BIF 701 LAB 9: Gene Expression Analysis

Christopher Eeles

Abstract— **In this lab we will be comparing gene expression data from human monocytes responding to *M. tuberculosis* infection in the presence of all-trans-retinoic acid (ATRA) and 1,2,5-dihydroxy vitamin D (1,25D3). Data was retrieved from GEO accession numbers GSE46268 and input into the GEO2R tool to generate the visualizations contained in this report. We hope to detect differences in induced cellular and genomic responses, particularly gene expression co-regulation, to better understand the mechanism by which these vitamins activate antimicrobial pathways in the human innate immune system.**

I. INTRODUCTION

All-trans retinoic acid (ATRA) and 1,2,5-dihydroxy vitamin D (1,25D3)—*i.e.,* biologically active Vitamin A and D3—have been implicated in enhancing the antimicrobial response of monocytes in both epidemiological and *in vitro* studies.[5] In this lab we will be utilizing the data from Millwright *et al.*'s 2014 study of *in vitro* ATRA induced antimicrobial activity of human monocytes during *M. tuberculosis* infection. We will replicate their analysis of gene expression using the web-based tool GEO2R to generate statistical measures of relative expression from Millwright's original dataset.[1] In conducting this analysis we will demonstrate the insights that bioinformatics can provide to the study molecular pathways; additionally, we will familiarize ourselves with some of the web-tools available for studying differential gene expression. We hope to identify co-regulation of expressed genes induced by ATRA, 1,25D3 or both to elucidate genes likely to be involved in regulation and catalysis of this antimicrobial cellular pathway.[1] Identified co-regulation will be compared with a subset of genes from Matthew *et al.* and should support the conclusions of that paper assuming they used an equivalent analytic tool.

II. METHODS

A. NCBI GEO[3]

The NCBI's Gene Expression Omnibus (GEO) is an international resource which stores micro-array, next-generations sequencing and other forms of high-throughput genomics data. The repository is available online at `https://www.ncbi.nlm.nih.gov/geo/` where it can be accessed publicly. All submissions to GEO must follow MIAME criteria as well as organize data into platform, sample and sequence records. GEO curators use records to assemble DataSets which are biologically and statistically comparable; these are used to derive profiles, which collate information across samples for an individual gene. Additionally, DataSets are the input for a range of analysis and visualization tools available within the GEO web-server.

B. GEO2R[4]

GEO2R is an interactive web tool which uses submitter-supplied and processed data tables for comparison of two or more sample. Unlike other GEO's other DataSet analysis tools, GEO does not rely on curated GEO data, instead affecting the original data file directly. Using the R packages GEOquery and limma from the Bioconductor project the tool outputs the results of statistical tests from the R language to identify deferentially expressed genes. In this laboratory we will be conducting a statistical comparison of gene expression between a control, ATRA and D3 group, each containing 4 samples.

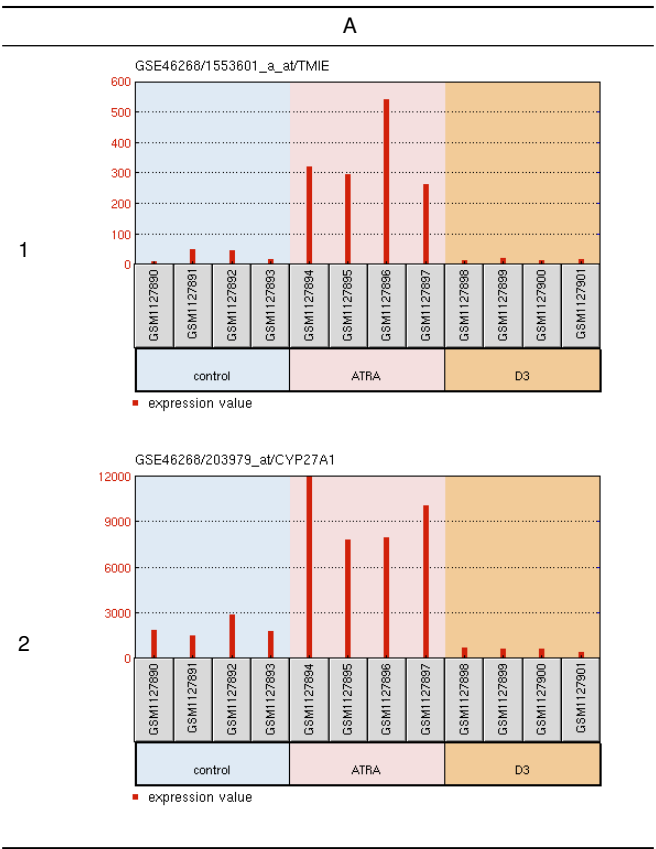## III. RESULTS

A. All-trans retinoic acid

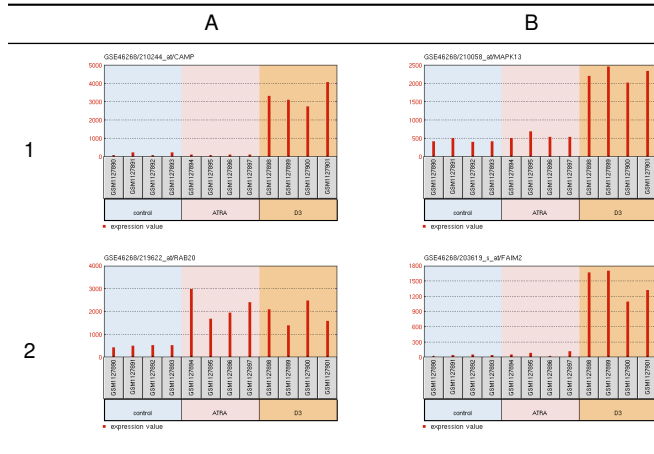TABLE I

B. 1,2,5 dihydroxy vitamin D3

TABLE II

IV. DISCUSSION AND CONCLUSION

A literature search was used to orientate our analysis of co-regulated gene expression data from GEO2R. Given the discussion of NPC2 and cathelidin as the two primary modulators of antimicrobial activity in ATRA and 1,25D3 respectively, the pathways of these two proteins were used to identify candidates for coregulation[3]. These pathways, as well as our observations, will be discussed presently.

A. Intracellular Cholesterol Pathway

Acidification of the cellular lysosome is a key factor in the effectiveness of monocytes at killing phagocytosed pathogens, *M. tuberculosis* (TB) included.[7] Without the ability to rapidly destroy injested TB cells or debris, the rate a monocyte's ability to reduce a pathogen is decreased; thus, less pathogens can be processed per monocyte per unit time. Cholestrol accumulation in the lysosome is a major inhibitor of acidification and so constitutes a major obstacle in innate immune response to infections.[7]

ATRA has been found to stimulate production of the NPC2 protein, a transporter protein essential to cholesterol egress from both the lysosome and larger cell.[3,7] In doing so it is able to enhance lysosomal acidification, thereby improving phagocyticity of monocytes, as well as prevent cholesterol accumulation in the cytosol—which may become a nutrient source for enclosed TB cells.[7] Careful inspection of the results from GEO2R did not indicate the presence of NPC2 transcripts directly. This could be explained by lower relative transcript abundance in the monocyte cytosol, therefore exluding it from the top 250 transcripts results from GEO2R. Moreover, were the lysosome to be lost in the experimental process—e.g., by incomplete lysis of these organelles—any NPC2 transcript within or associate with them could have been excluded from the microarray data.

Two proteins associated wtih fat metabolism were found to be positively correlated with the ATRA samples. Table I A1 displays transmembrane inner ear protien (TMIE), which the Uniprot databse indentified as playing a role in intracellular packaging and transport of vesciles. This transcript was positively correlated with ATRA treatment, likely due to the increased need for vesicular transport in exocytosis of lysome egressed cholesterol packages. Table I A2 displays CYP2A, identified using Uniprot as a cytochrome P450 enzyme involved in the digestion of sterols and therefore cholesterol. Both of these proteins were positively co-expressed, suggesting they may be co-regulated as part of the intracellular cholesterol pathway involving NPC2 which was identified by Millwright *et al*. While the protein array data was unable to detect expression of NPC2, qPCR results in the original study indicated a strong positive correlation of NPC2 with the ATRA treatment group.[3] This evidence suggests co-expressions, and therefore potential co-regulation, of these proteins by ATRA.

B. Cathelicidin Pathway

Studies prior to Millwright's identified cathelicidin as an important antimicrobial peptide. It plays an important role in triggering cellular apoptosis, which is necessary to break apart the tumour like granulomas formed by TB infections.[6] The key aim of the original study was to investigate if the cathelicidin protein, LL-37, was also involved in ATRA mediated antimicrobial effects.[3] Results however indicated this not to be the case, and light was shed on a possible new pathway of antimicrobial activation of immune cells. The LL-37 pathway is complex, involving a cascade of many proteins in the cell which converge to facilate apoptosis of its target. Interestingly, the mechanism of Vitamin D3 induced antimicrobial activation of the innate immune system via LL-37 has also been shown to play a role in immune-mediated destruction of cancer cells and this pathways is being investigated for potential targets for pharmaceutical development in the antimicrobial as well as antitumour drugs.[8]

Unlike NPC2, our analysis was able to directly identify LL-37, displayed in Table II A1, and confirm its correlation with the 1,25D3 treatment group. This protein was found to be negatively co-regulated with the proteins identified from the cholesterol pathway, supporting Millwright's supposition that these two pathways may interfere with the respective efficiency of their counterpart.[3] Table II B1 however, identifies a positive co-expression between the RAB20RAS protein in both ATRA and 1,25D3 treatments. Uniprot identifies this protein as ras-related protein Rab-20, which plays a role in endocytosis and recycling within the cell. Additionally it faciliates acidification of phagosomes such as the lysosome. These results suggest that this protein is invovled in innate immune respoinse more generally as all activated immune cells will have increased metabolic activity and therefore need efficient ways to manage waste accumulation.

Table II A2 shows a correlation between MAPK13 transcripts and the 1,25D3 treatment group. Uniprot identifies this protein as mitogen-activate protein kinase which plays a role in phosphorylation of other proteins during cascade reactions such as the cathelicidin pathway. It's activation

is correlated with vitamin D3 concentration, and therefore it is likely a kinase in the LL-37 antimicrobial pathway. This protein was positively co-expressed with LL-37 and therefore supports Millwright and others suggestion that LL-37 is regulated by vitamin D3. Table II B2 shows correlation between FAIM2 and the vitamin D3 treatment. Interestingly, a Unprot search identified this protein as an apoptotic inhibitor, seemly contradicting the suggested role of cathelicidin as an apotosis inducer. However, considering that monocytes are the preferred host for TB proliferation this is likely necessary to preventing premature apoptosis of infected host cells.[6] All three protein suspected to be involved in the cathelicidin pathway showed positive co-expression, supporting the hypothesis that vitamin D3 is indeed the regulator of this pathway.

C. Conclusion

The above analysis demonstrates the utility of bioinformatic analysis of cell transcript arrays in identifying novel molecular pathways. While our analysis was informed by existing literature, it is easy to see how a more rigorous comparison between the expression plots from our data may have been used to indepently eludicate the pathways regulated by ATRA and 1,25D3. The utlity of GEO2R and similar tools extends beyond novel identification, as identify co-expressed, and thus potentially co-regulated, transcripts may be used to support existing hypotheses or include new factors in kown pathways. As system biology becomes an increasingly necessary tool in biological research, it is likely that anlyses such as these will become more common and essential to the indentification of the complex, interacting web of pathways which sum to create all life and determine its interaction with the environment around it.

REFERENCES

[1] School of Biological Sciences and Applied Chemistry. (2018). BIF701 Lab 9: Gene Expression Profiles with Vitamin A and D3. Seneca College: Toronto, ON.

[2] School of Biological Sciences and Applied Chemistry. (2018). BIF701 Topic 9: Gene Expression Analysis. Seneca College: Toronto, ON.

[3] NCBI. (2016). GEO Overview. In *Gene Expression Omnibus* [Internet]. Retrieved from http://www.ncbi.nlm.nih.gov/geo/info/overview.html.

[4] NCBI. (2016). GEO2R. In *Gene Expression Omnibus* [Internet]. Retrieved from https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html.

[5] Matthew Wheelwright *et al*. (2014). All-trans retinoic acid triggered antimicrobial activity against *Mycobacterium tuberculosis* is dependent on NPC2. In *J Immunol*, 192(5), pp. 2280-2290.

[6] Fenton, M. J., Vermeulen, M. W. (1996). Immunopathology of tuberculosis: roles of macrophages and monocytes. Infection and immunity, 64(3), 683-90. Retrieved from url-https://www.ncbi.nlm.nih.gov/pmc/articles/PMC173823/pdf/640683.pdf.

[7] Infante, R., *et al*. (2008). NPC2 facilitates bidirectional transfer of cholesterol between NPC1 and lipid bilayers, a step in cholesterol egress from lysosomes. In *PNAS, 34*(40), pp. 15287-15292. Retrieved from https://doi.org/10.1073/pnas.0807328105

[8] Kuroda, K., *et al*. (2015). The human catehlicidin antimicrobial peptide LL-37 and mimics are potential anticancer drugs. In *Frint. Oncol.*. Retrieved from https://doi.org/10.3389/fonc.2015.00144.