Results of pathway analysis of Kimball ATVG dataset

Alex Ruan
alexruan@umich.edu

Jason Kwon
kwonju@umich.edu

Abstract

Here we found that only two genes, PGAM1 and CHCHD2, were significantly differentially expressed in DIO mice vs. ND mice when comparing the differential expression between the two mice of the difference in expression between Ly6C^{Hi} and Ly6C^{Lo} cells. Specifically, PGAM1 was upregulated and CHCHD2 was downregulated. It is known that upregulation of PGAM1 induces the upregulation of IL-1β expression and apoptotic cell death (Song et al., 2018). Similarly, it is known that when mitochondrial CHCHD2 is downregulated, apoptosis increases (Liu Y, 2015). It then makes sense that these two genes are most differentially expressed when comparing DIO and ND mice, as both of these gene regulation differences increase inflammation and apoptosis, which are associated with the initial tissue destruction phase of wound healing. As there are only two signifiicantly differentially expressed genes, this limits pathway analysis. If we however increase our FDR cutoff to 0.3 however, we now have 9 significantly differentially expressed genes that yield far more pathways relating to inflammation regulated to increase pro-inflammatory responses/chronic inflammation, such as the down regulation of lipoxin. These results however are less reliable due to the increased FDR cutoff.

1 Objective

We went ahead and went through the pathway analysis pipeline based on your pipeline as well as the workflow given by https://dockflow.org/workflow/rnaseq-gene-edgerql and have gotten results, however we're not entirely sure where to go from here. Through this writeup, we would like to show you all the results we obtained and how we obtained them to verify that we obtained them correctly, as well as ask what direction we might take from here.

2 Converting RNASeq data to read counts

Here we'd just like to list all the commands we used to verify that we're indeed using them correctly. This is the first time we've done any sort of trimming and alignment and so the commands we used mainly were copy-pasted from random forums online, trying to find the most concise versions closest to default settings.

2.1 FastQC and Trimming

We first ran through the RNASeq with fastqc to see what the quality was. We found that some adapter sequences still existed so we trimmed them off with Trimmomatic with the following command:

```
  java -jar trimmomatic-0.39.
↪ jar SE -phred33 in.fastq.gz
↪  trimmed.fq.gz ILLUMINACLIP:
↪ TruSeq3-SE:2:30:10 LEADING:3
↪ TRAILING:3 SLIDINGWINDOW:4:15
↪ MINLEN:36
```

After that, we ran it through fastqc again and generated the following report:

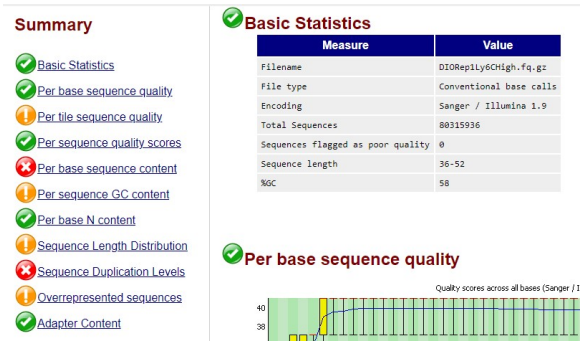

Figure 1: FastQC report from trimmed data

What was concerning was that Per base sequence content and Sequence Duplication Levels were

marked as X. It seems that this is experiment dependent, so we went ahead with the analysis with the assumption that this is expected. However we're curious whether this is ok.

2.2 Alignment & read counting

We then aligned the data using hisat, using the mm10 genome.

```
./hisat2-2.0.4/hisat2 -x mm10
↪ /genome -U trimmed.fq.gz -S
↪ aligned.sam
```

Then we converted it to bam format:

```
samtools view -bS aligned.sam >
↪  compressed.bam
```

Lastly, we counted it with the following featureCounts command with the **in-built** mm10 RefSeq exon annotation:

```
featureCounts -a annotation/
↪ mm10_RefSeq_exon.txt -o reads.
↪ txt ./*.bam -F SAF
```

This gave us our read count file.

3 Data Cleaning

Once we read in the read count data into a DGEList object, several other processes were performed in order to refine the data further. After mapping our Entrez Gene Ids to Gene Symbols using the NCBI database, we removed low count genes that didn't have a CPM above 0.5 for at least two libraries, and then normalized our data for composition bias.

	group	lib.size	norm.factors
DIO1HighReads	DIO.High	3.4e+07	0.98
DIO2HighReads	DIO.High	2.5e+07	0.96
DIO1LowReads	DIO.Low	3.5e+07	1.01
DIO2LowReads	DIO.Low	4.6e+07	0.98
ND1HighReads	ND.High	2.6e+07	1.00
ND2HighReads	ND.High	2.6e+07	1.03
ND1LowReads	ND.Low	2.2e+07	0.99
ND2LowReads	ND.Low	3.6e+07	1.04

Figure 2: Normalization Factors of Samples

4 Data Exploration

We then plotted our data to see if there were interesting patterns. Plotting the data with an MDS plot we found that one of the DIO Ly6C High replicates had a significant amount of upregulated genes. (Figure 3).

5 Dispersion Estimations

We found that our data had high dispersion (Figure 4), or at least higher than the workflow we

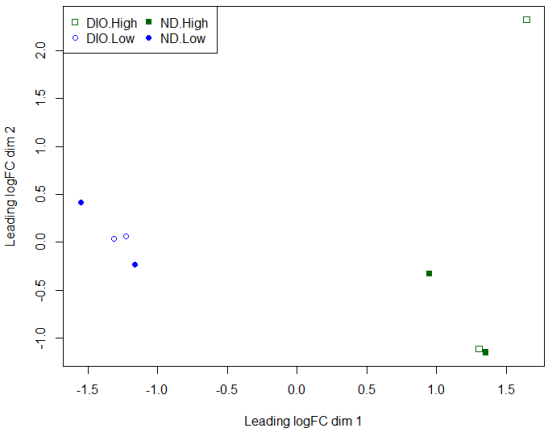

Figure 3: MDS Plot of Samples

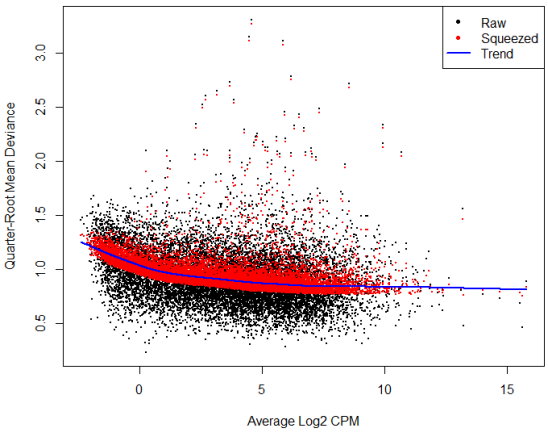

Figure 4: Dispersion Plot

referenced. Since the workflow we referenced mentioned that quasi-likelihood F-tests were better than the more popular likelihood-ratio tests for data with high dispersion, we decided to follow suit and use QLF-tests as well.

6 Differentially expressed data

We looked into the differential expression between (DIO Ly6C^{Hi} - DIO Ly6C^{Lo}) and (ND Ly6C^{Hi} - ND Ly6C^{Lo}) in order to remove the variability from just DIO vs. ND genes or Ly6C^{Hi} vs. Ly6C^{Lo}, and instead see how the change in expression was different between DIO and ND groups.

Running our data through glmQLF tests taking into account QL dispersion, we ended up finding that only two genes were significantly differentially

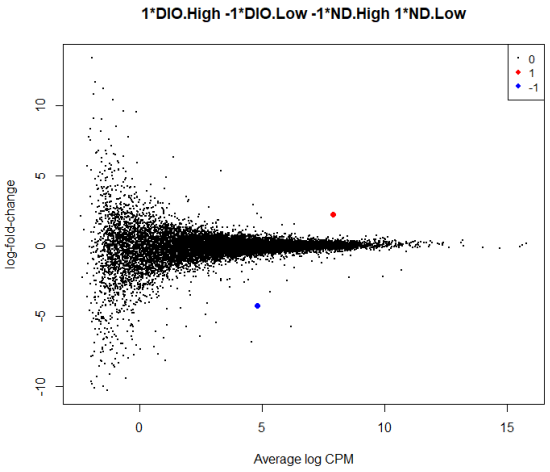

Figure 5: Differential expression MD plot

expressed under default conditions, shown in the MD plot (Figure 5). The next most significantly differentially expressed genes had FDR cutoffs above 0.05 (Figure 6).

```
Coefficient:  1*DIO.High -1*DIO.Low -1*ND.High 1*ND.Low
       Length Symbol logFC logCPM  F  PValue    FDR
18648    1832  Pgam1   2.3    7.9 98 2.4e-07 0.0031
14004     910 Chchd2  -4.2    4.8 82 2.7e-06 0.0177
22630    2110  Ywhaq  -3.2    5.3 54 3.0e-05 0.1295
18725    3464  Pira2   2.0    5.0 37 4.1e-05 0.1330
19326    6065 Rab11b  -1.7    4.6 32 7.6e-05 0.1948
170930    998  Sumo2  -1.8    4.2 28 1.5e-04 0.2817
11687    2414 Alox15  -6.4    2.5 29 1.6e-04 0.2817
15278    2448  Tfb2m   3.2    2.0 27 1.8e-04 0.2817
56807    3191 Scamp5   6.4    1.4 26 2.0e-04 0.2817
20775    2748   Sqle  -1.9    3.1 20 5.8e-04 0.7462
```

Figure 6: Top differentially expressed genes

Specifically, of the only two significantly differentially expressed genes, PGAM1 was upregulated and CHCHD2 was downregulated. It appears that upregulation of PGAM1 induces the upregulation of IL-1β expression and apoptotic cell death (Song et al., 2018). Similarly, it appears that when mitochondrial CHCHD2 is downregulated, apoptosis increases (Liu Y, 2015). It then makes sense that these two genes are most differentially expressed when comparing DIO and ND mice, as both of these gene regulation differences increase inflammation and apoptosis, which are associated with the initial tissue destruction phase of wound healing.

7 Heatmap clustering

Figure 7 is the heatmap showing the differential expression across all groups for the top differentially expressed genes between DIO and ND ($Ly6C^{Hi}$-$Ly6C^{Low}$).

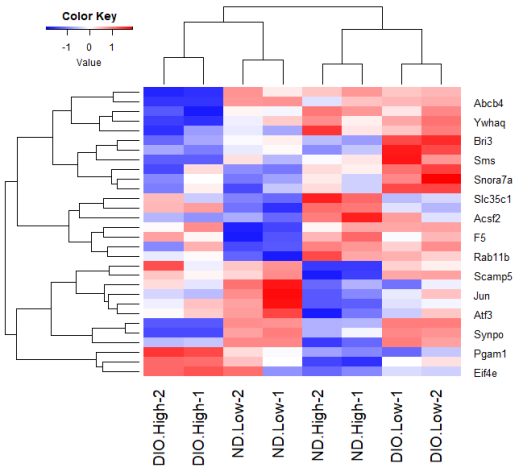

Figure 7: Heatmap

8 Pathway analysis

As there were only two differentially expressed genes, GO and KEGG pathway analysis yielded limited results as they were based on only two genes. The next most significantly differentially expressed gene had a FDR of 0.13, making it and all other genes with higher FDRs unviable for GO and KEGG analysis.

From the GO and KEGG pathway analysis (Figure 8 & 9) for these two genes, we see a variety of pathways that are affected by the up and down regulation of the PGAM1 and CHCHD2 genes.

```
                                                                  Term Ont  N Up Down    P.Up P.Down
GO:0004082                          bisphosphoglycerate mutase activity  MF  3  1    0 0.00023  1.000
GO:0004619                            phosphoglycerate mutase activity  MF  3  1    0 0.00023  1.000
GO:0043456                       regulation of pentose-phosphate shunt  BP  4  1    0 0.00031  1.000
GO:1902031                       regulation of NADP metabolic process  BP  5  1    0 0.00039  1.000
GO:0016868 intramolecular transferase activity, phosphotransferases  MF  9  1    0 0.00070  1.000
GO:0043471    regulation of cellular carbohydrate catabolic process  BP 11  1    0 0.00085  1.000
GO:0006098                                 pentose-phosphate shunt  BP 13  1    0 0.00101  1.000
GO:1900037               regulation of cellular response to hypoxia  BP 13  0    1 1.00000  0.001
GO:0006740                                 NADPH regeneration  BP 14  1    0 0.00109  1.000
GO:0051156                glucose 6-phosphate metabolic process  BP 19  1    0 0.00147  1.000
GO:0045730                                 respiratory burst  BP 23  1    0 0.00178  1.000
GO:0016866                        intramolecular transferase activity  MF 24  1    0 0.00186  1.000
GO:0044275                     cellular carbohydrate catabolic process  BP 28  1    0 0.00217  1.000
GO:0006739                                 NADP metabolic process  BP 33  1    0 0.00256  1.000
GO:0006110                           regulation of glycolytic process  BP 34  1    0 0.00264  1.000
```

Figure 8: Top GO results

In particular, from the GO results we see that the biophosphoglycerate mutase activity pathway is strongly upregulated. This main function of this pathway is the synthesis of 2,3-BPG from 1,3-BPG (an intermediate in glycolysis) which is found only in red-blood cells and placental cells. Specifically, 2,3-BPG binds with high affinity to Hemoglobin causing a release of oxygen. From the fact that this pathway is highly upregulated, and that the cells

	Pathway	N	Up	Down	P.Up	P.Down
path:mmu00260	Glycine, serine and threonine m...	23	1	0	0.0018	1
path:mmu00010	Glycolysis / Gluconeogenesis	47	1	0	0.0036	1
path:mmu05230	Central carbon metabolism in ca...	56	1	0	0.0043	1
path:mmu01230	Biosynthesis of amino acids	61	1	0	0.0047	1
path:mmu04922	Glucagon signaling pathway	76	1	0	0.0059	1
path:mmu01200	Carbon metabolism	98	1	0	0.0076	1
path:mmu01100	Metabolic pathways	1089	1	0	0.0845	1
path:mmu00785	Lipoic acid metabolism	1	0	0	1.0000	1
path:mmu00232	Caffeine metabolism	2	0	0	1.0000	1
path:mmu00290	Valine, leucine and isoleucine ...	2	0	0	1.0000	1
path:mmu00780	Biotin metabolism	3	0	0	1.0000	1
path:mmu04950	Maturity onset diabetes of the ...	3	0	0	1.0000	1
path:mmu00524	Neomycin, kanamycin and gentami...	3	0	0	1.0000	1
path:mmu00430	Taurine and hypotaurine metabol...	3	0	0	1.0000	1
path:mmu00471	D-Glutamine and D-glutamate met...	4	0	0	1.0000	1

Figure 9: Top KEGG results

	Term	Ont	N	Up	Down	P.Up	P.Down
GO:0047977	hepoxilin-epoxide hydrolase activity	MF	1	0	1	1.00000	0.000388
GO:0051120	hepoxilin A3 synthase activity	MF	1	0	1	1.00000	0.000388
GO:2001303	lipoxin A4 biosynthetic process	BP	1	0	1	1.00000	0.000388
GO:2001302	lipoxin A4 metabolic process	BP	1	0	1	1.00000	0.000388
GO:0034246	mitochondrial transcription factor activity	MF	2	1	0	0.00062	1.000000
GO:0006391	transcription initiation from mitochondrial promoter	BP	2	1	0	0.00062	1.000000
GO:0004052	arachidonate 12(S)-lipoxygenase activity	MF	2	0	1	1.00000	0.000775
GO:0050473	arachidonate 15-lipoxygenase activity	MF	2	0	1	1.00000	0.000775
GO:0016165	linoleate 13S-lipoxygenase activity	MF	2	0	1	1.00000	0.000775
GO:2001301	lipoxin biosynthetic process	BP	2	0	1	1.00000	0.000775
GO:2001300	lipoxin metabolic process	BP	2	0	1	1.00000	0.000775
GO:0004082	bisphosphoglycerate mutase activity	MF	3	1	0	0.00093	1.000000
GO:0004619	phosphoglycerate mutase activity	MF	3	1	0	0.00093	1.000000
GO:0000179	rRNA (adenine-N6,N6-)-dimethyltransferase activity	MF	3	1	0	0.00093	1.000000
GO:0035963	cellular response to interleukin-13	BP	3	0	1	1.00000	0.001163

Figure 10: Top GO results

	Pathway	N	Up	Down	P.Up	P.Down
path:mmu00591	Linoleic acid metabolism	8	0	1	1.00000	0.0031
path:mmu00260	Glycine, serine and threonine m...	23	1	0	0.00712	1.0000
path:mmu00590	Arachidonic acid metabolism	26	0	1	1.00000	0.0100
path:mmu04216	Ferroptosis	36	0	1	1.00000	0.0139
path:mmu04962	Vasopressin-regulated water rea...	37	0	1	1.00000	0.0143
path:mmu00010	Glycolysis / Gluconeogenesis	47	1	0	0.01450	1.0000
path:mmu05230	Central carbon metabolism in ca...	56	1	0	0.01726	1.0000
path:mmu01230	Biosynthesis of amino acids	61	1	0	0.01879	1.0000
path:mmu04662	B cell receptor signaling pathw...	74	1	0	0.02276	1.0000
path:mmu04726	Serotonergic synapse	60	0	1	1.00000	0.0231
path:mmu04922	Glucagon signaling pathway	76	1	0	0.02337	1.0000
path:mmu01200	Carbon metabolism	98	1	0	0.03006	1.0000
path:mmu04380	Osteoclast differentiation	112	1	0	0.03430	1.0000
path:mmu04114	Oocyte meiosis	92	0	1	1.00000	0.0352
path:mmu04152	AMPK signaling pathway	94	0	1	1.00000	0.0359

Figure 11: Top KEGG results

are taken from epithelial cells, we see that there is significantly more oxygen being released at the site of the wound.

Similarly we see the upregulation of pentose-phosphate shunt and NADP metapbolic processes are upregulated as well. Both of which are associated with increased metabolic activity, in particular those of red blood cells. In particular the pentose-phosphate shunt is primarily involved in anabolic reactions in red blood cells, producing NADPH and various nucleotides and nucleic acids.

The only down regulated pathway is the regulation of cellular response to hypoxia. This makes sense as because shown earlier, more oxygen is being released in this area.

Similarly, from the KEGG results we see that the Glycine, Serine and Threonine Metabolism pathway and Glycolysis pathway are upregulated. Both of these pathways are involved in glycolysis. In addition there is upregulation of the biosynthesis of amino acids and carbon metabolism which aligns with GO results, as well as upregulation of the glucagon signalling pathway, raising the concentration of glucose and fatty acids in the area.

As both oxygen release, increased sugar concentration, and glycolysis are associated with increased ATP production, this suggests that there is increased cell activity. This likely has to do with the fact DIO mice have an extended tissue destruction phase and chronic inflammation, which causes increased blood flow to deliver nutrients and white blood cells to the wound area. Similarly there seems to be upregulation of the production of amino acids.

If we however increase the FDR cutoff of genes included in the GO and KEGG analysis to 0.3, we now include 9 rather than 2 genes and receive the following results (Figure 10 & 11)

Though these results are more unreliable due to the increase of the FDR cutoff, we see the up/down regulation of more pathways related to inflammation.

From the GO analysis results, we see in particular,the down regulation of hepoxilin-epoxide hydrolase activity and hepoxilin A3 synthase activity. As hepoxilin A3 is associated with promoting neutrophil-based inflammatory response to certain bacteria in the intestines and lungs of rodents, downregulation of this pathway suggests a weakened immune system. Oddly enough, hepoxilin-epoxide hydrolase activity is also downregulated, which involves the conversion of hepoxilin to its trioxilin form, effectively deactivitating the biologically active hepoxilin form.

Similarly, we see the down regulation of the lipoxin A4 biosynthetic and metabolic processes. As lipoxin are anti-inflammatory, pro-resolving molecules that play critical roles in in reducing excessive tissue injury and chronic inflammation (Chandrasekharan JA, 2015), it then makes a lot of sense that these are significantly downregulated for DIO mice, as they have chronic inflammation.

In addition we see the upregulation of mitochondrial transcription factor activity. It is unclear whether this includes mitochondrial transcription factor A, but it is known that injecting TFAM upregulates circulating levels of pro-inflammatory cytokines and increases neutrophil infiltration into the lungs (Chaung WW, 2012).

Similarly, it seems arachidonate 12(S)-lipoxygenase activity was down regulated. As metabolite stimualtes the chemotaxis of human

neutrophils, this suggests that an impaired immune response.

Lastly, it seems that cellular response to interleukin-13 is downregulated. As interleukin-13 has an anti-inflammatory response, this makes sense as for DIO mice there is chronic inflammation.

From KEGG analysis results, we see similar results. First, we see downregulation of linoleic acid metabolism. Linoelic acid is an essential fatty acid, and it is found that in rats a diet deficient in linoleate is shown to cause poor wound healing (Ruthig and Meckling-Gill, 1999). It is also known that linoleic acid has anti-inflammatory properties when applied topically to the skin (LETAWE et al., 1998).

Just as in the previous analysis, we also see down regulation of arachidonic acid metabolism, upregulation of biosynthesis of amino acids glycolysis, glucagon signaling pathway, carbon metabolism etc.

9 Closing

This are the results we've received. Two significantly differentially expressed genes, both with a role in inflammation. The most significantly affected pathways seem to involve the increased release of oxygen, the upregulation of glycolysis, and the increased production of amino acids. If we expand our FDR cutoff to 0.3 however, we see far more interesting results, showing a variety of proinflammatory responses. These results however are less reliable to the higher FDR cutoff requried. We are now left with a few questions:

- Did we process the RNA-seq data to read counts correctly?
- Is it ok that Per base sequence content and Sequence Duplication Levels were marked as X for this experiment?
- Did we find differential expression correctly?
- Was it ok to use quasi-likelihood F-tests rather than likelihood ratio tests given the amount of dispersion?
- With only two genes significantly differentially expressed, would it be fine to increase the FDR cutoff to 0.13 (3 more genes) or 0.3 (7 more genes) for pathway analysis?
- Are we really able to do pathway analysis with just two significantly expressed genes?
- **What direction do we go from here?**

References

Sharma-Walia N. Chandrasekharan JA. 2015. Lipoxins: nature's way to resolve inflammation. *Inflamm Res.*

Ji Y Dong W Wang P. Chaung WW, Wu R. 2012. Mitochondrial transcription factor a is a proinflammatory mediator in hemorrhagic shock.. *Int J Mol Med.*

LETAWE, BOONE, and PIÉRARD. 1998. Digital image analysis of the effect of topically applied linoleic acid on acne microcomedones. *Clinical and Experimental Dermatology*, 23(2):56–58.

Leslie PL Di J Tollini LA He Y Kim TH Jin A Graves LM Zheng J Zhang Y. Liu Y, Clegg HV. 2015. Chchd2 inhibits apoptosis by interacting with bcl-x l to regulate bax activation. *Cell Death Differ.*

Derek J. Ruthig and Kelly A. Meckling-Gill. 1999. Both (n-3) and (n-6) Fatty Acids Stimulate Wound Healing in the Rat Intestinal Epithelial Cell Line, IEC-6. *The Journal of Nutrition*, 129(10):1791–1798.

Jinsoo Song, In-Jeoung Baek, Churl-Hong Chun, and Eun-Jung Jin. 2018. Dysregulation of the nudt7-pgam1 axis is responsible for chondrocyte death during osteoarthritis pathogenesis. *Nature Communications*, 9.