CANCER PERSISTER CELL PATHWAY ANALYSIS

By: Karthik Raj

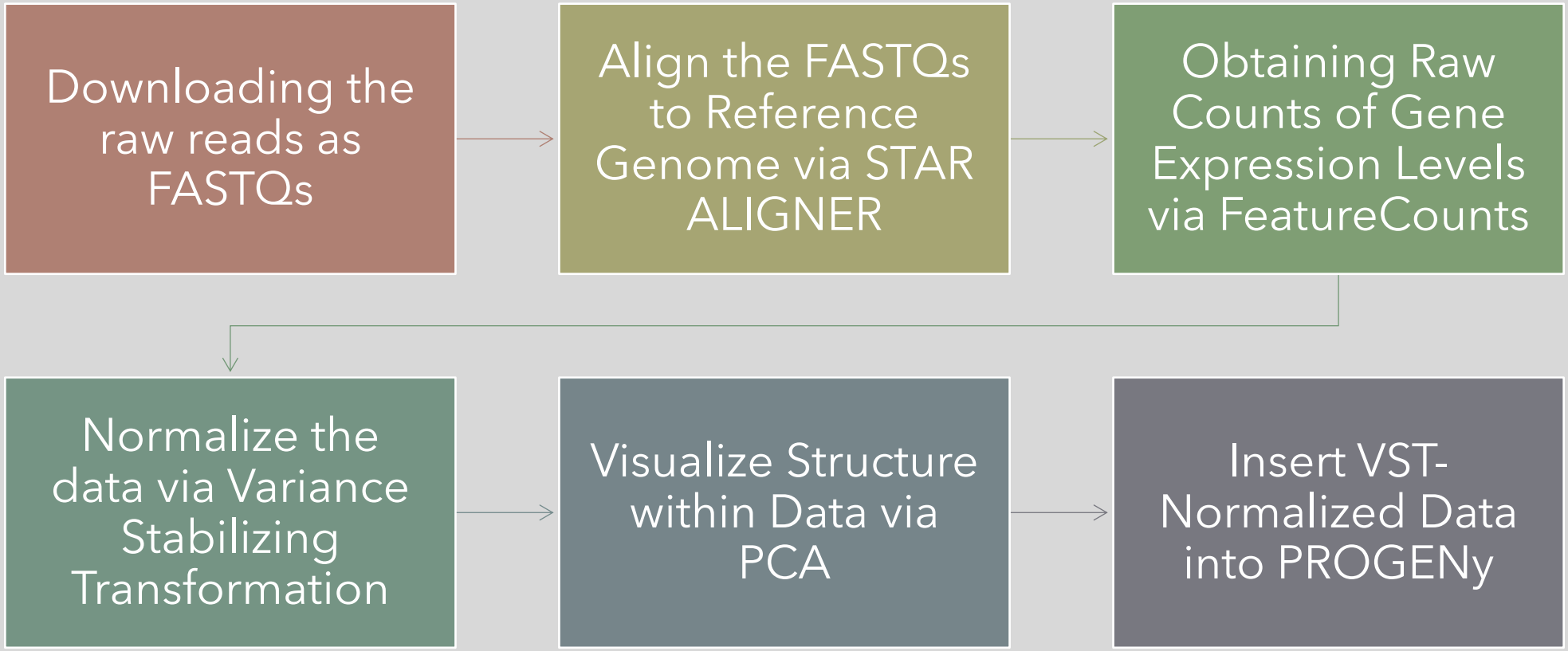
Persister Cells Background

- Cancer cell subpopulation **responsible for re-current cases of cancer** within initially-deemed cancer-free patients
- **Operates primarily in a dormant phase = escape** from treatments targeting common cell metabolism and cell growth processes
- Coexist with other non-stem cancer cells in a competitive environment
  - Non-stem cancer cells **inhibit** cell division of these dormant cells during **low-stress situations**
  - **BUT amplify** the cell division and activation of these dormant cells during **high-stress situations**
- To counter this persister cell population, treatments have been developed specifically to kill this population like inhibition of GPX4

PROGENy

- Pathway Analysis tool superior to other tools currently in use
  - Constructed from the largest dataset: 568 perturbed pathway experiments yielding 2652 microarrays
  - Grants scores to 14 major pathways: cell growth, cell survival, etc.
- Developed within multiple phases:
  1. Calculated z-scores for each gene from each of the perturbed pathway experiments
  2. Trained regression model with the modified pathway as input and gene expression as output
  3. Genes that were consistent outliers in these regression models are Core Responsive Genes (100 Core Genes for **EACH** of the 14 pathways)
  4. Z-scores for these pathways' genes were placed in a matrix
  5. Transformation of the user's normalized gene expression data by this matrix = the pathway's respective score

Data Analysis Workflow



Data Quality Control Checks

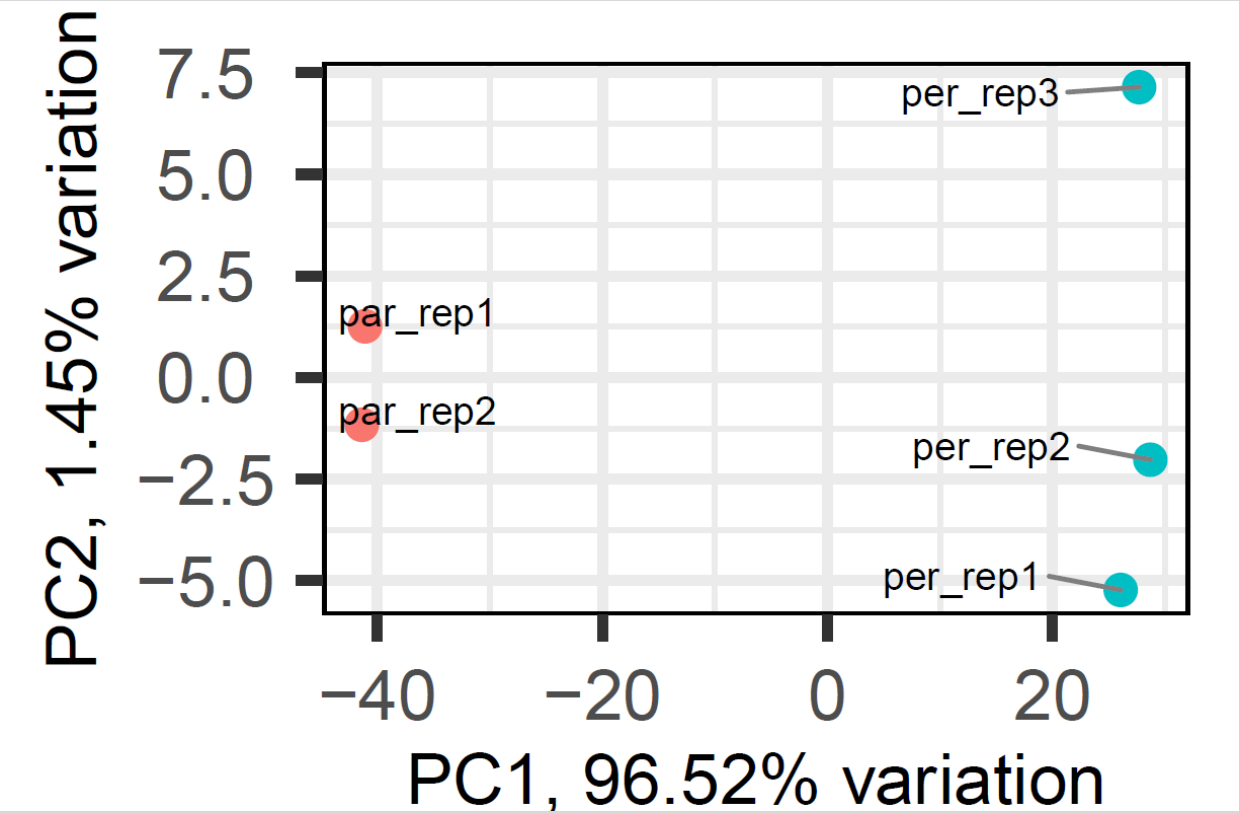
Read Alignment Conducted via STAR Aligner

Sample Name	% Uniquely Aligned Reads	Aligned Reads (in Millions)
par_rep_1	85.8%	135.7
par_rep_2	85.9%	135.4
per_rep_1	85.5%	130.0
per_rep_2	85.7%	142.6
per_rep_3	85.7%	130.5

Gene Counts obtained via FeatureCounts

Sample Name	% Assigned Alignments	Assigned Alignments (in Millions)
par_rep_1	59.6%	114.8
par_rep_2	59.7%	114.7
per_rep_1	56.7%	105.7
per_rep_2	56.7%	115.4
per_rep_3	57.1%	106.4

PCA → Structure within Data



Important Notes:

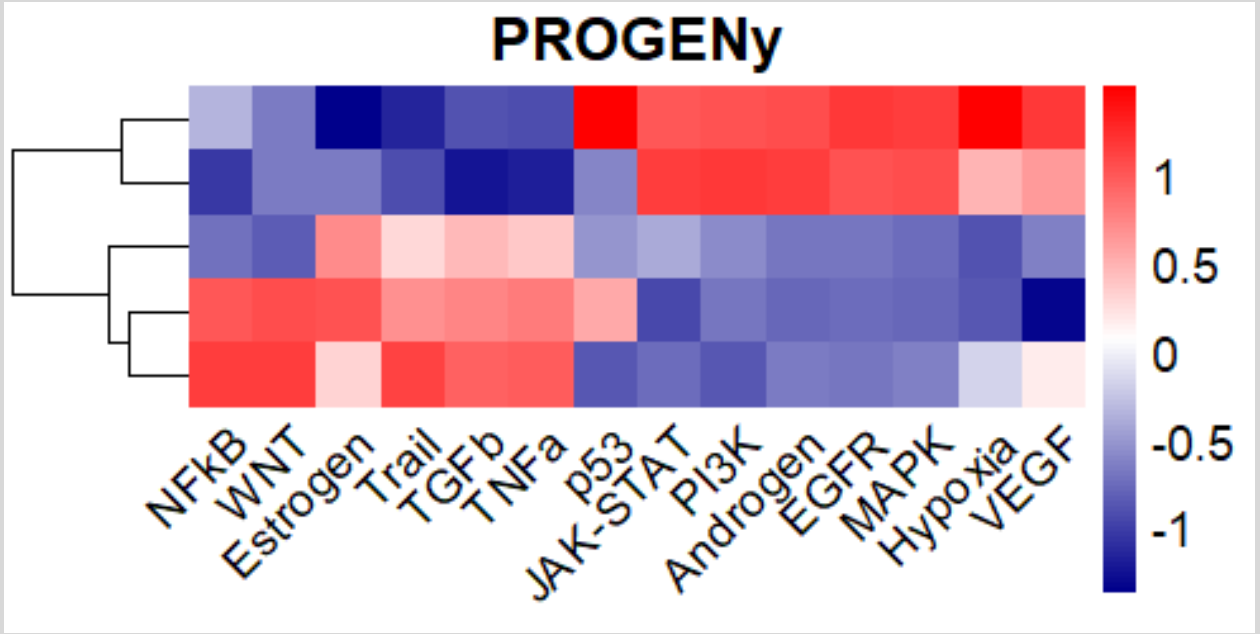
- VST-Normalized Data allows for truly high variant genes to reveal the underlying structure
- Parental Replicates cluster around each other
- Persister Replicates (#1 and #2) cluster **BUT** replicate #3 is clearly distant from the others
- Component #1 accounts for 96.52% of the variance so clearly separation is based off the parental/persister type

PROGENy Set-Up and Results

The 14 PROGENy Pathways and the # of Genes Used Per Pathway

Androgen: 98 (98%)	EGFR: 94 (94%)	Estrogen: 98 (98%)
Hypoxia: 98 (98%)	JAK-STAT: 99 (99%)	MAPK: 98 (98%)
NFkB: 98 (98%)	p53: 93 (93%)	PI3K: 88 (88%)
TGFb: 96 (96%)	TNFa: 95 (95%)	Trail: 73 (73%)
VEGF: 87 (87%)	WNT: 97 (97%)	

Visual Representation of Pathway Scores with the 2 Parental Replicates On The 2 Top Rows and below 3 Persister Replicates:



PROGENy Scores

Abc pathways.c... reps	# pathways.csv Androgen	# pathways.... Egfr	# pathways.csv Estrogen	# pathways.csv Hypoxia	# pathways.csv Jak-Stat	# pathways.c... MAPK	# pathways.... NFkb	# pathways.... p53	# pathways.... PI3K	# pathways.... TGFb	# pathways.... TNFa	# pathways.... Trail	# pathways.... Vegf	# pathways.... WNT
par_rep1	1.06024	1.16292	-1.37783	1.48279	0.99695	1.12675	-0.34676	1.49900	1.01400	-0.90223	-0.93372	-1.16565	1.16937	-0.68560
par_rep2	1.12771	1.02474	-0.66559	0.46913	1.14808	1.06105	-1.04591	-0.62878	1.16261	-1.23981	-1.19262	-0.92209	0.63242	-0.66247
per_rep1	-0.71450	-0.70825	0.71199	-0.89198	-0.42188	-0.75995	-0.72681	-0.52530	-0.58749	0.45601	0.36148	0.28750	-0.63377	-0.83693
per_rep2	-0.80047	-0.76782	1.02991	-0.86454	-0.97430	-0.77606	0.98405	0.54246	-0.71597	0.74410	0.78806	0.69045	-1.34191	1.05050
per_rep3	-0.67298	-0.71160	0.30152	-0.19540	-0.74885	-0.65179	1.13543	-0.88737	-0.87314	0.94192	0.97680	1.10979	0.17389	1.13450

Scaled along each Pathway such that each **Pathway's Mean: 0** and **Standard Deviation: 1**

Pros:

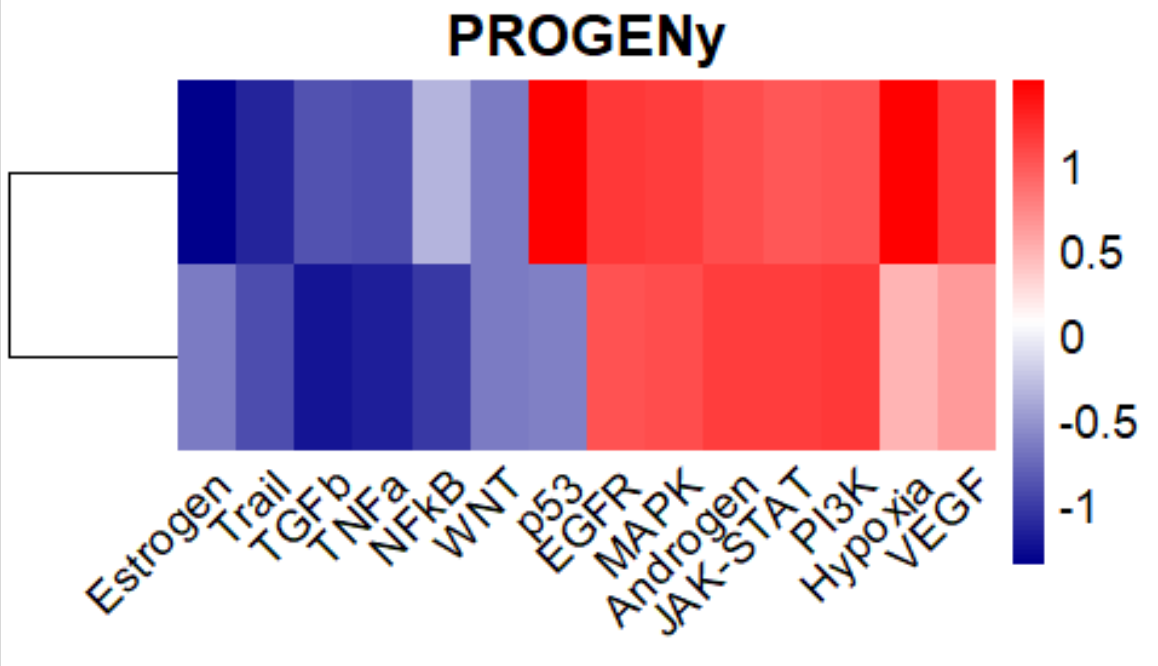
- Scores can give perspective on the usages of different pathways relative to different replicates

Cons:

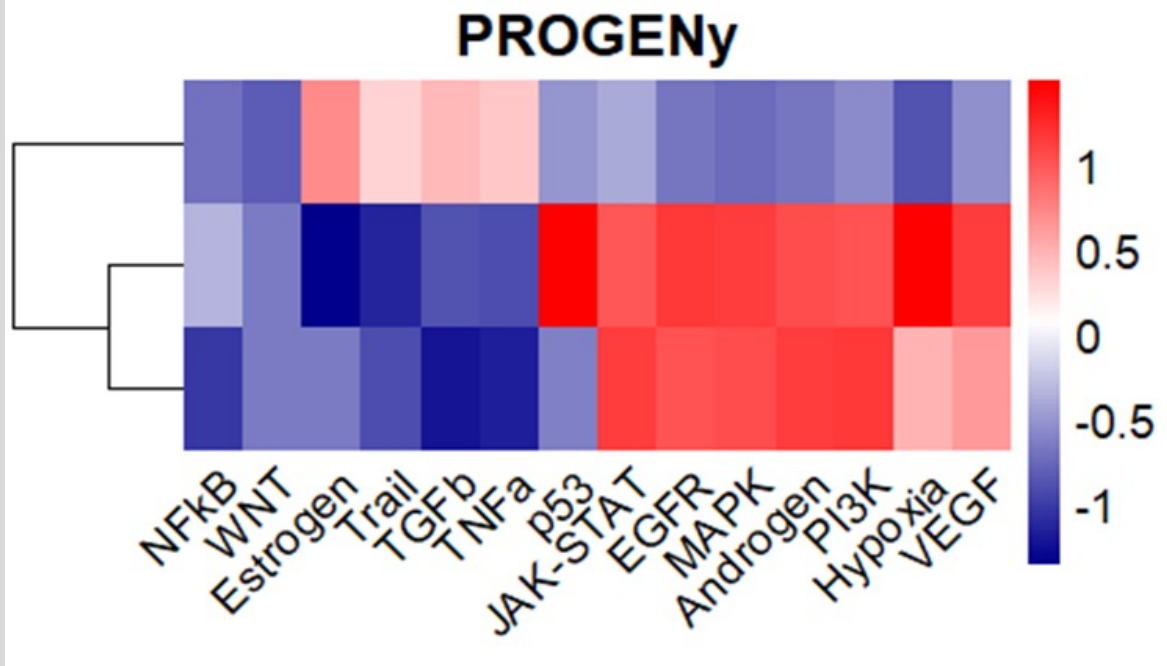
- Limited to these 14 pathways
- Can only determine whether pathways are relatively active or inhibited

Split Clustered Heatmaps

2 Parental Replicates

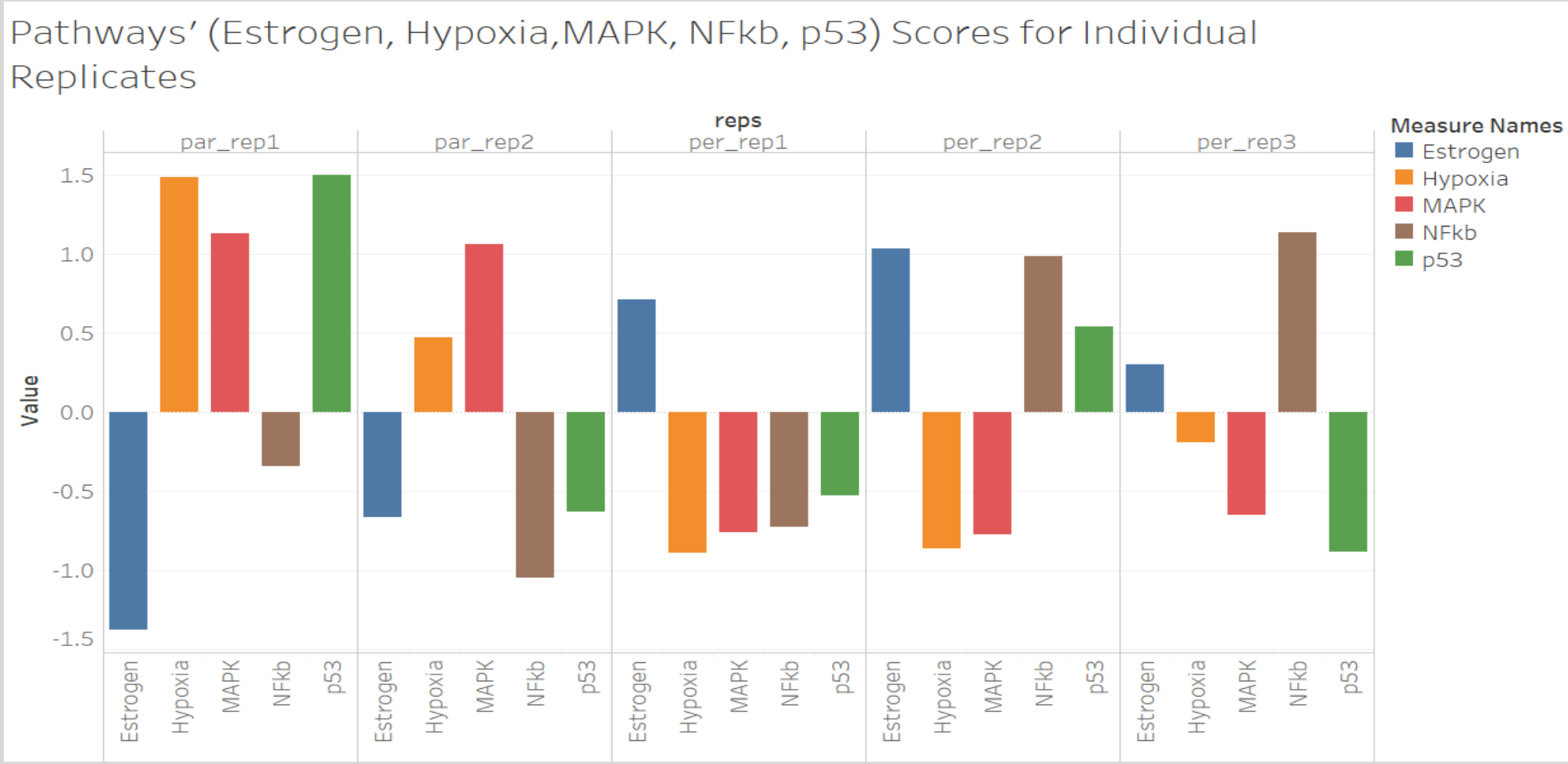


3 Persister Replicates

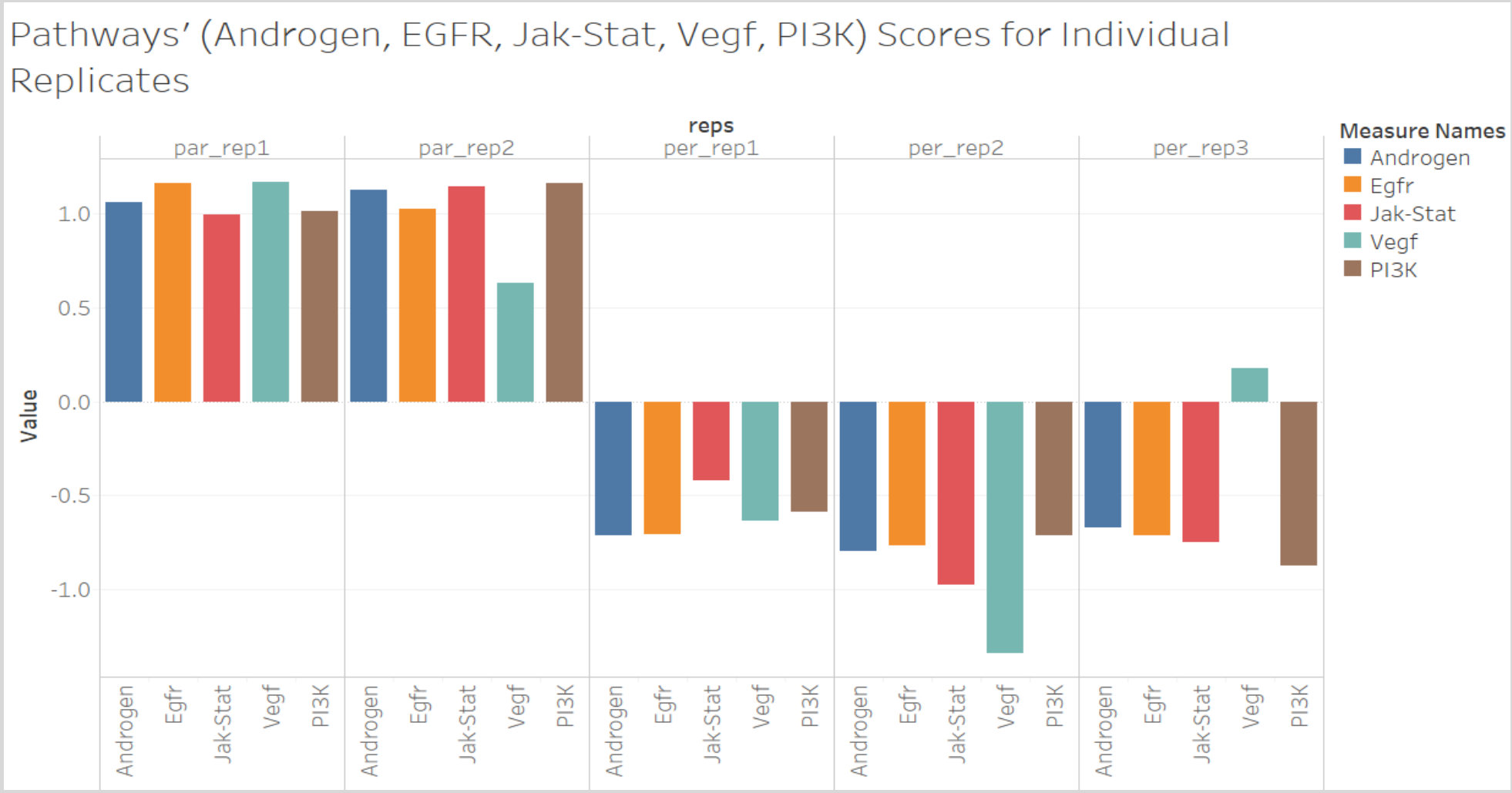


Differences within the similar replicates' pathways' scores potentially reflects distinct utility of the components of these pathways than the norm

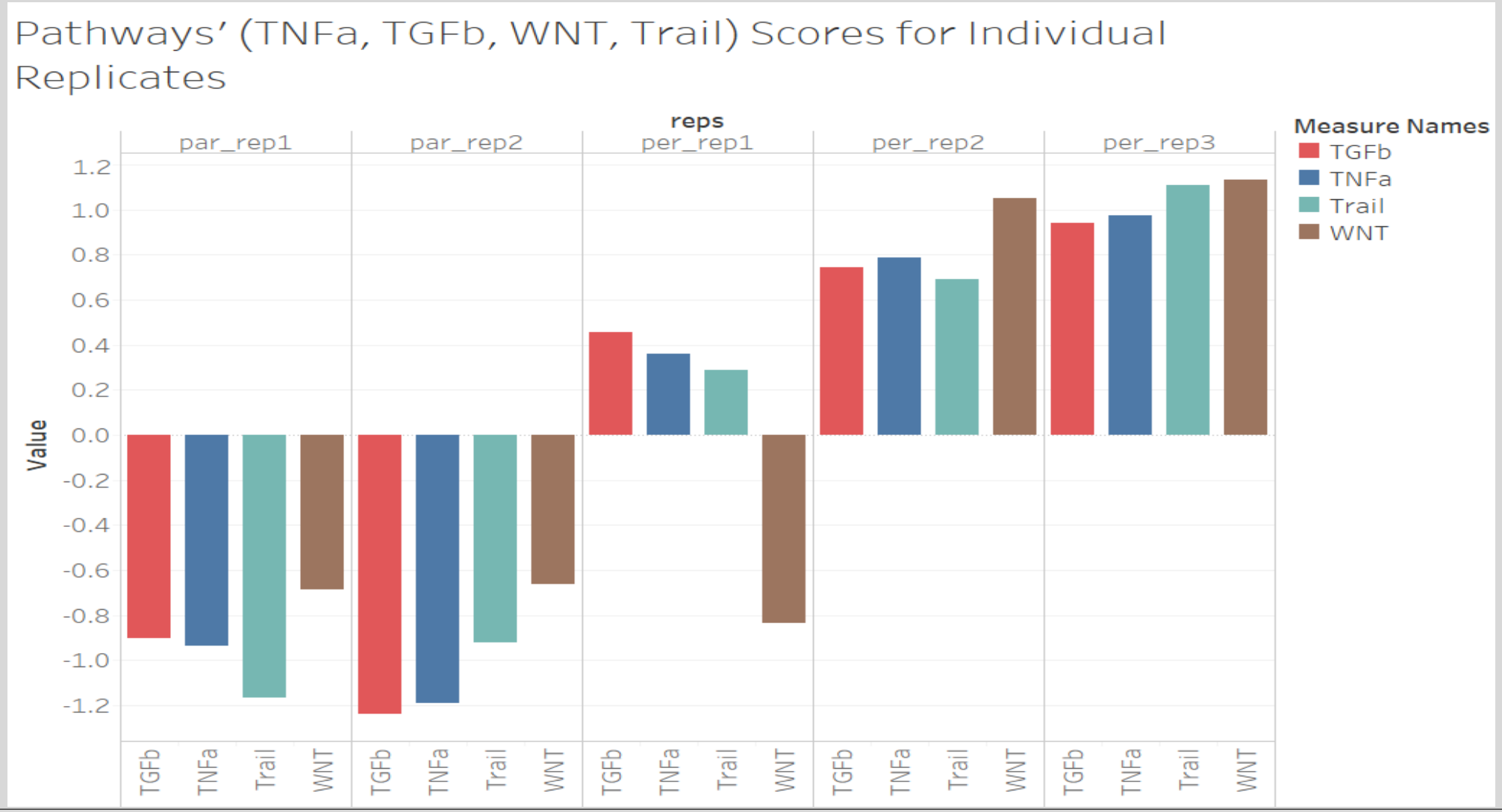
PROGENy Results on Pathway Level



PROGENy Results (Continued)



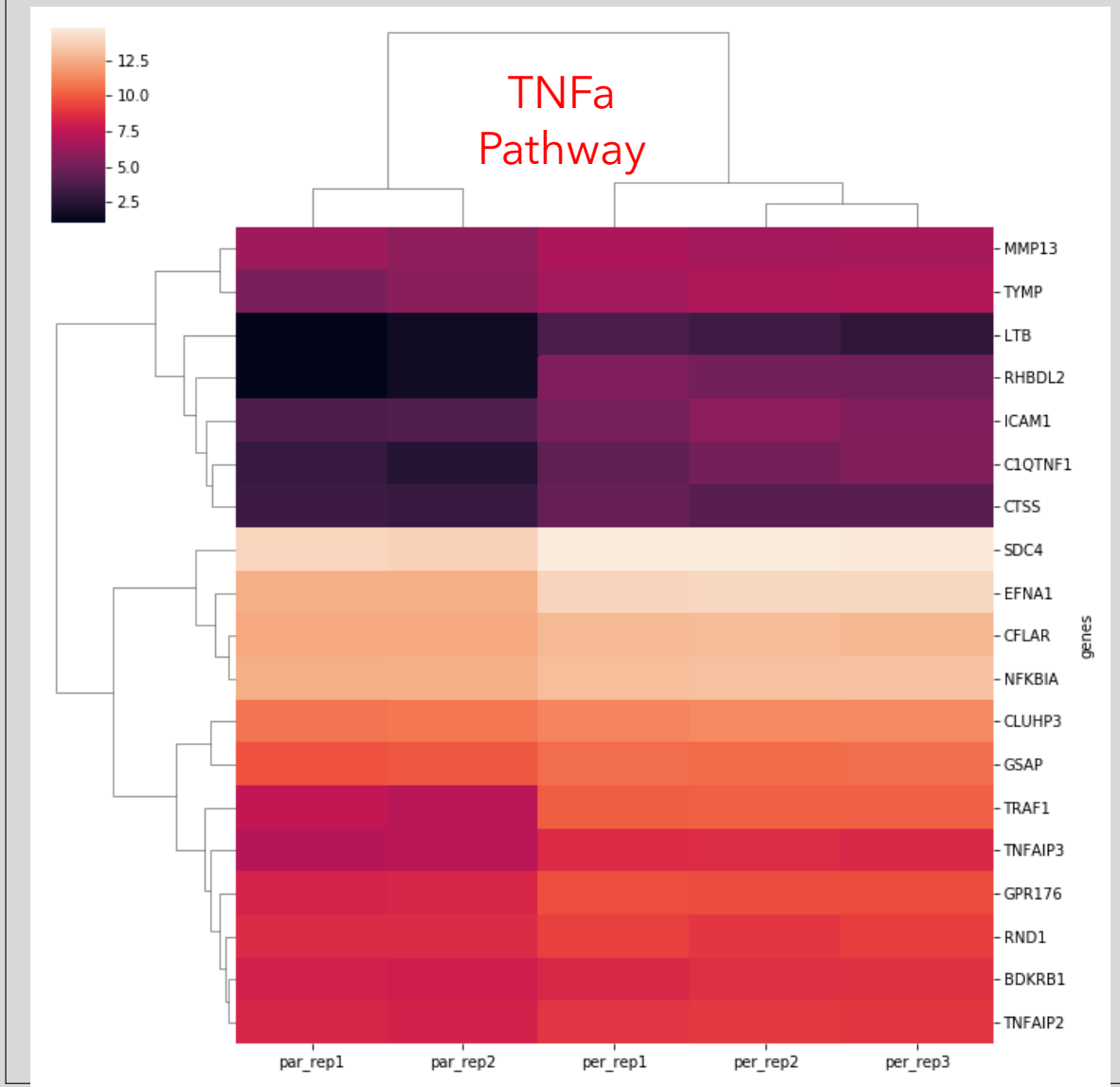
PROGENy Results (Continued)

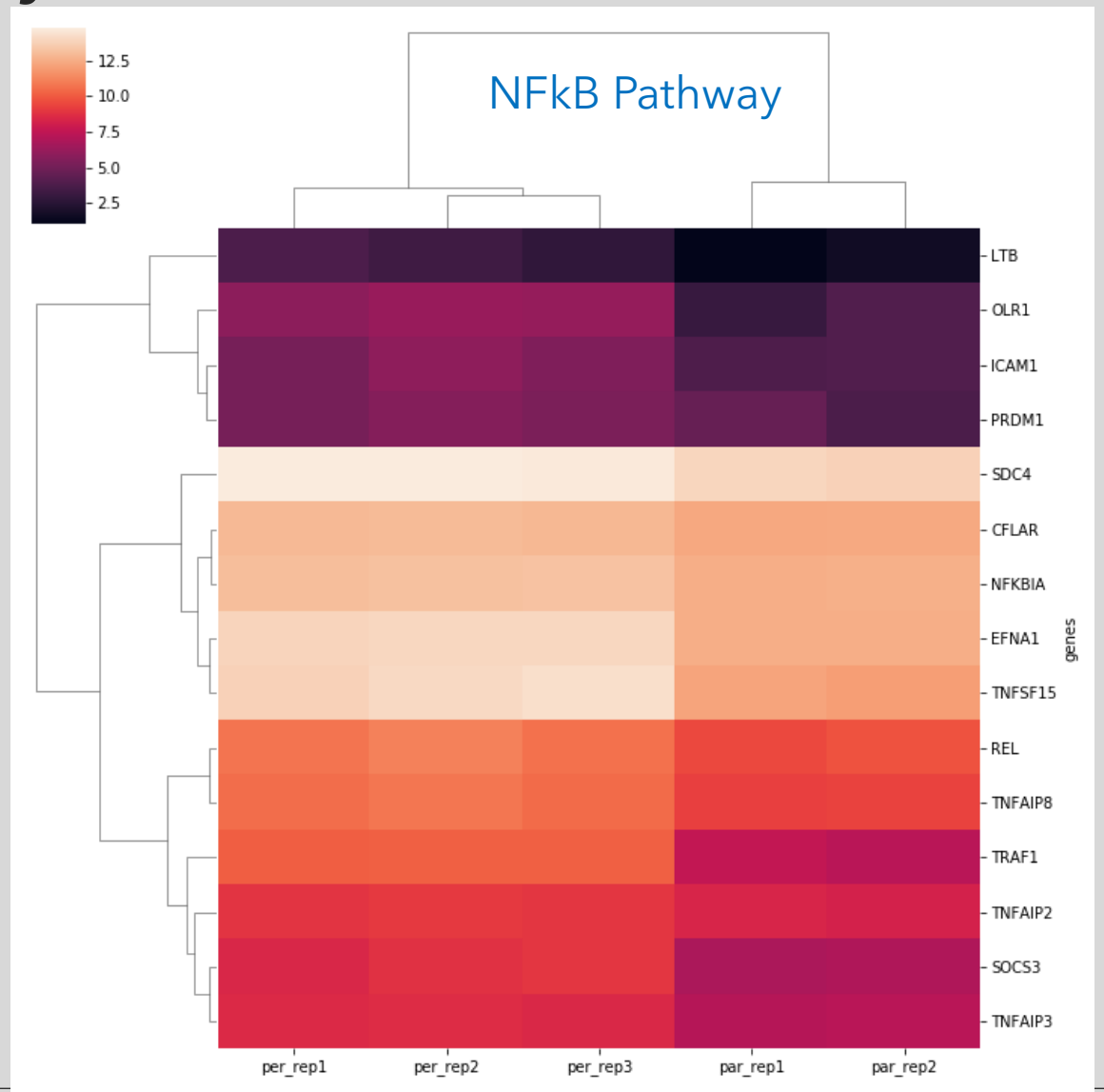


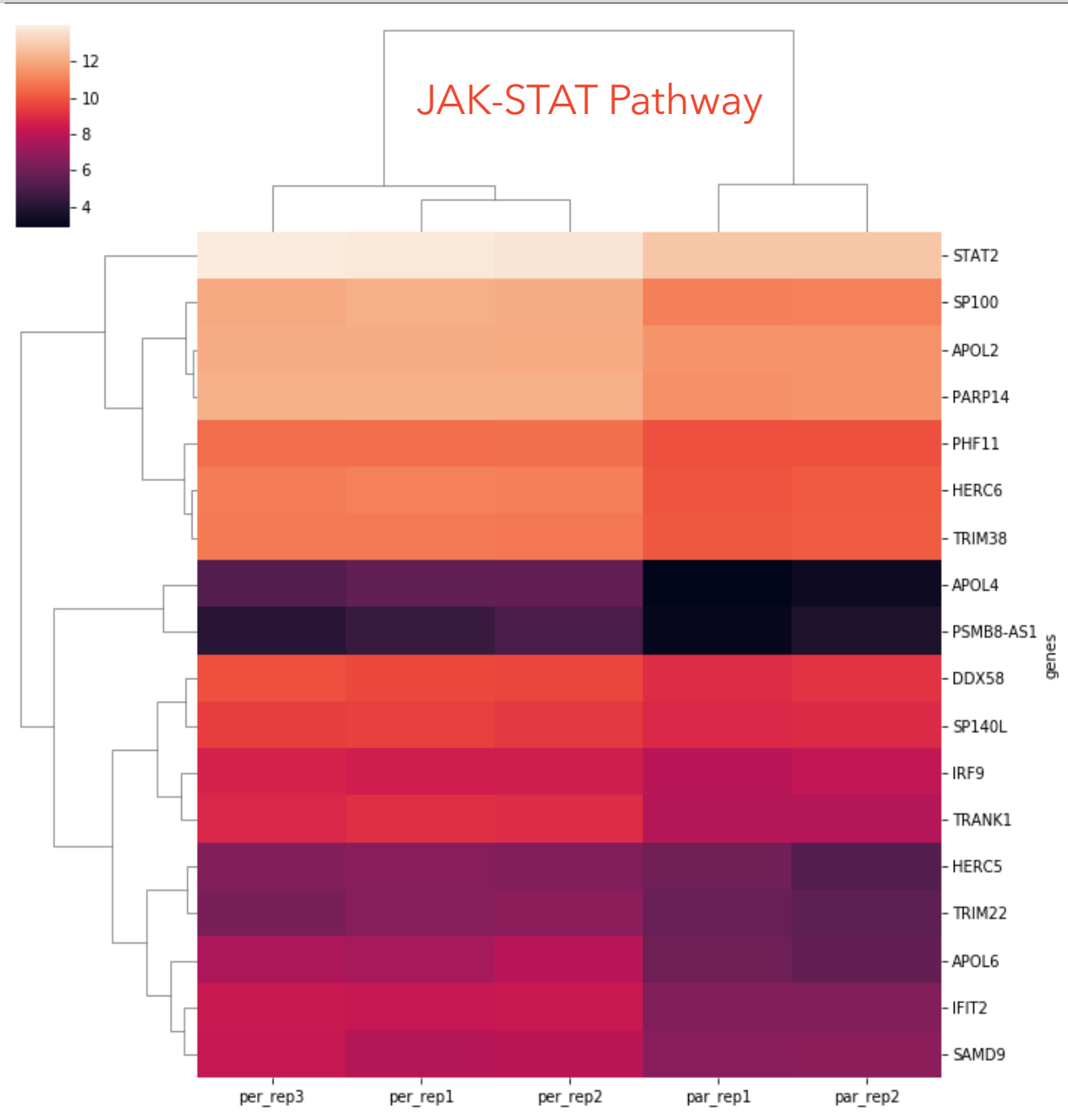
Identifying DE Genes Within PROGENy Model

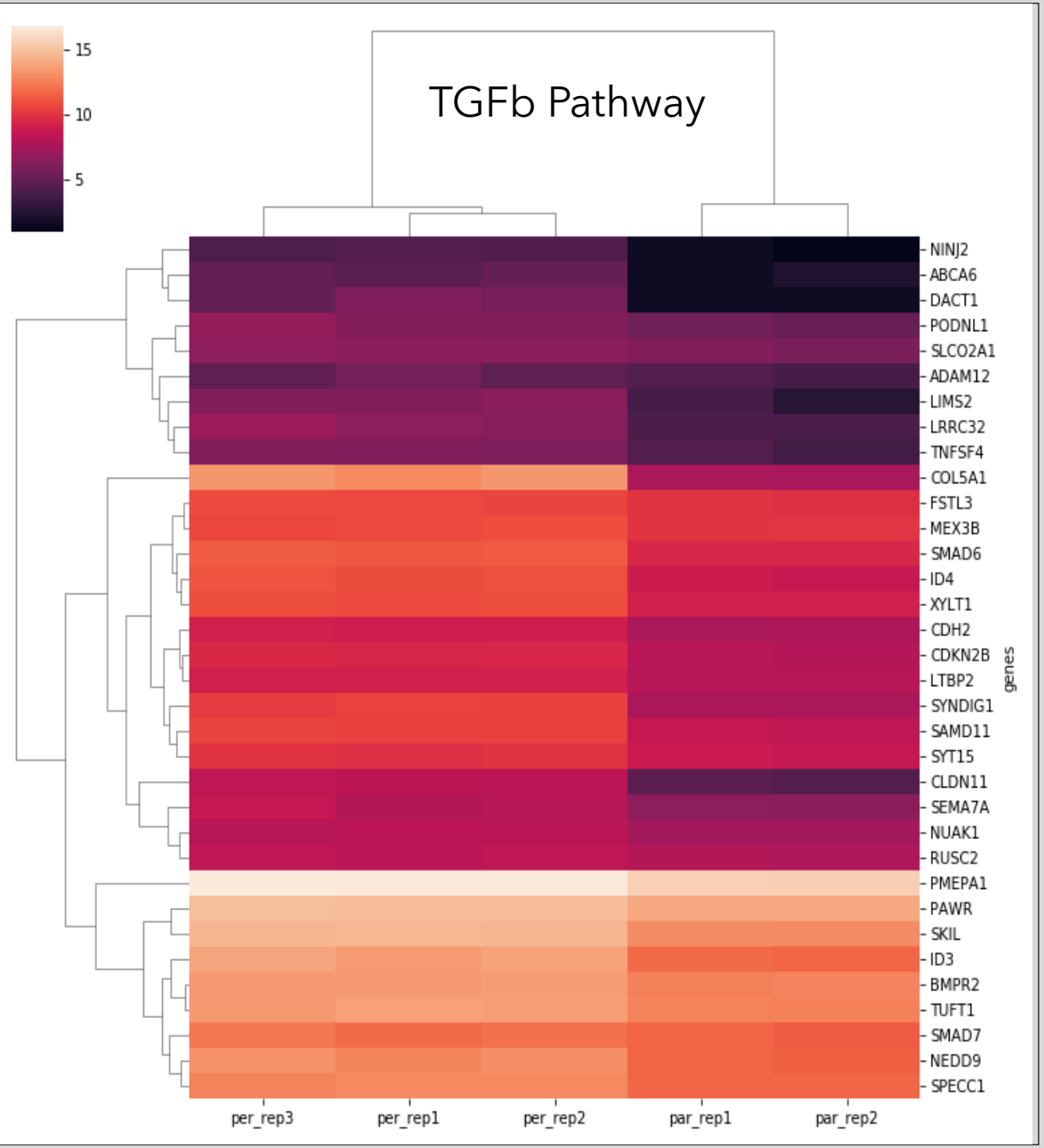
- Observing differences in pathway activity not only between the parental and persister replicates, but also within both the persister and parental replicates
- Wanted to identify the differential genes that are responsible for these observed differences in pathway activity especially among the persister replicates' cellular survival and stress response pathways
- Process of Identifying these Genes within the PROGENy Model:
  1. Conducted Hypothesis Testing on the log2 normalized dataset
  2. Used a 2-fold criteria to sort the differential genes: Log-Fold Change (LFC) min: 0.5 and Padj cutoff: 0.1
     1. Goal was to reduce the genes to a set of truly differential genes, but leave a loose enough constraint to allow for more genes to be kept in case they played prominent roles verified thru literature
  3. Selection process resulted in 4,045 genes being kept
  4. Determined which of these genes were present within the PROGENy's model of Pathway CORE Responsive Genes for each pathway
  5. Placed the identified genes' log2 normalized values in a clustered heatmap and made one for each of the 14 pathways

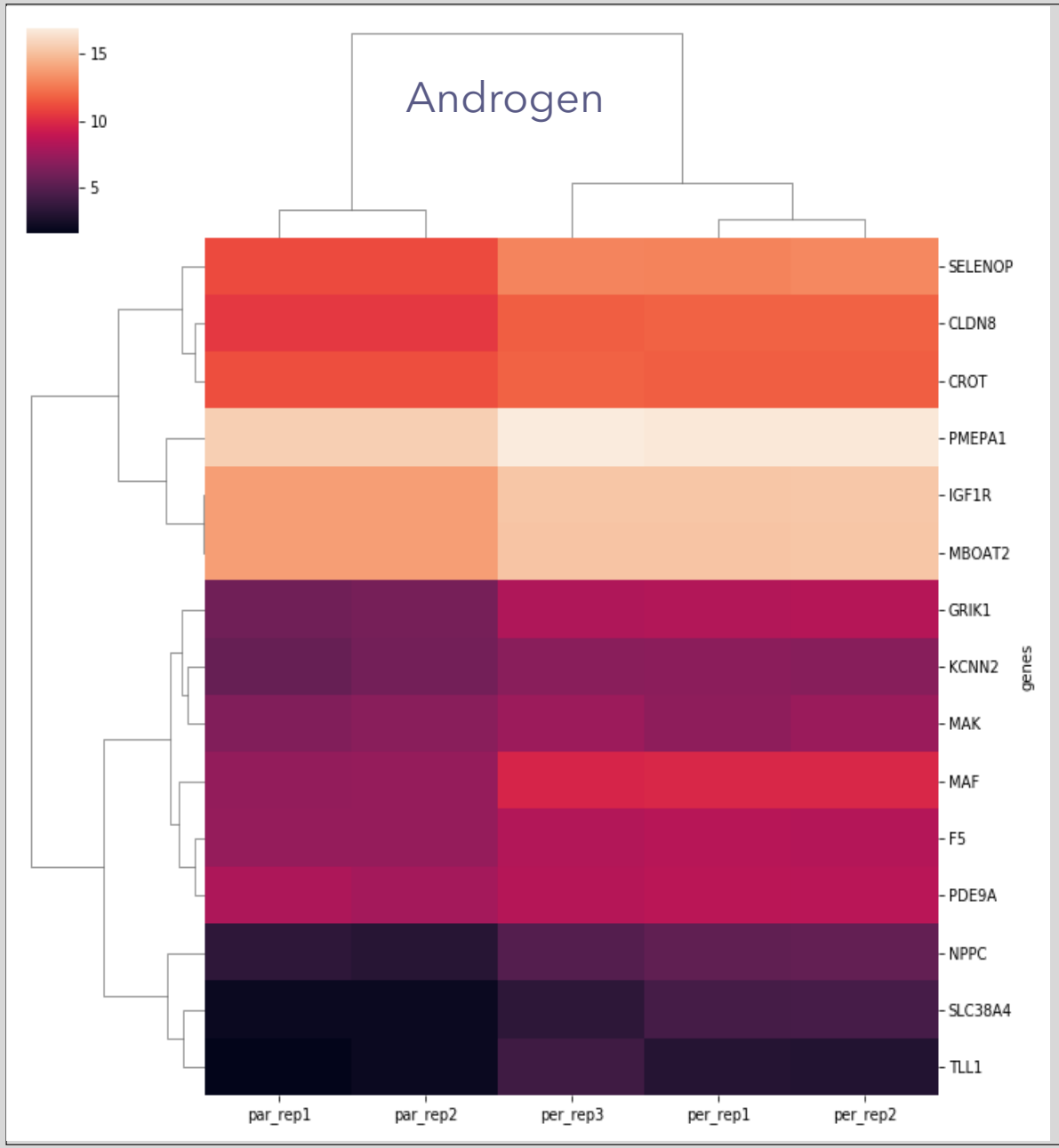
Cellular Stress Pathways: NFkB and TNFa

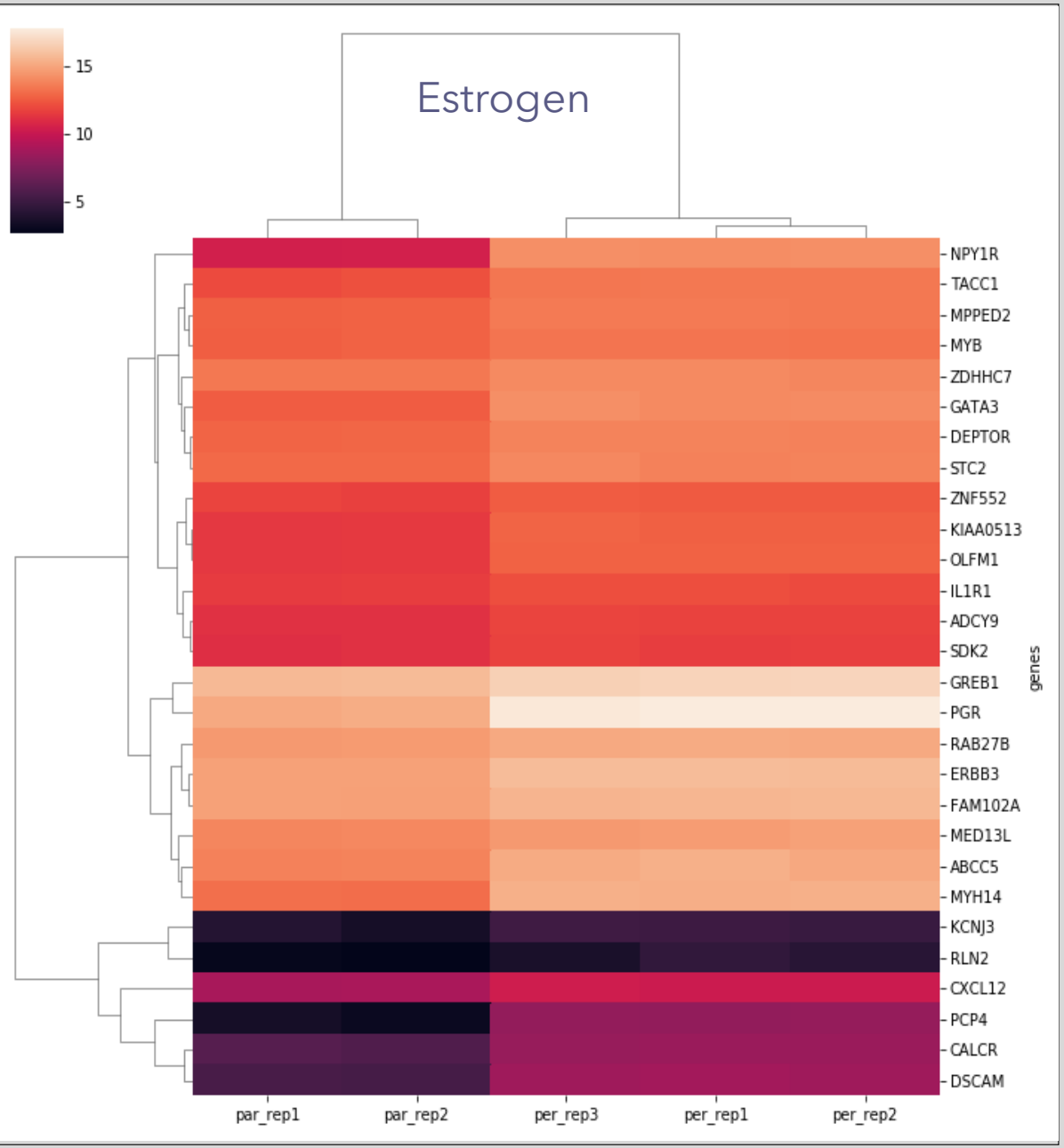


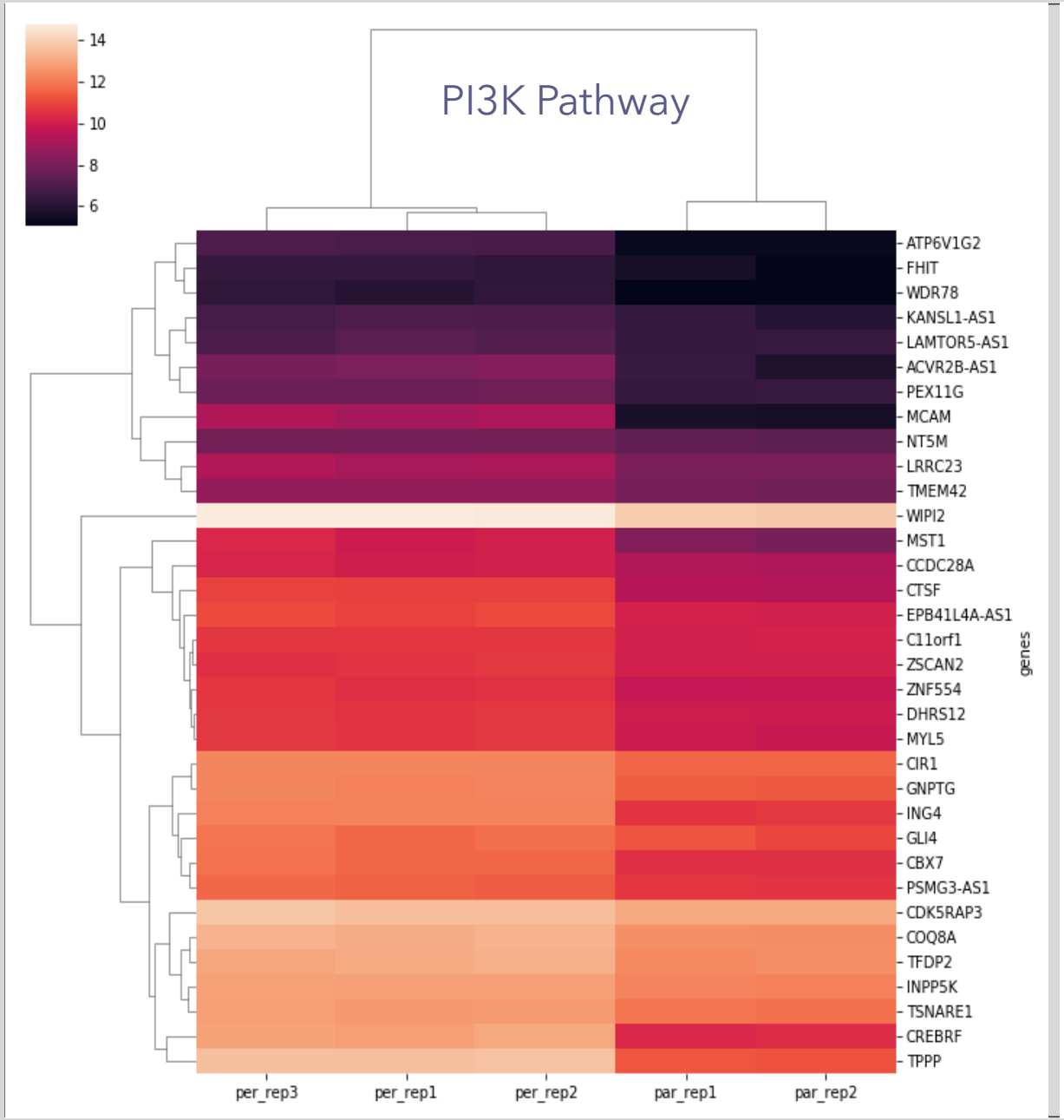


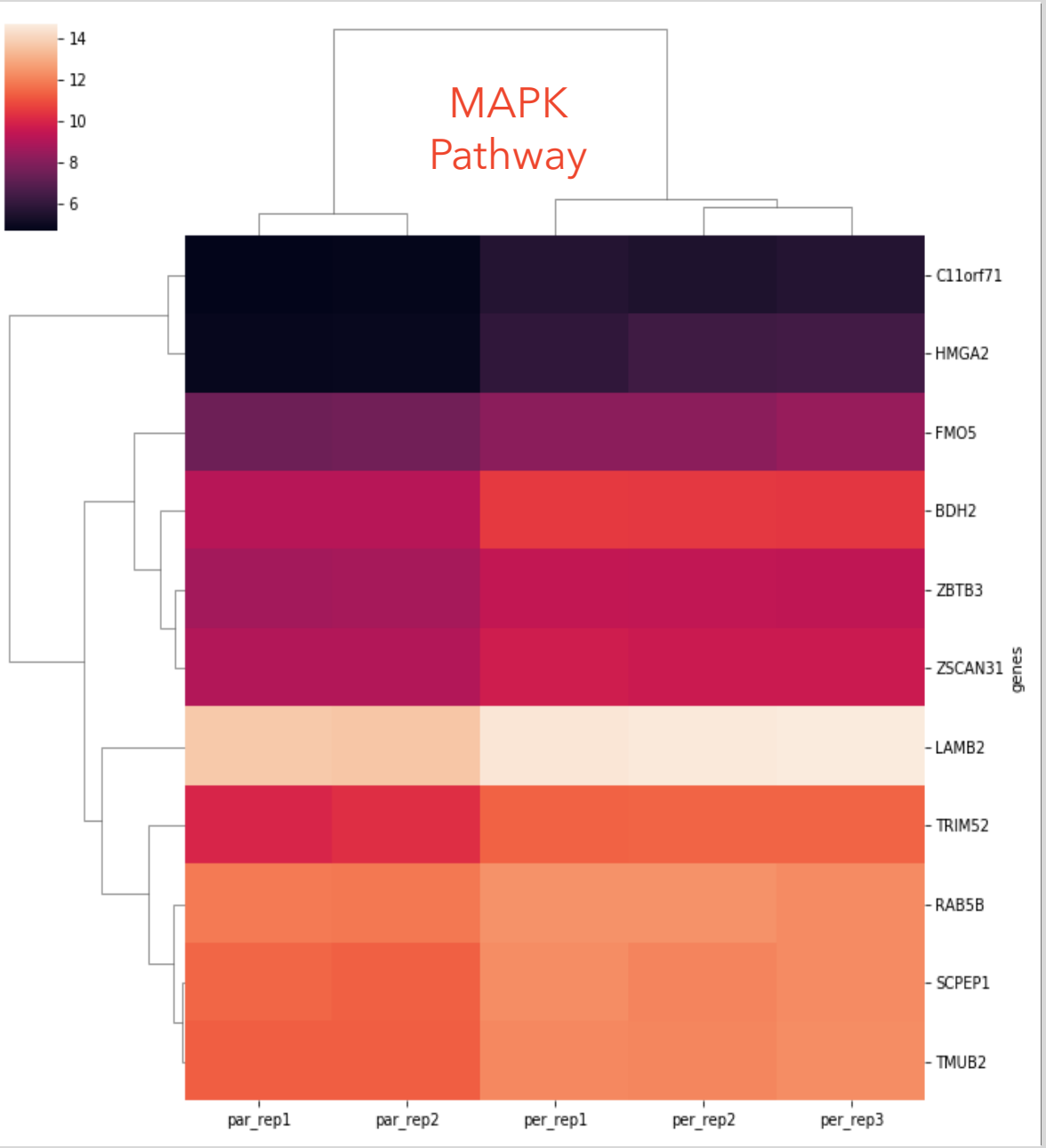

JAK-STAT Pathway
12
10
8
6
4
STAT2
SP100
APOL2
PARP14
PHF11
HERC6
TRIM38
APOL4
PSMB8-AS1
DDX58
SP140L
IRF9
TRANK1
HERC5
TRIM22
APOL6
IFIT2
SAMD9
genes
per_rep3
per_rep1
per_rep2
par_rep1
par_rep2

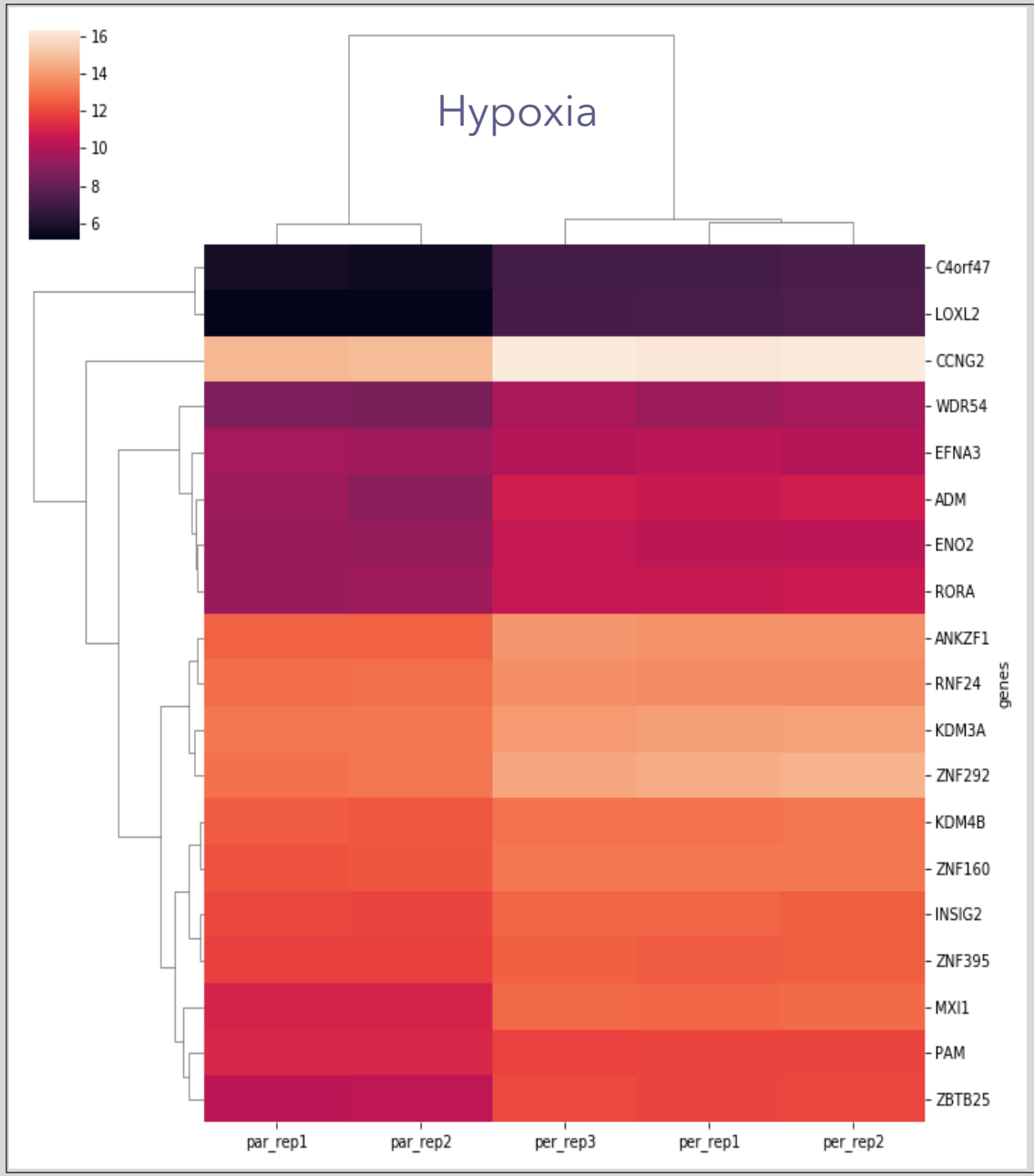

TGFb Pathway
15
10
5
NINJ2
ABCA6
DACT1
PODNL1
SLCO2A1
ADAM12
LIMS2
LRRC32
TNFSF4
COL5A1
FSTL3
MEX3B
SMAD6
ID4
XYLT1
CDH2
CDKN2B
LTBP2
SYNDIG1
SAMD11
SYT15
CLDN11
SEMA7A
NUAK1
RUSC2
PMEPA1
PAWR
SKIL
ID3
BMPR2
TUFT1
SMAD7
NEDD9
SPECC1
genes
per_rep3
per_rep1
per_rep2
par_rep1
par_rep2

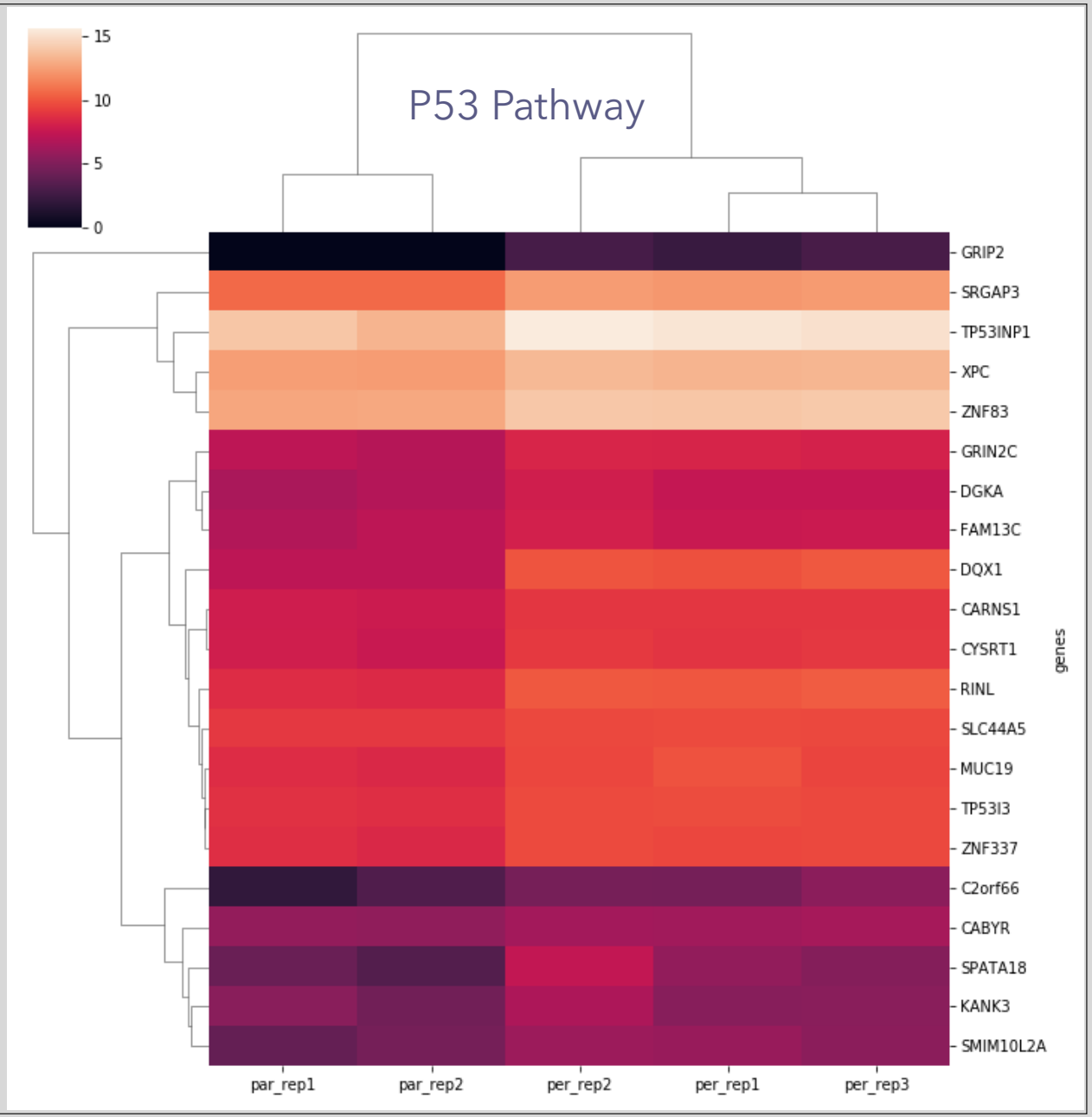

Androgen
15
10
5
SELENOP
CLDN8
CROT
PMEPA1
IGF1R
MBOAT2
GRIK1
KCNN2
MAK
MAF
F5
PDE9A
NPPC
SLC38A4
TLL1
genes
par_rep1
par_rep2
per_rep3
per_rep1
per_rep2

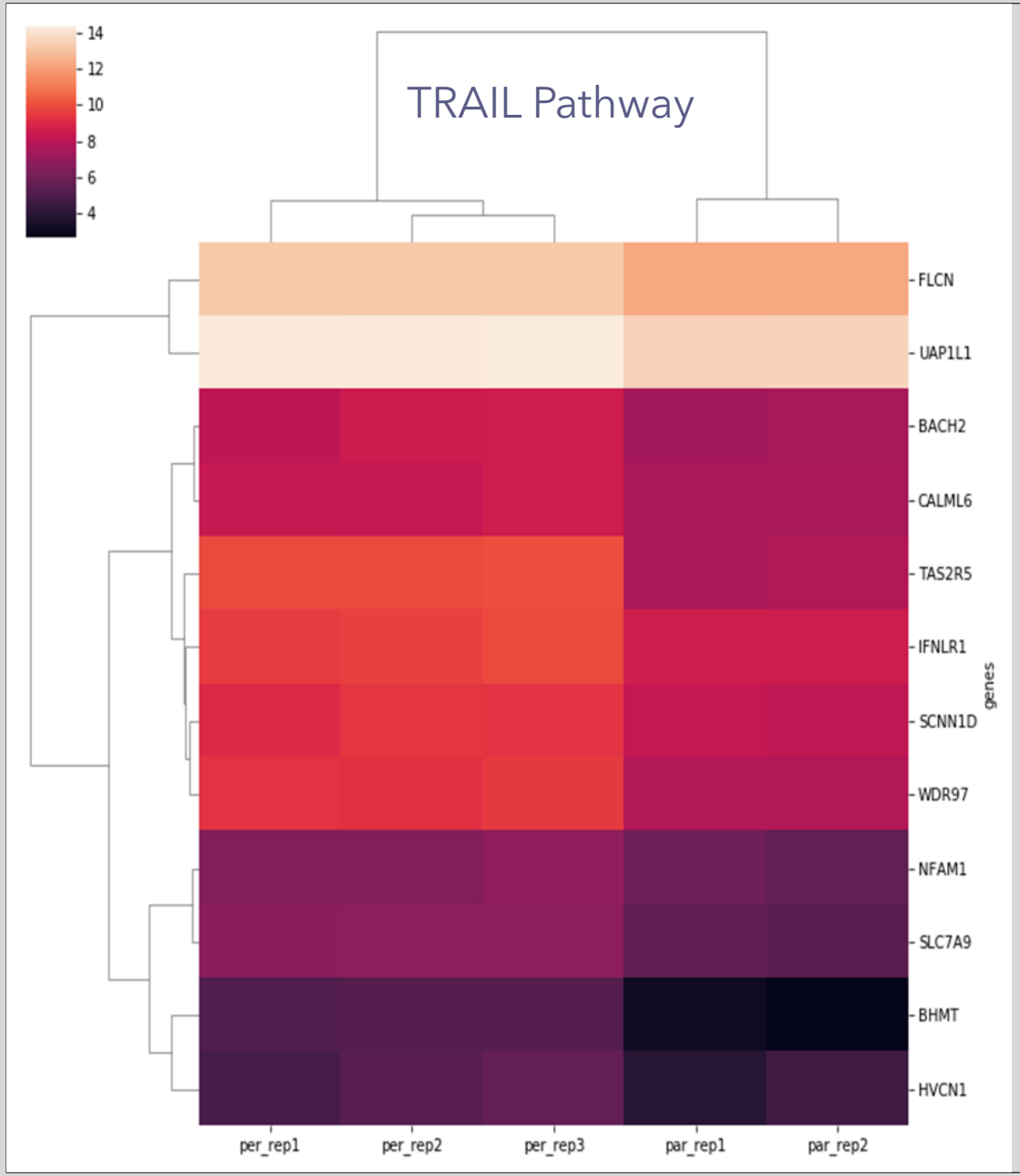

Estrogen
15
10
5
NPY1R
TACC1
MPPED2
MYB
ZDHHC7
GATA3
DEPTOR
STC2
ZNF552
KIAA0513
OLFM1
IL1R1
ADCY9
SDK2
GREB1
PGR
RAB27B
ERBB3
FAM102A
MED13L
ABCC5
MYH14
KCNJ3
RLN2
CXCL12
PCP4
CALCR
DSCAM
genes
par_rep1
par_rep2
per_rep3
per_rep1
per_rep2

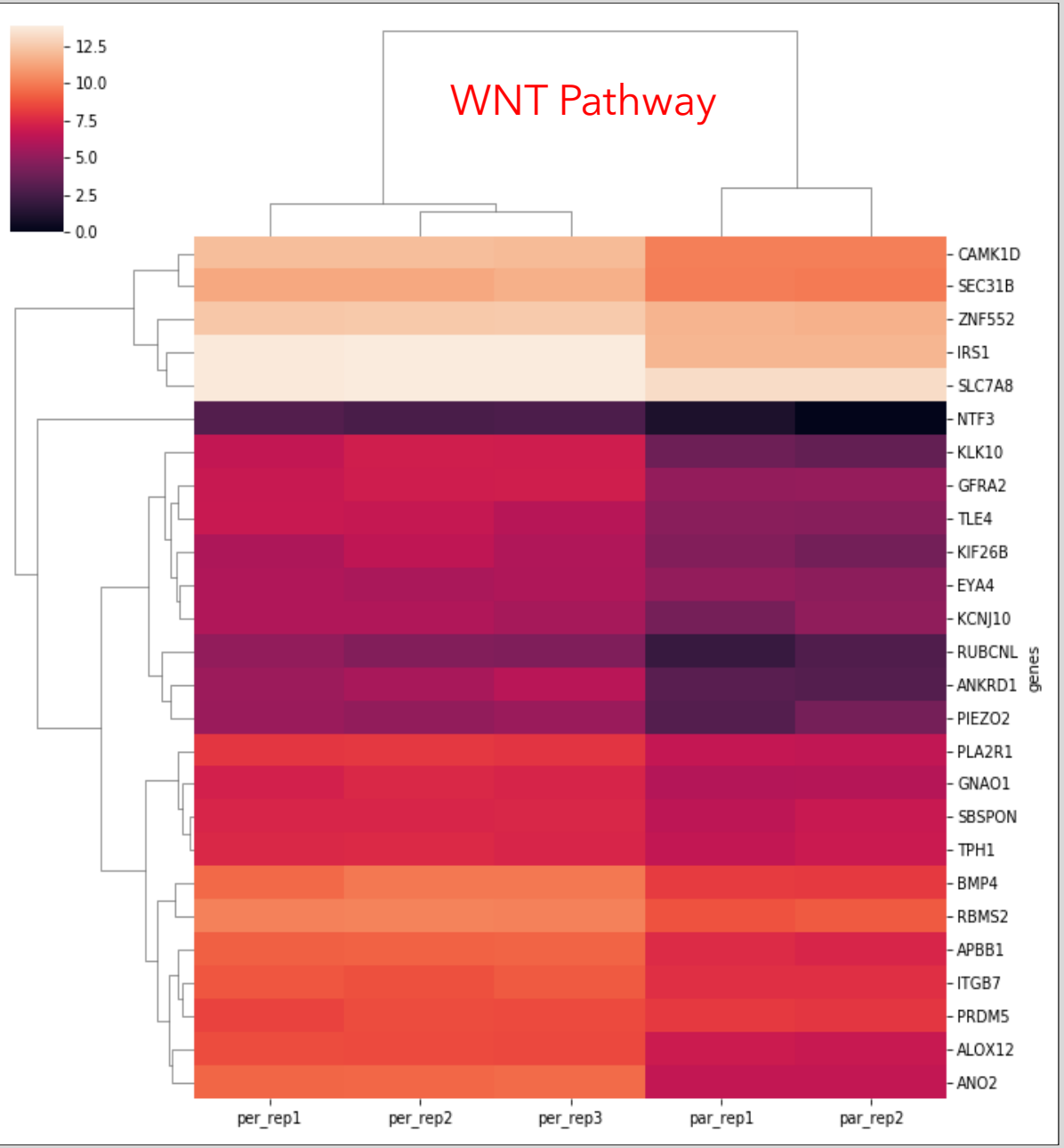

PI3K Pathway
14
12
10
8
6
ATP6V1G2
FHIT
WDR78
KANSL1-AS1
LAMTOR5-AS1
ACVR2B-AS1
PEX11G
MCAM
NT5M
LRRC23
TMEM42
WIPI2
MST1
CCDC28A
CTSF
EPB41L4A-AS1
C11orf1
ZSCAN2
ZNF554
DHRS12
MYL5
CIR1
GNPTG
ING4
GLI4
CBX7
PSMG3-AS1
CDK5RAP3
COQ8A
TFDP2
INPP5K
TSNARE1
CREBRF
TPPP
genes
per_rep3
per_rep1
per_rep2
par_rep1
par_rep2

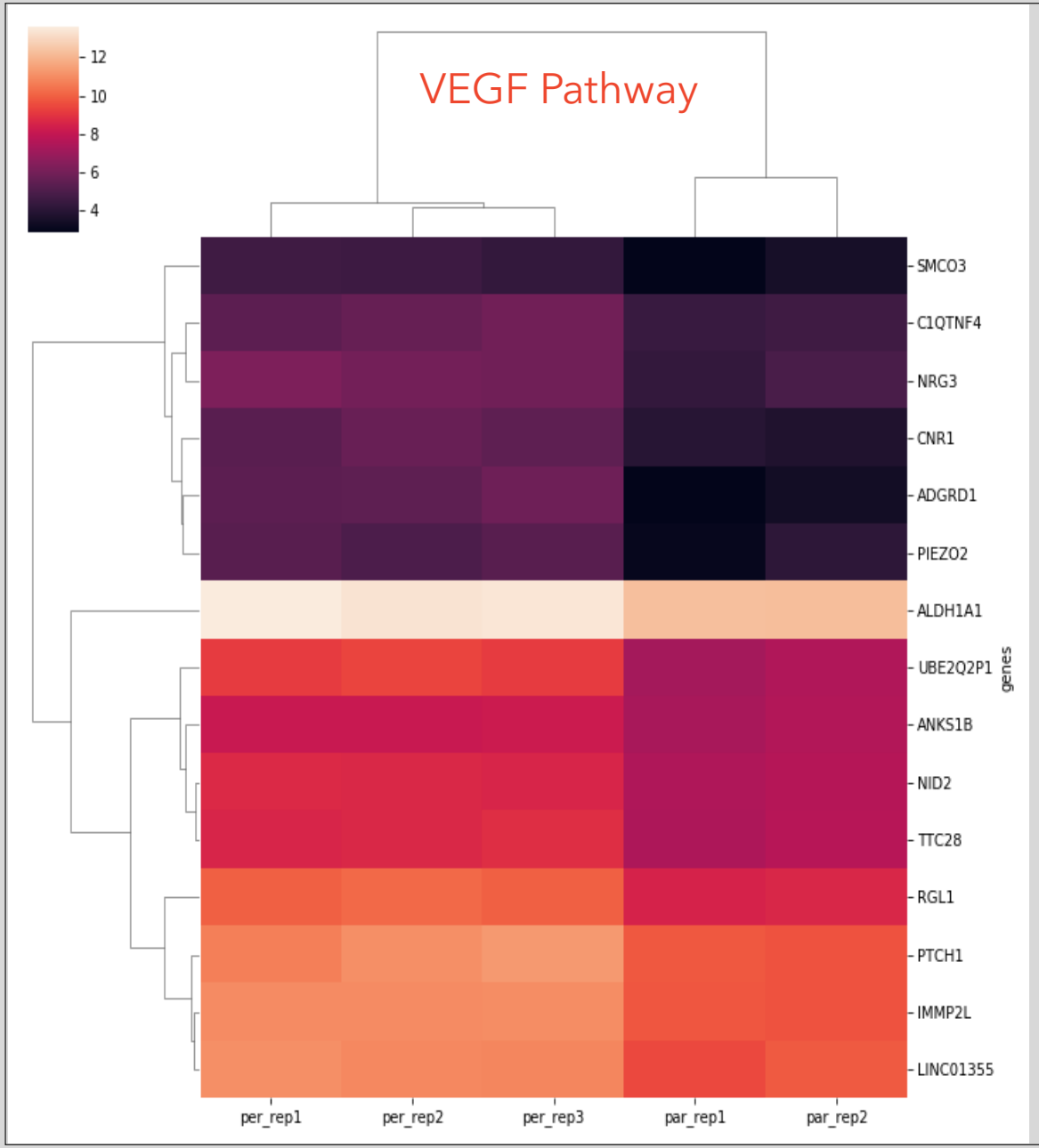

MAPK Pathway
14
12
10
8
6
C11orf71
HMGA2
FMO5
BDH2
ZBTB3
ZSCAN31
LAMB2
TRIM52
RAB5B
SCPEP1
TMUB2
genes
par_rep1
par_rep2
per_rep1
per_rep2
per_rep3

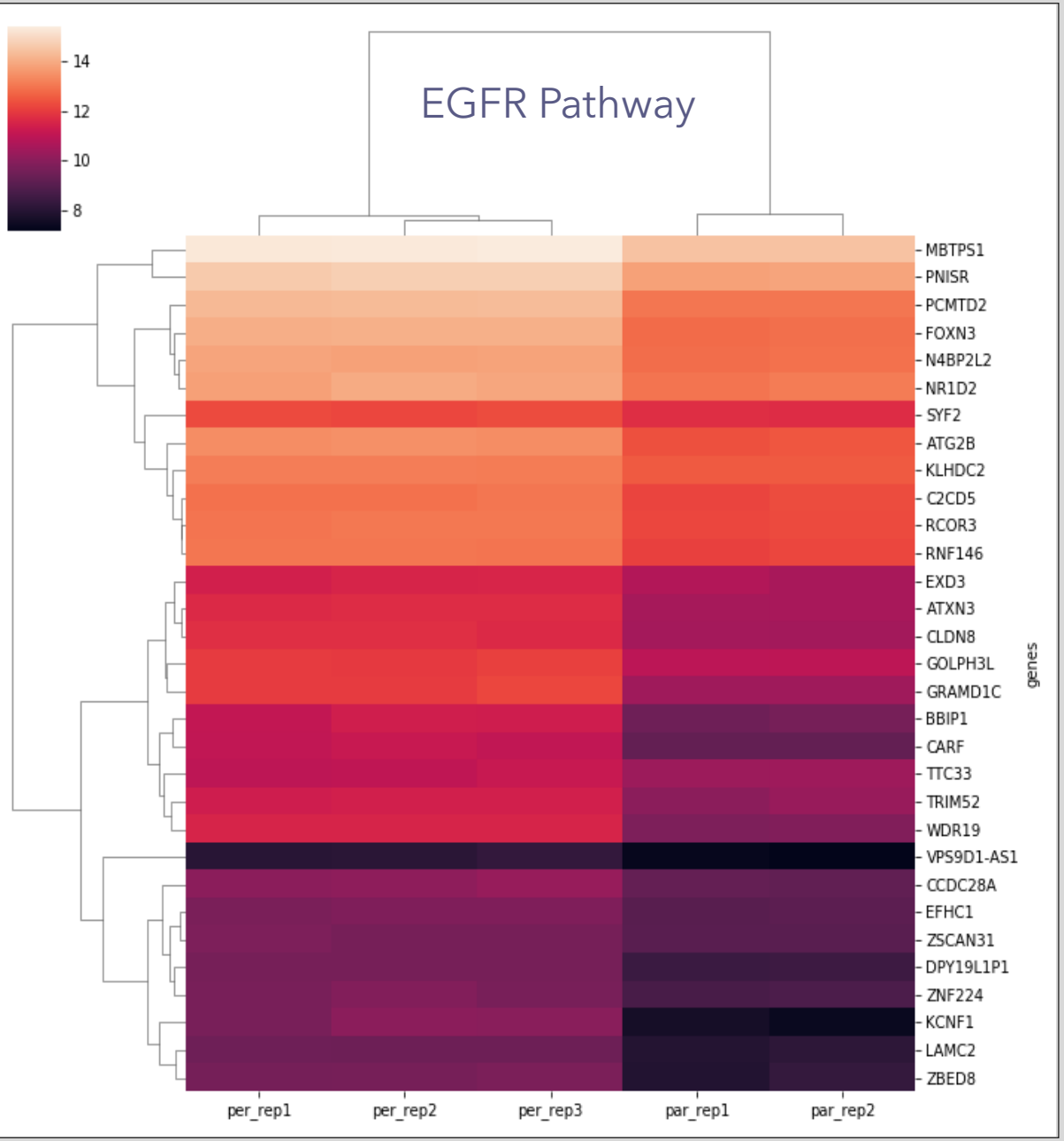

Hypoxia
16
14
12
10
8
6
C4orf47
LOXL2
CCNG2
WDR54
EFNA3
ADM
ENO2
RORA
ANKZF1
RNF24
KDM3A
ZNF292
KDM4B
ZNF160
INSIG2
ZNF395
MXI1
PAM
ZBTB25
genes
par_rep1
par_rep2
per_rep3
per_rep1
per_rep2


P53 Pathway
15
10
5
0
GRIP2
SRGAP3
TP53INP1
XPC
ZNF83
GRIN2C
DGKA
FAM13C
DQX1
CARNS1
CYSRT1
RINL
SLC44A5
MUC19
TP53I3
ZNF337
C2orf66
CABYR
SPATA18
KANK3
SMIM10L2A
genes
par_rep1
par_rep2
per_rep2
per_rep1
per_rep3


TRAIL Pathway
14
12
10
8
6
4
FLCN
UAP1L1
BACH2
CALML6
TAS2R5
IFNLR1
SCNN1D
WDR97
NFAM1
SLC7A9
BHMT
HVCN1
genes
per_rep1
per_rep2
per_rep3
par_rep1
par_rep2


WNT Pathway
12.5
10.0
7.5
5.0
2.5
0.0
CAMK1D
SEC31B
ZNF552
IRS1
SLC7A8
NTF3
KLK10
GFRA2
TLE4
KIF26B
EYA4
KCNJ10
RUBCNL
ANKRD1
PIEZO2
PLA2R1
GNAO1
SBSPON
TPH1
BMP4
RBMS2
APBB1
ITGB7
PRDM5
ALOX12
ANO2
genes
per_rep1
per_rep2
per_rep3
par_rep1
par_rep2


VEGF Pathway
12
10
8
6
4
SMCO3
C1QTNF4
NRG3
CNR1
ADGRD1
PIEZO2
ALDH1A1
UBE2Q2P1
ANKS1B
NID2
TTC28
RGL1
PTCH1
IMMP2L
LINC01355
genes
per_rep1
per_rep2
per_rep3
par_rep1
par_rep2


EGFR Pathway
14
12
10
8
MBTPS1
PNISR
PCMTD2
FOXN3
N4BP2L2
NR1D2
SYF2
ATG2B
KLHDC2
C2CD5
RCOR3
RNF146
EXD3
ATXN3
CLDN8
GOLPH3L
GRAMD1C
BBIP1
CARF
TTC33
TRIM52
WDR19
VPS9D1-AS1
CCDC28A
EFHC1
ZSCAN31
DPY19L1P1
ZNF224
KCNF1
LAMC2
ZBED8
genes
per_rep1
per_rep2
per_rep3
par_rep1
par_rep2