A Showcase of HIVE RNAseq Pipeline - Identifying Molecular Features of Neural Stem Cells in Varied Differentiation Stages

Sydney Fenstermaker[1], Ge Ma[2], Viswanadham Sridhara[1], Wei-Lun Alterovitz[1], Ilya Mazo[1], Luis Santana-Quintero[1], Brent McCright[2]
[1]Food and Drug Administration, Center for Biologics Evaluation and Research, OBE, HIVE
[2]Food and Drug Administration, Center for Biologics Evaluation and Research, OTAT, DCGT, CTTB

Introduction

Neural stem cells (NSCs) are being developed as cellular therapies aimed at treating neurodegenerative diseases. In this study, we evaluated the significance of Notch expression in Notch2+ and Notch2- NSCs using HIVE RNA-seq platform. HIVE platform allows an easy-to-use user interface to run the RNA-seq pipeline using multiple compute and data nodes.

NSCs isolated from fetal and pluripotent stem cell sources have been proposed for use in the treatment of Parkinson's disease, Alzheimer's disease, and spinal cord injury. However, culture and expansion of these cells can result in heterogenous cells of unknown differentiation status. Biomarkers that predict the potential of NSCs to provide clinical benefit are needed to advance the development of neural stem cell products. The purpose of this study is to show the uses of the HIVE RNASeq platform and how it enhances the analysis of RNA Sequencing.

Methodology

Preprocessing of the data for adapter trimming and removing low complexity regions was done by Fastp[1] (version 0.20.0). Preprocessed reads were aligned to the human reference genome (hg38) using Hisat2[2] (v2.2.0). The alignments were then used with featureCounts[3] (v2.0.0) to quantitate gene expression. 3 samples of each of Notch+, Notch- and control were pooled into different groups for differential expression analyses using DeSeq2[4]. The top differentially expressed genes were used in Ingenuity Pathway Analyses[5] (IPA) to identify upstream regulators, downstream effectors and the pathways that were enriched with the differentially expressed genes.

1A. Samples Used in Analysis

Ctrl Samples	AB- Samples	AB+ Samples
control 1	notch - (1)	notch + (1)
control 2	notch - (2)	notch + (2)
control 3	notch - (3)	notch + (3)

1B.

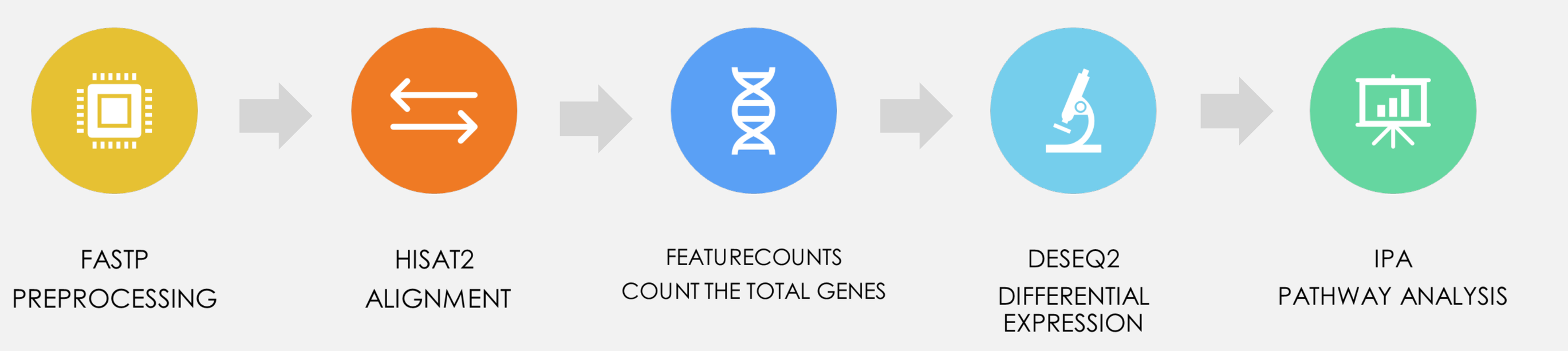


FIGURE 1. Pipeline used for HIVE RNASeq Analysis

A. Table containing all the samples used for the analysis for the control, notch+, and notch -.

B. The images above show the five basic steps when performing HIVE RNASeq analysis. The first step illustrated is using Fastp[1] for preprocessing of the data followed by the second step, show as the next bubble, which is performing the alignment using Hisat2[2]. The third step involved in this pipeline and illustrated next is using featureCounts[3] in order to compile the alignments and count the total number of genes expressed as well as the generation of volcano plots. The fourth step illustrated is using DeSeq2[4] for differential expression and the fifth step illustrated by the final bubble is using IPA[5] for pathway analysis.

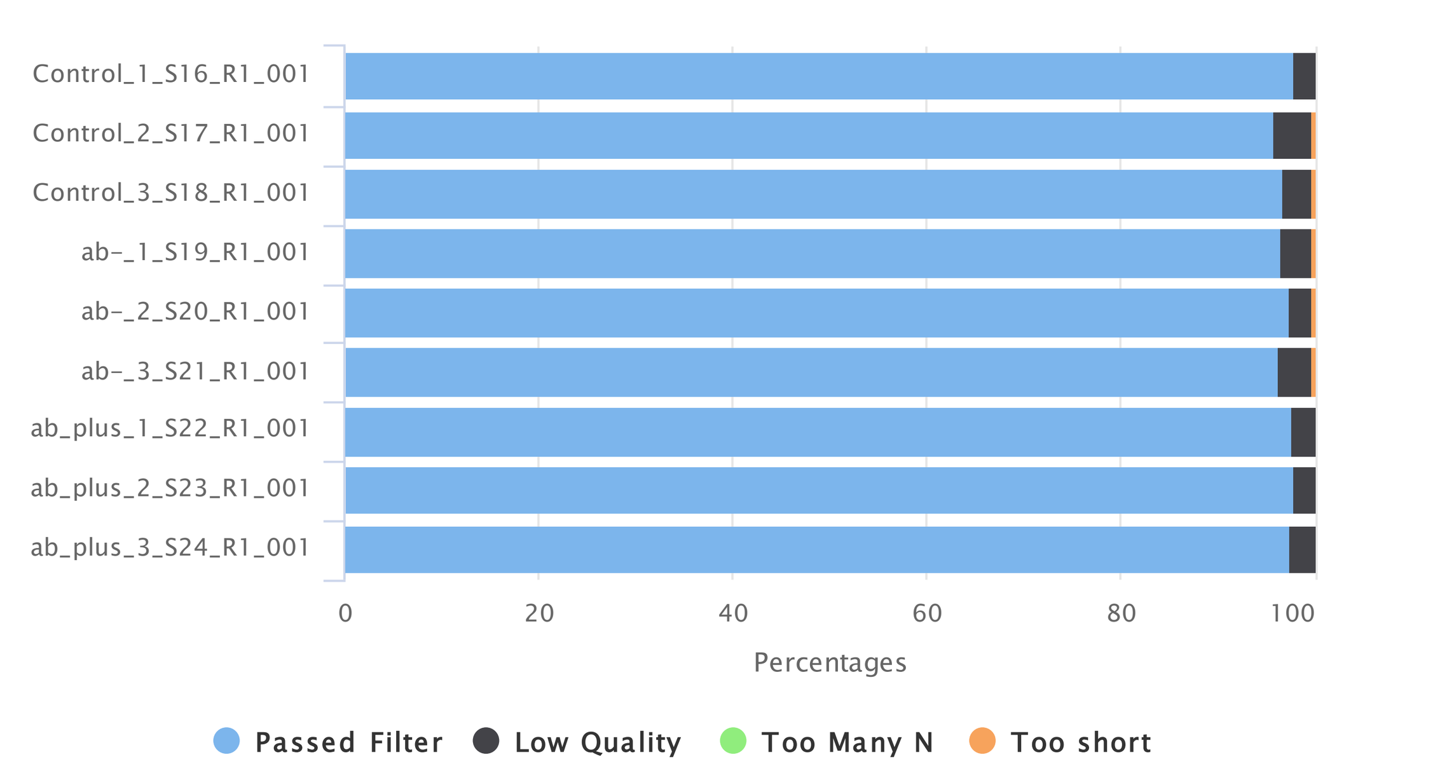
Results

Th2 pathway appeared to be impacted the most significantly from our IPA results. This project demonstrated the benefit of HIVE RNAseq when performing bioinformatic analysis with in-depth results in a clear and visual way. The HIVE platform constantly updates and follows the best practices in bioinformatics analysis, advancing the field on bioinformatics with new pipelines available for researchers and reviewers to advance on protecting the public health. Using HIVE RNA-seq pipeline, we identified differentially expressed genes between Notch2+ and Notch2- cells along with identifying the pathways that are up and down regulated.

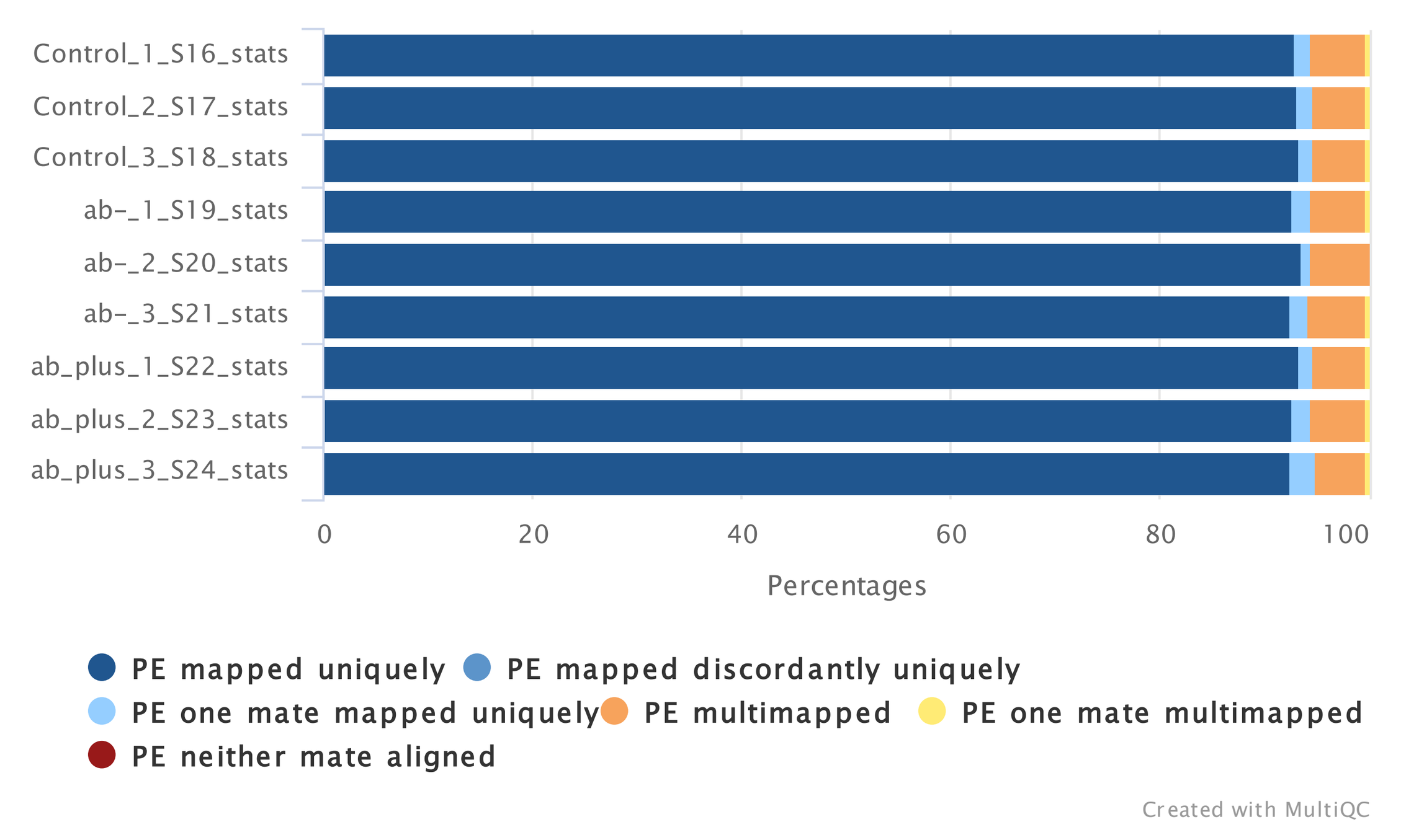
FIGURE 2. Detailed Results from the Pipeline used for HIVE RNASeq Analysis

A. Plot from FastP output using MultiQC showing the percentage of reads that pass the quality check.
B. Plot from Hisat2 generated using MultiQC showing the alignment scores of the samples.
C. Heatmap of the top 30 genes for each pairwise comparison for the 9 total samples (i.e. Control vs AB+, Control vs AB-, and AB+ vs AB-), so there are the non-redundant genes from the 90 genes of the comparisons shown, and samples are represented on the bottom.
D. Volcano plot generated using DESeq2 for the pairwise comparison of AB+ vs AB-. The log2 fold-change is > 1.5 and the adjusted P-value is <0.01.
E. Figure showing the Th2 pathway for AB+ vs AB- generated using IPA. Under expressed genes in AB+ vs AB- are highlighted in green and overexpressed genes in red.
F. Table associated with Figure E showing the molecules associated with the Th2 pathway and gene annotations (Z-score is -1.604 with 20 out of 136 molecules associated, P-value is 2.230E-1). Done on IPA.
G. Table result of the comparison analysis for the Th2 pathway for Ctrl vs AB+ and Ctrl vs AB-. Done on IPA.
H. Figure showing molecules and genes associated with advanced extracranial solid tumor, taken from IPA.

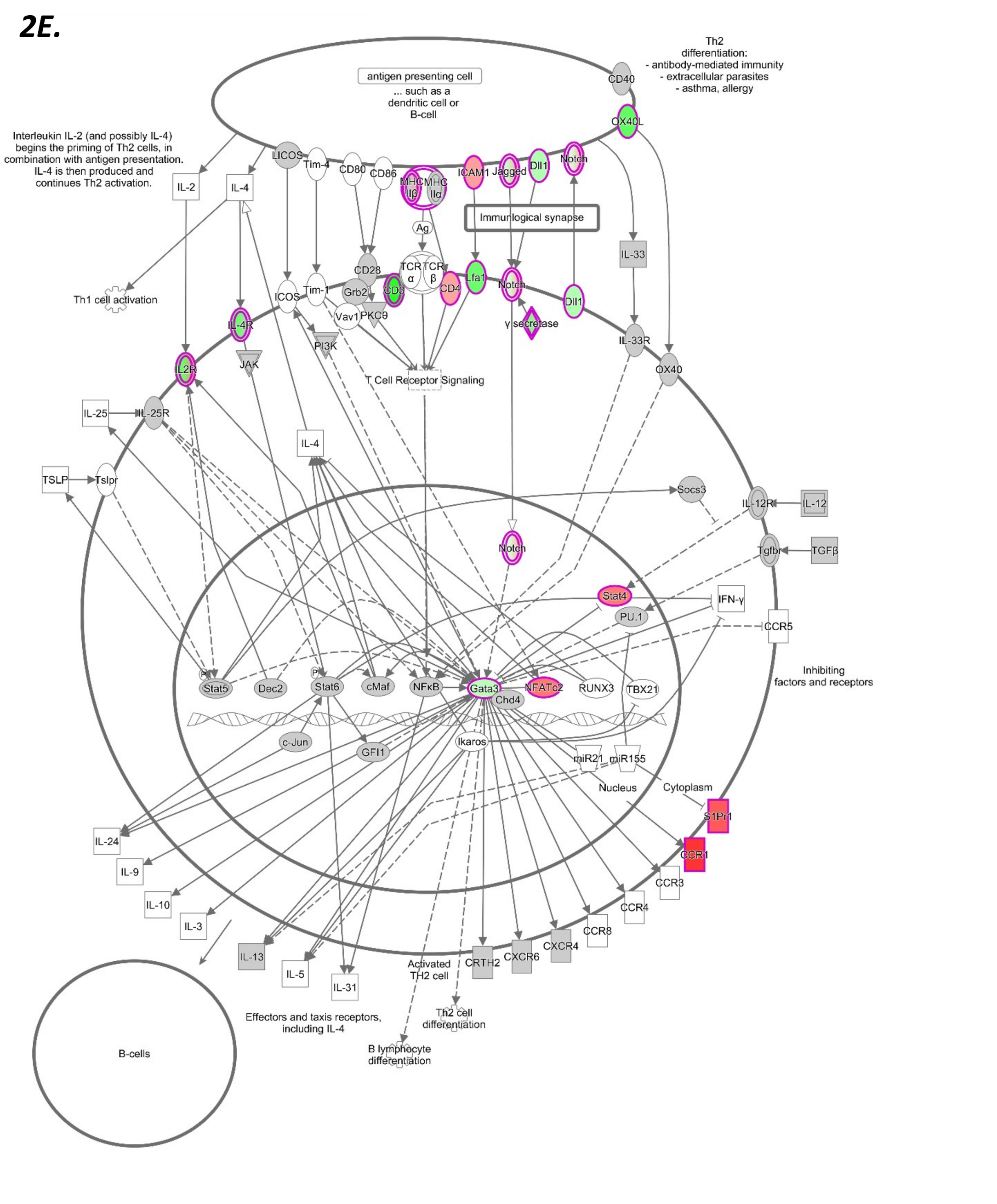
2A.



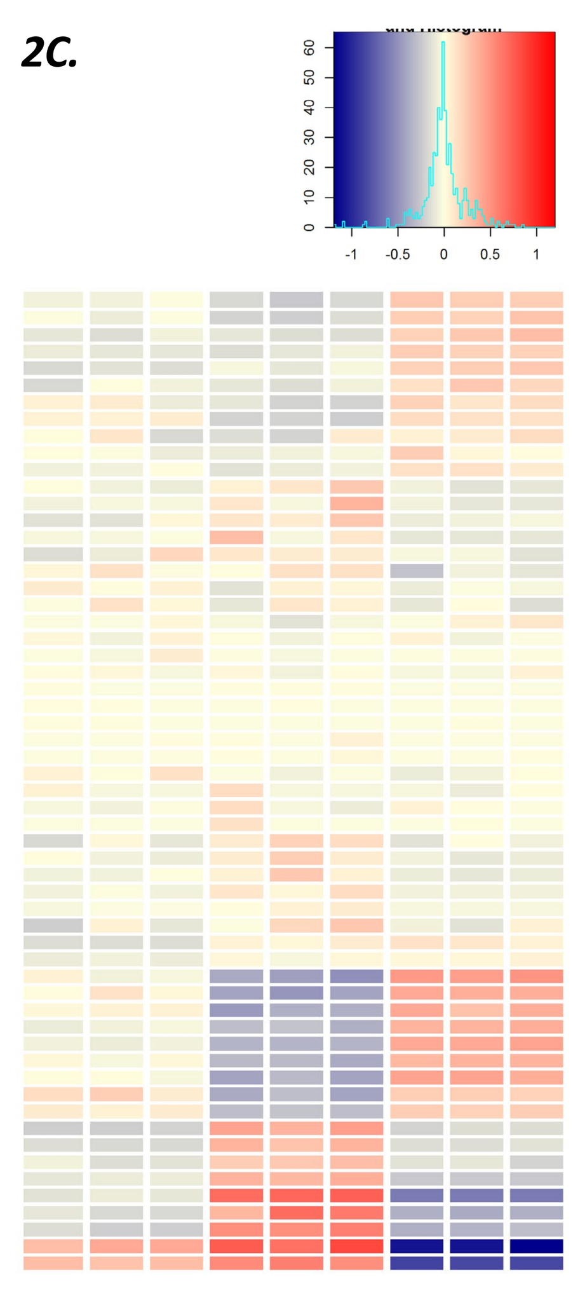
2B.



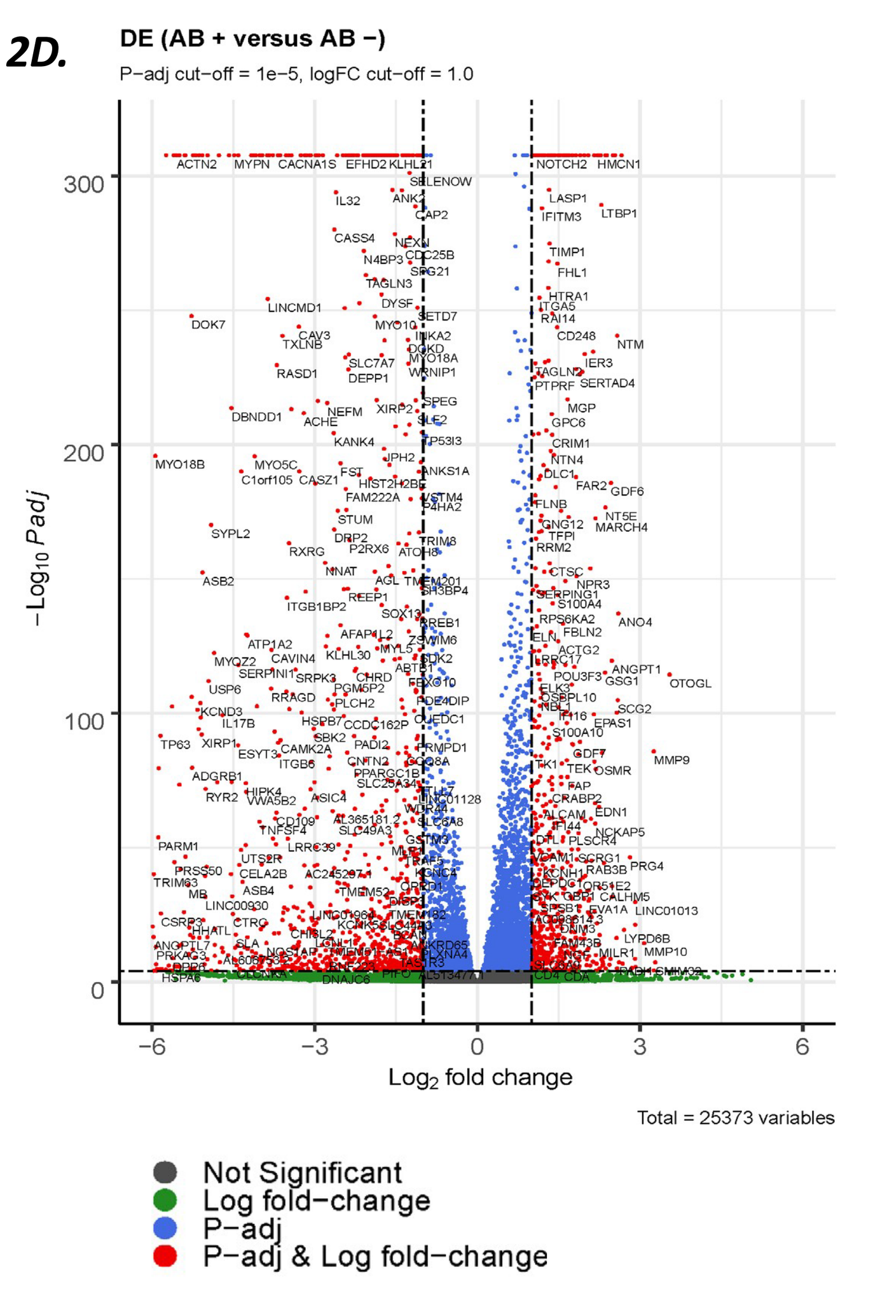
2E.



2C.



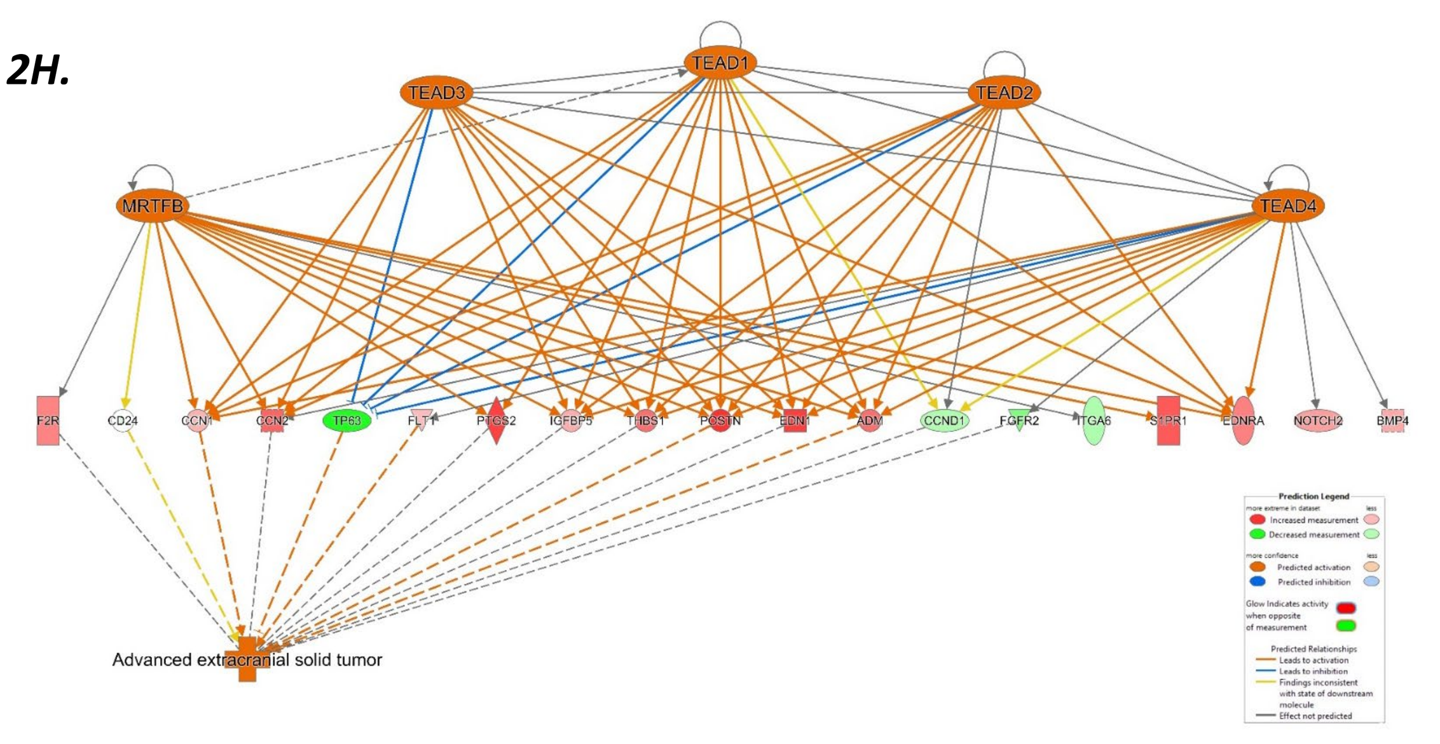
2D.



Conclusion

Tables *2F* and *2G* signify that the comparison results and the expression results agree, signifying that AB+ shows under expression for the genes involved in the Th2 pathway compared to AB- as well as the control and AB- shows overexpression compared to the control as well as AB+. The final analysis of interest was the third top regulator effect network, which was "advanced extracranial solid tumor". This network pathway showed predicted activation of this pathway with high confidence. The pathway is shown in figure *2H*, showing the molecules involved as well as the relationship between activation state. With use of the RNASeq pipeline in HIVE, bioinformatic analysis was made in a clear in visual way.

2H.



2F. Th2 Pathway for AB+ vs AB-

Symbol	Entrez Gene Name	Expr Log Ratio	Expr *P*-value	Expr False Discovery Rate (q-value)	Expected	Location	Type(s)	Biomarker Application(s)
CCR1	C-C motif chemokine receptor 1	4.129	0.000309	0.00111	Up	Plasma Membrane	G-protein coupled receptor	
S1PR1	sphingosine-1-phosphate receptor 1	1.873	2.89E-10	1.81E-09	Down	Plasma Membrane	G-protein coupled receptor	
NFATC2	nuclear factor of activated T cells 2	1.555	5.22E-13	3.87E-12		Nucleus	transcription regulator	
STAT4	signal transducer and activator of transcription 4	1.518	0.00152	0.00486	Down	Nucleus	transcription regulator	
IL2RB	interleukin 2 receptor subunit beta	1.388	0.00786	0.022	Up	Plasma Membrane	transmembrane receptor	efficacy
HLA-B	major histocompatibility complex, class I, B	1.15	3.91E-51	8.57E-50		Plasma Membrane	transmembrane receptor	safety
ICAM1	intercellular adhesion molecule 1	1.089	0.00000095	0.00000455		Plasma Membrane	transmembrane receptor	diagnosis,efficacy,prognosis,unspecified application
NOTCH2	notch receptor 2	1.088	0	0	Up	Plasma Membrane	transcription regulator	unspecified application
HLA-DPB1	major histocompatibility complex, class II, DP beta 1	1.084	0.0283	0.069		Plasma Membrane	transmembrane receptor	
CD4	CD4 molecule	1.053	0.000000319	0.0000016		Plasma Membrane	transmembrane receptor	diagnosis,efficacy,unspecified application
JAG1	jagged canonical Notch ligand 1	0.991	9.09E-81	2.93E-79	Up	Extracellular Space	growth factor	efficacy
NOTCH1	notch receptor 1	-0.835	4.59E-134	2.47E-132	Up	Plasma Membrane	transcription regulator	diagnosis,efficacy
JAG2	jagged canonical Notch ligand 2	-1.299	3.8E-142	2.14E-140	Up	Extracellular Space	growth factor	
GATA3	GATA binding protein 3	-1.746	0.0372	0.0874	Up	Nucleus	transcription regulator	diagnosis,unspecified application
DLL1	delta like canonical Notch ligand 1	-1.948	0	0	Up	Plasma Membrane	enzyme	
PSEN2	presenilin 2	-3.151	0	0	Up	Cytoplasm	peptidase	diagnosis
ITGB2	integrin subunit beta 2	-3.728	6.97E-63	1.81E-61		Plasma Membrane	transmembrane receptor	
IL2RG	interleukin 2 receptor subunit gamma	-3.769	0.0000191	0.0000803	Up	Plasma Membrane	transmembrane receptor	unspecified application
TNFSF4	TNF superfamily member 4	-4.017	1.06E-61	2.72E-60	Up	Extracellular Space	cytokine	
CD247	CD247 molecule	-5.559	0.000000369	0.00000183	Up	Plasma Membrane	transmembrane receptor	

2G. Comparison Analysis Th2 Canonical Pathway

		Ctrl vs AB-			Ctrl vs AB+				
Symbol	**Entrez Gene Name**	**Expr Log Ratio**	**Expr *P*-value**	**Expr False Discovery Rate (q-value)**	**Expr Log Ratio**	**Expr *P*-value**	**Expr False Discovery Rate (q-value)**	**Location**	**Type(s)**
CD247	CD247 molecule	0.597	0.187	0.34	-4.961	0.00000659	0.0000449	Plasma Membrane	transmembrane receptor
TNFSF4	TNF superfamily member 4	0.337	0.0188	0.0516	-3.68	4.01E-51	2.16E-49	Extracellular Space	cytokine
IL2RG	interleukin 2 receptor subunit gamma	1.164	0.0576	0.134	-2.605	0.0045	0.0178	Plasma Membrane	transmembrane receptor
ITGB2	integrin subunit beta 2	1.48	2.87E-18	4.24E-17	-2.248	7.94E-22	1.8E-20	Plasma Membrane	transmembrane receptor
PSEN2	presenilin 2	0.981	1.05E-94	1.03E-92	-2.17	1.44E-284	4.79E-282	Cytoplasm	peptidase
HLA-DOA	major histocompatibility complex, class II, DO alpha	-1.306	0.00147	0.00536	-1.413	0.000473	0.00235	Plasma Membrane	transmembrane receptor
HLA-DPA1	major histocompatibility complex, class II, DP alpha 1	-1.833	0.00306	0.0104	-1.128	0.0436	0.125	Plasma Membrane	transmembrane receptor
CD40	CD40 molecule	-1.458	0.0119	0.0349	-0.932	0.0837	0.21	Plasma Membrane	transmembrane receptor
DLL1	delta like canonical Notch ligand 1	1.186	4.82E-155	1.03E-152	-0.762	1.25E-52	7.15E-51	Plasma Membrane	enzyme
HLA-DPB1	major histocompatibility complex, class II, DP beta 1	-1.731	0.000382	0.00157	-0.647	0.124	0.285	Plasma Membrane	transmembrane receptor
HLA-A	major histocompatibility complex, class I, A	-0.957	5.4E-17	7.4E-16	-0.272	0.00995	0.0357	Plasma Membrane	other
IL2RB	interleukin 2 receptor subunit beta	-1.633	0.00188	0.00672	-0.245	0.584	0.774	Plasma Membrane	transmembrane receptor
JAG2	jagged canonical Notch ligand 2	1.068	1.19E-94	1.16E-92	-0.231	0.0000222	0.00014	Extracellular Space	growth factor
HLA-B	major histocompatibility complex, class I, B	-1.369	1.12E-70	7.4E-69	-0.219	0.00126	0.0057	Plasma Membrane	transmembrane receptor
STAT4	signal transducer and activator of transcription 4	-1.618	0.000855	0.00328	-0.099	0.807	0.913	Nucleus	transcription regulator
ICAM1	intercellular adhesion molecule 1	-1.085	0.0000017	0.00000981	0.004	0.983	0.994	Plasma Membrane	transmembrane receptor
NOTCH1	notch receptor 1	0.917	8.64E-155	1.83E-152	0.082	0.0193	0.0633	Plasma Membrane	transcription regulator
CCR1	C-C motif chemokine receptor 1	-3.732	0.00129	0.00475	0.397	0.519	0.726	Plasma Membrane	G-protein coupled receptor
NFATC2	nuclear factor of activated T cells 2	-0.793	0.000528	0.00212	0.762	0.000132	0.000729	Nucleus	transcription regulator
JUN	Jun proto-oncogene, AP-1 transcription factor subunit	0.681	4.14E-63	2.42E-61	0.86	1.86E-101	2.31E-99	Nucleus	transcription regulator
S1PR1	sphingosine-1-phosphate receptor 1	-0.828	0.00958	0.0288	1.045	0.000121	0.00067	Plasma Membrane	G-protein coupled receptor

References

1. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890 (2018). doi:10.1093/bioinformatics/bty560pmid:30423086
2. Kim D. et al. (2015) Hisat: a fast spliced aligner with low memory requirements. Nature Methods, 12, 357–360.
3. Liao Y, Smyth GK, Shi W: featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014, 30: 923-930. 10.1093/bioinformatics/btt656.
4. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA- seq data with DESeq2. Genome Biol 15, 550 (2014). https://doi.org/10.1186/s13059-014-0550-8
5. Andreas Krämer, Jeff Green, Jack Pollard, Stuart Tugendreich, Causal analysis approaches in Ingenuity Pathway Analysis, Bioinformatics, Volume 30, Issue 4, 15 February 2014, Pages 523–530, https://doi.org/10.1093/bioinformatics/btt703