390242	1.0816543	4.38062314	0.04639898	0.28812366	0.25540048	MYOM21	1
CC	GO:0001673	male germ cell nucleus	1/23	53/19880	0.01886792	6.3084495	3.79791295	0.05958483	0.28812366	0.25540048	TOP2A1	1
CC	GO:0001533	cornified envelope	1/23	60/19880	0.01666667	14.4057971	3.5932798	0.06719585	0.28812366	0.25540048	DSC11	1
CC	GO:0101003	ficollin-1-rich granule membrane	1/23	61/19880	0.01639344	14.1696365	3.50592163	0.06827832	0.28812366	0.25540048	DSC11	1
CC	GO:0043073	germ cell nucleus	1/23	68/19880	0.01470588	12.7109974	3.29221773	0.07582202	0.28812366	0.25540048	TOP2A1	1
MF	GO:0003950	NAD+ poly-ADP-ribosyltransferase activity	2/24	23/18639	0.086956526	7.5326087	11.4638116	0.00039542	0.02827274	0.0228929	PARP9/PARP14.2	2
MF	GO:1990404	NAD+ protein mono-ADP-ribosyltransferase activity	2/24	23/18639	0.08695652	67.5326087	11.4638116	0.00039542	0.02827274	0.0228929	PARP9/PARP142	2
MF	GO:0016763	pentosyltransferase activity	2/24	49/18639	0.04081633	31.6989796	7.72602376	0.00180085	0.04828082	0.03909378	PARP9/PARP142	2
MF	GO:0005080	protein kinase C binding	2/24	52/18639	0.03846154	29.8701923	7.48549377	0.00202577	0.04828082	0.03909378	TOP2A/FEZ12	2
MF	GO:0003725	double-stranded RNA binding	2/24	75/18639	0.02666667	20.71	6.04122813	0.00416365	0.08505734	0.06887234	EIF2AK2/IFIH12	2
MF	GO:0044389	ubiquitin-like protein ligase binding	3/24	322/18639	0.00931677	7.23563665	4.05280136	0.00790092	0.13716392	0.1110639	STAT2/PARP9/STAT1.3	3
MF	GO:0004382	GDP phosphatase activity	1/24	10/18639	0.1	77.6625	8.706868	0.01280495	0.13716392	0.1110639	GBP11	1
MF	GO:0016779	nucleotidyltransferase activity	2/24	136/18639	0.01470588	11.42095529	4.37955204	0.01313184	0.13716392	0.1110639	PARP9/PARP14.2	2
MF	GO:0015173	aromatic amino acid transmembrane transporter activity	1/24	11/18639	0.09090909	70.6022727	8.29106659	0.01407677	0.13716392	0.1110639	SLC16A101	1
MF	GO:0097675	TAT family protein binding	1/24	12/18639	0.08333333	64.71875	7.92793658	0.01534701	0.13716392	0.1110639	PARP91	1

Supplementary Table S5. KEGG pathway enrichment results.

category	ID	Description	GeneRatio	BgRatio	RichFactor	FoldEnrich	zScore	pvalue	p.adjust	qvaluegene	ID	Count
Human Diseases	hsa05162	Measles	4/11	139/8541	0.028777	22.34402	9.110516	2.03E-05	0.000611	0.000339	STAT2/EIF2AK2/STAT1/IFIH1	4
Human Diseases	hsa05164	Influenza A	4/11	173/8541	0.023121	17.95271	8.089134	4.80E-05	0.000611	0.000339	STAT2/EIF2AK2/STAT1/IFIH1	4
Human Diseases	hsa05168	Herpes simplex virus 1 infection	4/11	182/8541	0.021978	17.06494	7.866622	5.86E-05	0.000611	0.000339	STAT2/EIF2AK2/STAT1/IFIH1	4
Organismal Systems	hsa04621	NOD-like receptor signaling pathway	4/11	189/8541	0.021164	16.4329	7.704314	6.79E-05	0.000611	0.000339	STAT2/GBP1/STAT1/GBP5	4
Human Diseases	hsa05171	Coronavirus disease - COVID-19	4/11	238/8541	0.016807	13.04966	6.77012	0.000166	0.001197	0.000665	STAT2/EIF2AK2/STAT1/IFIH1	4
Human Diseases	hsa05165	Human papillomavirus infection	4/11	333/8541	0.012012	9.326781	5.565848	0.00602	0.003614	0.002008	STAT2/EIF2AK2/HES4/STAT1	4
Cellular Processes	hsa04217	Necroptosis	3/11	159/8541	0.018868	14.65009	6.238941	0.00936	0.004024	0.002236	STAT2/EIF2AK2/STAT1	3
Human Diseases	hsa05160	Hepatitis C	3/11	159/8541	0.018868	14.65009	6.238941	0.00936	0.004024	0.002236	STAT2/EIF2AK2/STAT1	3
Human Diseases	hsa05161	Hepatitis B	3/11	163/8541	0.018405	14.29057	6.152026	0.011006	0.004024	0.002236	STAT2/STAT1/IFIH1	3
Human Diseases	hsa05167	Kaposi sarcoma-associated herpesvirus infection	3/11	196/8541	0.015306	11.88451	5.535724	0.001714	0.006171	0.003428	STAT2/EIF2AK2/STAT1	3
Human Diseases	hsa05169	Epstein-Barr virus infection	3/11	204/8541	0.014706	11.41845	5.40834	0.001923	0.006293	0.004024	STAT2/EIF2AK2/STAT1	3
Organismal Systems	hsa04625	C-type lectin receptor signaling pathway	2/11	105/8541	0.019048	14.78961	5.105392	0.007659	0.0228	0.012667	STAT2/STAT1	2
Organismal Systems	hsa04620	Toll-like receptor signaling pathway	2/11	109/8541	0.018349	14.24687	4.998182	0.008234	0.0228	0.012667	STAT2/STAT1	2
Organismal Systems	hsa04919	Thyroid hormone signaling pathway	2/11	122/8541	0.016393	12.72876	4.685464	0.010231	0.026308	0.014616	SLC16A10/STAT1	2
Organismal Systems	hsa04380	Osteoclast differentiation	2/11	143/8541	0.013986	10.8595	4.269588	0.013867	0.033281	0.018489	STAT2/STAT1	2
Environmental Information Processing	hsa04630	JAK-STAT signaling pathway	2/11	168/8541	0.011905	9.243506	3.875046	0.018826	0.042358	0.023532	STAT2/STAT1	2
Organismal Systems	hsa04062	Chemokine signaling pathway	2/11	193/8541	0.010363	8.046161	3.555419	0.024432	0.0517390	.028744	STAT2/STAT1	12