The Development and Application of a Systems Biology Approach to Mapping Monocyte Gene Regulatory Networks

by
Mudra Choudhury

A THESIS

submitted to

Oregon State University

University Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in BioResource Research
(Honors Scholar)

Presented May 18th, 2016
Commencement June 2016

AN ABSTRACT OF THE THESIS OF

Mudra Choudhury for the degree of Honors Baccalaureate of Science in BioResource Research presented on May 18th, 2016. Title: The Development and Application of a Systems Biology Approach to Mapping Monocyte Gene Regulatory Networks.

Abstract approved:

_____

Stephen A. Ramsey

While monocytes are vital for host defense against pathogens, in cases of non-resolving inflammation, monocytes and their differentiated progeny can secrete tissue-damaging factors, present auto-antigens, and promote the inflammatory response. In this study, a systems biology approach was used to map human monocyte gene regulatory regions by integrating expression quantitative trait loci (eQTL) and transcription factor (TF) binding site data from previous studies. We hypothesized that monocyte-specific TFs will have binding sites that are concentrated in the vicinity of the monocyte eQTLs. A block permutation test was used to determine the significance of the overlap between monocyte eQTLs and TF binding sites. *P* values were obtained for each TF experiment, and TFs were ranked according to monocyte specificity. This approach offers an efficient method to identify monocyte-specific TFs and gain insight into the TF regulatory network controlling population variation in human monocyte gene expression. We have developed a putative network of interactions among top-ranked TFs identified in this study. Additionally, candidate monocyte TFs that were uncovered in this analysis include BCLAF1, SIN3A, TAF1, and TBP. This integrative approach could potentially yield novel targets for therapies for chronic inflammatory disorders. Moreover, it can be utilized to predict mammalian gene regulatory networks for various cell types.

Key Words: Monocytes, Transcription Factor Binding Sites, eQTLs, Gene Regulatory Network

Corresponding e-mail address: mudrachoudhury3@gmail.com

The Development and Application of a Systems Biology Approach to Mapping Monocyte Gene Regulatory Networks

by
Mudra Choudhury

A THESIS

submitted to

Oregon State University

University Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in BioResource Research
(Honors Scholar)

Presented May 18th, 2016
Commencement June 2016

Honors Baccalaureate of Science in BioResource Research project of Mudra Choudhury presented on May 18th, 2016.

APPROVED:

Stephen A. Ramsey, Mentor, representing the Department of Biomedical Sciences and the School of Electrical Engineering and Computer Science

Katharine G. Field, Committee Member, Director, BioResource Research

Thomas J. Sharpton, Committee Member, representing the Dept, of Microbiology

Toni Doolen, Dean, University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, University Honors College, and will become part of the Scholars Archive collection for BioResource Research. My signature below authorizes release of my project to any reader upon request.

Mudra Choudhury, Author

1 Introduction

Monocytes are white blood cells that play a central role in the innate immune system as circulating precursors to macrophages and some dendritic cells. Monocytes can be rapidly mobilized to peripheral tissues in response to infection or wounding, where they and their progeny carry out diverse immune functions such as phagocytosis, antigen presentation, and cytokine secretion [1]. These cells are controlled at the level of gene transcription by different classes of transcription factors (TFs), as well as several transcriptional co-regulators [2]. While monocytes are vital for host defense against pathogens, in cases of non-resolving inflammation, monocytes and their progeny can secrete tissue-damaging factors, present auto-antigens, and promote the inflammatory response. Because of their important roles in numerous diseases such as atherosclerosis and coronary artery disease, understanding the complex gene regulatory network that controls monocyte properties such as behavior, cell differentiation, and response to molecular signals is essential for the rational discovery of novel therapies and diagnostics.

Currently, mapping cell type-specific gene regulatory networks is a great challenge due to the large number (> 1,500) of TFs and the expansiveness of the noncoding human genome [3]. One way that this problem is being addressed is through tissue-specific expression quantitative trait loci (eQTL) studies, which identify expression single nucleotide polymorphisms (eSNPs) that are correlated with expression levels of one or more genes. Previous studies have identified transcriptome measurements obtained from individuals within a large population and then combined these measurements to identify common genetic variants that are

associated with differential gene expression in the cell type of interest. The genomic regions that are in linkage disequilibrium (LD) with these variants are enriched for *cis*–regulatory elements such as TF binding sites. Because of their potential biomedical significance as circulating biomarkers (via their gene expression or cell surface protein expression) and as targets for therapeutic targets, monocytes have been the subject of two large population studies, which have collectively mapped over 80,000 eQTLs [4, 5]. Additionally, the Encyclopedia of DNA Elements (ENCODE) project has provided measurements of binding sites for about 140 TFs derived from various human cell lines. These measurements were obtained via chromatin immunoprecipitation-to-tag-sequencing (ChIP-seq) analysis [6].

In this study, a systems biology approach was used to map human monocyte gene regulatory networks by integrating publicly available eQTL and TF location data. We hypothesized that monocyte-specific TFs will have binding sites that are concentrated in the vicinity of the monocyte eQTLs. If this hypothesis is consistent with available monocyte eQTL and human TF ChIP-seq data, it could lead to new understanding of the gene regulatory basis of chronic inflammatory diseases and to novel therapeutic approaches to these diseases involving the targeting of immune cell populations. Additionally, it could significantly increase the interest of the systems biology community in utilizing eQTLs for mapping cell-type specific, mammalian gene regulatory networks.

2 Methods

2.1 Data Extraction

A collection of 400 TF binding site (TFBS) experiments derived from ChIP−seq analysis of 68 human cell lines and 140 TFs were obtained in genome reference consortium human assembly 37 (GRCh37) coordinates from the public repository of the National Human Genome Research Institute (NHGRI) ENCODE project [6]. Additionally, human monocyte eQTL data, in the form of eSNPs, were retrieved from two published large-cohort studies of genotype information and peripheral blood monocyte gene expression in GRCh36 from 1,773 unrelated individuals [4, 5]. Monocyte eSNPs were mapped to their respective GRCh37 chromosomal coordinate locations using coordinate information from the Database of Single Nucleotide Polymorphisms (dbSNP) [7].

2.2 eSNP Expansion

Locations that are in linkage disequilibrium (LD) with each monocyte eSNP were mapped using the SNP Annotation and Proxy (SNAP) search tool (Broad Institute) based on haplotype information from the 1,000 Genomes Project, obtained from Utah residents with northern and western ancestry (CEU population) [8]. The eSNPs that are located on nuclear autosomes were each expanded to an LD block using SNAP, by obtaining the maximum allowed R^2 threshold (nominally 1.0) and a distance limit of 10 kilobase pairs (kb). The left-most and right-most proxy-SNPs for each eQTL were employed to define the boundaries of the LD blocks. Monocyte LD blocks collectively yielded approximately 14% genome coverage. This genome coverage excludes sex chromosomes. LD blocks located on nuclear autosomes were

used in analysis due to the limited proxy−SNPs identified to be in LD with eSNPs on the X and Y chromosomes.

2.3 Overlap Tabulation and Statistical Analysis

2.3.1 Computer Software and Overlap Tabulation

The R statistical computing language integrated development environment, RStudio version 3.2.2 (R Foundation for Statistical Computing, Vienna, Austria), was used to tabulate overlap of TF binding sites within monocyte LD blocks and for statistical analysis [9]. To analyze the TFBS frequencies within monocyte specific LD block locations, TFBS peak locations were treated as point locations in the genome. For each TFBS experiment, the number of peaks that directly overlapped monocyte eSNP LD blocks was tabulated.

2.3.2 Statistical Analysis

For the purpose of testing the significance of each TFBS experiment's overlap count, a block permutation approach was used. A concatenated file of genome locations was created to function as a background model for randomization. This file was constructed by collapsing genomic locations in the 400 TFBS experiments and the LD block locations. Peak locations that were within 1.9 kb of one another were merged together. The value of 1.9 kb was determined empirically by ensuring that all LD blocks could be accommodated within the background model during randomization. This peak-merging procedure yielded genomic features covering 26% of the genome. Monocyte eQTL LD blocks were assigned to random, non-overlapping locations in the genome restricted by the background model. For each ChIP-seq experiment, counts of the binding sites within the randomly placed LD

blocks were tabulated. The LD block randomization procedure was repeated 1,000 times for each ChIP-seq experiment.

The actual overlap count for each ChIP-seq experiment within monocyte LD blocks was compared to the counts tabulated from the 1,000 randomizations. As a preliminary analysis, *Z*-scores were computed for each TFBS ChIP-seq experiment. TFs were ranked according to descending *Z*-scores. Gene Set Enrichment Analysis (GSEA) was conducted on various cell types, such as the myelogenous K562 cell line [10]. This analysis yielded GSEA enrichment scores (*E* scores), between -1.0 and 1.0, to describe the concentration of a specific cell type towards the top of the list of TFs when ranked according to *Z*-score [11]. The rationale for the GSEA was that it would quantify the extent to which the K562 cell line, a cell type with known immune function, is concentrated towards the top of the preliminary ranked list.

P values and adjusted *P* values were obtained for each TFBS ChIP-seq experiment using a one-tailed binomial test. For each TFBS ChIP-seq experiment, a "background" probability that a randomly selected binding site would be located within a randomized monocyte LD block was calculated by computing the average overlap count from 1,000 randomizations and dividing this value by the total number of binding sites for the TFBS ChIP-seq experiment,

$$Background\ Probability\ of\ TFBS\ experiment = \frac{Average\ overlap\ count\ of\ randomizations}{Total\ no.of\ binding\ sites}.$$

2.4 Ranking of Transcription Factors According to Significance

In the ENCODE TFBS ChIP-seq dataset, in many cases there are multiple TFBS ChIP-seq experiments (usually from different cell lines) for the same TF. To condense the TFBS ChIP-seq experiments to include only one representative experiment for each TF, a ratio value was calculated for each ChIP-seq experiment by taking the ratio of the actual overlap, the total number of binding sites for the TF, and the background probability of each experiment (i.e., if in a given experiment, transcription factor X has an actual overlap of 50 binding sites, 1,000 total binding sites, and a background probability of 0.05, the ratio would be calculated as 50 / 1,000 / 0.05 = 1). For each TF, the ChIP-seq experiment with the highest ratio for the TF was used to represent the TF. This finalized list of 140 unique TFs was then ranked in order of ascending adjusted *P* value.

2.5 Analysis of Significant TFs

A biomedical literature review was conducted in order to obtain a list of TFs that are previously known to have regulatory functions in monocytes. TFs obtained through literature analysis were compared to the 140 unique TFs that were ranked by their adjusted *P* values. The density of known monocyte-regulating TFs near the top of the ranked list of 140 TFs was analyzed. A *P* value cut off of $P < 1.00 \times 10^{-7}$ was selected to obtain a list of the top 20% of TFs. GSEA analysis was conducted on the top-ranked (i.e., smallest *P* value) TFs in the list to reinforce that the TFs were highly concentrated near the top of the list when preliminarily ranked according to *Z*-score.

A predicted network of interactions between the top 20% of the TFs was identified using GeneMANIA, a web-based bioinformatics tool [12]. Mouse

orthologs expressed in monocytes and other immune cell types of the top TFs were identified using the Immunological Genome (ImmGen) project database [13]. The gene annotation portal, BioGPS [14], was used to obtain human microarray-based gene expression measurements for the top TFs. Additional literature review was conducted on TFs that had not been identified to have regulatory function in monocytes in order to ascertain their previously known functions in immunity, if any.

3 Results and Discussion

3.1 GSEA of K562 cell line

GSEA was conducted for the K562 cell line (which is a myeloblast cell line that can differentiate towards monocyte lineage). This yielded a positive *E* score of 0.359, reinforcing that ChIP-seq experiments derived from the K562 cell line are concentrated towards the top of the preliminary list of TFs ranked according to *Z*-score. Analysis of the H1-hESC human embryonic stem cell line as a negative control yielded an *E* score of -0.316. The GSEA plots of both cell lines are displayed in Fig. 1.

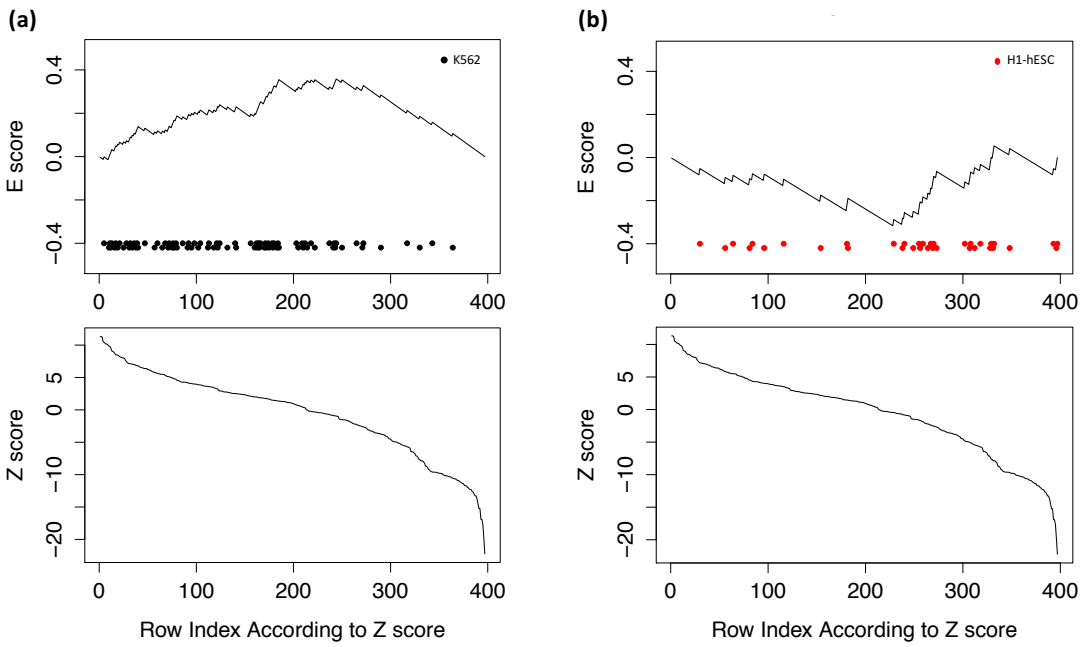

Fig. 1 (**a**) GSEA plot of K562 cell line with an *E* score of 0.359 (**b**) GSEA plot of H1-hESC cell line with an *E* score of -0.316. Black and red points on uppermost graphs describe locations of each cell line within the list of ChIP-seq experiments when ranked by *Z*-score. Distribution of *Z*-scores is shown directly below GSEA plots.

3.2 Comparison to Previous Literature

3.2.1 Literature Analysis

A biomedical literature review was carried out in order to identify TFs previously known to have regulatory functions in monocytes, as summarized in Table 1. TFs identified from the literature review were joined to form a combined list of monocyte–regulating TFs, which were used to assess the accuracy of our results.

Year	Authors	Transcription Factors	Reference No.
2002	A. Friedman	CBF, CEBPε, EGR1, IRF8, JUN, MAFB, MYB, PU.1, RARA, SP1	[15]
1998	A. Valledor, F. Borras, M. Cullell-Young et al.	EGR1, FOS, GATA1, GATA2, HOXB7, IRF1, JUN, JUNB, MYB, MYC, PU.1, RUNX1, SCL, STAT1, STAT3,	[16]
2014	B. Fairfax, P. Humburg, S. Makino et al.	IRF2, IRF8, IRF9, POU2F1, STAT1	[17]

2013	C. Dong, G. Zhao, M. Zhong et al	AR, C/EBPα, C/EBPδ, E2F4, EGR1, P300, FOS, FOXA2, FOXO1, GATA1, GATA2, HNF4A, RARA, RB1, RUNX1, RUNX3, SOX2, STAT3, VDR	[18]
2004	C. Shi, X. Zhang, Z. Chen et al.	C/EBPα, C/EBPβ, FOXP1, PU.1, RUNX1	[19]
2013	D. Kurotaki, N. Osato, A. Nishiyama et al.	KLF4	[20]
2007	F. Rosenbauer, D. Tenan	BMI1, C/EBPα, CEBPε, ELF1, GATA1, GFI1, IRF8, PU.1, RARA, RUNX1, SCL, TCF	[21]
2011	H. Kikuchi, F. Kuribayashi, N. Kiwaki et al.	GCN5	[22]
2013	J. Larabee, S. Shakir, S. Barua et al.	HES1, HEY1, IL2RA, IL7R	[23]
2012	J. Van den Bossche, B. Malissen, A. Mantovani et al.	NF-κB	[24]
2009	K. Khan, A. Coaquette, C. Davrinche et al.	BCL3	[25]
2006	K. Resendes, A. Rosmarin	GABP	[26]
2011	K. Weigelt, L. Carvalho, R. Drexhage et al.	ATF3, EGR3, PU.1	[27]
2010	M. Marullo, M. Valenza, C. Mariotti et al.	NRSF	[28]
2011	P. Li, J. Wong, C. Sum et al.	IRF3	[29]
2014	R. Huber, D. Pietsch, J. Gunther et al.	C/EBPα, C/EBPβ, C/EBPε, IRF4, JUNB, PU.1, STAT1, STAT3, VDR	[30]
2014	T. Nikolic, D. Movita, M. Lambers et al.	CTCF	[31]
2012	T. Suzuki, M. Nakano-Ikegaya, H. Yabukami-Okuda et al.	BTG2, CEBPα, CREG1, EGR2, FOS, FOSB, HCLS1, IRF8, JUNB, LMO2, LYL1, MAF, MAFB, MNDA, MXD1, MYB, MYC, PU.1, STAT5A, ZFP36	[32]

Table 1 Preliminary analysis of previous literature yielding a list of monocyte-specific TFs.

3.2.2 Previous Literature vs. Results Obtained

A ranked list of about 140 unique TFs according to adjusted *P* value was obtained after the statistical analysis of Sec. 2.4 was completed. Among all ENCODE TFs ranked by monocyte eQTL overlap *P* value, known monocyte TFs are biased toward more significant *P* values (Fig. 2).

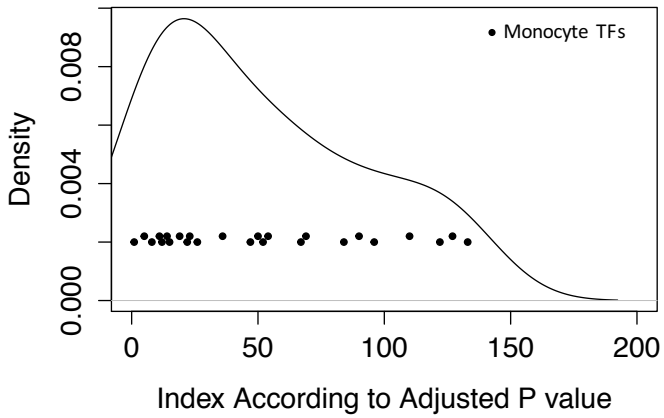

Fig. 2 Density plot displaying the concentration of monocyte-specific TFs from previous studies near the top of the list of TFs ranked according to adjusted *P* value.

3.3 Analysis of Ranked List of TFs

The top 20% of TFs in the ranked list was obtained by selecting an adjusted *P* value cutoff of $P < 1.00 \times 10^{-7}$ (Table 2). GSEA was conducted for all TFs that were not identified to be monocyte-specific in previous literature, referred to as "Novel Monocyte TFs" (NMTFs). This analysis was performed to test whether the NMTFs were concentrated near top of the TFBS list when their individual ChIP-seq experiments are ranked according to *Z*-score. Specifically, for each NMTF shown in Table 2, the *Z*-scores for all ChIP-seq experiments for that TF (within the global set of ChIP-seq experiments ranked by *Z*-score) were analyzed by GSEA. As shown in Table 2, NMTFs in the top 20% of TFs (by *P* value) appear to have high *E* scores. The average *E* score of the top 20% of NMTFs is 0.823, while the average *E* score of the middle 20% of TFs is 0.099.

Rank	Transcription Factor Name(s)	Adjusted *P* value	*E* score
1	**HEY1**	1.37×10^{-38}	--
2	BCLAF1	4.10×10^{-35}	0.899
3	SIN3A	7.17×10^{-35}	0.719
4	TAF1	4.45×10^{-27}	0.778
5	**GABP/ GABPA**	3.19×10^{-23}	--
6	TBP	2.10×10^{-22}	0.849
7	GTF2F1	1.97×10^{-21}	0.866
8	**GATA1**	2.91×10^{-18}	--
9	ELK4	3.07×10^{-17}	0.714
10	MXI1	6.11×10^{-17}	0.922
11	**BCL3**	3.26×10^{-16}	--
12	**IRF1**	1.49×10^{-15}	--
13	**NFKB/ NFKB1**	8.35×10^{-14}	--
14	**ELF1**	2.02×10^{-13}	--
15	SP2	2.06×10^{-13}	0.932
16	NRF1	8.30×10^{-13}	0.870
17	GTF2B	1.04×10^{-12}	0.924
18	**MYC**	2.19×10^{-12}	--
19	YY1	1.53×10^{-11}	0.690
21	**P300/EP300**	2.26×10^{-10}	--
22	**FOS**	3.89×10^{-10}	--
23	THAP1	1.41×10^{-09}	0.859
24	ETS1	2.81×10^{-09}	0.858
25	**E2F4**	3.06×10^{-09}	--
27	TAF7	3.53×10^{-08}	0.823
28	TR4/NR2C2	1.03×10^{-07}	0.602

Table 2 Top 20% TFs ranked according to adjusted *P* value with cut off of $P < 1.00 \times 10^{-7}$. Bolded names are TFs that have been identified to be monocyte-specific in previous literature.

GSEA plots of two top-ranked TFs, TAF1 and TBP, are shown in Fig. 3. These TFs had high GSEA *E* scores based on multiple TFBS experiments (ten and five, respectively).

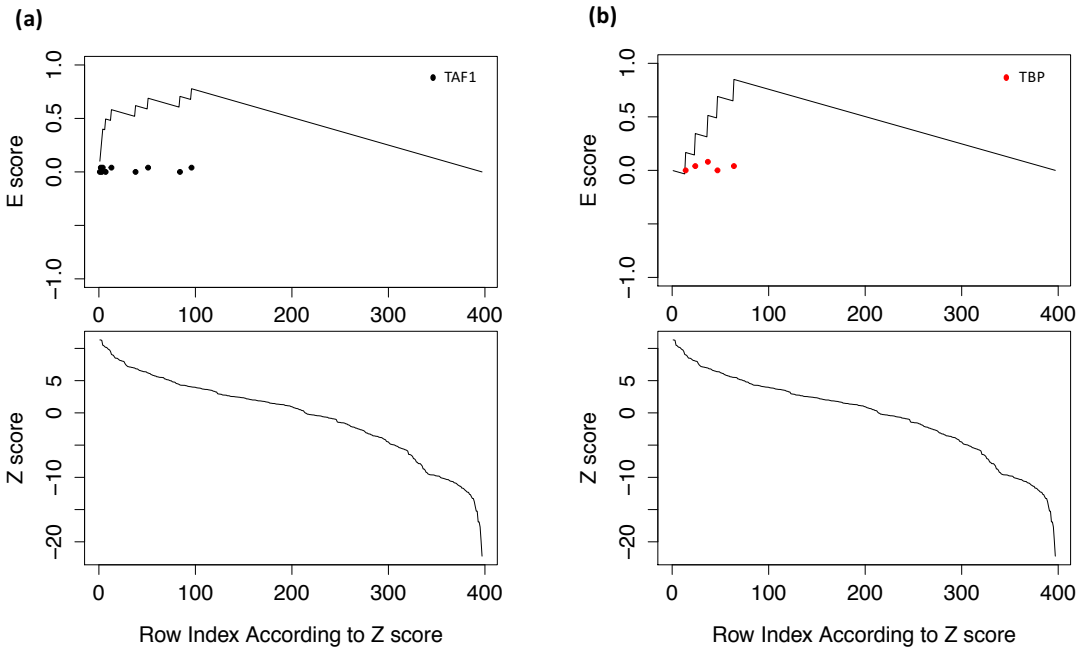

Fig. 3 GSEA plots of highly ranked TFs, (**a**) TAF1 and (**b**) TBP. Black and red points on uppermost graphs describe locations of each TF within the list of ChIP-seq experiments when ranked by *Z*-score. Distribution of *Z*–scores is shown directly below GSEA plots.

3.4 Predicting Relationships in the Gene Regulatory Network

A network of predicted relationships between the top 20% of TFs was obtained using the GeneMANIA bioinformatics web tool, which identifies relationships among proteins by using a large set of functional association data [12]. Predicted relationships between the top TFs are displayed in Fig. 4 as a network of proteins connected by physical interactions, co-expression, and predicted interactions.

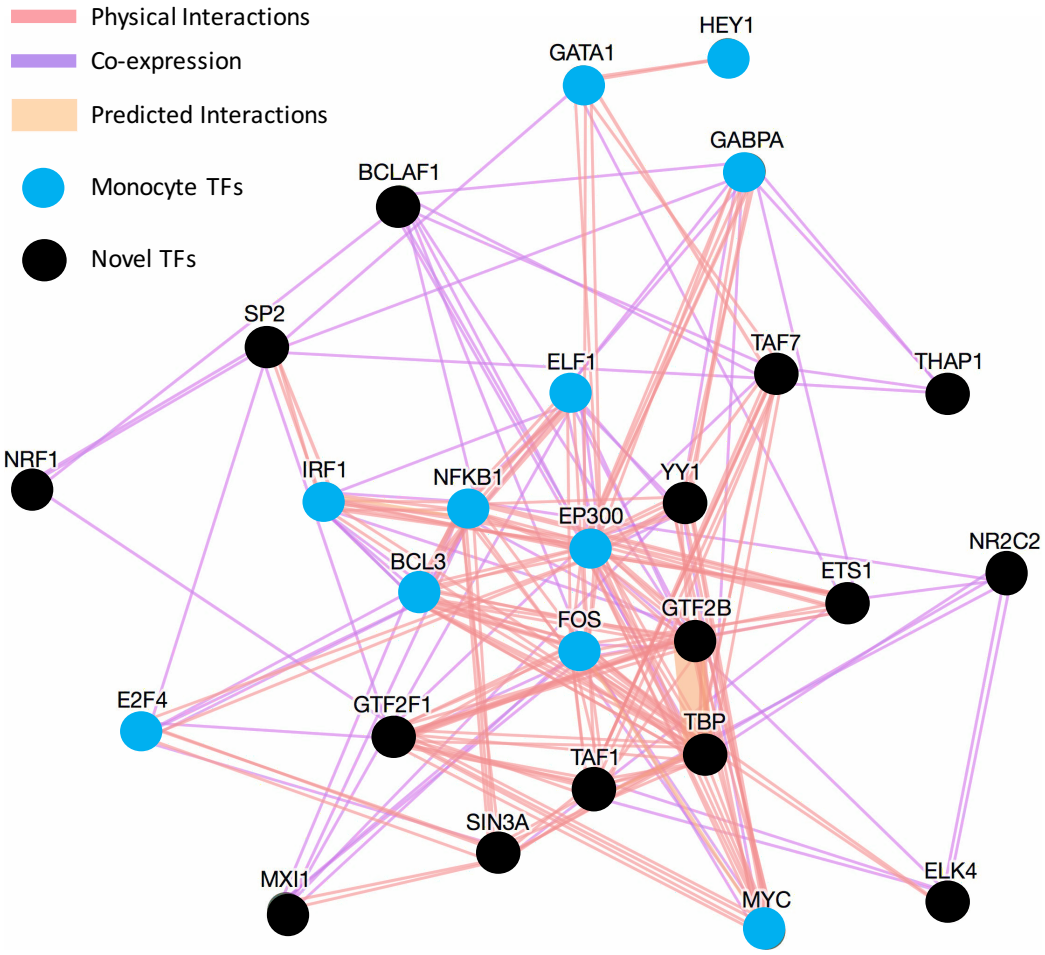

Fig. 4 TF network predicted by GeneMANIA. Each node represents a TF from Table 2. Monocyte-specific TFs from previous studies (blue circles) appear to have high connectedness to many novel TFs (black circles) that have been identified to be in the top 20% of the ranked list of TFs.

3.5 Analysis of Novel Monocyte-specific TFs

TFs in the top 20% of the ranked list that were not previously identified to be monocyte-specific were investigated in order to examine their known molecular and/or cellular functions. Mouse orthologs of the NMTF genes were identified using the ImmGen project database in order to obtain expression levels for the TFs in murine monocytes and other immune cell types [13]. Human microarray-based gene

expression measurements were also reviewed for each novel TF, using data from the gene annotation portal, BioGPS [14].

Of the NMTFs identified in this study, many have been previously identified to have functions related to immunity or have relationships with previously identified monocyte-specific TFs. The NMTF identified with the most significant *P* value in Table 2, BCLAF1, is highly expressed in B cells, T cells, natural killer (NK) cells, and monocytes in mice [13, 14]. It has also been found to play a crucial role in transducing the senescence-inducing signal between NF-κB and C/EBPβ during therapeutic drug senescence in multiple cancer cells [33]. Both NF-κB and C/EBPβ are high-scoring TFs in Table 1, and this relationship with BCLAF1 may provide insight into its function in monocytes. Another significant NMTF, SIN3A, has been identified along with the monocyte-specific TF, MYC, as playing a role in a signaling circuit which may be the underlying mechanism for the pathogenesis of acute myeloid leukemia [34]. Although limited data is available about SIN3A expression in monocytes and its function in immunity, our study shows that it may have a more prominent role in monocytes than previously recognized.

The NMTFs TAF1, TAF7, and TBP may have dependencies on each other in their function in monocytes, as they are components of the Transcription Factor II D (TFIID), which is the first protein to bind DNA during the initiation of transcription [35]. TBP, or the TATA binding protein, has high expression in both human and mouse monocyte cells [13, 14]. Additionally, its expression levels across various human cell lines correlate with the expression of other monocyte-regulating TFs, such as PU.1, C/EBPβ and NF-κB [33]. TAF7 appears to have a significantly

high expression in monocytes, and has the highest expression in immune cells when compared to all other cell types [13, 14]. It also has been recognized that in human HeLa cells the level of expression of TAF7 directly correlates with the level of expression of TAF1, a NMTF that has not been well studied for its role in immunity [36].

Many other novel TFs were identified to have functions in immunity. Evaluation of mouse liver proteins after host defense response against the parasitic protozoan *Toxoplasma gondii* has shown that the NMTF SP2 (along with other monocyte-regulating TFs, STAT1, STAT3, IRF1, and EGR1) is among the top ten TFs used by upregulated genes [37]. YY1 is known to significantly regulate the activity of the promoter of murine IRF-3, a TF which plays a key role in inflammation and immune response [38]. Furthermore, the highly-ranked NMTF, GTF2B, is known to be a target of the YY1 TF gene in bovine muscle tissue [39].

THAP1, ETS1, TR4, and NRF1 are also highly ranked NMTFs that have been previously identified to have specific immune regulatory functions. THAP1 is known to play a role in inducing T-cell apoptosis [40]. STAT3, a monocyte-regulating TF, is a known regulator of ETS1 in inflammation control in mouse macrophages [41]. In addition, TR4 has been found to be a target of mouse miR-133a, microRNA that is expressed in macrophages and that is known to play a key role in the vascular inflammatory disease, atherosclerosis [42]. Lastly, NRF1 has been identified as an essential transcription factor in regulating gene expression in human integrin associated protein, which is expressed in hematopoietic cells [43].

The remaining top NMTFs, which includes GTF2F1, ELK4, and MXI1, appear to be expressed in monocytes to a lesser extent than other cell types or have not been studied in relation to the immune system. ELK4 and MXI1 have medium-to-low expression in mouse monocytes [13]. However, in human LNCaP (prostate adenocarcinoma) cells, the NMTF ELK4 has been found to be regulated by CTCF, a known monocyte-regulating TF [44]. These TFs require more in-depth analysis in order to determine their functions, if any, in monocytes.

4 Conclusion

Our novel systems biology approach offers an efficient method to identify monocyte-regulating TFs and gain insight into the monocyte gene regulatory network, leveraging publicly available human molecular datasets. It appears that TFs with known immune functions and monocyte-specificity tend to have binding sites located in the vicinity of monocyte eQTLs. Of the candidate novel monocyte-regulating TFs that were identified in our analysis, the most significant (by our statistical enrichment score) were BCLAF1, SIN3A, TAF1, and TBP. As revealed by protein network reconstruction, the top 20% of TFs identified by our approach are highly interconnected. We speculate that the network is enriched for TF pairs that cooperatively regulate monocyte gene expression (a prediction that could be tested in a future study, through pairwise binding site proximity analysis).

Further experimentation on the TFs identified will provide greater insight on their role in monocytes and the monocyte gene regulatory network. This can yield novel information on genes to target when developing therapies for chronic

inflammatory disorders, such as atherosclerosis. Moreover, this approach can be utilized to identify cell-type specific TFs in order to predict mammalian gene regulatory networks for various cell types.

Acknowledgments

I extend my sincere thanks to Dr. Stephen A. Ramey for the opportunity to work in his lab and for his mentorship throughout this project. I am also grateful to Alvin Yu and all of my fellow lab-mates for providing me with help through various steps of this project. I thank the Director of the BioResource Research Interdisciplinary Program, Dr. Katharine G. Field, and Dr. Thomas J. Sharpton for serving on my thesis committee. I also extend my appreciation to Wanda Crannell, my academic advisor, who greatly supported me throughout my undergraduate studies. I sincerely thank my friends and family for standing behind me and encouraging me throughout my academics. Lastly, I would like to acknowledge our financial support. This work was funded by the NIH (award HL098807 to SAR), the NSF (awards 1557605-DMS and 1553728-DBI to SAR), the PhRMA Foundation (Research Starter Grant award to SAR), the Medical Research Foundation of Oregon (New Investigator Grant to SAR), and Oregon State University (Division of Health Sciences Interdisciplinary Research Grant to SAR), the OSU DeLoach Work Scholarship Program, the OSU Undergraduate Research Innovation Scholarship, and the OSU Honors Experience Award.

References

1. Shi C, Pamer EG (2011) Monocyte recruitment during infection and inflammation. Nat Rev Immunol 11:762–74. doi: 10.1038/nri3070
2. Medzhitov R, Horng T (2009) Transcriptional control of the inflammatory response. Nat Rev Immunol 9:692–703. doi: 10.1038/nri2634
3. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. Nat Rev Genet 10:252–63. doi: 10.1038/nrg2538
4. Fairfax BP, Makino S, Radhakrishnan J, et al. (2012) Genetics of gene expression in primary immune cells identifies cell type–specific master regulators and roles of HLA alleles. Nat Genet 44:502–510. doi: 10.1038/ng.2205
5. Zeller T, Wild P, Szymczak S, et al. (2010) Genetics and Beyond – The Transcriptome of Human Monocytes and Disease Susceptibility. PLoS One 5:e10693. doi: 10.1371/journal.pone.0010693
6. Landt SG, Marinov GK, Kundaje A, et al. (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 22:1813–31. doi: 10.1101/gr.136184.111
7. Sherry ST, Ward MH, Kholodov M, et al. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29:308–11.
8. Johnson AD, Handsaker RE, Pulit SL, et al. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24:2938–2939. doi: 10.1093/bioinformatics/btn564
9. R Development Core Team (2011) R: A Language and Environment for Statistical Computing.
10. Lozzio BB, Lozzio CB, Bamberger EG, Feliu AS (1981) A Multipotential Leukemia Cell Line (K-562) of Human Origin. Exp Biol Med 166:546–550. doi: 10.3181/00379727-166-41106
11. Subramanian A, Tamayo P, Mootha VK, et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102:15545–50. doi: 10.1073/pnas.0506580102
12. Warde-Farley D, Donaldson SL, Comes O, et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 38:W214–20. doi: 10.1093/nar/gkq537

13. Heng TSP, Painter MW (2008) The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol 9:1091–4. doi: 10.1038/ni1008-1091

14. Wu C, Jin X, Tsueng G, et al. (2015) BioGPS: building your own mash-up of gene annotations and expression profiles. Nucleic Acids Res 44:D313–316. doi: 10.1093/nar/gkv1104

15. Friedman AD (2002) Transcriptional regulation of granulocyte and monocyte development. Oncogene 21:3377–90. doi: 10.1038/sj.onc.1205324

16. Valledor AF, Borràs FE, Cullell-Young M, Celada A (1998) Transcription factors that regulate monocyte/macrophage differentiation. J Leukoc Biol 63:405–17.

17. Fairfax BP, Humburg P, Makino S, et al. (2014) Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science 343:1246949. doi: 10.1126/science.1246949

18. Dong C, Zhao G, Zhong M, et al. (2013) RNA sequencing and transcriptomal analysis of human monocyte to macrophage differentiation. Gene 519:279–87. doi: 10.1016/j.gene.2013.02.015

19. Shi C, Zhang X, Chen Z, et al. (2004) Integrin engagement regulates monocyte differentiation through the forkhead transcription factor Foxp1. J Clin Invest 114:408–18. doi: 10.1172/JCI21100

20. Kurotaki D, Osato N, Nishiyama A, et al. (2013) Essential role of the IRF8-KLF4 transcription factor cascade in murine monocyte differentiation. Blood 121:1839–49. doi: 10.1182/blood-2012-06-437863

21. Rosenbauer F, Tenen DG (2007) Transcription factors in myeloid development: balancing differentiation with transformation. Nat Rev Immunol 7:105–17. doi: 10.1038/nri2024

22. Kikuchi H, Kuribayashi F, Kiwaki N, et al. (2011) GCN5 regulates the superoxide-generating system in leukocytes via controlling gp91-phox gene expression. J Immunol 186:3015–22. doi: 10.4049/jimmunol.1000364

23. Larabee JL, Shakir SM, Barua S, Ballard JD (2013) Increased cAMP in monocytes augments Notch signaling mechanisms by elevating RBP-J and transducin-like enhancer of Split (TLE). J Biol Chem 288:21526–36. doi: 10.1074/jbc.M113.465120

24. Van den Bossche J, Malissen B, Mantovani A, et al. (2012) Regulation and function of the E-cadherin/catenin complex in cells of the monocyte-macrophage lineage and DCs. Blood 119:1623–33. doi: 10.1182/blood-2011-10-384289

25. Khan KA, Coaquette A, Davrinche C, Herbein G (2009) Bcl-3-regulated transcription from major immediate-early promoter of human cytomegalovirus in monocyte-derived macrophages. J Immunol 182:7784–94. doi: 10.4049/jimmunol.0803800

26. Resendes KK, Rosmarin AG (2006) GA-binding protein and p300 are essential components of a retinoic acid-induced enhanceosome in myeloid cells. Mol Cell Biol 26:3060–70. doi: 10.1128/MCB.26.8.3060-3070.2006

27. Weigelt K, Carvalho LA, Drexhage RC, et al. (2011) TREM-1 and DAP12 expression in monocytes of patients with severe psychiatric disorders. EGR3, ATF3 and PU.1 as important transcription factors. Brain Behav Immun 25:1162–9. doi: 10.1016/j.bbi.2011.03.006

28. Marullo M, Valenza M, Mariotti C, et al. (2010) Analysis of the repressor element-1 silencing transcription factor/neuron-restrictive silencer factor occupancy of non-neuronal genes in peripheral lymphocytes from patients with Huntington's disease. Brain Pathol 20:96–105. doi: 10.1111/j.1750-3639.2008.00249.x

29. Li P, Wong JJ-Y, Sum C, et al. (2011) IRF8 and IRF3 cooperatively regulate rapid interferon-β induction in human blood monocytes. Blood 117:2847–54. doi: 10.1182/blood-2010-07-294272

30. Huber R, Pietsch D, Günther J, et al. (2014) Regulation of monocyte differentiation by specific signaling modules and associated transcription factor networks. Cell Mol Life Sci 71:63–92. doi: 10.1007/s00018-013-1322-4

31. Nikolic T, Movita D, Lambers MEH, et al. (2014) The DNA-binding factor Ctcf critically controls gene expression in macrophages. Cell Mol Immunol 11:58–70. doi: 10.1038/cmi.2013.41

32. Suzuki T, Nakano-Ikegaya M, Yabukami-Okuda H, et al. (2012) Reconstruction of monocyte transcriptional regulatory network accompanies monocytic functions in human fibroblasts. PLoS One 7:e33474. doi: 10.1371/journal.pone.0033474

33. Adamik J, Wang KZQ, Unlu S, et al. (2013) Distinct mechanisms for induction and tolerance regulate the immediate early genes encoding interleukin 1β and tumor necrosis factor α. PLoS One 8:e70622. doi: 10.1371/journal.pone.0070622

34. Jiang X, Hu C, Arnovitz S, et al. (2016) miR-22 has a potent anti-tumour role with therapeutic potential in acute myeloid leukaemia. Nat Commun 7:11452. doi: 10.1038/ncomms11452

35. Bieniossek C, Papai G, Schaffitzel C, et al. (2013) The architecture of human general transcription factor TFIID core complex. Nature 493:699–702. doi: 10.1038/nature11791

36. Devaiah BN, Lu H, Gegonne A, et al. (2010) Novel functions for TAF7, a regulator of TAF1-independent transcription. J Biol Chem 285:38772–80. doi: 10.1074/jbc.M110.173864

37. He J-J, Ma J, Elsheikha HM, et al. (2016) Proteomic Profiling of Mouse Liver following Acute Toxoplasma gondii Infection. PLoS One 11:e0152022. doi: 10.1371/journal.pone.0152022

38. Xu H-G, Liu L, Gao S, et al. (2016) Cloning and characterizing of the murine IRF-3 gene promoter region. Immunol Res. doi: 10.1007/s12026-015-8780-8

39. Moisá SJ, Shike DW, Meteer WT, et al. (2013) Yin yang 1 and adipogenic gene network expression in longissimus muscle of beef cattle in response to nutritional management. Gene Regul Syst Bio 7:71–83. doi: 10.4137/GRSB.S11783

40. Lu C, Li J-Y, Ge Z, et al. (2013) Par-4/THAP1 complex and Notch3 competitively regulated pre-mRNA splicing of CCAR1 and affected inversely the survival of T-cell acute lymphoblastic leukemia cells. Oncogene 32:5602–13. doi: 10.1038/onc.2013.349

41. Zhang H, Hu H, Greeley N, et al. (2014) STAT3 restrains RANK- and TLR4-mediated signalling by suppressing expression of the E2 ubiquitin-conjugating enzyme Ubc13. Nat Commun 5:5798. doi: 10.1038/ncomms6798

42. Peng X-P, Huang L, Liu Z-H (2016) miRNA-133a attenuates lipid accumulation via TR4-CD36 pathway in macrophages. Biochimie. doi: 10.1016/j.biochi.2016.04.012

43. Chang W-T, Huang A-M (2004) Alpha-Pal/NRF-1 regulates the promoter of the human integrin-associated protein/CD47 gene. J Biol Chem 279:14542–50. doi: 10.1074/jbc.M309825200

44. Qin F, Song Y, Zhang Y, et al. (2016) Role of CTCF in Regulating SLC45A3-ELK4 Chimeric RNA. PLoS One 11:e0150382. doi: 10.1371/journal.pone.0150382