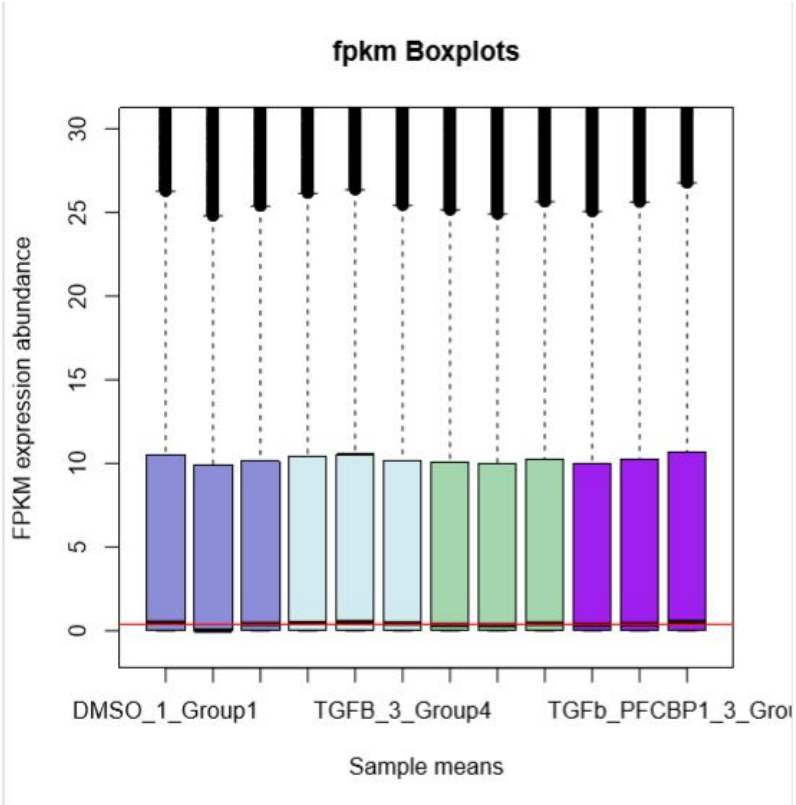
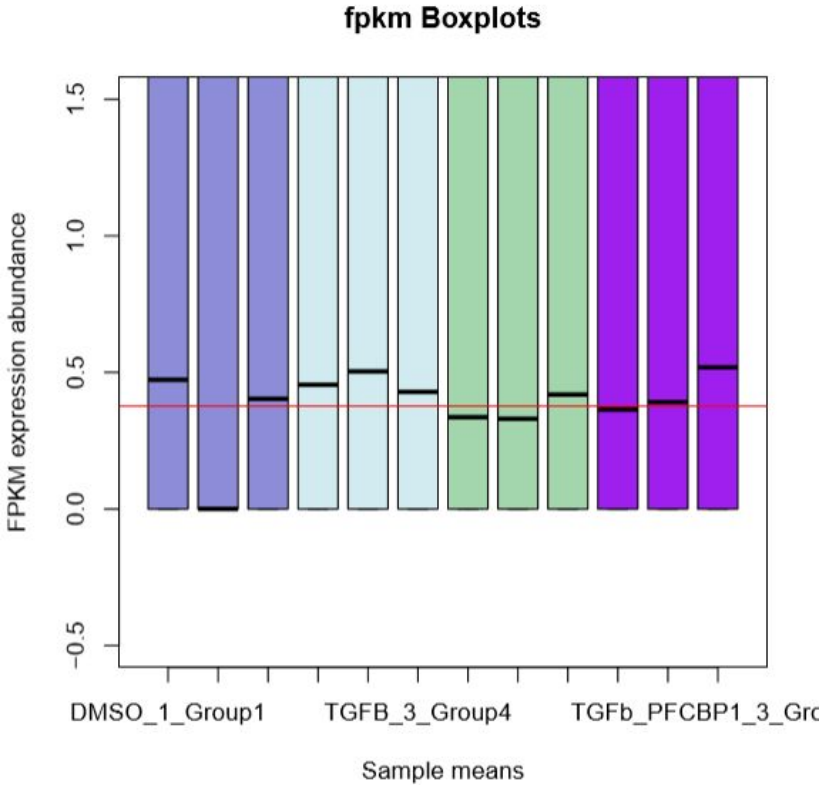
TGFb p2 update

Harrison

12/2/19

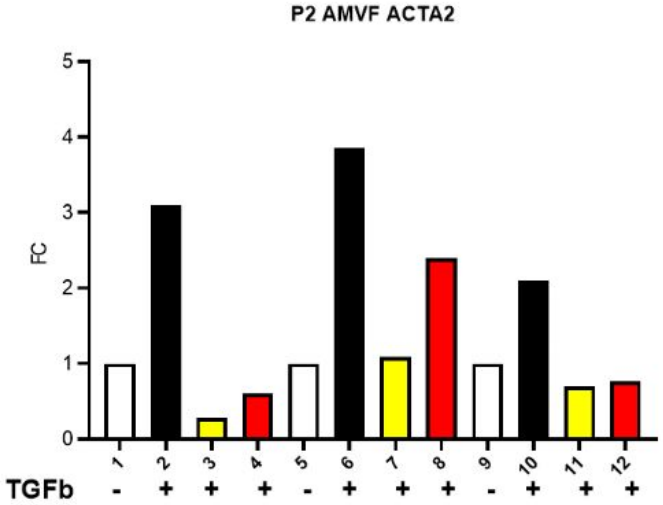
Boxplots of library size

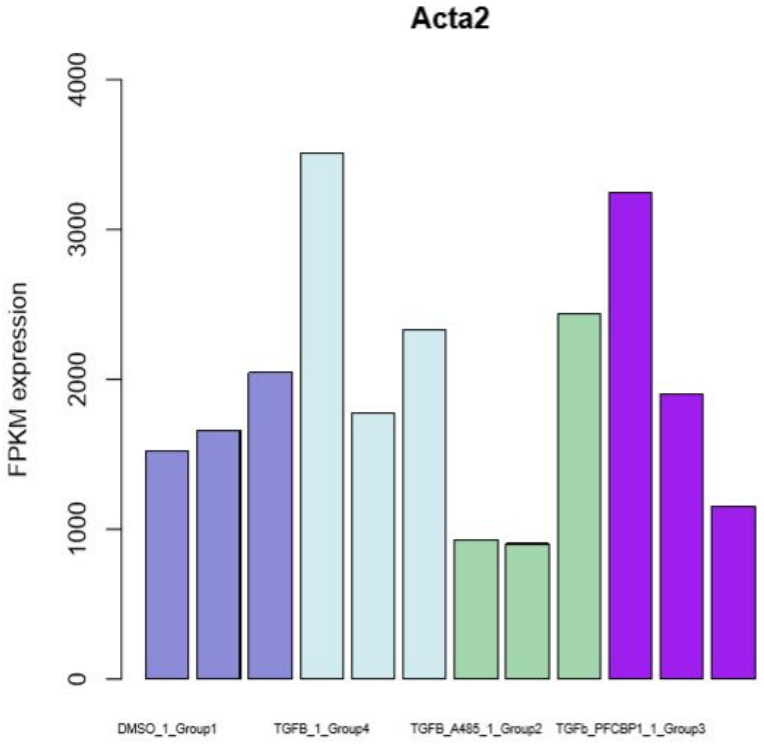




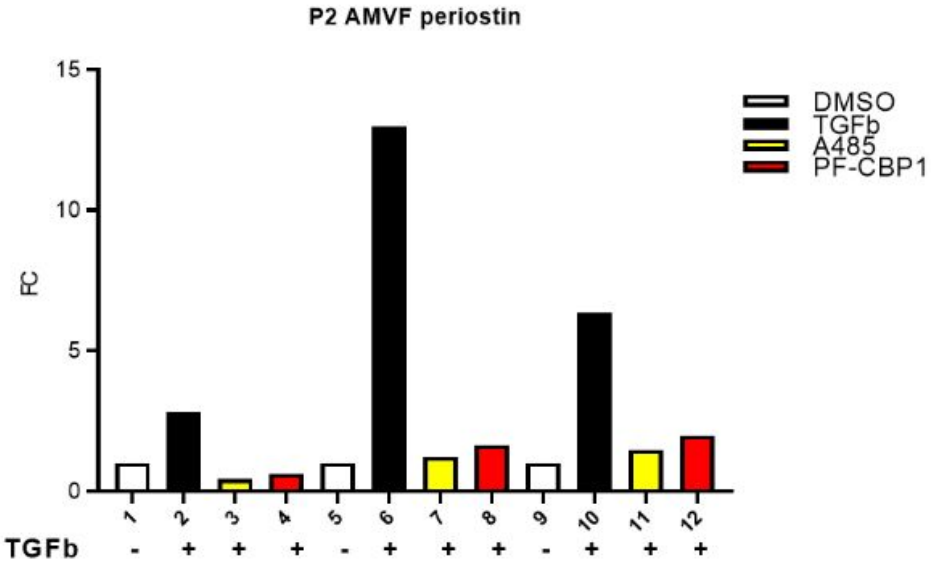
Validation

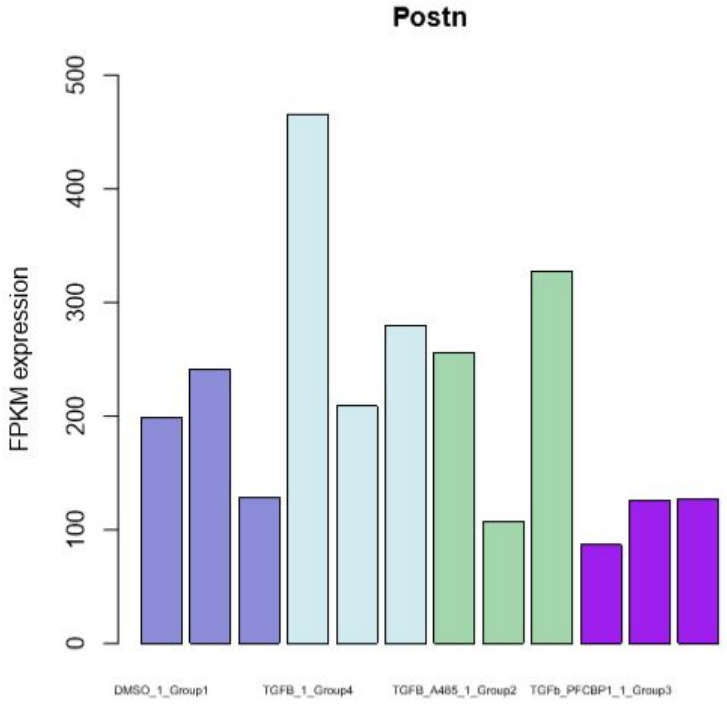
Although the replicates are in a different order I believe we see some similarity enough to ensure that the samples are labeled properly.



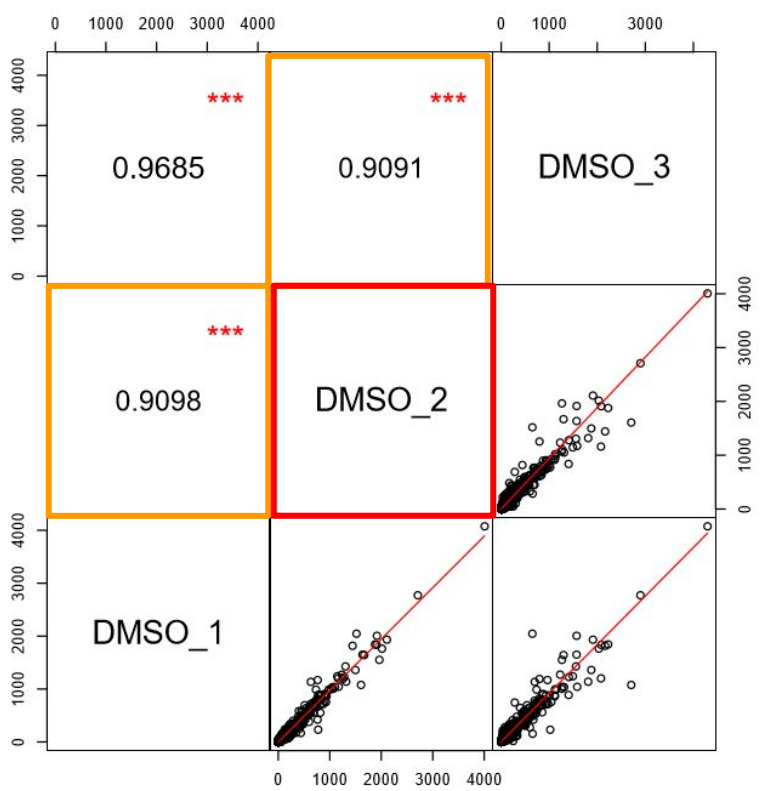


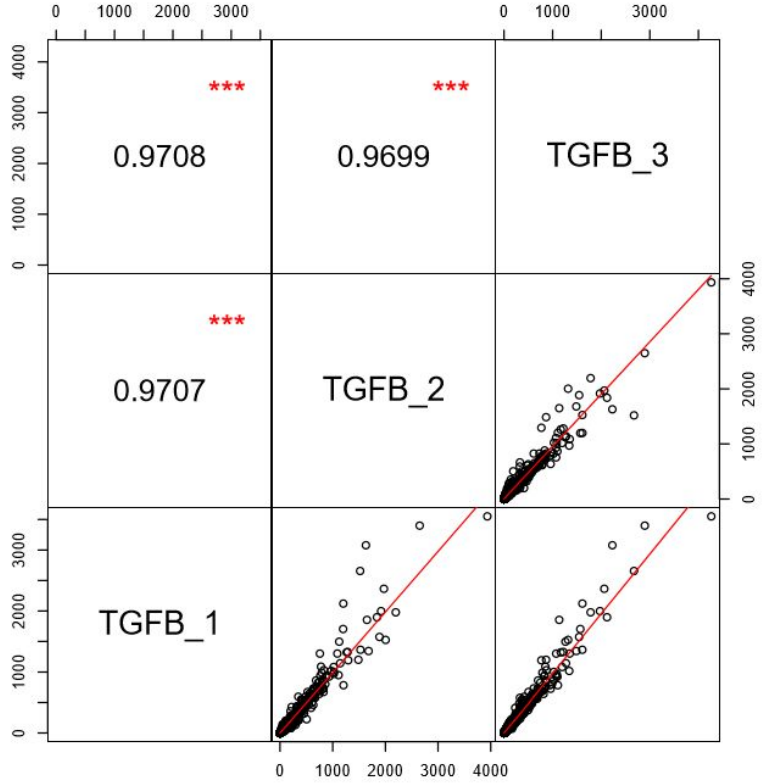
Validation


P2 AMVF periostin
FC
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12
TGFb - + + + - + + + - + + +
DMSO
TGFb
A485
PF-CBP1

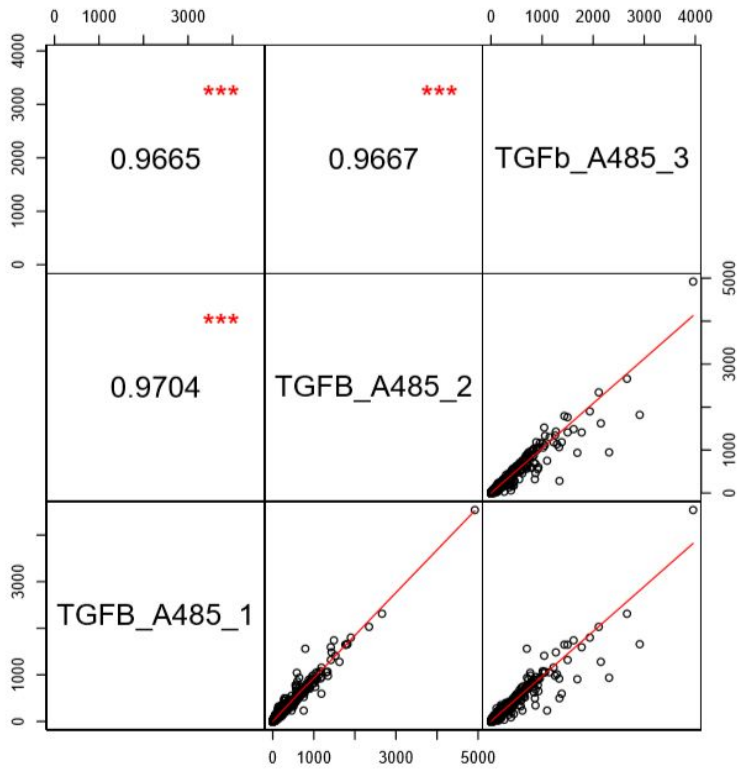

Postn
FPKM expression
0
100
200
300
400
500
DMSO_1_Group1
TGFB_1_Group4
TGFB_A485_1_Group2
TGFb_PFCBP1_1_Group3

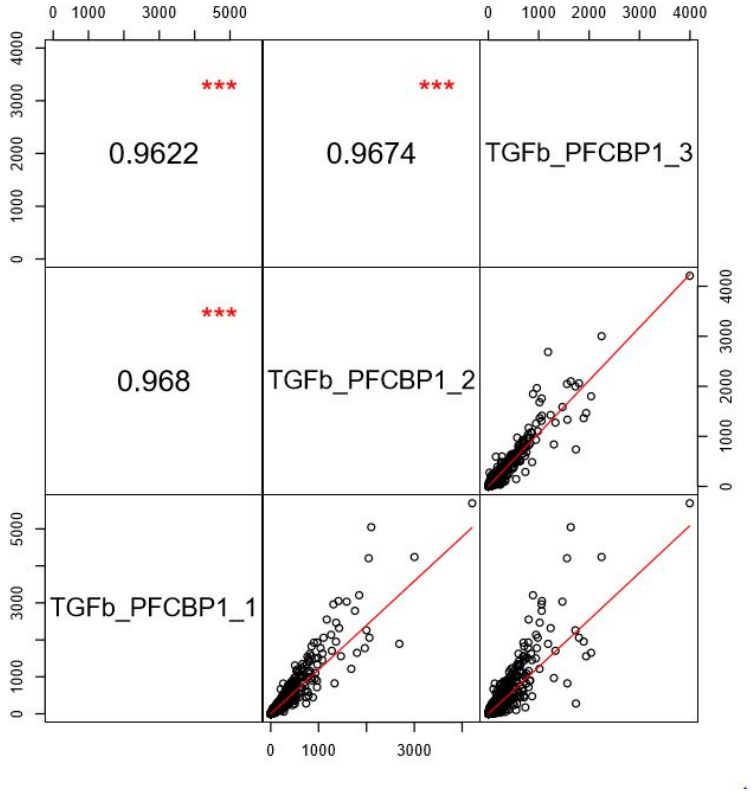
Pairwise Replicate Correlations (FPKM)





Pairwise Replicate Correlations (FPKM)



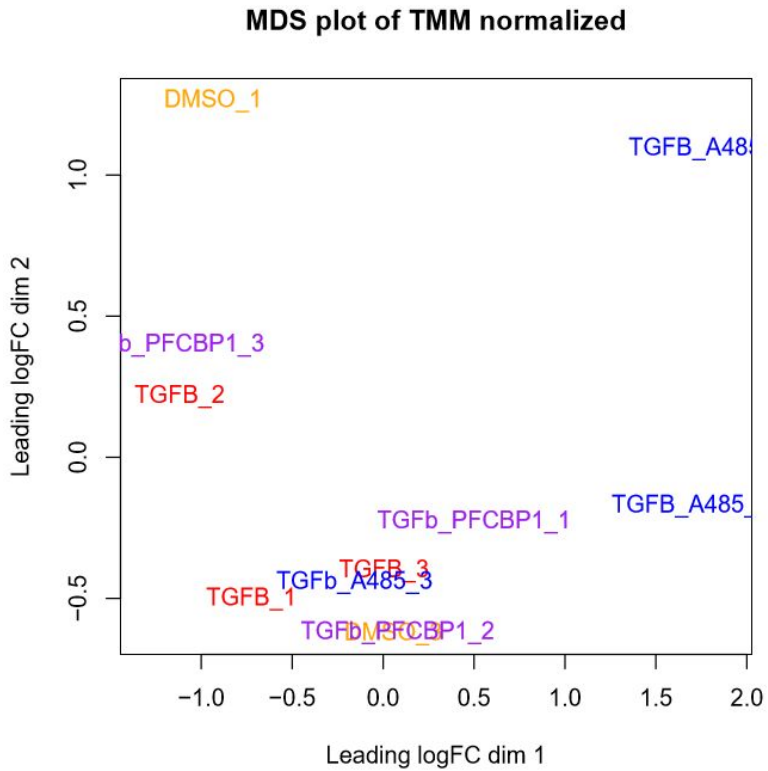


Multidimensional Scaling (MDS)

TMM normalized reads of the leading logFC genes among replicates.

Visualizes clustering of samples.

Post fixing of the samples and their barcodes we still see high variability between replicates.



DESeq2 analysis

Filtered genes for at least 10 counts summed across samples

Post dropping of sample #2 (DMSO_rep2)

Log Transformed counts

FDR < .05

Log2FC > |1|

As you can see, the data likely varies too greatly to get “truly” significant results.

At this point I would examine the pre-adjusted P-values or select an arbitrarily low number of genes that are close as possible to being significant.

The 1 gene downregulated in the TGFb vs TGFb+ PFCBP1 is **Postn.** This, we can be confident about.

DMSO vs TGFb

0 DEG

0 up regulated

0 downregulated

TGFb vs PFCBP1

0 DEG

0 up regulated

1 downregulated

TGFb vs A485

0 DEG

0 up regulated

0 downregulated

A485 vs PFCBP1

32 DEG

23 up regulated

9 downregulated

EdgeR

TMM normalization

Filtering for 10 counts summed across all samples

FDR < .05

Log2fc > |1|

Here we capture a few more DE genes under EdgeR's slightly less stringent approach.

Postn Is also seen in the TGFb vs TGFb+PFCBP1 comparison.

DMSO vs TGFb

5 DEG

2 up regulated

3 downregulated.

TGFb vs A485

0 DEG

0 up regulated

0 downregulated.

TGFb vs PFCBP1

3 DEG

0 up regulated

3 downregulated.

A485 vs PFCBP1

32 DEG

23 up regulated

9 downregulated

DESeq2 analysis (P-value)

Filtered genes for at least 10 counts summed across samples

Post dropping of sample #2 (DMSO_rep2)

Log Transformed counts

P-value < .05

Log2FC > |1|

These genes are identified as significant with a fold change equivalent to 1 fold higher or double the expression value inbetween experimental groups.

There are more genes that express more subtle changes

DMSO vs TGFb

106 DEG

64 up regulated

42 downregulated

TGFb vs PFCBP1

98 DEG

35 up regulated

63 downregulated

TGFb vs A485

438 DEG

50 up regulated

388 downregulated

A485 vs PFCBP1

422 DEG

347 up regulated

75 downregulated

DESeq2 analysis (FDR)

Filtered genes for at least 10 counts summed across samples

Post dropping of sample #2 (DMSO_rep2)

Log Transformed counts

FDR < **.20** (less reliable)

Log2FC > **|.5|** (small change)

From these genes I can create a heatmap, but I would make the A485 vs PFCBP1 in a separate heatmap or it can be or combined with the rest if you'd like.

For now, I will use the top DE genes for a heatmap.

There are potentially more interesting genes with slightly higher FDR however that's not something I can assume or consider myself the authority to do so before consulting you.

DMSO vs TGFb

15 DEG

7 up regulated
8 downregulated

TGFb vs A485

0 DEG

but

FDR .30-.31 gave 3 genes (lowest FDR)

2 up and 1 down

TGFb vs PFCBP1

3 DEG

0 up regulated
3 downregulated

A485 vs PFCBP1

132 DEG

104 up regulated
28 downregulated

Volcano Plots

DMSO vs TGFb

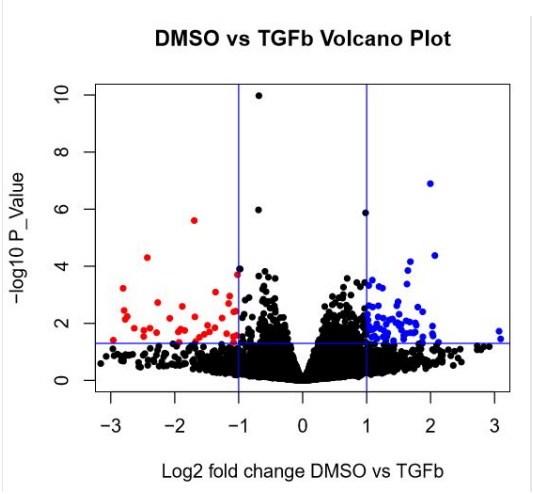
Here we can see the resulting volcano plot of our data and a histogram to view the distribution of values.

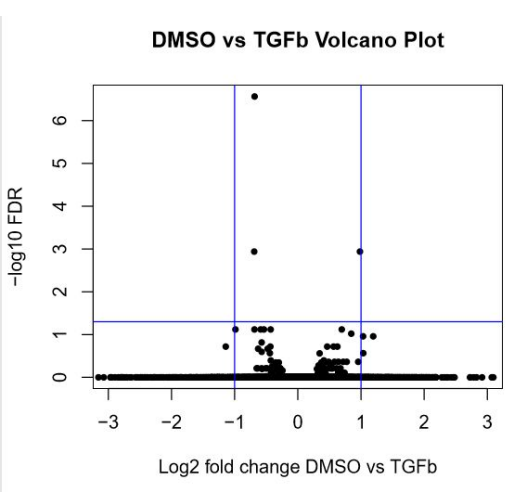
The volcano plots reflect the differential expression selection we made for genes examining both P-values and FDR values. The blue lines indicate the cutoffs we are using.

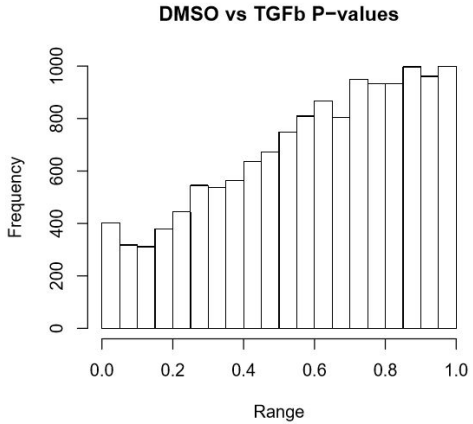
Log2FC > |1| and P-value or FDR < .05

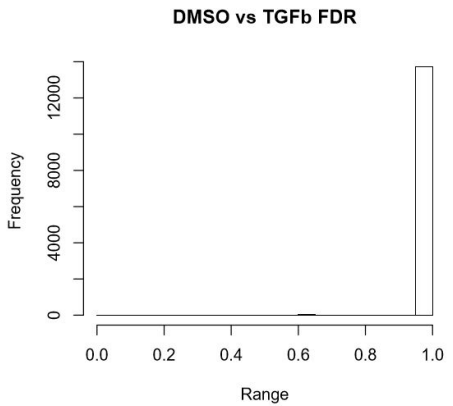
-log10 of the P-values makes them easier to observe

All the plots are more of less the same.





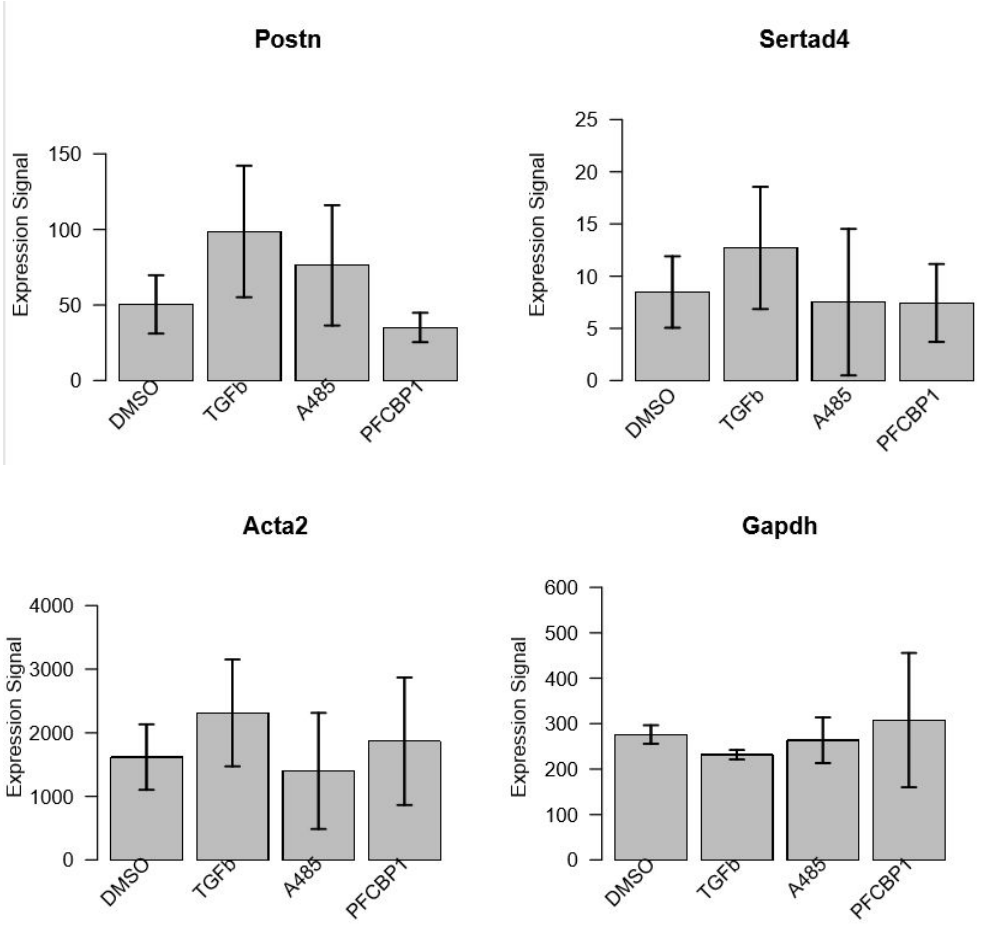




Expression Plots

Here we can see clear upregulation of genes we would expect to change as an indication of fibroblast activation.

We can also examine more genes that are related to TGFb either by canonical or non canonical pathways which might also provide indication of TGFb treatment efficacy.

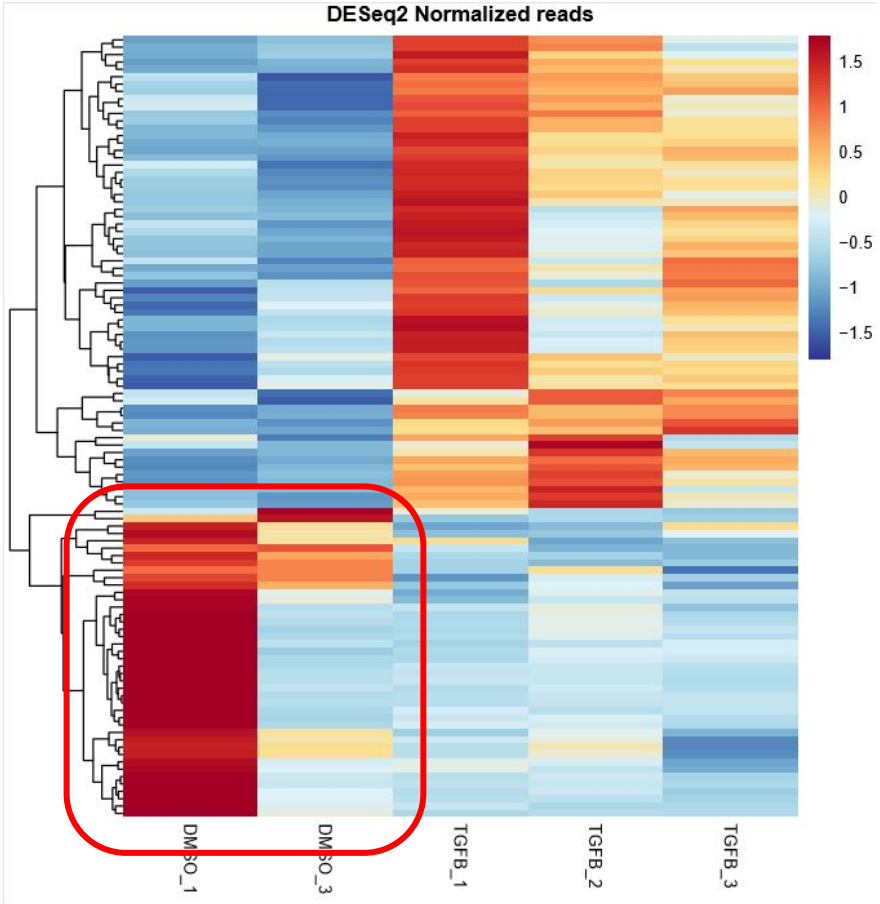


HeatMap

For P-value < .05 and LogFC > |1|

This represents the 106 genes changing in the TGFb compared to the DMSO group.

We can see clear distinction however the replicate does not show consistency among all changing genes.

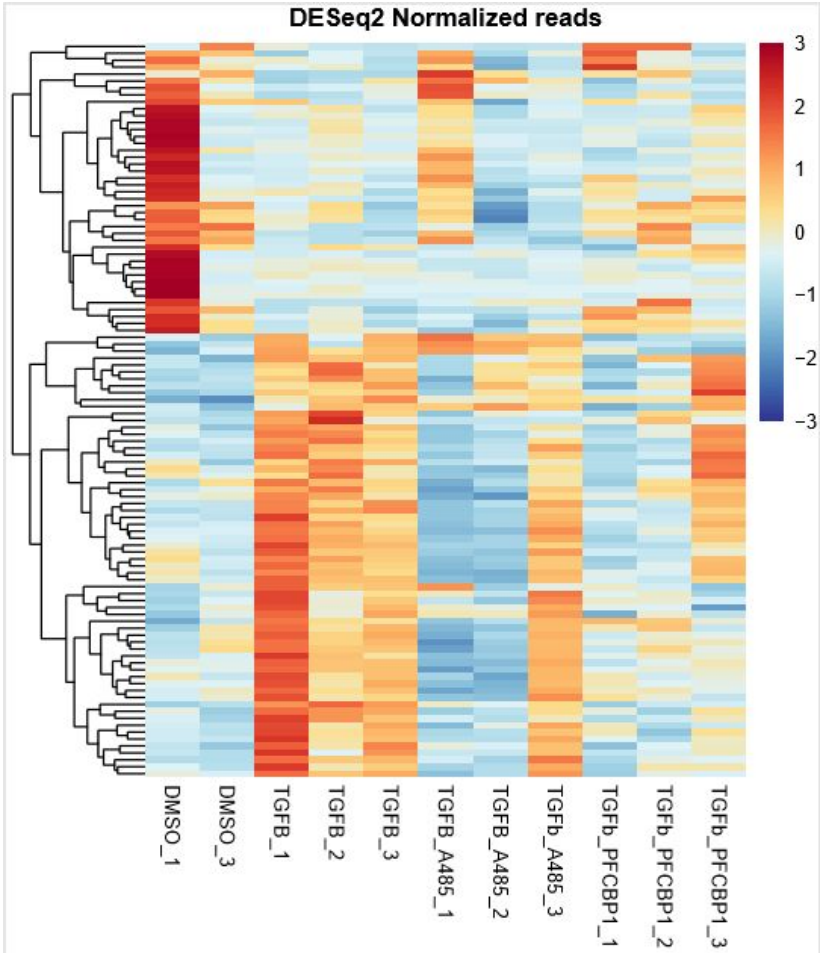


HeatMap

For P-value < .05 and LogFC > |1|

This represents the 106 genes changing in the TGFb compared to the DMSO group but we are now viewing the same genes in all experimental groups.

We may examine any subset of these genes if desired.



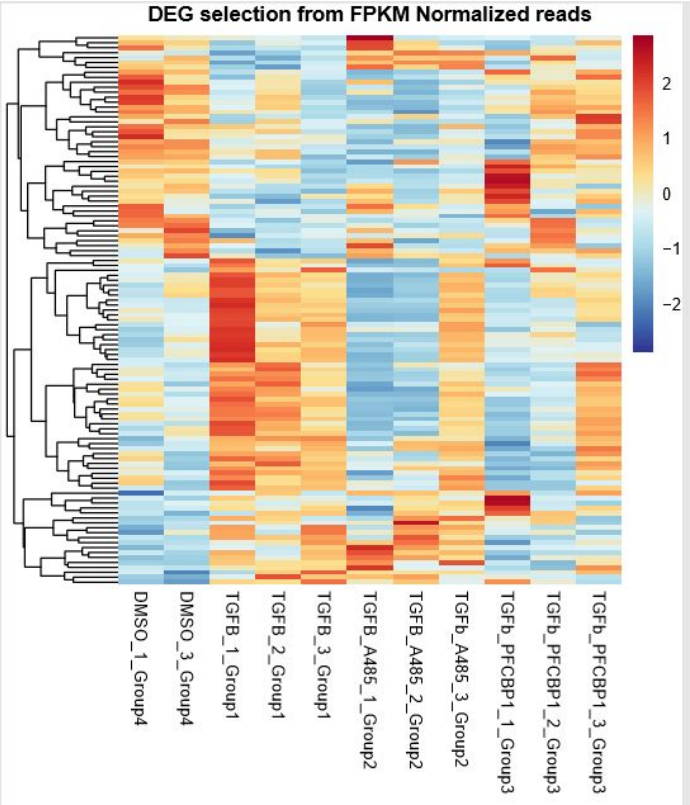
Here I performed DE analysis using FPKM normalized reads.

I found 111 genes using :

Pvalue < .15

Log2fc > |.58|

Similar looking heatmap however ~75% do not overlap with the DESeq2 found genes. We can view downstream results as well but I think it would be wiser to use the DESeq2 selected genes.



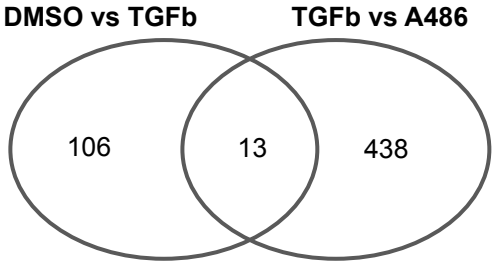
HeatMap (A485 effects)

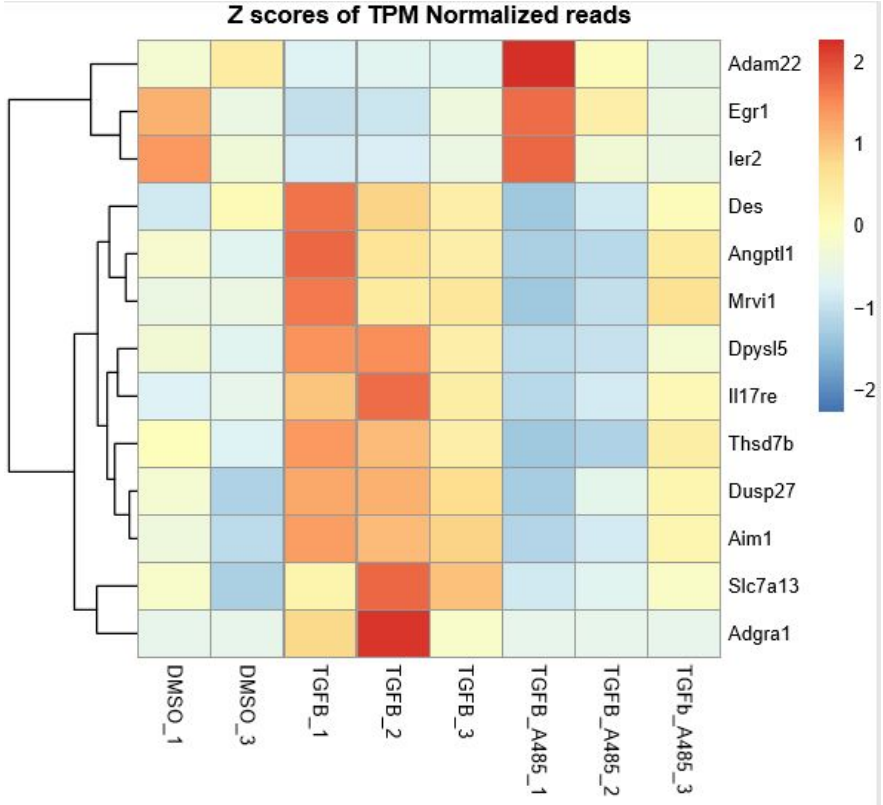
For P-value < .05 and LogFC > |1|

Here I took the genes changing in the A485 treatment and intersected the list with the genes changing in the DMSO vs TGFb comparison

This is to focus on genes that affected by TGFb and also affected by the A485 inhibitor.

I retain all other affected genes as “side effect genes” and it is likely that there or more with genes with subtle changes.



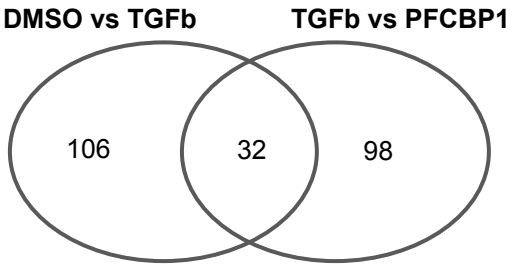


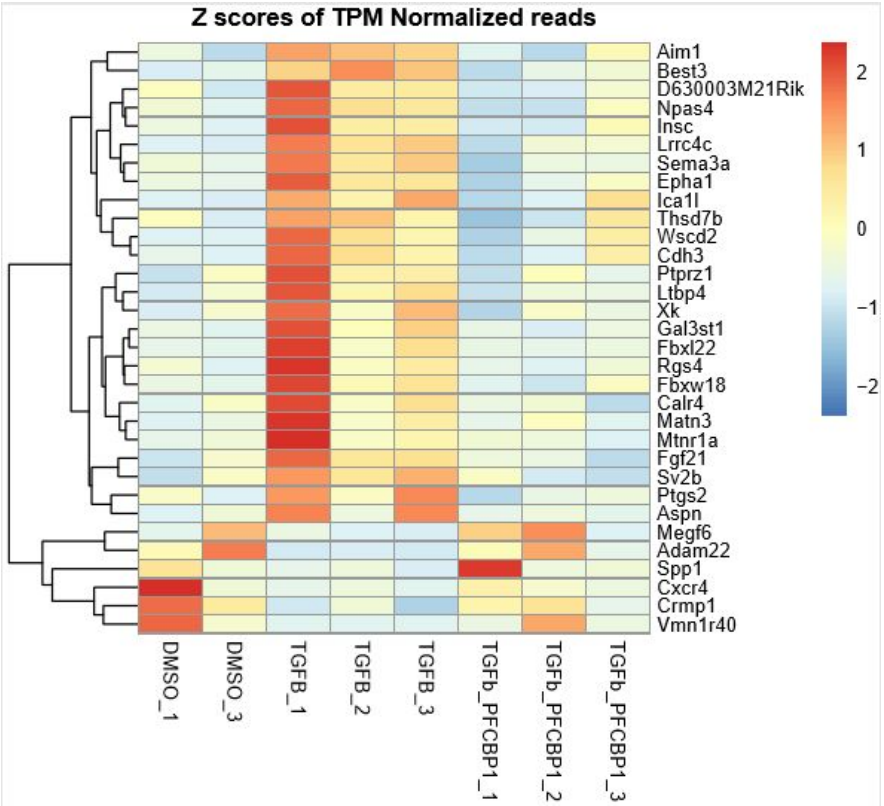
HeatMap (PFCBP1 effects)

For P-value < .05 and LogFC > |1|

Here I took the genes changing in the PFCBP1 treatment and intersected the list with the genes changing in the DMSO vs TGFb comparison

This is to focus on genes that affected by TGFb and also affected by the PFCBP1 inhibitor.





HeatMap (effective genes of PFCBP1 and A485)

Utilizing the same parameters we then view the the same genes that are affected by both inhibitors

4 of which overlap between inhibitor groups

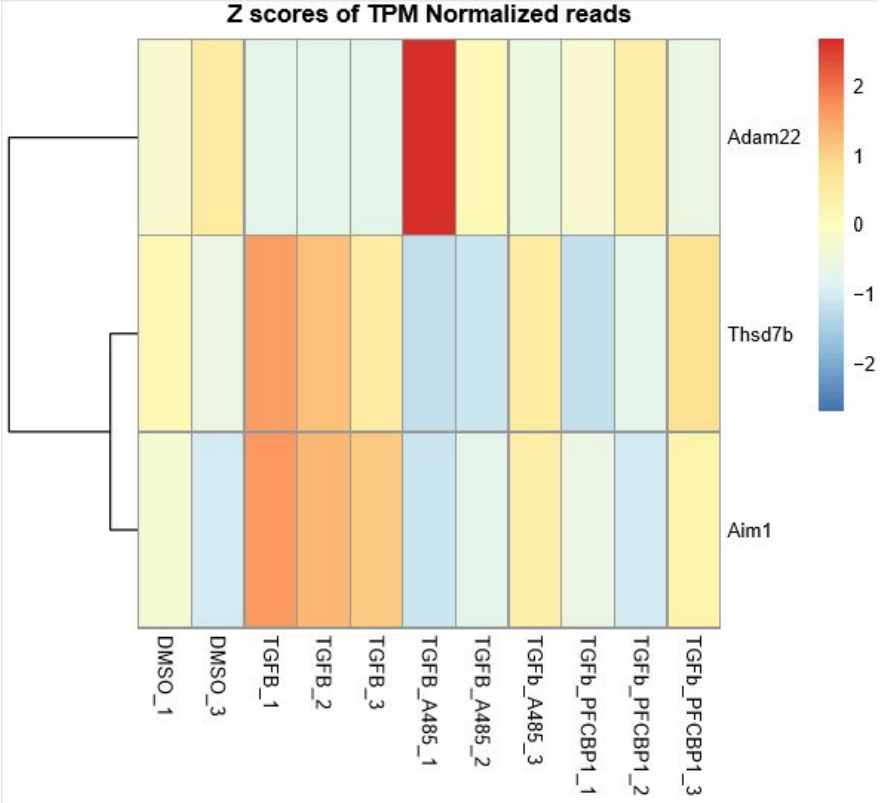
3 of these are also found to be DE in the TGFb treated cells compared to the DMSO group.

Adam22

Aim1

Thsd7b

Osr1

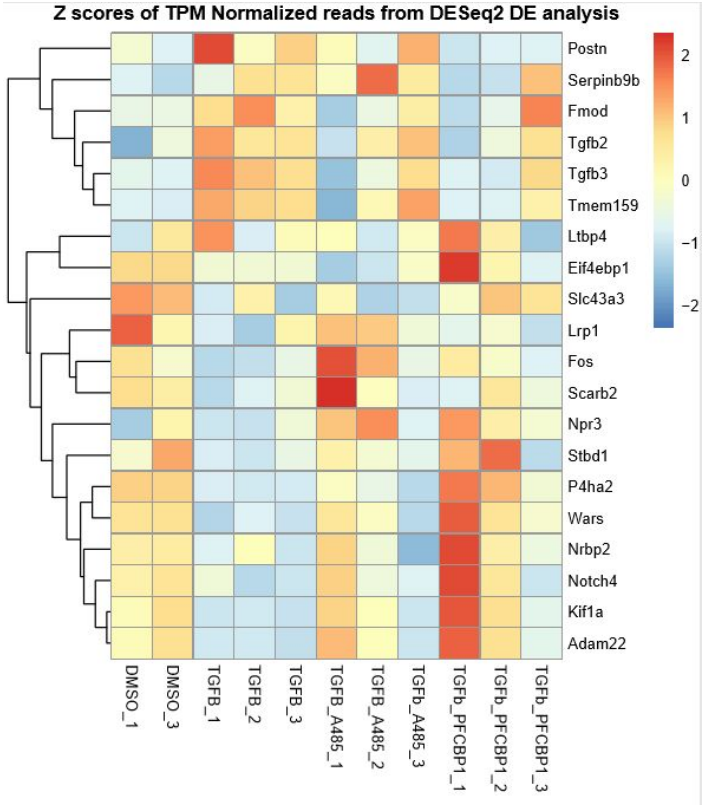


HeatMap

For FDR < .2 and LogFC > |.5|

(except for the top 3 choices in the TGFb vs A485 comparison)

These are the genes identified as the most significant selection we can make based on the more stringent statistical FDR metric



GSEA results

Here we can take quick look at the GSEA results.

DMSO vs TGFb
We can see that a few more gene sets are enriched for pathways in the DMSO group.
Doesn't mean these are all negative pathways just ones that may be standard with genes more upregulated with this group.

TGFb vs A485
We can see that in there is quite the strong correlation with genes enriched in TGFB as opposed to the A485 group. My assumption is that this is because of how many disease state changing genes and more are blunted due the broad inhibition.

TGFb vs PFCBP1
Here we can see a fairly equal amount of pathways enriched between these groups but is worth noting the difference from how much of an impact the A485 inhibitor had.

Enrichment in phenotype: DMSO (1 samples)

- 1897 / 3266 gene sets are upregulated in phenotype **DMSO**
- 79 gene sets are significant at FDR < 25%
- 116 gene sets are significantly enriched at nominal pvalue < 1%
- 298 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: TGFB (1 samples)

- 1369 / 3266 gene sets are upregulated in phenotype **TGFB**
- 19 gene sets are significantly enriched at FDR < 25%
- 63 gene sets are significantly enriched at nominal pvalue < 1%
- 167 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

GSEA results

Enrichment in phenotype: TGFB (1 samples)

- 1567 / 3195 gene sets are upregulated in phenotype **TGFB**
- 547 gene sets are significant at FDR < 25%
- 231 gene sets are significantly enriched at nominal pvalue < 1%
- 453 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: TGFB (1 samples)

- 2911 / 3195 gene sets are upregulated in phenotype **TGFB**
- 1629 gene sets are significant at FDR < 25%
- 862 gene sets are significantly enriched at nominal pvalue < 1%
- 1255 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: PFCBP1 (1 samples)

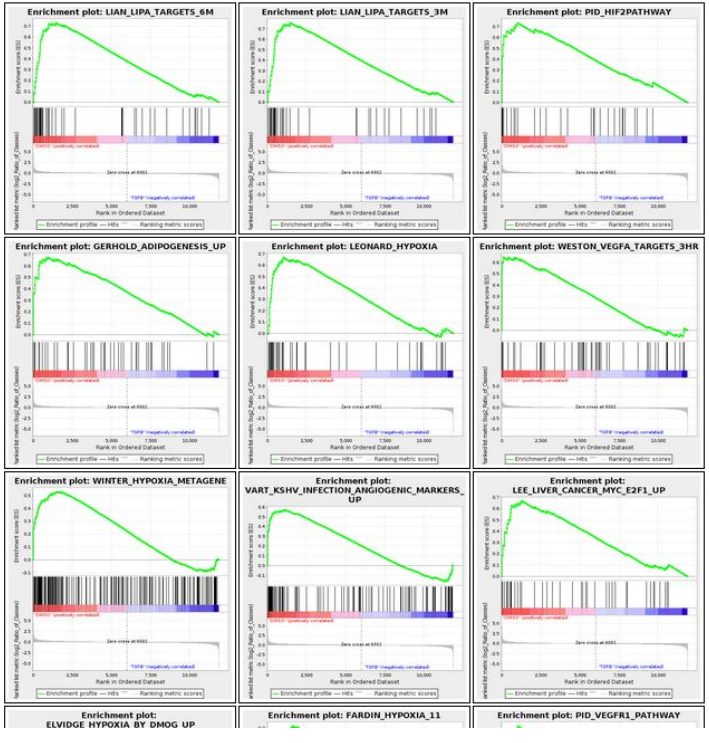
- 1628 / 3195 gene sets are upregulated in phenotype **PFCBP1**
- 550 gene sets are significantly enriched at FDR < 25%
- 263 gene sets are significantly enriched at nominal pvalue < 1%
- 464 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

Enrichment in phenotype: A485 (1 samples)

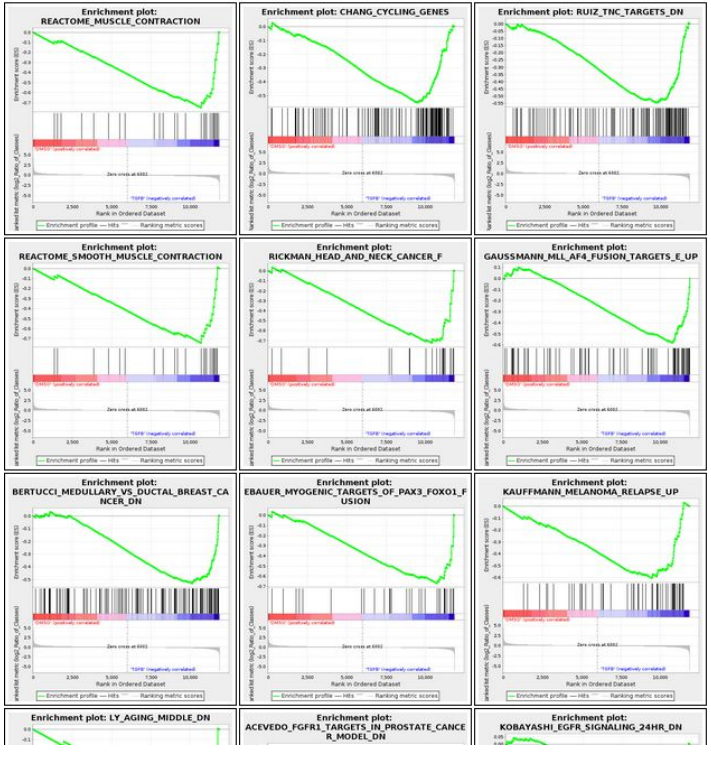
- 284 / 3195 gene sets are upregulated in phenotype **A485**
- 34 gene sets are significantly enriched at FDR < 25%
- 32 gene sets are significantly enriched at nominal pvalue < 1%
- 46 gene sets are significantly enriched at nominal pvalue < 5%
- Snapshot of enrichment results
- Detailed enrichment results in html format
- Detailed enrichment results in excel format (tab delimited text)
- Guide to interpret results

GSEA (Post dropping DMSO replicate 2)

DMSO

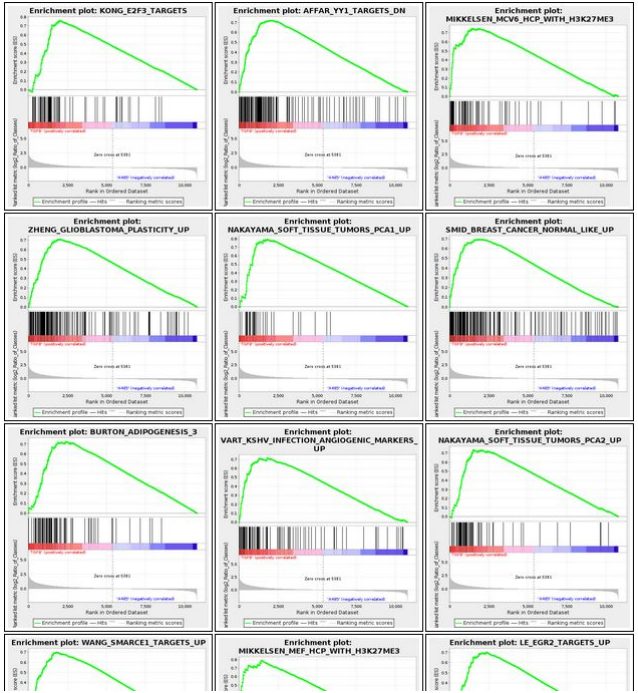


TGFB

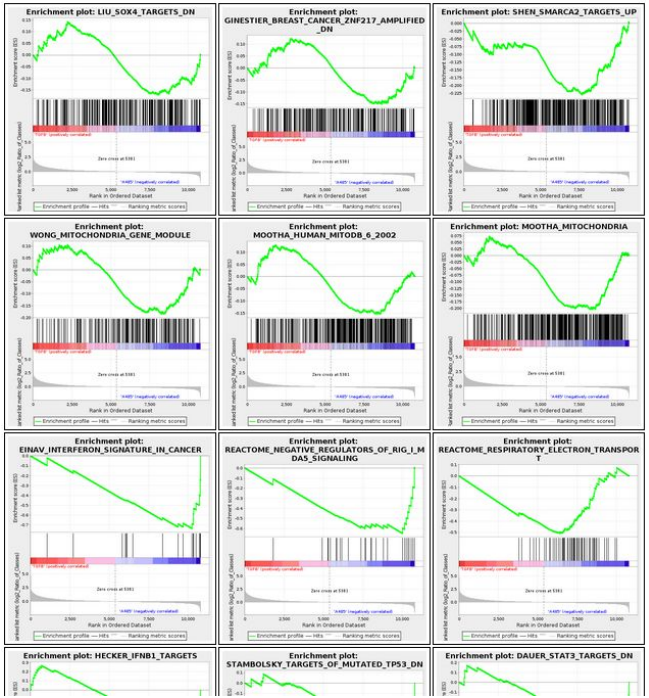


GSEA

TGFb

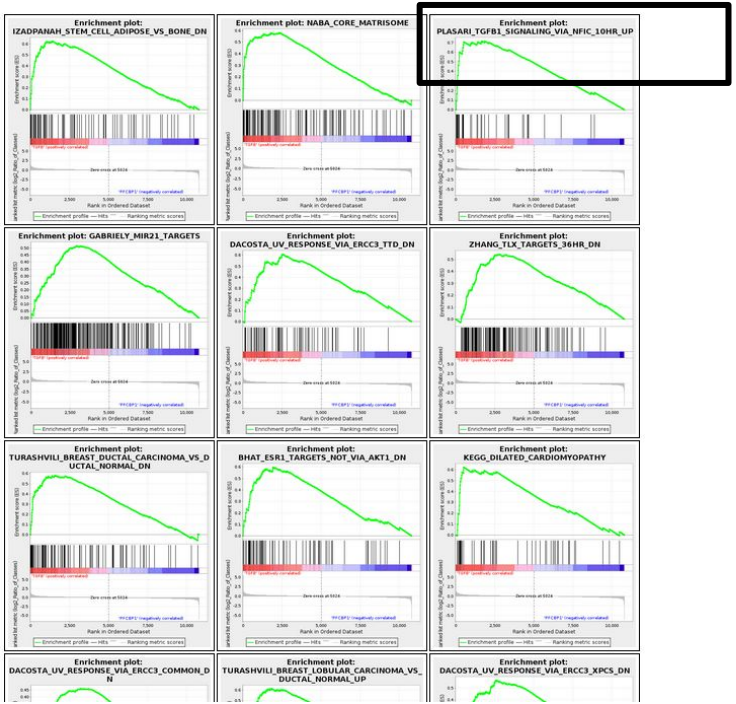


TGFb + A485

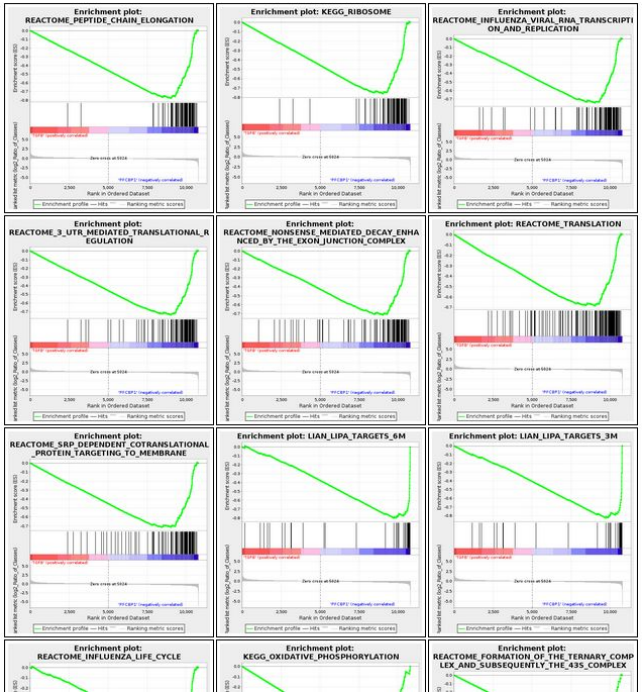


GSEA

TGFb



TGFb + PFCBP1



GSEA (analysis of DMSO vs TGFb)

We want to validate that the treatment is working. We can start by identifying the most obvious indication as per identified pathways and then broaden our selection to related pathways that involve cell proliferation, inflammation, invasiveness and mesenchymal transition.

Here we view all pathways in the results with TGFb in the title of the pathway in all comparisons.

This does not tell us the efficacy of the treatment but rather a quick sign that we have an increase in TGFb related pathways in that sample.

The next thing we can do is examine the top pathways in both groups and decide on criteria that indicates what we are looking for.

Ultimately we can then study all pathways involved in the top enriched groups and decide if there is a clear distinction between samples that support effective treatment.

Pathways that just have TGFB in the title:

The # indicates the ranking of the gene set among all found.

DMSO- 14 total ->236,394,464, 472, 591, 647, 664, 985,992, 1017, 1048, 1091, 1384,1669

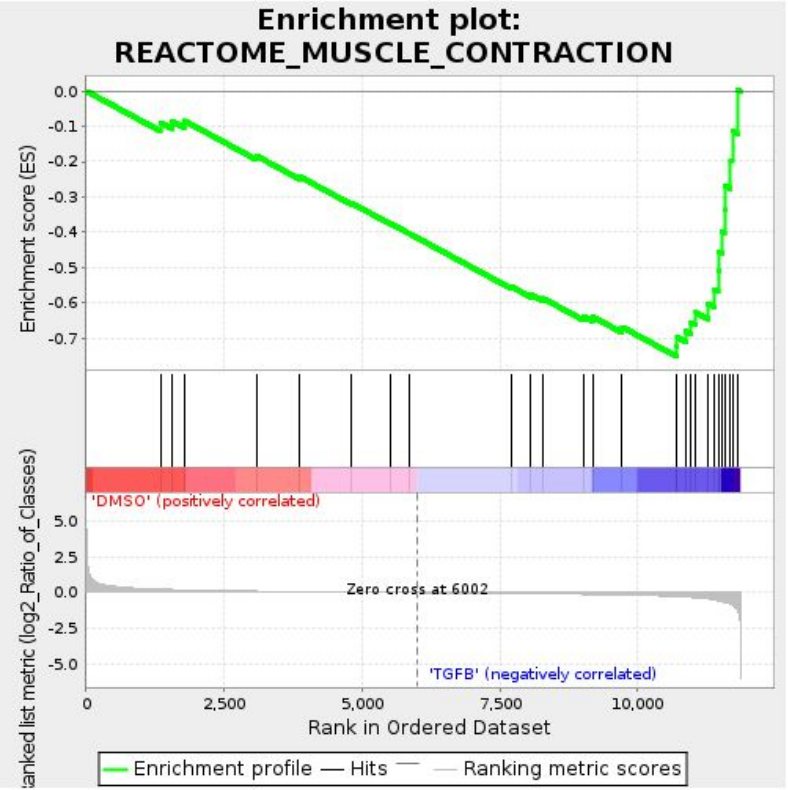
TGFB - 15 total ->107,135, 199, 326, 330, 368, 583, 698, 785, 837, 998, 1029, 1071,1119,

Top GSEA Enrichment Plots

Reactome Muscle Contraction

This is the top most significantly enriched pathways in the TGFb sample vs DMSO.

Here we can see there is a distinct enrichment for genes in this muscle contraction pathway.



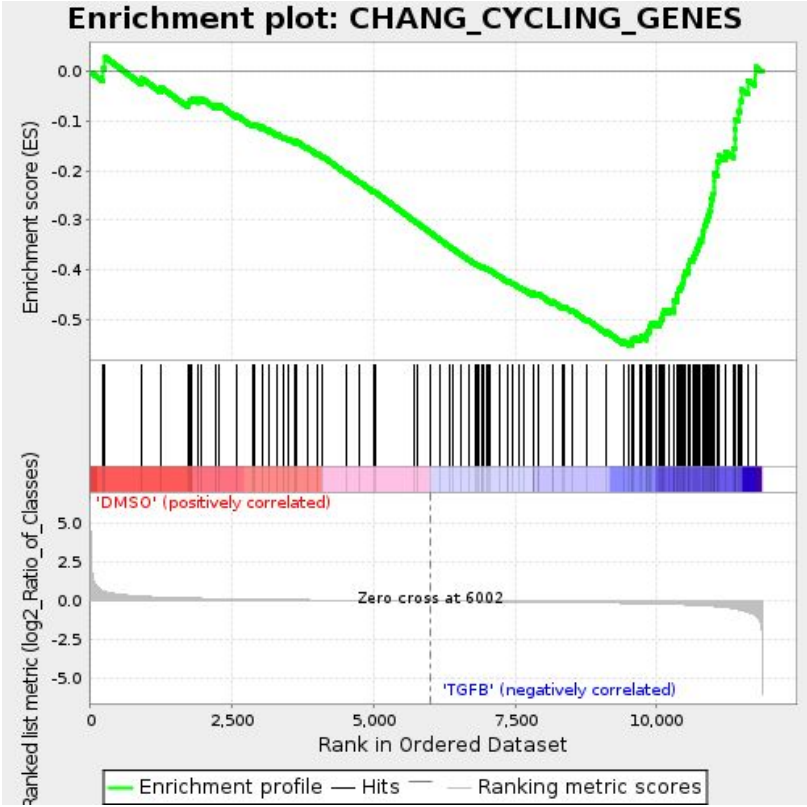
Enrichment Plots

Chang Cycling Genes

The is the second top enriched pathway for TGFb.

As per the GSEA description, this pathway is represented as Fibroblast serum response genes.

These genes are related to the pathway that reflects the multifaceted role of fibroblasts in wound healing.



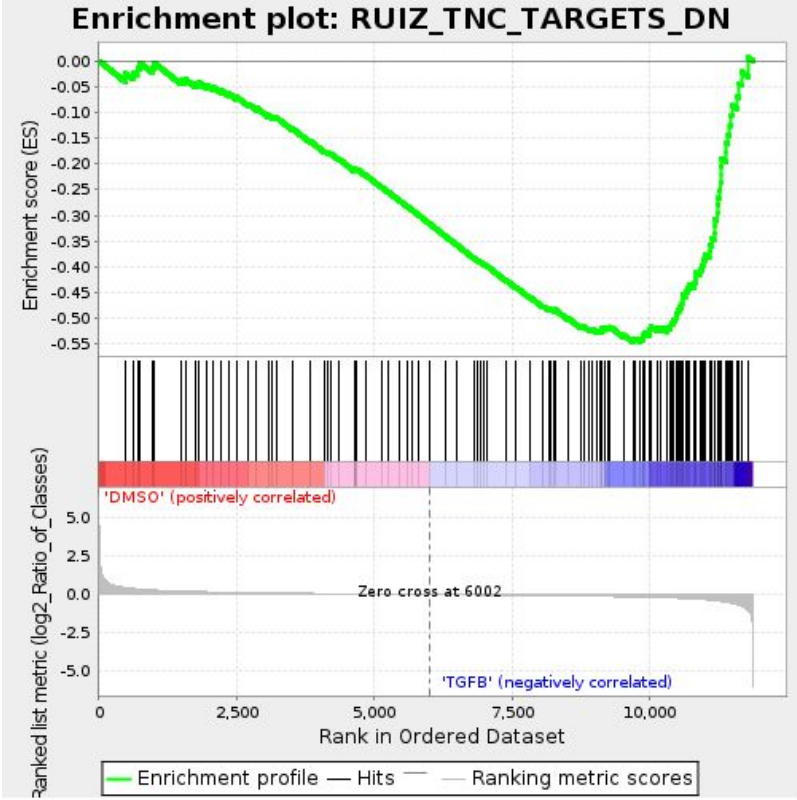
Enrichment Plots

RUIZ TNC Targets DN

The 3rd most enriched pathway.

This pathways is denoted by genes that were found to induce extracellular matrix and tumor cell proliferation.

We can explore all other sets. We can also take the same approach used to produce figures created for the givinostat project if need be.



Gene Ontology (GOrilla) Biological Process

For GO analysis I enjoy the simplicity of the GOrilla web based app.

It allows you to view pathways identified from a subset of genes in a relatively simple fashion.

I used the 106 DE genes identified in the from the TGFb treatment and used the remaining ~13,500 or so genes as background.

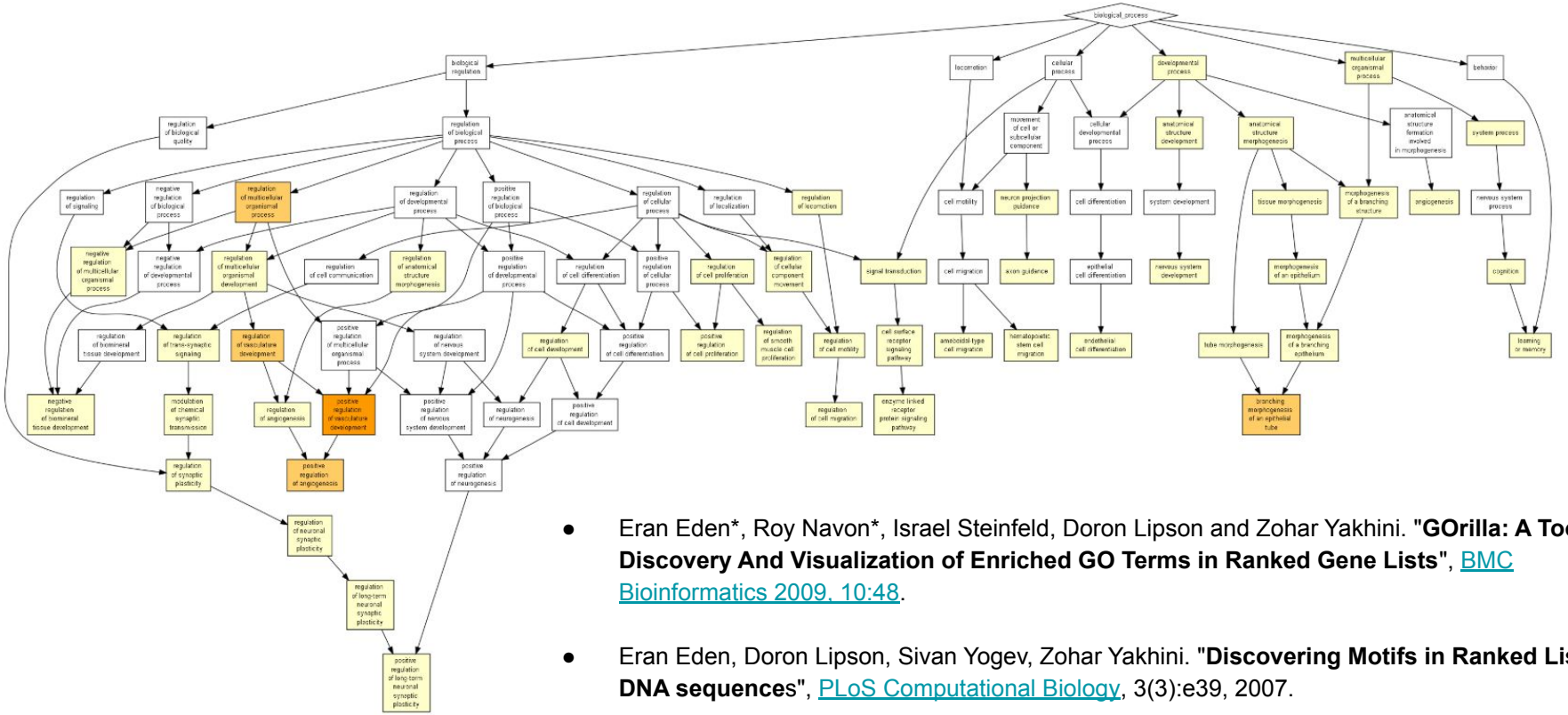
Here we are examining the top biological processes associated with the subsets. We will also examine cellular component and molecular function domains.

N -> Total background genes
B -> All genes associated with process
n -> Genes identified from DE analysis
b -> Intersection of B and n

Enrichment score = (b/n)/(B/N)

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:1904018	positive regulation of vasculature development	7.18E-8	1.04E-3	8.42 (13215,166,104,11)	[-] Hide genes Kdr - kinase insert domain protein receptor Pgf - placental growth factor Cx3cr1 - chemokine (c-x3-c) receptor 1 Ptgs2 - prostaglandin-endoperoxide synthase 2 Nos3 - nitric oxide synthase 3, endothelial cell Epha1 - eph receptor a1 Cxcr4 - chemokine (c-x-c motif) receptor 4 Sfrp2 - secreted frizzled-related protein 2 Kit - kit oncogene Egr1 - early growth response 1 Hmga2 - high mobility group at-hook 2
GO:1901342	regulation of vasculature development	1.9E-6	1.38E-2	5.47 (13215,279,104,12)	[-] Hide genes Vash1 - vasohibin 1 Kdr - kinase insert domain protein receptor Pgf - placental growth factor Cx3cr1 - chemokine (c-x3-c) receptor 1 Ptgs2 - prostaglandin-endoperoxide synthase 2 Nos3 - nitric oxide synthase 3, endothelial cell Epha1 - eph receptor a1 Cxcr4 - chemokine (c-x-c motif) receptor 4 Kit - kit oncogene Sfrp2 - secreted frizzled-related protein 2 Egr1 - early growth response 1 Hmga2 - high mobility group at-hook 2
GO:0045766	positive regulation of angiogenesis	2.7E-6	1.31E-2	7.62 (13215,150,104,9)	[-] Hide genes Kdr - kinase insert domain protein receptor Pgf - placental growth factor Cx3cr1 - chemokine (c-x3-c) receptor 1 Ptgs2 - prostaglandin-endoperoxide synthase 2 Nos3 - nitric oxide synthase 3, endothelial cell Epha1 - eph receptor a1 Cxcr4 - chemokine (c-x-c motif) receptor 4 Sfrp2 - secreted frizzled-related protein 2 Hmga2 - high mobility group at-hook 2
GO:0048754	branching morphogenesis of an epithelial tube	5.47E-6	1.99E-2	8.26 (13215,123,104,8)	[-] Hide genes Lama1 - laminin, alpha 1 Kdr - kinase insert domain protein receptor Pgf - placental growth factor Cxcr4 - chemokine (c-x-c motif) receptor 4 Sfrp2 - secreted frizzled-related protein 2 Rasip1 - ras interacting protein 1 Hmga2 - high mobility group at-hook 2 Dchs1 - dachsous 1 (drosophila)
GO:0051239	regulation of multicellular organismal process	9.77E-6	2.84E-2	2.01 (13215,2341,104,37)	[+] Show genes
GO:0048729	tissue morphogenesis	1.03E-5	2.5E-2	4.63 (13215,329,104,12)	[+] Show genes
GO:0061138	morphogenesis of a branching epithelium	1.42E-5	2.95E-2	7.26 (13215,140,104,8)	[+] Show genes
GO:0001525	angiogenesis	1.58E-5	2.87E-2	5.41 (13215,235,104,10)	[+] Show genes
GO:0001763	morphogenesis of a branching structure	2.13E-5	3.44E-2	6.87 (13215,148,104,8)	[+] Show genes
GO:0045765	regulation of angiogenesis	2.69E-5	3.92E-2	5.08 (13215,250,104,10)	[+] Show genes
GO:0042127	regulation of cell proliferation	2.86E-5	3.78E-2	2.44 (13215,1248,104,24)	[+] Show genes

Gene Ontology (GOrilla) Biological Process

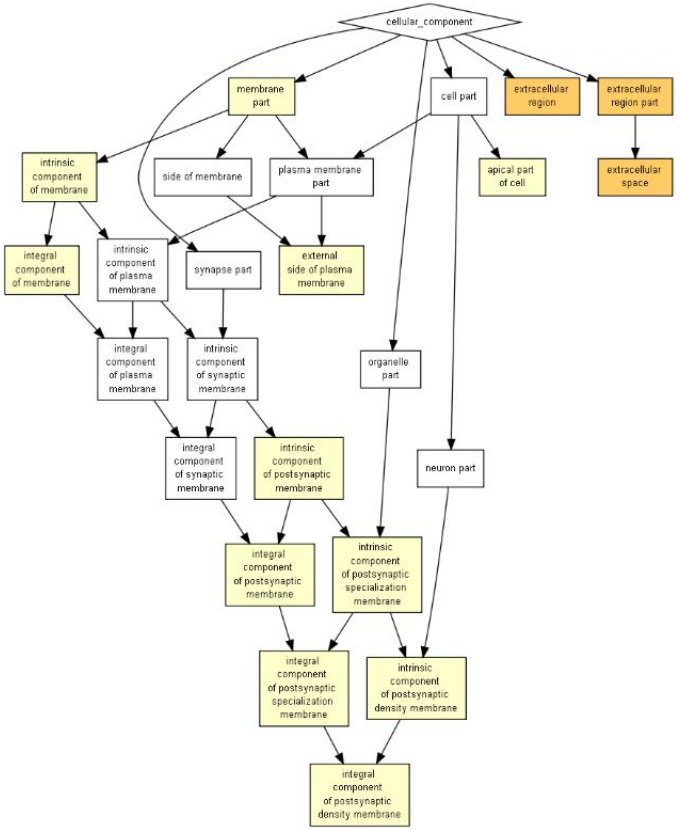


- Eran Eden*, Roy Navon*, Israel Steinfeld, Doron Lipson and Zohar Yakhini. "**GOrilla: A Tool For Discovery And Visualization of Enriched GO Terms in Ranked Gene Lists**", BMC Bioinformatics 2009, 10:48.

- Eran Eden, Doron Lipson, Sivan Yogev, Zohar Yakhini. "**Discovering Motifs in Ranked Lists of DNA sequences**", PLoS Computational Biology, 3(3):e39, 2007.

Gene Ontology (GOrilla) Cellular Component

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0005615	extracellular space	5.16E-7	9.78E-4	3.44 (13215,775,104,21)	[-] Hide genes Lrrc32 - leucine rich repeat containing 32 Ltbp4 - latent transforming growth factor beta binding protein 4 Angptl1 - angiopoietin-like 1 Sfrp2 - secreted frizzled-related protein 2 Kit - kit oncogene Fgf21 - fibroblast growth factor 21 Serpinb9b - serine (or cysteine) peptidase inhibitor, clade b, member 9b Apln - apelin Matn3 - matrilin 3 Vash1 - vasohibin 1 Lama1 - laminin, alpha 1 Adamts4 - a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 4 Stc1 - stanniocalcin 1 Pgf - placental growth factor Ptprz1 - protein tyrosine phosphatase, receptor type z, polypeptide 1 Prl3a1 - prolactin family 3, subfamily a, member 1 Spp1 - secreted phosphoprotein 1 Serpinb1c - serine (or cysteine) peptidase inhibitor, clade b, member 1c Sparcl1 - sparc-like 1 Clec11a - c-type lectin domain family 11, member a Sema3a - sema domain, immunoglobulin domain (ig), short basic domain, secreted, (semaphorin) 3a
GO:0044421	extracellular region part	2.68E-6	2.54E-3	2.90 (13215,1008,104,23)	[+] Show genes
GO:0005576	extracellular region	4.71E-6	2.97E-3	3.11 (13215,816,104,20)	[+] Show genes
GO:0099060	integral component of postsynaptic specialization membrane	2.08E-5	9.85E-3	14.78 (13215,43,104,5)	[+] Show genes
GO:0031224	intrinsic component of membrane	2.12E-5	8.02E-3	1.82 (13215,2929,104,42)	[+] Show genes
GO:0098948	intrinsic component of postsynaptic specialization membrane	3.59E-5	1.13E-2	13.24 (13215,48,104,5)	[+] Show genes



Gene Ontology (GOrilla) Molecular Function

GO term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
GO:0030545	receptor regulator activity	2.11E-5	8.73E-2	5.23 (13215,243,104,10)	[-] Hide genes Gpihbp1 - gpi-anchored hdl-binding protein 1 Stc1 - stanniocalcin 1 Pgf - placental growth factor Prl3a1 - prolactin family 3, subfamily a, member 1 Spp1 - secreted phosphoprotein 1 Sfrp2 - secreted frizzled-related protein 2 Fgf21 - fibroblast growth factor 21 Sema3a - sema domain, immunoglobulin domain (ig), short basic domain, secreted, (semaphorin) 3a Clec11a - c-type lectin domain family 11, member a Apln - apelin
GO:0048018	receptor ligand activity	6.03E-5	1.25E-1	5.17 (13215,221,104,9)	[+] Show genes
GO:0038023	signaling receptor activity	6.48E-5	8.94E-2	3.58 (13215,461,104,13)	[+] Show genes
GO:0060089	molecular transducer activity	1.71E-4	1.77E-1	3.25 (13215,508,104,13)	[+] Show genes
GO:0004888	transmembrane signaling receptor activity	4.58E-4	3.79E-1	3.60 (13215,353,104,10)	[+] Show genes
GO:0019955	cytokine binding	4.63E-4	3.19E-1	7.75 (13215,82,104,5)	[+] Show genes
GO:0004896	cytokine receptor activity	8.46E-4	5E-1	9.41 (13215,54,104,4)	[+] Show genes