Title
Computational Methods for the Analysis of DNA Methylation and Gene Expression Data

Author
Lam, Larry Tao

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Computational Methods for the Analysis of DNA Methylation and Gene Expression Data

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy in Bioinformatics

by

Larry Tao Lam

2016

ABSTRACT OF THE DISSERTATION

Analytical Strategies for the Interpretation of DNA Methylation

and Gene Expression Data

by

Larry Tao Lam

Doctor of Philosophy in Bioinformatics

University of California, Los Angeles, 2016

Professor Matteo Pellegrini, Chair

RNA expression profiling and DNA methylation analysis have been essential tools in understanding genomic mechanisms underlying human health and disease. Although many annotation databases are publically available, alternative data resources may be overlooked. This work focuses on the development of computational tools and strategies that incorporate results from both the leading functional annotation tools as well as working directly with publicly available expression and methylation datasets. Chapter 1 outlines the leading approaches for interpreting DNA methylation and RNA expression analyses. In addition, chapter 1 provides a brief background of Burkitt's lymphoma and amyotophic lateral sclerosis (ALS) for studies discussed in later chapters. In chapter 2, we developed a set of methylation characterization and visualization tools for bisulfite sequencing data. These tools also characterize methylation levels at genomic features, like gene bodies as well as transcription factor targets. We provide a means

to detect epigenetic regulation of transcription factor binding sites. Chapter 3 describes a multi-omics approach to understand an epigenetic mechanism for chemoresistance in Burkitt's lymphoma. Burkitt's lymphoma cell lines were cultured with drugs and developed increasing levels of resistance to chemotherapy. By analyzing transcriptional profiles of the chemoresistant cell lines with healthy B-cells at different stages of maturation as well as subsequent integration of DNA methylation and ChiP-Seq data from the chemoresistant cell lines, we were able to propose a novel mechanism of drug resistance in which E2a and PRC2 drive changes in the B-cell epigenome. In chapters 3 and chapter 4, we focused on the transcriptional and DNA methylation analysis of peripheral blood mononuclear cells (PBMCs) of patients affected with amyotrophic lateral sclerosis (ALS). Using transcriptional data of monocytes stimulated by different molecules, we were able to categorize our samples into inflammatory and non-inflammatory groups. A pathway enrichment analysis of the differentially expressed genes reveals potential targets of immune based treatments for ALS. In chapter 5, we investigated the differences in DNA methylation profiles in PBMCs from a pair of monzygotic twins discordant in the diagnosis for ALS. We developed a cell type abundance analysis method which suggest that the affected twin loses T-cells and gains monocytes during the course of the disease. Our direct use of reference data sets highlights the potential for understanding RNA-Seq and BS-Seq data and provides the groundwork for development of generalized transcription or methylation analysis tools, like CEllFi. Chapter 6 outlines the implementation of CEllFi, a bisulfite sequencing based method that allows for cellular deconvolution of heterogenous samples.

The dissertation of Larry Tao Lam is approved.

Thomas G Graeber

Jason Ernst

Dan Ruderman

Matteo Pellegrini, Committee Chair

University of California, Los Angeles

2016

This work is dedicated to my wife, Ivy Tang, for her love, support, and encouragement.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Chapter two is a version of the article by Liao, Wen-Wei, Ming-Ren Yen, Evaline Ju, Fei-Man Hsu, Larry Lam, and Pao-Yang Chen. 2015. “MethGo: A Comprehensive Tool for Analyzing Whole-Genome Bisulfite Sequencing Data.” *BMC Genomics* 16 (12): 1–8. doi:10.1186/1471-2164-16-S12-S11. PC conceived the project. WL, EJ and LL implemented the software. MY wrote the manuscript. FH, EJ, LL and PC edited the manuscript. All authors have read and approved the final manuscript.

Chapter three is a version of the article by Flinders, Colin, Larry Lam, Liudmilla Rubbi, Roberto Ferrari, Sorel Fitz-Gibbon, Pao-Yang Chen, Michael Thompson, Heather Christofk, David B Agus, Daniel Ruderman, Parag Mallick, and Matteo Pellegrini. 2016. “Epigenetic Changes Mediated by Polycomb Repressive Complex 2 and E2a Are Associated with Drug Resistance in a Mouse Model of Lymphoma.” Genome Medicine 8 (1). Genome Medicine: 1-11. doi:10.1186/s13073-016-0305-0. CF, RF, and LR performed experiments, CF, LL, RF, SFG, PYC, DR, and MP analyzed results, MT performed methylation analysis of the Cancer Genome Atlas data, CF and LL wrote the manuscript. DBA provided clinical input. DR, PM, and MP supervised the project. All authors read and approved the final manuscript.

Chapter four is a version of the article by Lam, Larry, Ramesh C Halder, Dennis J Montoya, Liudmilla Rubbi, Arturo Rinaldi, Bien Sagong, Sarah Weitzman, Rachel Rubattino, Ram Raj Singh, Matteo Pellegrini, and Milan Fiala. 2015. “Anti-Inflammatory Therapies of Amyotrophic Lateral Sclerosis Guided by Immune Pathways.” American Journal of Neurodegenerative Disease 4 (2): 28-39. http://europepmc.org/articles/PMC4700124. MF and MP designed research. LL and DM performed bioinformatics analysis. RH, LR, AR performed

molecular studies. MF, MP, LL, BS, SW, and RR analyzed the data. MF, MP, LL, RRS wrote the paper. All authors read and approved the final manuscript.

Chapter five is a version of the article by Lam, Larry, Lydia Chin, Ramesh C. Halder, Bien Sagong, Sam Famenini, James Sayre, Dennis Montoya, Liudmilla Rubbi, Matteo Pellegrini, and Milan Fiala. 2016. "Epigenetic Changes in T-Cell and Monocyte Signatures and Production of Neurotoxic Cytokines in ALS Patients." Federation of American Societies for Experimental Biology, 1-13. doi:10.1096/fj.201600259RR. M. Fiala and M. Pellegrini designed the research; L. Lam, L. Chin, R. C. Halder, B. Sagong, S. Famenini, and L. Rubbi performed the research; L. Lam, J. Sayre, D. Montoya, M. Pellegrini, and M. Fiala analyzed the data; and M. Fiala, L. Lam, and M. Pellegrini wrote the paper.

The work in chapter 2 was supported by a grant from Academia Sinica, and grants from MOST-103-2313-B-001-003-MY3 and MOST-103-2633-B-001-002 and NHRI-EX104-10324SC. The work presented in chapter 3 was funded by grant H47213-5 U54 CA143907-02, Physical Sciences in Oncology by the National Cancer Institute. Funding sources in chapter four by LL was supported by 5P01GM099134-03, DM was supported by 1P50AR063020-01, and RRS was supported by R01 AI80778. The study presented in chapter five was provided by U.S. National Institutes of Health (NIH) National Institute of Human Genome Research Grant T32 HG008553-01 (to L.M.) and NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases Grant 3P50AR063020-03S1 (to D.M.). This dissertation was funded in part by the UCLA Biomedical Big Data Training Grant 1 T32 CA201160-01.

VITA

Education:

University of California, Los Angeles — Los Angeles, CA
Bioinformatics Ph.D. Candidate — 2011 - 2016

California State University, Los Angeles — Los Angeles, CA
B.S. Biology — 2011

University of Southern California — Los Angeles, CA
B.S. Computer Science — 2004

Grants and Awards:

Biomedical Big Data Training Program — 2015 - 2016
University of California, Los Angeles

Publications:

Lam L., Hong F., Morselli M., Rubbi L., Pellegrini M., Montoya D. BS-Deconvolution: A Bisulfite Sequencing Based Method for Cellular Deconvolution of Heterogeneous Samples. *In Preparation*

Lam L., Chin L., Halder R.C., Sagong B., Famenini S., Sayre J., Montoya D., Rubbi L., Pellegrini M., Fiala M. (2016). Epigenetic changes in T-cell and monocyte signatures and production of neurotoxic cytokines in ALS patients. *Federation of American Societies for Experimental Biology*. 1–13. http://doi.org/10.1096/fj.201600259RR

Brumm, A. J., Nunez, S., Doroudchi, M. M., Kawaguchi, R., Duan, J., Pellegrini, M., **Lam L.**, Carmichael S.T., Deb A., Hinman, J. D. (2016). Astrocytes Can Adopt Endothelial Cell Fates in a p53-Dependent Manner. *Molecular Neurobiology*. 1–13. http://doi.org/10.1007/s12035-016-9974-3

Flinders C., **Lam L.**, Rubbi L., Ferrari R., Fitz-Gibbon S., Chen P.Y., Thompson M., Christofk H.B., Agus D., Ruderman D., Mallick P., Pellegrini M. (2016). Epigenetic changes mediated by polycomb repressive complex 2 and E2a are associated with drug resistance in a mouse model of lymphoma. *Genome Medicine*. 8(1):1-11.

Lam L., Halder R.C., Montoya D.J., Rubbi L., Rinaldi A., Sagong B., Weitzman S., Rubattino R., Singh R.R., Pellegrini M., Fiala M. (2015). Anti-inflammatory therapies of amyotrophic lateral sclerosis guided by immune pathways. *American Journal of Neurodegenerative Disease*. 4(2), 28-39.

Liao W-W., Yen M-R., Ju E., Hsu F-M., **Lam L**., Chen P-Y. (2015). MethGo: a comprehensive tool for analyzing whole-genome bisulfite sequencing data. *BMC Genomics*, 16(12):1–8.

Presentations and Posters:

BS-Deconvolution: A Bisulfite Sequencing Based Method for Cellular Deconvolution of Heterogeneous Samples. (Poster). Intelligent Systems for Molecular Biology. Orlando, FL. July 2016

BS-Analyze: A feature-rich post-processing analysis and visualization tool for bisulfate sequencing. (Talk). Moscone Center, Tri-Con. San Francisco, CA. February 2015.

A Cancer Evolution Space-Time Machine. (Talk). UCLA Conference Center, MCDB Annual Research Conference. Lake Arrowhead, CA. December 2013.

Chapter 1:

Introduction

1.1 Introduction to interpreting RNA expression

RNA expression profiling has been an essential tool in understanding biological activity at the genomic level. One specific tool for quantifying expression levels, RNA-Seq, provides transcriptome wide measurements with high levels of precision compared to alternative high throughput methods (Wang, Gerstein, & Snyder, 2009). While RNA-Seq provides quantification of tens of thousands of genes, analyzing and understanding expression dynamics remains a challenge.

Typical studies using RNA-Seq seek to identify genes with differential levels of expression between case control studies. A common approach in producing a list of differentially expressed genes, is to test that the fold change between the two groups of samples are significantly different from the null. Some of the most commonly used tools to test for differential expression are edgeR and DESeq2 (M. I. Love, Huber, & Anders, 2014; Robinson, McCarthy, & Smyth, 2009). Once a list of differentially expressed genes are produced, tools like DAVID can identify pathways or gene families enriched in the differentially expressed genes (Huang, Lempicki, & Sherman, 2009). Another approach for interpreting gene expression levels of interest is to test for enrichment of expression levels against different sets of genes stored in databases, like MSigDB (Liberzon et al., 2011; Subramanian et al., 2005). Although many tools and data sources have been developed to understand RNA-Seq results, summarizing expression results remains cumbersome.

1.2 Introduction to DNA methylation analysis

An important epigenetic mark that is related to gene expression is the methylation of cytosine. DNA methylation levels can be affected by the expression of genes and in the promoter are often associated with gene silencing (Laird, 2003). Studies have also shown that DNA

methylation levels can change during cancer progression (Landan 2012). The gold standard in performing genome wide methylation profiling, whole genome bisulfite sequencing (WGBS), couples sodium bisulfite conversion of unmethylated cytosines with PCR amplification and short read sequencing, providing methylation quantitation at single nucleotide resolution. An alternative and cost effective approach to WGBS, is reduced representation bisulfite sequencing (RRBS). RRBS utilizes restriction enzymes to enrich for DNA fragments high in CpG dinucleotides (Meissner et al., 2005). Using bisulfite sequencing alignment tools, like BSSeeker2, one can produce methylation level calls for further post-alignment analyses (Guo et al., 2013).

Many DNA methylation analyses focus on identifying and annotating loci or fragments that are differentially methylated between groups. One metric for detecting differentially methylated cytosines is the difference in the percent of methylated cytosines in one group compared to another. As changes in methylation often occur in hotspots that correspond to CpG islands, it is useful in aggregating the methylation levels into regions or fragments (Lister et al., 2011). In order to annotate lists of differentially methylated regions, regions may be associated to genes based on proximity to transcriptional start sites. Lists of gene families may be subsequently tested for enrichment against pathway databases (McLean et al., 2010).

1.3 Epigenetics of Burkitt's Lymphoma

Burkitt's lymphoma is a B-cell lymphoma characterized by the rapid proliferation of malignant cells and deregulation of the *MYC* gene via translocation to one of the immunoglobulin loci (C. Love et al., 2012). Multiagent chemotherapy can lead to complete remission in 90-95% of children and 60-70% of adults. Patients that show primary chemoresistance frequently die early. Although rescue chemotherapy is effective to children that

relapse, success has not been reported for adults. Previous studies have shown that the epigenetic modifier 5-aza-2-deoxycytidine is able to repress *MYC* expression. Yet, understanding the mechanism of epigenetic repression as well as the molecular basis for chemoresistance remains an important clinical goal (Guan et al., 2013; Richter-Larrea et al., 2010).

1.4 ALS and Autoimmunity

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disorder that leads to upper and lower motor neuron loss and is associated with frontotermporal dementia. The median survival of ALS is 3 to 5 years after symptomatic onset with 20% surviving beyond 5 years. It has been theorized that autoimmunity is part of the pathogenesis or mediated in the apparent sporadic nature across affected patients (Turner, Talbot, & Goldacre, 2013). With increasing evidence associating autoimmune disease and ALS, there is greater interest in modulating inflammation to mediate the effects of ALS.

1.5 Project overview

In this dissertation, multiple analytical tools and strategies were developed and employed for the interpretation of DNA methylation and expression data. One key aim of this dissertation is to develop the computational tools to aid in the interpretation of both expression and methylation analyses. The second aim is to identify genes and genomic regions that change in expression and DNA methylation in concert with the changes in cell line resistance. Using a mouse model for Burkitt's lymphoma, several lines of Eμ- myc /Arf-/- cells were generated with varying levels of resistance to mafosfamide treatment. A multi-omics approach was employed to identify the mechanisms leading to the change in resistance across the cell lines. Data sets generated from the chemoresistant cells include microarray expression, RRBS, Chip-Seq, DNA-Seq, and metabolomics. Chapter 3 of this dissertation reports methylation fragments with

monotonic changes across the cell lines, and a covariance analysis is performed to identify the extreme genes and DNA fragments whose expression or methylation vary with the cell lines in concert with resistance.

The third aim is to understand the immune activity of ALS patients. RNA-Seq was performed on the peripheral blood mononuclear cells (PBMCs) of 9 ALS patients and one unaffected identical twin. Using an unsupervised clustering analysis of the 10 samples, two groups were identified as inflammatory and non-inflammatory using expression signatures derived from perturbed PBMCs from published expression datasets. Chapter 4 of this work reports the immune pathways enriched in the differentially expressed genes between the inflammatory and non-inflammatory samples. The fourth aim is to identify the epigenetic and expression differences between the PBMCs of monozygotic twins discordant for ALS. Both RRBS and RNA-Seq was performed on the patient PBMCs. Chapter 5 of this dissertation reports the differentially expressed genes and differentially methylated fragments between the identical twins using the RRBS analysis tool CEllFi. CEllFi's implementation is described in Chapter 6.

1.2 References

Guan, H., Xie, L., Klapproth, K., Weitzer, C. D., Wirth, T., & Ushmorov, A. (2013). Decitabine represses translocated MYC oncogene in Burkitt lymphoma. *J Pathol*, *229*. http://doi.org/10.1002/path.4164

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., & Zhang, M. Q. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*, *14*. http://doi.org/10.1186/1471-2164-14-774

Huang, D. W., Lempicki, R. a, & Sherman, B. T. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. http://doi.org/10.1038/nprot.2008.211

Laird, P. W. (2003). The power and the promise of DNA methylation markers. *Nature Reviews. Cancer*, *3*(April), 253–266. http://doi.org/10.1038/nrc1045

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, *27*(12), 1739–1740. http://doi.org/10.1093/bioinformatics/btr260

Lister, R., Pelizzola, M., Kida, Y. S., Hawkins, R. D., Nery, J. R., Hon, G., … Ecker, J. R. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*, *471*(7336), 68–73. http://doi.org/10.1038/nature09798

Love, C., Sun, Z., Jima, D., Li, G., Zhang, J., Miles, R., … Dave, S. S. (2012). The genetic landscape of mutations in Burkitt lymphoma. *Nature Genetics*, *44*(12), 1321–5. http://doi.org/10.1038/ng.2468

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. http://doi.org/10.1186/s13059-014-0550-8

Martens, J. H. A., & Stunnenberg, H. G. (2013). BLUEPRINT: Mapping human blood cell epigenomes. *Haematologica*, *98*(10), 1487–1489. http://doi.org/10.3324/haematol.2013.094243

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., … Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology*, *28*(5), 495–501. http://doi.org/10.1038/nbt.1630

Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., & Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res*, *33*. http://doi.org/10.1093/nar/gki901

Richter-Larrea, J. A., Robles, E. F., Fresquet, V., Beltran, E., Rullan, A. J., & Agirre, X. (2010).

Reversion of epigenetically mediated BIM silencing overcomes chemoresistance in Burkitt lymphoma. *Blood*, *116*. http://doi.org/10.1182/blood-2010-02-268003

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. http://doi.org/10.1093/bioinformatics/btp616

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. a, … Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–50. http://doi.org/10.1073/pnas.0506580102

Turner, M. R., Talbot, K., & Goldacre, M. J. (2013). Autoimmune ALS.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, *10*(1), 57–63. http://doi.org/10.1038/nrg2484

Chapter 2:

MethGo: a comprehensive tool for analyzing whole-genome bisulfite sequencing data

Liao et al. BMC Genomics 2015, 16(Suppl 12):S11
http://www.biomedcentral.com/1471-2164/16/S12/S11

BMC Genomics

RESEARCH Open Access

MethGo: a comprehensive tool for analyzing whole-genome bisulfite sequencing data

Wen-Wei Liao[1†], Ming-Ren Yen[1†], Evaline Ju[2], Fei-Man Hsu[1], Larry Lam[3], Pao-Yang Chen[1*]

From Joint 26th Genome Informatics Workshop and Asia Pacific Bioinformatics Network (APBioNet) 14th International Conference on Bioinformatics (GIW/InCoB2015)
Tokyo, Japan. 9-11 September 2015

Abstract

Background: DNA methylation is a major epigenetic modification regulating several biological processes. A standard approach to measure DNA methylation is bisulfite sequencing (BS-Seq). BS-Seq couples bisulfite conversion of DNA with next-generation sequencing to profile genome-wide DNA methylation at single base resolution. The analysis of BS-Seq data involves the use of customized aligners for mapping bisulfite converted reads and the bioinformatic pipelines for downstream data analysis.

Results: Here we developed MethGo, a software tool designed for the analysis of data from whole-genome bisulfite sequencing (WGBS) and reduced representation bisulfite sequencing (RRBS). MethGo provides both genomic and epigenomic analyses including: 1) coverage distribution of each cytosine; 2) global cytosine methylation level; 3) cytosine methylation level distribution; 4) cytosine methylation level of genomic elements; 5) chromosome-wide cytosine methylation level distribution; 6) Gene-centric cytosine methylation level; 7) cytosine methylation levels at transcription factor binding sites (TFBSs); 8) single nucleotide polymorphism (SNP) calling, and 9) copy number variation (CNV) calling.

Conclusions: MethGo is a simple and effective tool for the analysis of BS-Seq data including both WGBS and RRBS. It contains 9 analyses in 5 major modules to profile (epi)genome. It profiles genome-wide DNA methylation in global and in gene level scale. It can also analyze the methylation pattern around the transcription factor binding sites, and assess genetic variations such as SNPs and CNVs. MethGo is coded in Python and is publically available at http://paoyangchen-laboratory.github.io/methgo/.

Background

Cytosine methylation is a crucial epigenetic modification involved in numerous biological processes, including transcriptional regulation, cell differentiation, and X-chromosome inactivation [1]. It is very important in the development in plants, animals, and human [2,3]. Many human diseases and cancers have been found to be associated with abnormal DNA methylation [4].

To evaluate DNA methylation, bisulfite treatment of genomic DNA has been widely used to convert unmethylated cytosines (Cs) to uracils while methylated Cs remain unconverted [5]. After PCR amplification, the converted Cs (to uracils) will be replaced by thymines (Ts). By comparing the bisulfite-converted sequences with the unconverted, the methylation status can be revealed.

Sodium bisulfite treatment coupling with high throughput sequencing (BS-seq) makes it possible to profile genome-wide DNA methylation in single base resolution [5,6]. The two major sequencing strategies for BS-seq are, reduced representation bisulfite sequencing (RRBS) which uses restriction enzymes to digest genomic DNA and size selects CpG-rich areas of genome [7], and whole-genome bisulfite sequencing (WGBS) which investigates all cytosines in the genome, and is state-of-the-art

* Correspondence: paoyang@gate.sinica.edu.tw
† Contributed equally
[1]Institute of Plant and Microbial Biology, Academia Sinica, Taipei 11529, Taiwan
Full list of author information is available at the end of the article

profiling method for genome-wide DNA methylation [8,9]. Both methods are used to profile the epigenomes of cell lines and tissues by large consortiums such as the ENCODE project [10], NIH Roadmap Epigenomics project [11], and The Cancer Genome Atlas (TCGA) [12].

The first step to process BS-seq data is to align the BS reads to the reference genome. Aligners such as Bowtie2 [13] and SOAP [14] are not applicable since the C-to-T conversion in the BS reads are incorrectly treated as mismatches for mapping penalty. Customized bisulfite aligners such as BS-Seeker 2 [15] were introduced to efficiently perform genome indexing, read mapping and methylation level calling.

After alignment, further bioinformatics steps are required for extracting biologically meaningful information. Several tools for such post-alignment analysis including Kismeth [16], Bis-SNP [17], GBSA [18], Repitools [19], and ReadDepth [20]. As these tools are designed for specific analyses, there is a lack of platforms providing a comprehensive overview of the BS-Seq data covering both genomic and epigenomic analyses.

In this paper we present MethGo, a post-alignment tool consisting of 9 analyses in 5 functional modules for processing and analyzing BS-Seq alignments. MethGo provides coverage distribution across all methylation sites, global methylation states and methylation levels according to several defined regions, such as promoter, gene body or transcription factor binding sites (TFBSs). In addition to DNA methylation, MethGo also provides the information of genetic variations including CNV calling and SNP calling. MethGo produces high quality figures and tables for data presentation that are ready for scientific publication.

Implementation

MethGo is a Python software that takes the alignment file from both WGBS and RRBS as the input data. It consists of three modules for methylation analysis and two modules to detect genetic variations (Figure 1, and see Additional file 1 and Additional file 2 for the description of the modules): COV module generates coverage distribution for methylation sites. MET module provides global cytosine methylation levels, cytosine methylation level distributions, cytosine methylation levels of genomic elements, chromosome-wide cytosine methylation level distributions, and gene-centric cytosine methylation levels. TXN module plots the methylation level relative to TFBSs. The SNP module detects SNPs and the CNV module detects CNVs across the genome.

Here we describe the 5 modules in more details:

COV: coverage distribution of methylation sites

Coverage of the methylation sites is a factor for evaluating the quality of sequencing data. Sites with high coverage are likely to provide accurate methylation status estimation. The COV module extracts the coverage for each cytosine from post-alignment data (i.e., CGmap) and generates reverse cumulative plot for methylation sites by the genomic contexts (CG, CHG and CHH, H refers to A, C, or T). For example, Figure 2 shows the coverage distribution of two *Arabidopsis* methylomes. In the WT methylome approximately 20% of the genome are covered with 20× depth of sequencing, whereas the *met1* methylome shows ~70% of the genome are covered. The coverage plot helps user to evaluate the quality of sequencing data and defines the cutoff for reads depth.

MET: methylation profiling

The MET module takes methylation calls generated from the bisulfite aligner such as BS Seeker 2, and gene annotation file for analyses. Five analyses are carried out in MET module. First, global cytosine methylation level of CG/CHG/CHH are calculated and plotted (Figure 3A). Second, methylation level of sites in each context is calculated and plotted into cytosine methylation level distribution plot (Figure 3B). Third, a genome is further divided into promoter, gene body, exon, intron and intergenic non-coding region (IGN), referring to genomic elements, and generated cytosine methylation level of genomic elements plot (Figure 3C). The promoter is defined as the region 1,000 bp upstream of transcription start site (TSS), and gene body is defined as the region between TSS and transcription termination sites (TTS). Fourth, chromosome-wide cytosine methylation level distribution are generated so that user could visualize the methylation level dynamics across each chromosome (Figure 4A). Fifth, MET module profiles gene-centric cytosine methylation level (Additional file 3), a gene whose methylation levels of promoter, gene body, intron and exon are listed in a summary table.

TXN: evaluating DNA methylation level at transcription factor binding sites

DNA methylation at the TFBSs can interfere with the binding of proteins and hence affects the activation of transcription [21]. The TXN module aims to reveal such a relationship by plotting the DNA methylation level at the TFBS of specific transcriptional factors. The TXN module processes methylation calls and transcription factor binding positions. The methylation levels within 1,500 bp of the TFBSs are averaged over tiling windows (30 bp) and reported in a scatter plot to reveal the methylation pattern around the TFBS (Figure 4B). By comparing the methylation patterns between transcription factors, alternated methylation level around the TFBS suggests the binding of the transcription factor may be associated with DNA methylation.

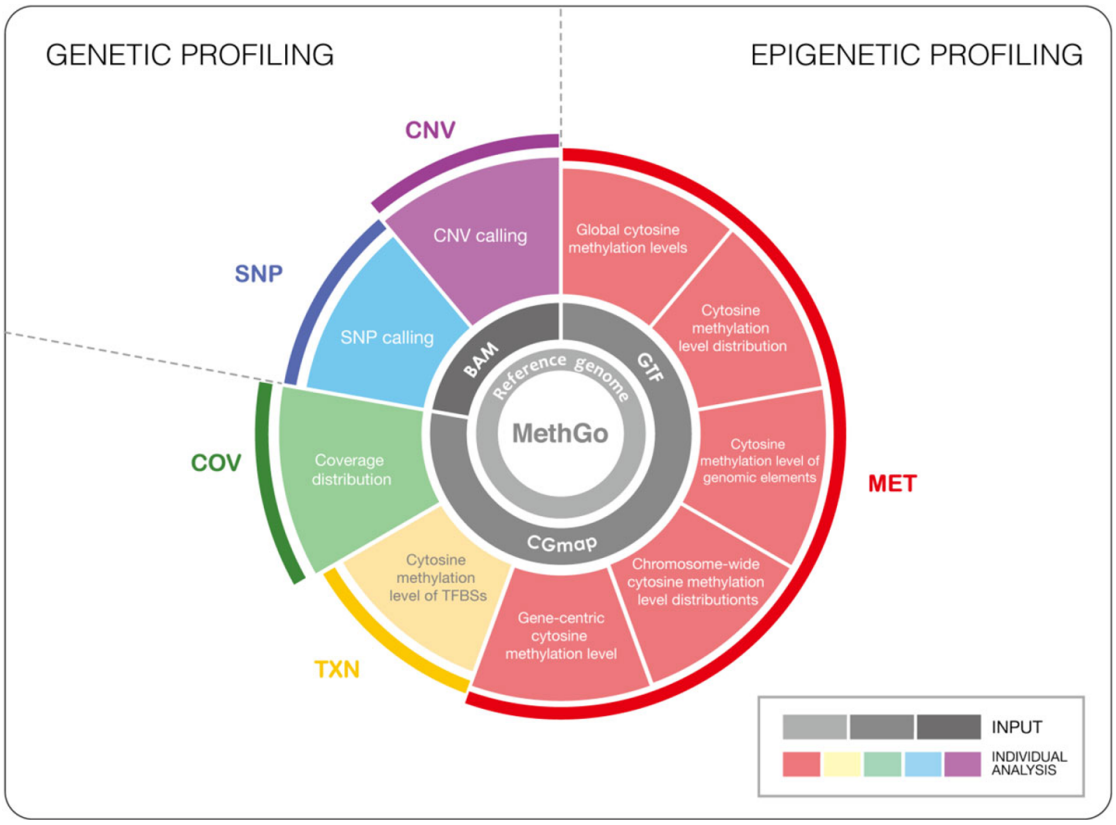


Figure 1 Overview of MethGo software pipeline. MethGo is a comprehensive tool for analyzing post-alignment bisulfite reads. MethGo consists of five modules for investigating DNA methylation as well as genetic variations.

CNV: CNV calling

Since BS-seq is DNA-based sequencing, the coverage of the reads (i.e. depth) can be used as a proxy for assessing CNVs. The CNV module extracts the read coverage from the alignments and plots the coverage across the genome (Figure 3C). Genomic regions with large-scale rearrangement such as duplication and deletion are likely to show in coverage plot, amplification and depletion respectively. Therefore, the CNV module is able to detect genome abnormality such as aneuploidy. Regions of continuous depletion or amplification, indicative of genome duplication or deletion, are reported in a text

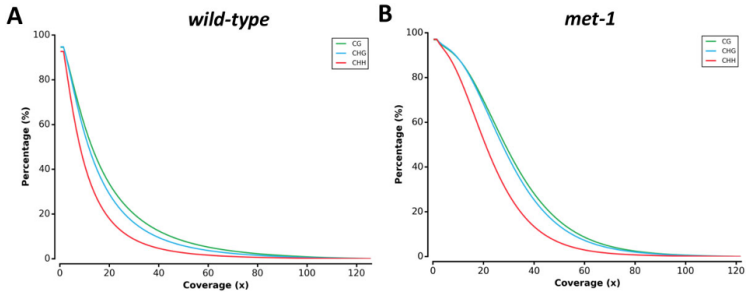


Figure 2 Coverage distribution of *Arabidopsis* WGBS. A. wild-type. B. *met-1*. Reverse cumulative plot with x-axis representing the coverage and y-axis representing the percentage of sites across genome. CG/CHG/CHH contexts are indicated by green/blue/red.

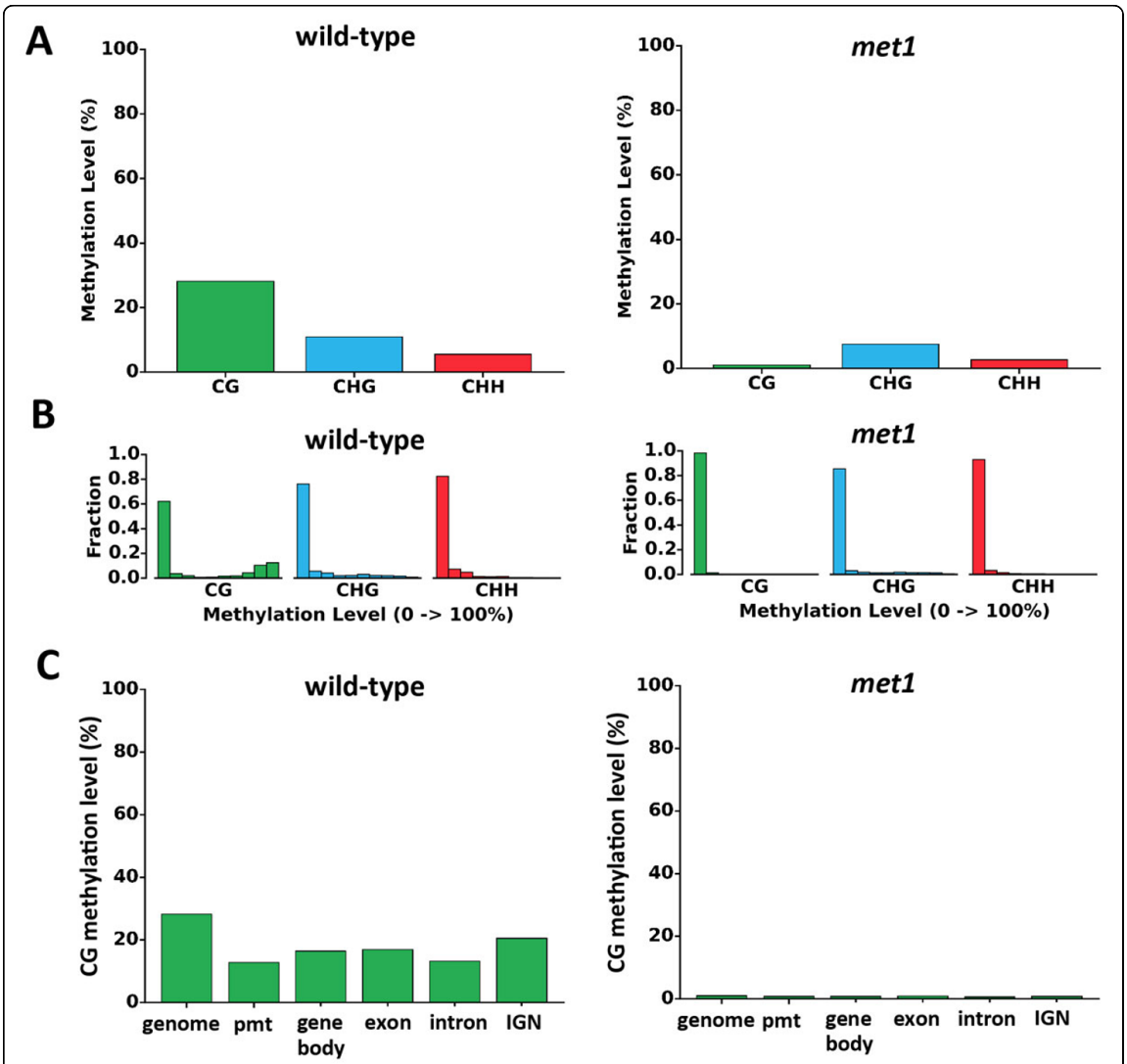


Figure 3 DNA methylation profiles in wild-type and *met1* mutant *Arabidopsis*. A. Global cytosine methylation levels. B. Distribution of cytosine methylation. The x-axis represents methylation levels binned in ten increment of 10% (i.e. 0%-10%, 10%-20%, *etc.*); y-axis is the fraction of total CG/CHG/CHH. C. Plot of CG methylation levels in different genomic elements.

file and represented in a plot of genome-wide copy number.

SNP: SNP calling

The SNP module identifies both homozygous and heterozygous SNPs from the alignment. The homozygous SNPs are polymorphisms where the majority reads show one dominant allele, which is different from the allele on the reference genome. The heterozygous SNPs are the ones where reads show two major alleles, potentially reflecting the two parental alleles (Additional file 4). In BS-seq, the alignment on the genomic C is not applicable due to the C-to-T conversion, so instead the alignment on the other strand G is used for SNP calling.

Results and discusssion

Feature evaluation with other major analyzers

We examined the functional features of MethGo together with five major post-alignment tools for BS-seq analysis, namely Bis-SNP, Kismeth, GBSA, Repitools, and ReadDepth. (see Table 1 for a summary of their functional features). Bis-SNP is written in Java, provides methylation

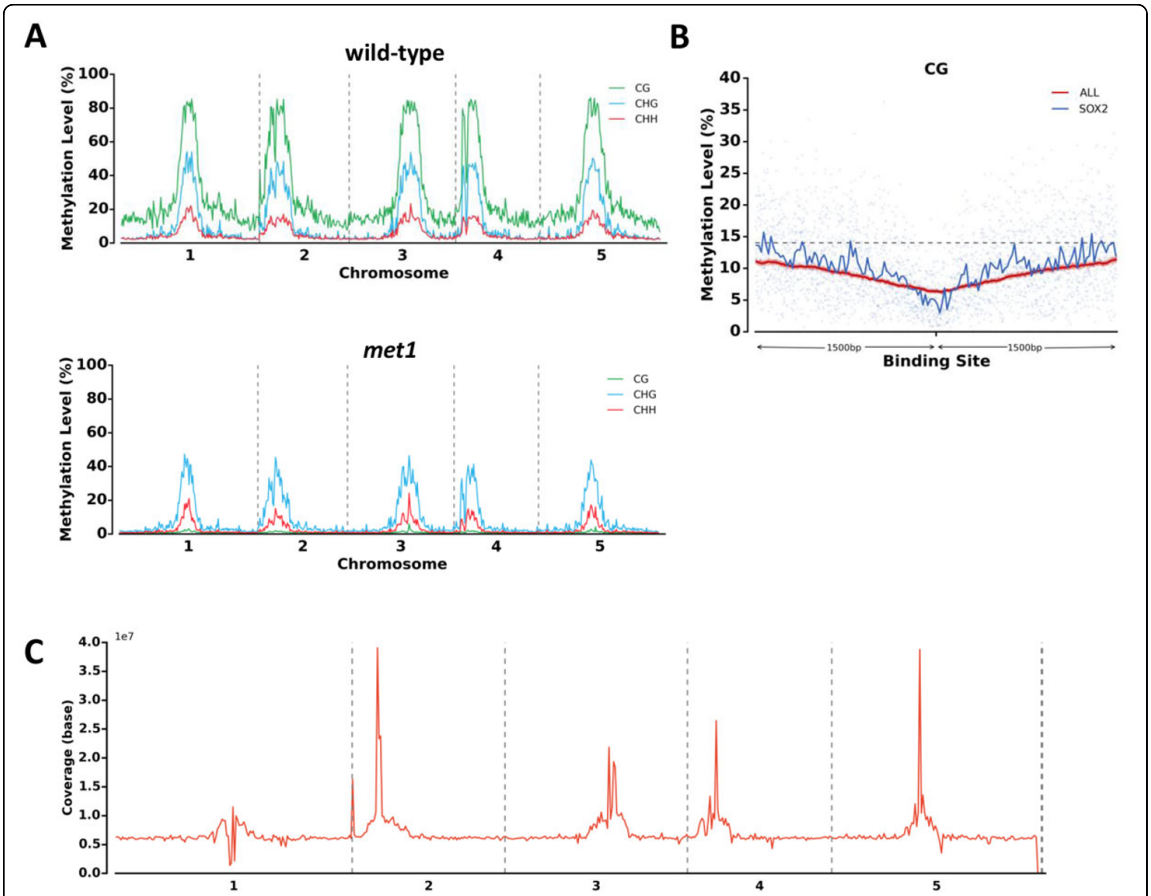


Figure 4 MethGo profiling of *Arabidopsis* and mouse WGBS. **A**. Chromosome-wide distribution of cytosine methylation levels in wild-type (top) and *met1* mutant (bottom) *Arabidopsis*. **B**. Methylation level around TFBS in mouse WGBS. The length of each TFBS is variant, and MethGo takes the center of the sequence for analysis. **C**. CNV on all chromosomes in *Arabidopsis* WGBS.

levels for each cytosine and calls SNPs from BS-seq data. Kismeth is a web-based tool, which calculates global methylation levels and provides platform for nucleotide-resolution methylation status visualization. GBSA is a tool written in Python and provides sequencing quality assessment, gene-centric methylation level, functional data management and visualization of methylation in nucleotide resolution. Repitools is an R package for the analysis of enrichment-based assay and displays the distribution of enriched DNA across the genome followed by visualizing and summarizing the interaction between epigenetic mark and gene expression. ReadDepth is also an R package to detect CNVs by measuring the depth of coverage in the sequencing data. MethGo provides 9 analyses for both epigenetic and genetic profiling, including coverage distribution, global cytosine methylation level, cytosine methylation level distribution, chromosome-wide cytosine methylation level distribution, cytosine methylation level of genomic elements, gene-centric cytosine methylation level, cytosine methylation level of TFBSs, CNV calling, and SNP calling. Altogether, MethGo includes the functions such as cytosine methylation level distribution, cytosine methylation level of genomic elements, chromosome-wide cytosine methylation level distribution and cytosine methylation level of TFBSs, which are not included in Bis-SNP, GBSA, Kismeth, Repitools, and ReadDepth. In addition, MethGo is the only tool to profile both SNPs and CNVs.

Demonstrating COV, MET, CNV and SNP modules with *Arabidopsis* WGBS data

In order to demonstrate MethGo on real data, we downloaded and processed WGBS data of wild-type and *met1* mutant of *Arabidopsis* [22]. MET1 is methyltransferase 1, which controls faithful maintenance of cytosine methylation primarily at CG sites. After mapping with BS aligner,

Table 1. Summary of bioinformatic tools for data analysis using aligned BS-seq

	MethGo	Kismeth[16]	Bis-SNP [17]	GBSA[18]	Repitools[19]	ReadDepth [20]
Programming Language	Python	unknown	Java	Python	R	R
Operating System	Windows/ Unix	web	Windows/ Unix	Windows/ Unix	Windows/ Unix	Unix
Interface	Command-line	GUI	Command-line	GUI/ Command-line	Command-line	Command-line
Coverage distribution	Yes (*F)	-	-	-	Yes	-
Global cytosine methylation level	Yes (*F)	Yes	-	-	-	-
Cytosine methylation level distribution	Yes (*F)	-	-	-	-	-
Cytosine methylation level of genomic elements	Yes (*F)	-	-	-	-	-
Chromosome-wide cytosine methylation level distribution	Yes (*F)	-	-	-	-	-
Gene-centric cytosine methylation level	Yes (*T)	-	-	Yes	-	-
Cytosine methylation level of TFBSs	Yes (*F)	-	-	-	-	-
SNP calling	Yes (*T)	-	Yes	-	-	-
CNV calling	Yes (*F)	-	-	-	-	Yes
Main functions of tool	Methylation profiling and extracting genetic variation information from bisulfite sequencing data	Global methylation levels calculation and visualization at nucleotide resolution	SNP calling	Gene-centric methylation level scoring and visualizetion	Enrichment based epigenomic data analysis such as coverage distribution of CpG sites	CNV calling

*F: the output file is figure; *T: the output file is table.

BS-Seeker2, the output was loaded into MethGo for processing. COV module outputs reverse cumulative plot of coverage distribution. Different sequencing samples show different coverage distribution due to sequencing depth of data. (Figure 2A, B).

As for DNA methylation profiling with MET module, the CG methylation in *met1* mutant is much lower compared to WT (Figure 3A). The *met1* mutant almost abolishes the CG methylation with relatively less effect on CHG and CHH contexts. The cytosine methylation distribution plots show the methylation distribution of cytosine sites in all three contexts. As shown in Figure 3B, the CG methylation shows a bimodal distribution where most sites are either in low or high methylation. The CHG and CHH sites are generally weakly methylated. The cytosine methylation level of genomic elements plots showed the average methylation level in genome, promoter, gene body, exon, intron, and intergenic regions by CG, CHG, and CHH contexts (Figure 3C and Additional file 5). Compared with other regions, the methylation level in promoters is lower due to facilitation of protein binding. The chromosome-wide cytosine methylation level distribution showed the landscape of DNA methylation throughout a genome (Figure 4A). The plots showed that in *Arabidopsis*, the methylation levels are higher near the pericentromeric regions in all contexts. The MET module also generates gene-centric cytosine methylation levels for each gene for wild-type *Arabidopsis* (Additional file 3). The CNV module profiles genome-wide CNVs (Figure 4C). Peri-centromeric regions of all 5 chromosomes show high coverage due to the presence of repetitive sequence. The SNP calling module generates tabular file of homozygous and heterozygous SNPs, which helps researchers to investigate potential mutations or serves as a marker for genotyping (Additional file 6 and Additional file 7).

TXN module demonstration with mouse WGBS data

We downloaded WGBS data of mouse primordial germ cells to demonstrate TXN module of MethGo [23]. The accessibility of TFBS is important for gene regulation

and TFBS should be exploited of DNA methylation. TXN module plots methylation levels of specific transcription factor of interest and an average methylation level for comparison. In Figure 4B, SOX2 is a transcription factor regulating cell pluripotency [24], and its corresponding TFBS shows significant decrease of methylation relative to average of all TFBSs.

Conclusions

We presented MethGo, specifically for analyzing post-alignment from BS-seq. In comparison with other popular similar tools, MethGo is a streamlined tool capable of profiling both genome-wide DNA methylation and genetic variations. It also generates high resolution plots. MethGo comes with a user-friendly manual and tutorials with examples for biologists to evaluate DNA methylation. The MethGo installation guide and module requirements can be found in Additional file 8. We have made this tool publicly available for the community.

Additional material

Additional file 1: Detailed description of implementation. This file contains information on the implementation for all the modules.

Additional file 2: File format of input and output. Table of file format of input and output required for MethGo modules.

Additional file 3: DNA methylation of genes in wild-type *Arabidopsis*. Average DNA methylation levels of regions related to genes, including promoter, gene body, exon, and intron.

Additional file 4: Heterozygous and homozygous SNP. A. The illustration of heterozygous and homozygous SNP. B. Screenshot of heterozygous SNP. There are two different alleles comparing to reference genome (bottom). C. Screenshot of homozygous SNP. There is one allele different from the reference genome (bottom).

Additional file 5: Non-CG methylation levels in different genomic elements in *Arabidopsis*. A. wild-type. B. *met1*.

Additional file 6: Homozygous SNPs. This file contains homozygous SNP calling of WT *Arabidopsis*.

Additional file 7: Heterozygous SNPs. This file contains heterozygous SNP calling of WT *Arabidopsis*.

Additional file 8: Software installation guide and requirements. This file contains the MethGo installation guide and module requirements.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PC conceived the project. WL, EJ and LL implemented the software. MY wrote the manuscript. FH, EJ, LL and PC edited the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

This work was supported by a grant from Academia Sinica, and grants from MOST-103-2313-B-001-003-MY3 and MOST-103-2633-B-001-002 and NHRI-EX104-10324SC.

Declarations

Publication charges for this article have been funded by a grant from Academia Sinica to P.-Y. C.

This article has been published as part of *BMC Genomics* Volume 16 Supplement 12, 2015: Joint 26th Genome Informatics Workshop and 14th International Conference on Bioinformatics: Genomics. The full contents of the supplement are available online at http://www.biomedcentral.com/bmcgenomics/supplements/16/S12.

Authors' details

[1]Institute of Plant and Microbial Biology, Academia Sinica, Taipei 11529, Taiwan. [2]Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [3]Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095, USA.

Published: 9 December 2015

References

1. Goto T, Monk M: **Regulation of X-chromosome inactivation in development in mice and humans.** *Microbiol Mol Biol Rev* 1998, **62(2)**:362-378.
2. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, *et al*: **Conservation and divergence of methylation patterning in plants and animals.** *Proc Natl Acad Sci USA* 2010, **107(19)**:8689-8694.
3. Gkountela S, Zhang KX, Shafiq TA, Liao WW, Hargan-Calvopina J, Chen PY, Clark AT: **DNA Demethylation Dynamics in the Human Prenatal Germline.** *Cell* 2015, **161(6)**:1425-1436.
4. Feinberg AP: **Phenotypic plasticity and the epigenetics of human disease.** *Nature* 2007, **447(7143)**:433-440.
5. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL: **A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands.** *Proc Natl Acad Sci USA* 1992, **89(5)**:1827-1831.
6. Laird PW: **Principles and challenges of genomewide DNA methylation analysis.** *Nat Rev Genet* 2010, **11(3)**:191-203.
7. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R: **Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis.** *Nucleic Acids Res* 2005, **33(18)**:5868-5877.
8. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haudenschild CD, Pradhan S, Nelson SF, Pellegrini M, Jacobsen SE: **Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning.** *Nature* 2008, **452(7184)**:215-219.
9. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR: **Highly integrated single-base resolution maps of the epigenome in Arabidopsis.** *Cell* 2008, **133(3)**:523-536.
10. ENCODE project Consortium: **A user's guide to the encyclopedia of DNA elements (ENCODE).** *PLoS Biol* 2011, **9(4)**:e1001046..
11. [http://www.roadmapepigenomics.org].
12. [http://cancergenome.nih.gov/].
13. Langmead B, Salzberg SL: **Fast gapped-read alignment with Bowtie 2.** *Nat Methods* 2012, **9(4)**:357-359.
14. Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J: **SOAP2: an improved ultrafast tool for short read alignment.** *Bioinformatics* 2009, **25(15)**:1966-1967.
15. Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, Chen PY, Pellegrini M: **BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data.** *BMC Genomics* 2013, **14**:774.
16. Gruntman E, Qi Y, Slotkin RK, Roeder T, Martienssen RA, Sachidanandam R: **Kismeth: analyzer of plant methylation states through bisulfite sequencing.** *BMC Bioinformatics* 2008, **9**:371.
17. Liu Y, Siegmund KD, Laird PW, Berman BP: **Bis-SNP: combined DNA methylation and SNP calling for Bisulfite-seq data.** *Genome Biol* 2012, **13(7)**:R61.
18. Benoukraf T, Wongphayak S, Hadi LH, Wu M, Soong R: **GBSA: a comprehensive software for analysing whole genome bisulfite sequencing data.** *Nucleic Acids Res* 2013, **41(4)**:e55.
19. Statham AL, Strbenac D, Coolen MW, Stirzaker C, Clark SJ, Robinson MD: **Repitools: an R package for the analysis of enrichment-based epigenomic data.** *Bioinformatics* 2010, **26(13)**:1662-1663.
20. Miller CA, Hampton O, Coarfa C, Milosavljevic A: **ReadDepth: a parallel R package for detecting copy number alterations from short sequencing reads.** *PLoS One* 2011, **6(1)**:e16327.

21. Siegfried Z, Eden S, Mendelsohn M, Feng X, Tsuberi BZ, Cedar H: **DNA methylation represses transcription in vivo.** *Nat Genet* 1999, **22(2)**:203-206.
22. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE: **Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome.** *Cell* 2013, **152(1-2)**:352-364.
23. Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, Yanagisawa E, Sakashita A, Wakai T, Suzuki Y, Ito T, *et al*: **High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice.** *Genome Res* 2013, **23(4)**:616-627.
24. Kamachi Y, Uchikawa M, Kondoh H: **Pairing SOX off: with partners in the regulation of embryonic development.** *Trends Genet* 2000, **16(4)**:182-187.

doi:10.1186/1471-2164-16-S12-S11
Cite this article as: Liao *et al.*: MethGo: a comprehensive tool for analyzing whole-genome bisulfite sequencing data. *BMC Genomics* 2015 **16**(Suppl 12):S11.

Chapter 3:

Epigenetic changes mediated by polycomb repressive complex 2 and E2a are associated with drug resistance in a mouse model of lymphoma

Flinders et al. Genome Medicine (2016) 8:54
DOI 10.1186/s13073-016-0305-0

Genome Medicine

RESEARCH Open Access

Epigenetic changes mediated by polycomb repressive complex 2 and E2a are associated with drug resistance in a mouse model of lymphoma

Colin Flinders[1,6†], Larry Lam[2†], Liudmilla Rubbi[2], Roberto Ferrari[2], Sorel Fitz-Gibbon[2], Pao-Yang Chen[2], Michael Thompson[2], Heather Christofk[3], David B Agus[4,6], Daniel Ruderman[4,6], Parag Mallick[5,6*] and Matteo Pellegrini[2*]

Abstract

Background: The genetic origins of chemotherapy resistance are well established; however, the role of epigenetics in drug resistance is less well understood. To investigate mechanisms of drug resistance, we performed systematic genetic, epigenetic, and transcriptomic analyses of an alkylating agent-sensitive murine lymphoma cell line and a series of resistant lines derived by drug dose escalation.

Methods: Dose escalation of the alkylating agent mafosfamide was used to create a series of increasingly drug-resistant mouse Burkitt's lymphoma cell lines. Whole genome sequencing, DNA microarrays, reduced representation bisulfite sequencing, and chromatin immunoprecipitation sequencing were used to identify alterations in DNA sequence, mRNA expression, CpG methylation, and H3K27me3 occupancy, respectively, that were associated with increased resistance.

Results: Our data suggest that acquired resistance cannot be explained by genetic alterations. Based on integration of transcriptional profiles with transcription factor binding data, we hypothesize that resistance is driven by epigenetic plasticity. We observed that the resistant cells had H3K27me3 and DNA methylation profiles distinct from those of the parental lines. Moreover, we observed DNA methylation changes in the promoters of genes regulated by E2a and members of the polycomb repressor complex 2 (PRC2) and differentially expressed genes were enriched for targets of E2a. The integrative analysis considering H3K27me3 further supported a role for PRC2 in mediating resistance. By integrating our results with data from the Immunological Genome Project (Immgen.org), we showed that these transcriptional changes track the B-cell maturation axis.

Conclusions: Our data suggest a novel mechanism of drug resistance in which E2a and PRC2 drive changes in the B-cell epigenome; these alterations attenuate alkylating agent treatment-induced apoptosis.

Keywords: Burkitt's lymphoma, Mafosfamide, Resistance, Histone, Methylation, Epigenetics

* Correspondence: paragm@stanford.edu; matteop@mcdb.ucla.edu
†Equal contributors
[5]Canary Center, Stanford University, Palo Alto, CA 94305, USA
[2]Department of Molecular, Cellular and Developmental Biology, University of California, Los Angeles, CA 90095, USA
Full list of author information is available at the end of the article

Background

Many anticancer therapies lose effectiveness over time as tumors acquire resistance. Despite significant study [1], acquired resistance remains a major obstacle to improving remission rates and achieving prolonged disease-free survival. A number of explanations for resistance have been proposed, including the presence of cancer stem cells [2] and mutations that confer drug resistance [3].

Although Burkitt's lymphoma is extremely aggressive [4], 90–95 % of children receiving the standard-of-care therapy, a combination treatment of rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone, enter complete remission. Treatment of adults is not as successful [5], in part due to acquired resistance. In both children and adults, salvage treatment has a poor success rate, with only one-third of children and very few adults obtaining positive outcomes from salvage therapy.

The mechanisms underlying acquisition of resistance in Burkitt's lymphoma are only partially understood. Numerous genetic mechanisms have been hypothesized, including up-regulation of expression of proteins involved in drug efflux, such as the ATP-binding cassette transporter family, cyclophosphamide inactivation through aldehyde dehydrogenase up-regulation, increased expression of DNA repair enzymes, and deregulation of apoptosis through the loss of *Tp53* [1]. Genetic mutations are unable to explain cases of acquired resistance that arise rapidly or that reverse in response to a drug holiday [6, 7].

Alterations in histone modifications and DNA methylation that lead to an altered transcriptional program have been proposed to lead to acquired drug resistance in B-cell lymphoma [8, 9]. Recent work in an in vitro model of Burkitt's lymphoma has shown that treatment with the DNA methylation inhibitor 5-azacytidine reactivates expression of *Id2*, which encodes a repressor of translocated Myc, resulting in the inhibition of proliferation [10]. Similarly, treatment of lymphoma cell lines with the histone deacetylase inhibitor suberoylanilide hydroxamic acid has been shown to re-sensitize lymphoma cell lines to various therapeutic agents [11].

Studies of clinical specimens have revealed that tumors are both genetically and epigenetically heterogeneous [8, 12]. The role of genetic heterogeneity within tumors and its effect on treatment response and outcome has been extensively studied, but less is known about how epigenetic heterogeneity impacts disease progression and clinical outcome. Previous studies have shown that drug treatment generates selective pressure on heterogeneous populations that leads to the enrichment of specific genetically distinct subpopulations [13, 14]. These subpopulations can ultimately become the dominant population in a tumor, resulting in resistance to the therapeutic agent. It is possible that similar mechanisms of selection act at the epigenetic level. Recent research in prostate cancer has documented the inherent heterogeneity of DNA methylation in patient tumor samples [15], though the selection of epigenetically distinct subpopulations has yet to be shown.

Here we used dose escalation with mafosfamide on parental *Eμ-Myc Cdkn2a*$^{-/-}$ non-Hodgkin's B-cell lymphoma cells to generate a series of drug-resistant cell lines. We then investigated the mechanism by which these cells acquired resistance. In our experiment, parental cells were cultured with increasing doses of mafosfamide, from 100 nM to 4 μM, over 5 weeks. At four steps in the dose escalation, resistant clones were isolated and molecularly profiled. Whole-genome sequencing of the parental and resistant cell lines did not reveal any genetic alterations that might explain the acquired resistance. However, analyses of transcriptomic and DNA methylation profiling suggested that the acquired resistance is associated with genes and promoters targeted by transcription factor E2a and by members of the polycomb repressive complex 2 (PRC2), Suz12 and Pcl2. An integrative analysis considering histone H3 lysine 27 trimethylation (H3K27me3) further supported a role for PRC2 in mediating resistance. Comparison of the transcriptomic data from resistant lines and from B cells at different developmental stages [16] suggested that resistance is associated with epigenetic changes that are found along the B-cell maturation axis. E2a is a master regulator of B-cell development in mice. An analysis of human diffuse B-cell methylation data revealed that the methylation status of genes regulated by the human ortholog of E2a, TCF3, is associated with treatment failure. Thus, our study implicates epigenetic factors in evolution of acquired resistance.

Methods

Creation of resistant *Eμ-Myc Cdkn2a*$^{-/-}$ lines

The parental *Eμ-Myc Cdkn2a*$^{-/-}$ lymphoma line was generated from a C57BL/6 J mouse as described in Schmitt et al. [17]. Lymphoma cells from these mice were cultured in vitro to generate the cell line. Resistant strains were generated from this parental line via dose escalation from 100 nM to 4 μM of mafosfamide (Cell Signaling Technology) over a 34-day period.

Cell viability and cell cycle analysis

Cell viability was measured using the Perkin-Elmer Operetta platform with 2.5 μM Draq5 (Abcam) for nuclear detection and 5 μg/mL of propidium iodide (Sigma-Aldrich, St. Louis, MO, USA) to detect dead cells. For cell cycle analysis, cells were fixed in ethanol and placed in solution with propidium iodide. Cells were gated for G0/G1, S, and M phase on a Beckman LSRII.

Genome sequencing

Genomes of cell lines were sequenced to a minimum of 8× average coverage using Illumina HiSeq 2000 sequencers. The reads were aligned to the mm9 (MGSCv37) *Mus musculus* reference genome using BWA version 0.6.2-r126 (backtrack) [18] with default parameters. Duplicate reads were removed using PICARD version 1.85(1345) with default parameters (Additional file 1). The whole-genome sequencing data are available via the Sequence Read Archive under accession number SRP071753.

Oligonucleotide microarray analysis

Oligonucleotide microarray analysis was carried out using Affymetrix GeneChip Mouse Gene ST 1.0. The resulting data are publically available via Gene Expression Omnibus accession GSE60342. Data were quantified and processed with robust multi-array averaging using the justRMA function of the 1.40.0 affy R package [19]. Expression values were $\log_2$ transformed for further downstream analysis. Probe sets were annotated using the Affymetrix MoGene-1_0-st-v1.na33.2.mm9.probeset.csv file. We selected the top 1000 probe sets ranked by their covariance to identify differentially regulated genes (Additional file 1).

Transcription factor analysis

Targets for 64 murine transcription factors were identified from ChIPBase (http://deepbase.sysu.edu.cn/chipbase, downloaded August 1, 2013) [20] and limited to genes with binding events within 5 kilobases (kb) of transcriptional start sites. To identify potential upstream regulators, we identified the overlap of chromatin immunoprecipitation-sequencing (ChIP-seq) data with predicted transcription factor targets and used a one-sided Fisher's exact test to determine significance.

ChIP-seq

Chromatin was immunoprecipitated as described previously [21]. Briefly, cells were grown to 50 % confluency. Formaldehyde was added for 10 min at room temperature and 100 μl of the lysate (5×10^6 cells) was used for each immunoprecipitation with anti-H3K27me3 (Active Motive, catalogue number 39155). Libraries were sequenced using an Illumina HiSeq 2000 to obtain 50-bp-long reads.

Peaks were called by comparing counts in the immunoprecipitated libraries with input libraries in windows tiling the genome using Poisson statistics as previously described [21]. Combinatorial clustering of data was achieved by determining significant enrichment for the histone mark in each condition within 5 kb upstream of transcription start sites (at least three 50-bp bins with $p < 1.0e\text{-}6$). A binary distribution was created based on a promoter being enriched (1) or not enriched (0) and a combinatorial matrix was created with all possible combinations across all conditions. H3K27me3 data were plotted based on the combinatorial clustering and visualized by a Cluster 3.0-generated CDT file loaded on Java-Tree view to produce a heat map. The ChIP-seq datasets are publically available via Gene Expression Omnibus accession GSE78939.

Bisulfite sequencing

Reduced representation bisulfite (RRBS) libraries were generated following the standard RRBS protocol [22]. The genome was digested with the methylation-insensitive restriction enzyme MspI and fragments from 100 to 300 bases were selected. The fragments were ligated with Illumina adaptors, denatured, and treated with sodium bisulfite. The libraries were sequenced using Illumina HiSeq 2000 sequencers. The reads were aligned using the bisulfite aligner BS Seeker2 [23] to determine where fragments uniquely mapped allowing for three mismatches to the reference genome (mouse mm9). The RRBS datasets are publically available via Gene Expression Omnibus accession GSE78939.

DNA methylation analysis

To identify DNA methylation changes correlated with resistance, we computed RRBS fragment CpG methylation levels and calculated the covariance between the fragment methylation score and sample order (ordered from least to most resistant; Additional file 1).

Principal component analysis

Principal component analysis of expression profiles was performed by applying the R *prcomp* function with the scaled option to the expression microarray values of the resistant cell lines and B cells at different stages of development (NCBI Gene Expression Omnibus accession GSE15907). See Additional file 1 for samples used.

Results

Generation of resistant lines and assessment of cell cycle characteristics

To investigate the factors driving acquisition of resistance to chemotherapy, we employed a widely used cell line derived from an *Eμ-Myc* mouse model of Burkitt's lymphoma [17]. We refer to this line as our "parental" line. This model has two genetic alterations: (1) a translocation in the *c-Myc* oncogene that causes its expression to be controlled by an immunoglobulin heavy chain enhancer, thus restricting its expression to B cell lineage cells; and (2) a deletion in *Cdkn2a* that recapitulates a common mutation seen in human tumors [24]. *Eμ-Myc* $Cdkn2a^{-/-}$ mice develop lymphomagenesis with highly invasive tumors with apoptotic defects [17]. Resistant lines were generated by gradually exposing the parental line to increasing concentrations from 100 nM to 4 μM of mafosfamide (an in vitro active form of cyclophosphamide) in cell culture (Fig. 1a, b; Additional file 1: Figure S1). Four

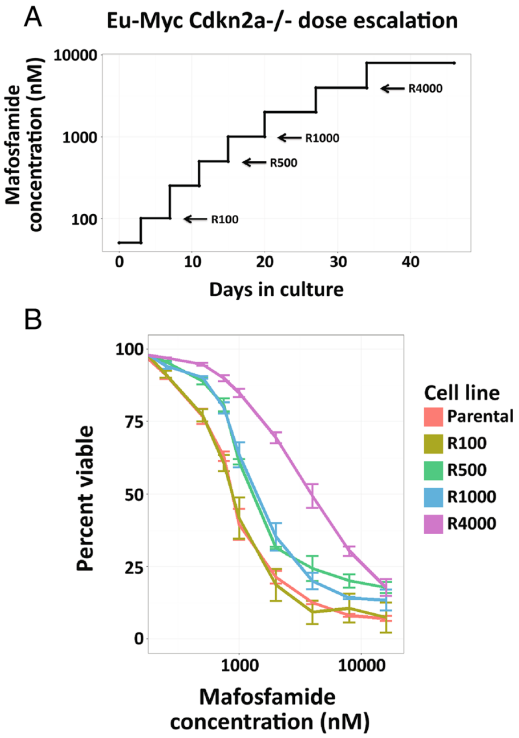


Fig. 1 Dose escalation of *Eμ-Myc Cdkn2a*$^{-/-}$ cells and resistance to mafosfamide. **a** *Eμ-Myc Cdkn2a*$^{-/-}$ cells were cultured with increasing doses of mafosfamide. After 10, 18, 23, and 37 days, corresponding to cells cultured in 100, 500, 1000, and 4000 nM of mafosfamide, surviving cells were harvested. **b** The percentage viability of the resistant cells was measured after 24-h exposure to mafosfamide

resistant sub-clones were isolated during the 5-week period of dose escalation (R100, R500, R1000, and R4000) for subsequent analysis. We refer to these as the "resistant" lines. The cycling rate of our parental and resistant lines in the absence of mafosfamide was very similar (Additional file 1: Table S1, Figure S2), suggesting that the resistance is not associated with large changes in cell cycle phases.

Mafosfamide treatment led to a greater percentage of apoptotic cells in the parental line relative to the resistant lines (Fig. 1b; Additional file 1: Table S1, Figures S1 and S2). In all lines, non-apoptotic cells showed an equivalent cell cycle delay as assessed by the decrease in percentage of cells in G1 and a concurrent increase in G2 percentage upon treatment. As all the lines showed approximately the same amount of cell cycle delay upon treatment in the non-apoptotic fraction, resistance is most likely not a result of an increased cell cycle delay.

Resistance is unlikely to have arisen by genetic mechanisms

Given the four-fold difference in drug sensitivity between the parental and resistant cell lines (parental half maximal effective concentration, EC_{50}, 857 ± 68 nM; R4000 EC_{50} 3446 ± 138 nM, mean ± standard error of the mean), we initially hypothesized that acquired resistance was mediated by genetic variants that were enriched in the population over the course of the dose escalation. To identify possible resistance-associated variants, we performed whole-genome sequencing and identified single-nucleotide variants (SNVs) that showed an increase in allele frequency that correlated with increased resistance. Twenty-three SNVs displayed this pattern (Additional file 1: Figure S3). Of these, 13 had variant alleles in only one sample (the most resistant line) and manual inspection of read alignments showed that the SNVs all occurred solely at the ends of sequencing reads or in repetitive regions (Additional file 1: Table S2) and were, therefore, likely sequencing artifacts. To validate that these were in fact false positives, we performed Sanger sequencing on regions surrounding five of the potential SNVs. No mutations were observed using this sequencing technique (Additional file 1: Figure S4). Additionally, no novel large structural variations or potential copy number variations were found in any of the lines. Together, these data suggest that genetic alterations are not the principal explanation for the acquired resistance.

Alterations in gene expression of mafosfamide-resistant cell lines

To determine whether the decreased sensitivity to mafosfamide observed in the resistant lines was associated with altered transcription, we measured the changes in gene expression for the parental and each resistant line. A subset of these gene expression changes were confirmed by quantitative PCR (Additional file 1: Figure S5). We expected to identify genes known to be involved in drug metabolism, transport of mustard alkylating agents, and DNA repair [25, 26]. We found that the expression of most genes involved in these processes did not increase significantly in our resistant lines, suggesting that none of these processes are likely to play a major role in mediating resistance (Additional file 1: Figure S6).

To identify transcripts with expression patterns that were perturbed with increasing resistance, we computed the covariance of the expression of each gene with the EC_{50} of the resistant cell line. The 1000 genes with highest and lowest covariance were clustered (Fig. 2a). Pathway analysis of those genes that showed a positive covariance with resistance (i.e., an increase in gene expression with increasing resistance) indicated that they encode proteins involved in B-cell maturation and development (Fig. 2b). Genes that showed a negative covariance were enriched for those associated with EIF2-mediated signaling and T helper cell differentiation (Fig. 2c).

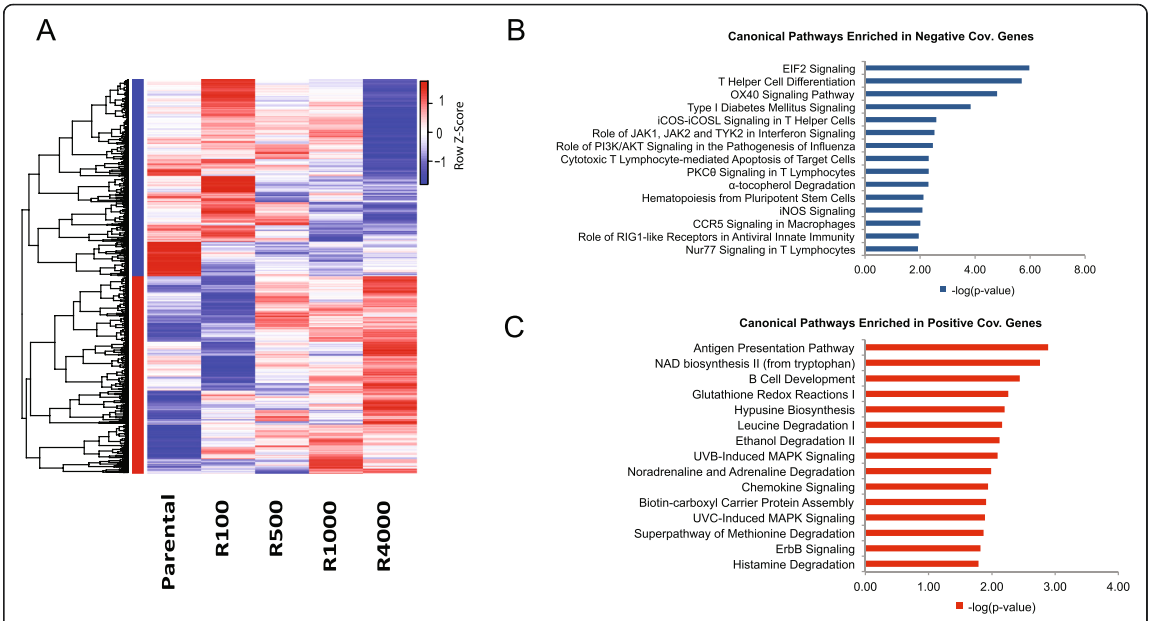


Fig. 2 Microarray expression profile of untreated cell lines. **a** Heat map of the probe sets. Positive covariance with increasing resistance to mafosfamide is indicated by the *red cluster*, negative covariance with resistance is indicated by the *blue cluster*. **b** The top 15 pathways enriched in negative covariance genes. **c** The top 15 pathways enriched in positive covariance genes

DNA methylation changes with drug resistance

DNA methylation is a central mechanism of transcriptional regulation and alterations in the methylome have previously been shown to contribute to lymphomagenesis and drug resistance [27–29]. To gain further insights into epigenetic changes that occurred during the acquisition of drug resistance, we measured DNA methylation using reduced representation bisulfite sequencing (RRBS) of the parental and resistant lines. This approach allowed us to quantify the methylation of approximately one million CpGs. To identify methylation sites associated with drug resistance, we computed the covariance between the methylation state of each CpG and the EC_{50} of the resistant cell line. Covarying sites were clustered (Fig. 3a). High covariance RRBS fragments had lower levels of gene expression compared with overall expression based on all annotated probes, whereas low covariance RRBS fragments had higher levels of gene expression (Additional file 1: Figure S7). The top 1000 high covariance RRBS fragments were located in gene regions that were functionally enriched for DNA binding, transcription factor activity, and cell differentiation (Additional file 1: Table S3).

We next asked whether the genes that were proximal to these sites were associated with specific transcription factors. Each transcription factor from ChIPBase [20] was tested for association with the 1000 genes with the highest and lowest DNA methylation covariance with respect to resistance (Fig. 3b). From this analysis, we identified the transcription factors E2a and PU.1; these transcription factors are involved in B-cell development [30–33]. We also found that members of the polycomb repressive complex 2 (PRC2) Suz12 and Pcl2 were very strongly enriched over positively covarying genes, suggesting an involvement of this complex in mediating epigenetic changes during the acquisition of drug resistance. The expression of PRC2 members was similar in parental and resistant lines; however, *Uty*, which encodes one of three known histone H3 lysine 27 demethylases [34], was down-regulated in the resistant lines, suggesting that Uty may be involved in the epigenome alteration that confers drug resistance (Additional file 1: Figure S8a, b).

Resistant cell lines show altered H3K27me3 occupancy

As PRC2 has been shown to be involved in placement of the H3K27me3 mark [35], we performed ChIP-seq for H3K27me3 in all lines (grown in the absence of mafosfamide) to verify the involvement of PRC2 in the transition to drug resistance. We found widespread acquisition of H3K27me3 at the transcriptional start sites of a large group of genes in the least resistant (R100, R500) lines (Fig. 4a). These effects appeared to be stable across biological replicates, as seen by the high correlation between replicate profiles (Additional file 1: Table S4). However, the lines with increased resistance (R1000, R4000) had

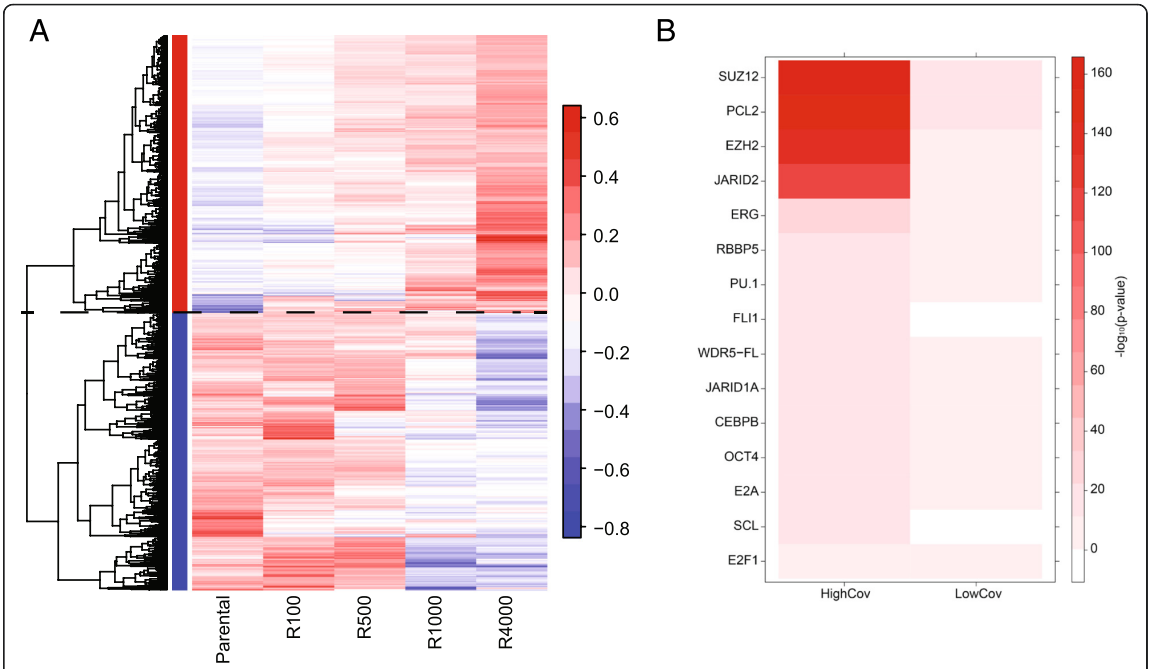


Fig. 3 DNA methylation analysis. **a** Heat map of the top and bottom 1000 RRBS fragments based on covariance. **b** Heat map of the –log10(*p*) values for the overlap of high and low covariance genes with transcription factor targets

decreased levels of H3K27me3 with respect to the parental line at many loci.

As most H3K27me3 peaks are near the transcription start sites, we analyzed promoters by means of combinatorial clustering (described in "Methods") and identified three major groups of genes: those with a high level of H3K27me3 throughout the time course (K1), those with rapid and widespread H3K27me3 acquisition followed by a gradual decrease (K2), and those with little to no H3K27me3 (K3). Functional analysis of these groups showed enrichment for developmentally regulated genes in K1 and cell cycle-regulated genes in K2 (Fig. 4b, c). The level of H3K27me3 across all genes bound by E2a revealed that the average profile followed a similar pattern of change to that of cluster K2 (Fig. 4d). Furthermore, E2a-bound genes were enriched in cluster K2 and were depleted from K1 (Fig. 4e). These results suggest that H3K27me3 occupancy may regulate expression of E2a-bound genes.

Previously published literature shows a linkage between Ezh2-mediated H3K27 methylation and DNA methylation. Ezh2 is the catalytic subunit of PRC2. Analysis of our RRBS data revealed that DNA methylation levels proximal to K2 genes increased between the parental and resistant cell lines (Additional file 1: Figure S9a); we did not observe an increase in DNA methylation in genes in K1 or K3 subsets compared with parental levels (Additional file 1: Figure S9b, c). Additionally, we found a significant overlap between the high covariance RRBS fragments and the K1 and K2 clusters (Additional file 1: Figure S10). These results suggest a linkage between H3K27me3 and DNA methylation at loci with altered expression during acquisition of resistance.

Principal component analysis of basal gene expression indicates alterations in B-cell maturation

The observation that changes in expression, DNA methylation, and H3K27me3 levels in the resistant lines involve genes that encode regulators of B-cell development (e.g., E2a and PU.1) led us to hypothesize that gradual dose escalation results in epigenetic changes associated with a B-cell maturation axis. To identify the transcriptional profile characteristic of the B-cell maturation axis, we used the Immunological Genome Project (Immgen.org) dataset [36], which contains expression data collected from B-cell progenitors, cells at multiple stages during maturation, and mature B cells.

To associate our expression data with the B-cell developmental axis, we performed a combined principle component analysis of gene expression data obtained from parental and resistant cell lines and B cells at different stages of development gathered from the Immgen dataset. We found that the second principal

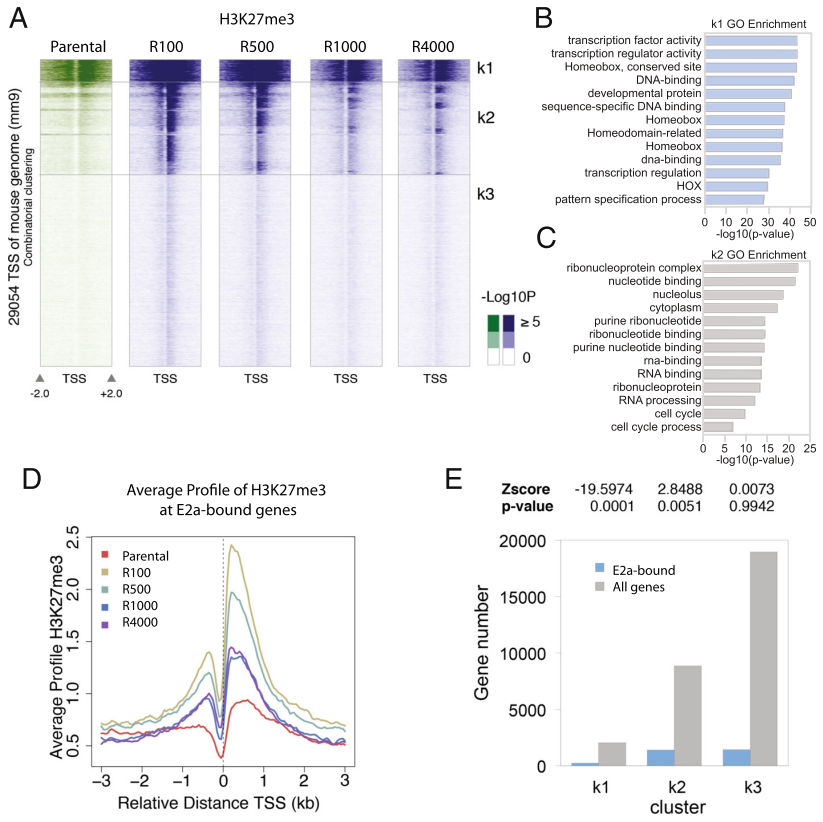


Fig. 4 H3K27me3 ChIP-seq analysis. **a** Clusters of promoters based on their H3K27me3 levels in the parental and resistant lines ($n = 2$). *TSS* transcription start site. **b** Group K1 shows enrichment for developmental genes. *GO* gene ontology. **c** Group K2 shows enrichment for cell cycle-regulated genes. **d** Average H2K27me3 levels across E2a-bound genes. **e** Analysis of E2a-bound genes in the three clusters

component, PC.2, captures the developmental state of B cells (Fig. 5a) with less differentiated progenitor cells toward the negative direction of the PC.2 axis and the more mature states in the positive. Our four resistant lines are ordered by their differentiation, with the least resistant line more differentiated than the most resistant line. This result suggests that the transcriptomes of our cells vary monotonically along this B-cell developmental axis. Additionally, functional annotation of the top principal component genes showed enrichment for the B-cell receptor signaling pathway in the top 1000 PC.2 genes (Additional file 1: Table S5).

To identify factors associated with this B-cell developmental axis, we identified genes that had a positive correlation with development (i.e., an increase in expression during differentiation). We then used ChIPBase data to identify factors with targets enriched within this list. Genes positively correlated with development were often targets of E2a and PU.1 (Fig. 5b).

Differential methylation status of E2a target genes is associated with treatment failure in patients with aggressive B-cell lymphoma

To confirm that E2a mediates resistance development, we sought to determine the methylation status of genes bound by TCF3, the human ortholog of E2a, in a clinically relevant dataset. We obtained DNA methylation data from diffuse large B-cell lymphoma (DLBCL) patients collected from the Cancer Genome Atlas project. DLBCL differs from Burkitt's lymphoma but is the most similar lymphoma for which we were able to access data on clinical samples.

Of 19 DLBCL samples available, 14 corresponded to patients who were tumor-free after their initial treatment course and five were from patients who experienced disease progression. For all methylation sites, we computed the Kolmogorov–Smirnov statistic and an associated p value between these two groups. From 482,421 total CG sites on the Illumina 450 K array, 5541 had a Kolmogorov–Smirnov

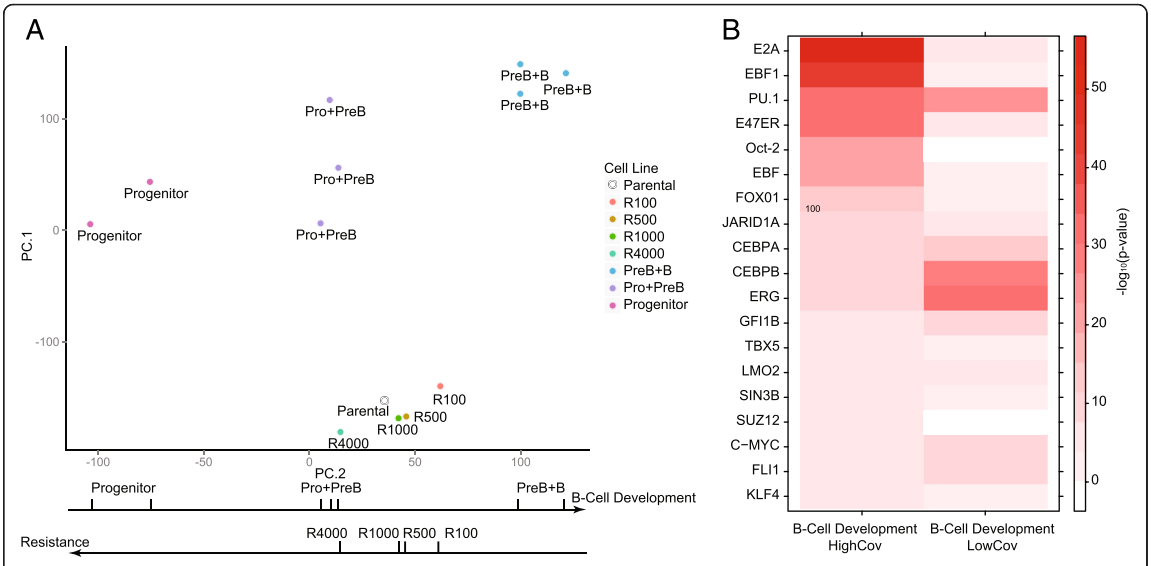


Fig. 5 B-cell development principal component analysis. **a** Scatter plot of the principal components of the Immunological Genome Project data and the data on resistant lines. **b** Heat map of the –log10(*p*) values for the overlap of high and low covariance genes with transcription factor target genes

test with $p < 0.05$. We linked CpG sites to TCF3 binding regions if both the CpG and binding region occurred within the gene body or within 10 kb of the gene's transcription start site. We then identified the intersection of genes that had both evidence of TCF3 binding and at least two CpG sites with significant methylation changes between the two patient populations. From a total of 17,744 genes used in the linking calculations, we derived Table 1, which shows a strong enrichment of TCF3-bound genes associated with differentially methylated sites (p value ≤ 0.0001). This analysis suggests that TCF3-bound genes are differentially methylated and are associated with treatment failure in a lymphoma subtype.

To further test the clinical significance of differential methylation, we compared lists compiled from our gene expression and DNA methylation analyses with a 1458-gene signature of DLBCL survival obtained from the Precog data set [37]. Fifty-eight genes from the top 449 high expression covariance gene list and an additional 58 genes from the top 597 high methylation covariance significantly overlapped with the 1458 genes associated with DLBCL survival ($p = 0.0001$ and $p = 0.0006$, respectively; Additional file 1: Figure S11a, b). Furthermore, those genes that overlap between our H3K27me3 cluster K1 and the high covariance RRBS fragments (Additional file 1: Figure S10) similarly significantly overlapped with the 1458 genes associated with DLBCL survival (Additional file 1: Figure S11c).

Table 1 TCF3 target gene methylation

	TCF3 target	Not TCF3 target
Differential methylation	208	80
No differential methylation	5086	12,370

DNA methylation data obtained from The Cancer Genome Atlas shows an increase in differential methylation in TCF3 target genes in patients with DLBCL who showed stable or progressive disease compared with individuals who had complete response

Discussion

By treating *Eμ-Myc Cdkn2a*$^{-/-}$ cells with escalating doses of mafosfamide, we developed four cell lines with increasing resistance to the drug. We hypothesized that if a mutation provided a proliferative or survival advantage, then the frequency of that mutation would increase over time in culture. We observed no verifiable mutations that displayed this pattern.

Reasoning that measuring gene expression changes across cell lines might reveal the resistance mechanism, we calculated the covariance of gene expression with respect to increasing resistance. Pathway enrichment analysis showed an enrichment for genes encoding proteins involved in B-cell development. Given the role that DNA methylation plays in B-cell lineage commitment [28, 38], we characterized the DNA methylome of our resistant lines. Analysis of DNA methylation changes suggested that expression of genes regulated by the histone methyltransferase complex PRC2 were altered in the most resistant lines.

PRC2 plays a critical role in B-cell development by repressing genes necessary for hematopoietic differentiation

through the addition of tri-methylation to lysine 27 of histone H3. During hematopoietic stem cell differentiation, this repressive mark is lost from genes that commit hematopoietic stem cells to differentiation [30, 31, 39]. Numerous studies have shown that overexpression of, or activating mutations in, PRC2 components, particularly Ezh2, contribute to proliferation and lymphomagenesis in DLBCL patients [40, 41]. To evaluate how the presence of the H3K27me3 mark is correlated with resistance, we performed ChIP-seq on H3K27me3 in the parental and resistant cell lines. A large fraction of promoters in the resistant lines showed rapid and widespread H3K27me3 acquisition followed by a gradual reduction. This pattern of rapid H3K27 methylation followed by demethylation has been reported in patients: loss of H3K27me3 is a predictor of poor outcome [42]. These results suggest that PRC2 is rapidly activated by the DNA damage induced by mafosfamide, which is consistent with previous literature suggesting the PRC2 is targeted to sites of DNA damage [43, 44]. Furthermore, these results suggest that this initial response is attenuated as cells become adapted to higher doses of mafosfamide and the most resistant cells have an H3K27me3 state that is similar to that of the parental cells.

The observation that H3K27me3 is rapidly gained but then lost suggests that it may not be the only mechanism leading to resistance in our setting. The removal of H3K27me3 at many loci may lead to its replacement with DNA methylation, a more permanent mark. It has been suggested that histone methylation by PRC2 may recruit Dnmt3l, an inactive homolog of DNA methyltransferase, resulting in the inhibition of CpG methylation [32]. The gradual removal of H3K27me3 could lead to the replacement of Dnmt3l by its active counterparts, Dnmt3a and Dnmt3b, resulting in an increase in DNA methylation. In agreement with a potential coupling between PRC2 and DNA methyltransferases, we observed that genes that gained H3K27me3 early and then lost this mark (cluster K2) showed a gradual increase in DNA methylation. Thus, we conclude that the addition of H3K27me3 to these loci results in the acquisition of DNA methylation, leading to a stable, drug-resistant epigenome.

Our analyses also indicate the involvement of B-cell developmental regulators in the epigenetic transition to drug resistance. Among these regulators, which include E2a and PU.1, we believe that E2a plays a central role. Clustering of the H3K27me3 ChIP-seq data showed that E2a is strongly associated with genes that gain and subsequently lose H3K27me3 that also gradually gain DNA methylation. Genes with significant levels of differential DNA methylation were also enriched in those bound by E2a. *E2a* encodes two proteins, E12 and E47, that are known to be master regulators involved in the process of B-cell lineage commitment [33]. Principle component analysis of gene expression from the resistant lines and from various stages of B-cell development indicated that the expression of E2a target genes is strongly correlated with B-cell maturation states. In addition, the *E2a* locus is repressed early in B-cell development but it becomes transcriptionally active during B-cell commitment [45]. The central role of E2a-mediated regulation of gene expression in Burkitt's lymphoma is made evident by the fact that it is the fifth-most mutated gene in patients with Burkitt's lymphoma and all mutations affect the DNA binding domain [46]. Inhibition of *E2a* expression using small interfering RNA results in lower levels of *Cdkna1* (*p21*) and higher *PUMA* expression, impairment of cell cycle arrest, and increased Tp53-dependent apoptosis [47]. Our data suggest that E2a may be playing a critical role in mediating the resistance phenotype as it appears to be a key regulator of the response to mafosfamide.

Taken together, our results suggest that PRC2 and known regulators of B-cell development induce epigenetic changes in genes encoding key hematopoietic developmental and pluripotency genes that result in the resistant phenotype. Upon treatment with mafosfamide, the cells we studied underwent rapid and widespread acquisition of the repressive mark H3K27me3. This methylation is coupled to changes in the activity of B-cell developmental regulators, such as E2a, ultimately resulting in changes in transcription of genes involved in B-cell development. As the cells became more resistant, the H3K27me3 response was attenuated and replaced by the more permanent mark 5-methylcytosine. We speculate that these changes led the *Eμ-Myc* cell lines to epigenetically transition along the B-cell developmental axis, leading to a suppression of the apoptotic response upon exposure to mafosfamide. Evidence for this transition is provided by our integrative analysis of Immgen data that captures a principal component of the B-cell developmental axis. As the cells became more resistant to mafosfamide, they appear to move backwards along this axis. The diminished apoptotic response of cells with stem-like characteristics compared with that of differentiated cells has been described in many other systems and may thus represent a general mechanism for the acquisition of drug resistance [48, 49].

Finally, we have shown that similar mechanisms may also occur in human lymphomas. Our analysis of samples from patients with diffuse large B-cell lymphoma suggests that there are epigenetic differences between patients who remain disease-free versus those who have disease progression. Moreover, the sites of differential methylation are significantly associated with genes that are bound by Tcf3, the human ortholog of E2a. Thus, our murine model may capture epigenetic resistance mechanisms that are also present in human disease,

suggesting that epigenetic plasticity impairs therapeutic regimens in the clinic.

Conclusions

Our results indicate that resistance to DNA alkylating agents in B-cell lymphoma is associated with alterations in both CpG methylation and H3K27me3 but not with genetic changes. We found that repressive epigenetic markers are preferentially altered in the promoter regions of genes bound by the B-cell development regulator E2a, suggesting that this transcription factor plays a key role in mediating the resistance phenotype, possibly by suppressing apoptosis. We also showed that differential methylation of genes bound by TCF3, the human ortholog of E2a, is associated with treatment failure in diffuse large B-cell lymphoma patients.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the Gene Expression Omibus repository, accession numbers GSE60342, GSE60342, and GSE78939, as well as the Sequence Read Archive, accession number SRP071753.

Additional file

Additional file 1: Supplementary Results and Methods. All supplementary figures, tables, legends, results, methods, and references are contained in Additional file 1. (ZIP 2461 kb)

Abbreviations

ChIP-seq: chromatin immunoprecipitation followed by sequencing; DLBCL: diffuse large B-cell lymphoma; EC_{50}: half maximal effective concentration; H3K27me3: histone H3 lysine 27 trimethylation; PRC2: polycomb repressive complex 2; RRBS: reduced representation bisulfite sequencing; SNV: single nucleotide variant.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CF, RF, and LR performed experiments, CF, LL, RF, SFG, PYC, DR, and MP analyzed results, MT performed methylation analysis of the Cancer Genome Atlas data, CF and LL wrote the manuscript. DBA provided clinical input. DR, PM, and MP supervised the project. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the Lowe lab for providing the initial cell line as well as guidance on aspects of culture conditions. We thank Laura Barsky of the USC Flow Cytometry Core Facility for technical assistance. The results shown here are in part based upon data generated by the TCGA Research Network (http://cancergenome.nih.gov/). We also thank Dr. Shannon Mumenthaler for her help and support.

Funding

This work was funded by grant H47213-5 U54 CA143907-02, Physical Sciences in Oncology by the National Cancer Institute.

Author details

[1]Department of Biological Chemistry, University of California, Los Angeles, CA 90095, USA. [2]Department of Molecular, Cellular and Developmental Biology, University of California, Los Angeles, CA 90095, USA. [3]Department of Molecular and Medical Pharmacology, University of California, Los Angeles, CA 90095, USA. [4]Department of Medicine, University of Southern California, Los Angeles, CA 90033, USA. [5]Canary Center, Stanford University, Palo Alto, CA 94305, USA. [6]Center for Applied Molecular Medicine, University of Southern California, Los Angeles, CA 90033, USA.

Received: 20 August 2015 Accepted: 13 April 2016
Published online: 04 May 2016

References

1. Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. Nat Rev Cancer. 2013;13(10):714–26. doi:10.1038/nrc3599.
2. Weigert O, Weinstock DM. The evolving contribution of hematopoietic progenitor cells to lymphomagenesis. Blood. 2012;120(13):2553–61. doi:10.1182/blood-2012-05-414995.
3. Mumenthaler SM, Foo J, Leder K, Choi NC, Agus DB, Pao W, et al. Evolutionary modeling of combination treatment strategies to overcome resistance to tyrosine kinase inhibitors in non-small cell lung cancer. Mol Pharm. 2011;8(6): 2069–79. doi:10.1021/mp200270v.
4. Choi MK, Jun HJ, Lee SY, Kim KH, do Lim H, Kim K, et al. Treatment outcome of adult patients with Burkitt lymphoma: results using the LMB protocol in Korea. Ann Hematol. 2009;88(11):1099–106. doi:10.1007/s00277-009-0729-3.
5. Richter-Larrea JA, Robles EF, Fresquet V, Beltran E, Rullan AJ, Agirre X, et al. Reversion of epigenetically mediated BIM silencing overcomes chemoresistance in Burkitt lymphoma. Blood. 2010;116(14):2531–42. doi:10.1182/blood-2010-02-268003.
6. Wilting RH, Dannenberg JH. Epigenetic mechanisms in tumorigenesis, tumor cell heterogeneity and drug resistance. Drug Resist Updat. 2012;15(1-2):21–38. doi:10.1016/j.drup.2012.01.008.
7. Knoechel B, Roderick JE, Williamson KE, Zhu J, Lohr JG, Cotton MJ, et al. An epigenetic mechanism of resistance to targeted therapy in T cell acute lymphoblastic leukemia. Nat Genet. 2014;46(4):364–70. doi:10.1038/ng.2913.
8. Chambwe N, Kormaksson M, Geng H, De S, Michor F, Johnson NA, et al. Variability in DNA methylation defines novel epigenetic subgroups of DLBCL associated with different clinical outcomes. Blood. 2014. doi:10.1182/blood-2013-07-509885.
9. Hiraga J, Tomita A, Sugimoto T, Shimada K, Ito M, Nakamura S, et al. Down-regulation of CD20 expression in B-cell lymphoma cells after treatment with rituximab-containing combination chemotherapies: its prevalence and clinical significance. Blood. 2009;113(20):4885–93. doi:10.1182/blood-2008-08-175208.
10. Guan H, Xie L, Klapproth K, Weitzer CD, Wirth T, Ushmorov A. Decitabine represses translocated MYC oncogene in Burkitt lymphoma. J Pathol. 2013; 229(5):775–83. doi:10.1002/path.4164.
11. Valdez BC, Nieto Y, Murray D, Li Y, Wang G, Champlin RE, et al. Epigenetic modifiers enhance the synergistic cytotoxicity of combined nucleoside analog-DNA alkylating agents in lymphoma cell lines. Exp Hematol. 2012; 40(10):800–10. doi:10.1016/j.exphem.2012.06.001.
12. Murawski N, Pfreundschuh M. New drugs for aggressive B-cell and T-cell lymphomas. Lancet Oncol. 2010;11(11):1074–85. doi:10.1016/S1470-2045(10)70210-2.
13. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010;141(1):69–80. doi:10.1016/j.cell.2010.02.027.
14. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481(7382):506–10. doi:10.1038/nature10738.
15. Aryee MJ, Liu W, Engelmann JC, Nuhn P, Gurel M, Haffner MC, et al. DNA methylation alterations exhibit intraindividual stability and interindividual heterogeneity in prostate cancer metastases. Sci Transl Med. 2013;5(169): 169ra10. doi:10.1126/scitranslmed.3005211.
16. Painter MW, Davis S, Hardy RR, Mathis D, Benoist C. Transcriptomes of the B and T lineages compared by multiplatform microarray profiling. J Immunol. 2011;186(5):3047–57. doi:10.4049/jimmunol.1002695.
17. Schmitt CA, McCurrach ME, de Stanchina E, Wallace-Brodeur RR, Lowe SW. INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. Genes Dev. 1999;13(20):2670–7.
18. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60. doi:10.1093/bioinformatics/btp324.

19. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics. 2003;19(2):185–93.
20. Yang JH, Li JH, Jiang S, Zhou H, Qu LH. ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. Nucleic Acids Res. 2013;41(Database issue):D177–87. doi:10.1093/nar/gks1060.
21. Ferrari R, Su T, Li B, Bonora G, Oberai A, Chan Y, et al. Reorganization of the host epigenome by a viral oncogene. Genome Res. 2012;22(7):1212–21. doi:10.1101/gr.132308.111.
22. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. 2005;33(18):5868–77. doi:10.1093/nar/gki901.
23. Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ, et al. BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics. 2013;14(1):774. doi:10.1186/1471-2164-14-774.
24. Maggi Jr LB, Winkeler CL, Miceli AP, Apicelli AJ, Brady SN, Kuchenreuther MJ, et al. ARF tumor suppression in the nucleolus. Biochim Biophys Acta. 2014; 1842(6):831–9. doi:10.1016/j.bbadis.2014.01.016.
25. Panasci L, Xu ZY, Bello V, Aloyz R. The role of DNA repair in nitrogen mustard drug resistance. Anticancer Drugs. 2002;13(3):211–20.
26. Sarkaria JN, Kitange GJ, James CD, Plummer R, Calvert H, Weller M, et al. Mechanisms of chemoresistance to alkylating agents in malignant glioma. Clin Cancer Res. 2008;14(10):2900–8. doi:10.1158/1078-0432.CCR-07-1719.
27. Mayle A, Yang L, Rodriguez B, Zhou T, Chang E, Curry CV, et al. Dnmt3a loss predisposes murine hematopoietic stem cells to malignant transformation. Blood. 2015;125(4):629–38. doi:10.1182/blood-2014-08-594648.
28. Esteller M, Gaidano G, Goodman SN, Zagonel V, Capello D, Botto B, et al. Hypermethylation of the DNA repair gene O(6)-methylguanine DNA methyltransferase and survival of patients with diffuse large B-cell lymphoma. J Natl Cancer Inst. 2002;94(1):26–32.
29. Clozel T, Yang S, Elstrom RL, Tam W, Martin P, Kormaksson M, et al. Mechanism-based epigenetic chemosensitization therapy of diffuse large B-cell lymphoma. Cancer Discov. 2013;3(9):1002–19. doi:10.1158/2159-8290.CD-13-0117.
30. Aloia L, Di Stefano B, Di Croce L. Polycomb complexes in stem cells and embryonic development. Development. 2013;140(12):2525–34. doi:10.1242/dev.091553.
31. Majewski IJ, Ritchie ME, Phipson B, Corbin J, Pakusch M, Ebert A, et al. Opposing roles of polycomb repressive complexes in hematopoietic stem and progenitor cells. Blood. 2010;116(5):731–9. doi:10.1182/blood-2009-12-260760.
32. Neri F, Krepelova A, Incarnato D, Maldotti M, Parlato C, Galvagni F, et al. Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs. Cell. 2013;155(1):121–34. doi:10.1016/j.cell.2013.08.056.
33. Murre C. Regulation and function of the E2A proteins in B cell development. Adv Exp Med Biol. 2007;596:1–7. doi:10.1007/0-387-46530-8_1.
34. Shpargel KB, Sengoku T, Yokoyama S, Magnuson T. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. PLoS Genet. 2012;8(9):e1002964. doi:10.1371/journal.pgen.1002964.
35. Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K. The polycomb group protein Suz12 is required for embryonic stem cell differentiation. Mol Cell Biol. 2007;27(10):3769–79. doi:10.1128/MCB.01432-06.
36. Heng TS, Painter MW. The Immunological Genome Project: networks of gene expression in immune cells. Nat Immunol. 2008;9(10):1091–4. doi:10.1038/ni1008-1091.
37. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat Med. 2015;21(8):938–45. doi:10.1038/nm.3909.
38. Shaknovich R, Cerchietti L, Tsikitas L, Kormaksson M, De S, Figueroa ME, et al. DNA methyltransferase 1 and DNA methylation patterning contribute to germinal center B-cell differentiation. Blood. 2011;118(13):3559–69. doi:10.1182/blood-2011-06-357996.
39. Majewski IJ, Blewitt ME, de Graaf CA, McManus EJ, Bahlo M, Hilton AA, et al. Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity. PLoS Biol. 2008;6(4):e93. doi:10.1371/journal.pbio.0060093.
40. Velichutina I, Shaknovich R, Geng H, Johnson NA, Gascoyne RD, Melnick AM, et al. EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. Blood. 2010;116(24): 5247–55. doi:10.1182/blood-2010-04-280149.
41. Yap DB, Chu J, Berg T, Schapira M, Cheng SW, Moradian A, et al. Somatic mutations at EZH2 Y641 act dominantly through a mechanism of selectively altered PRC2 catalytic activity, to increase H3K27 trimethylation. Blood. 2011; 117(8):2451–9. doi:10.1182/blood-2010-11-321208.
42. Wei Y, Xia W, Zhang Z, Liu J, Wang H, Adsay NV, et al. Loss of trimethylation at lysine 27 of histone H3 is a predictor of poor outcome in breast, ovarian, and pancreatic cancers. Mol Carcinog. 2008;47(9):701–6. doi:10.1002/mc.20413.
43. Campbell S, Ismail IH, Young LC, Poirier GG, Hendzel MJ. Polycomb repressive complex 2 contributes to DNA double-strand break repair. Cell Cycle. 2013; 12(16):2675–83. doi:10.4161/cc.25795.
44. Chou DM, Adamson B, Dephoure NE, Tan X, Nottke AC, Hurov KE, et al. A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. Proc Natl Acad Sci U S A. 2010;107(43):18475–80. doi:10.1073/pnas.1012946107.
45. Hystad ME, Myklebust JH, Bo TH, Sivertsen EA, Rian E, Forfang L, et al. Characterization of early stages of human B cell development by gene expression profiling. J Immunol. 2007;179(6):3662–71.
46. Schmitz R, Young RM, Ceribelli M, Jhavar S, Xiao W, Zhang M, et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature. 2012;490(7418):116–20. doi:10.1038/nature11378.
47. Andrysik Z, Kim J, Tan AC, Espinosa JM. A genetic screen identifies TCF3/E2A and TRIAP1 as pathway-specific regulators of the cellular response to p53 activation. Cell Rep. 2013;3(5):1346–54. doi:10.1016/j.celrep.2013.04.014.
48. Bakker ST, Passegue E. Resilient and resourceful: genome maintenance strategies in hematopoietic stem cells. Exp Hematol. 2013;41(11):915–23. doi:10.1016/j.exphem.2013.09.007.
49. Zhou BB, Zhang H, Damelin M, Geles KG, Grindley JC, Dirks PB. Tumour-initiating cells: challenges and opportunities for anticancer drug discovery. Nat Rev Drug Discov. 2009;8(10):806–23. doi:10.1038/nrd2137.

Chapter 4:

Anti-inflammatory therapies of amyotrophic lateral sclerosis guided by immune pathways

Am J Neurodegener Dis 2015;4(2):28-39
www.AJND.us /ISSN:2165-591X/AJND0017490

Original Article

Anti-inflammatory therapies of amyotrophic lateral sclerosis guided by immune pathways

Larry Lam[2], Ramesh C Halder[1], Dennis J Montoya[2], Liudmilla Rubbi[2], Arturo Rinaldi[2], Bien Sagong[1], Sarah Weitzman[1], Rachel Rubattino[1], Ram Raj Singh[3,4], Matteo Pellegrini[2], Milan Fiala[1]

[1]Department of Surgery, UCLA School of Medicine, Los Angeles, CA 90095-7022; [2]Department of Molecular, Cell and Developmental Biology, UCLA School of Medicine, Los Angeles, CA 90095-1606; [3]Department of Medicine, UCLA School of Medicine, Los Angeles, CA 90095-1670; [4]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA 90095-1670

Received October 8, 2015; Accepted December 7, 2015; Epub December 28, 2015; Published December 30, 2015

Abstract: Sporadic ALS patients display heterogeneous immune pathways in peripheral blood mononuclear cells (PBMCs). We tested nine sALS patients and one unaffected identical twin of an index case by RNA-Seq of PBMCs. The inflammatory patients (n = 3) clustered into a subset with an inflammatory Th1/Th17 signature and the non-inflammatory patients (n = 7) into another subset with a B cell signature. The inflammatory subset was remarkable for granulocyte and agranulocyte diapedesis, hepatic fibrosis, roles of cytokines and metalloproteases. The non-inflammatory subset was highlighted by degradation of vitamin E, serotonin and nucleotides, altered T cell and B cell signaling, agranulocyte diapedesis, and up regulation of B cell genes. Identification of these differentially regulated pathways in sALS patients may guide the choice of anti-inflammatory therapies.

Keywords: Amyotrophic lateral sclerosis, immune pathways, tocilizumab, hepatic fibrosis, vitamin E

Introduction

The presence of inflammation in the spinal cord of sporadic amyotrophic lateral sclerosis (sALS) patients is documented by immunohistochemical demonstration of activated microglia, astrocytes, and activated complement components [1], along with the infiltration by dendritic cells, macrophages [2], and IL-17A-positive CD4 and CD8 T cells and mast cells [3]. Inflammatory macrophages phagocytize both normal and apoptotic neurons in the ALS spinal cord [4] and infiltrate the spinal cord of sALS patients [3] and animal models of ALS, contributing to disease progression [5, 6]. Others claim that only T cells cross the blood-brain barrier and activate microglia [7], but this is inconsistent with the results in experimental models [5, 6] and infiltration by monocyte/macrophages in the ALS spinal cord [8]. The inflammatory cytokines IL-1 and IL-6 are induced in peripheral blood mononuclear cells (PBMCs) of sALS patients by mutant or aggregated wild-type superoxide dismutase-1 (SOD-1) [3]. IL-6 and TGF-β are present in the serum of sALS patients early in the disease and IL-17A is found in the mid-course of the disease [4]. Th17 cells have been increasingly found in ALS patients [3, 9, 10], and ALS has been associated with prior history of autoimmune diseases, including multiple sclerosis, myasthenia gravis and systemic lupus erythematosus [11].

Thus inflammation, including autoimmune inflammation, is documented in sALS but anti-inflammatory therapy has not yet been proven successful. A recent study identified heterogeneous, Th1, Th17, and IL-6 driven inflammatory pathways in ALS patients [10]. Specific neurotoxic mechanisms include Th17 cell disruption of the blood-brain barrier [12] and IL-6 trans-signaling [13]. Although the clinical trial of celecoxib failed [14], recent anti-inflammatory approaches against ALS are promising [15, 16]. In addition, the lipidic mediator resolvin D1 (RvD1) attenuated IL-6 and TNF-α production in ALS macrophages, suggesting beneficial role of omega-3 supplementation [4].

The recognition of IL-6-driven inflammation in inflammatory sALS patients [16] has stimulated a pending clinical trial of the IL-6 receptor antibody tocilizumab (Actemra^R). As shown in a previous cross-sectional study, one subgroup of "inflammatory" sALS patients had a highly increased expression of inflammatory cytokines, in particular IL1, IL6, and IL8, chemokines, metalloproteinases and transcription factors, whereas the "non-inflammatory" group had a near normal expression, except for increased chemokines CXCL9, CXCL10 and CXCL-11 [17]. Two "inflammatory" patients responded to Actemra^R therapy by down regulation of inflammatory genes [16]. To develop a scheme for sub grouping sALS patients for an appropriate anti-inflammatory therapy, we have investigated the transcriptome of ALS patients' PBMCs for the signatures of signaling pathways.

Materials and methods

Patients and controls

The immune and genetic investigation had institutional and ethical review board approval. The study population included ALS patients from a previous study [16] and a pair of twins discordant in the diagnosis of ALS (**Figure 1A**). The assignment to the previously described inflammatory group was based on RT PCR expression of *IL-1β* and *IL-6* in PBMCs [16] with the threshold value (Ct) at most time-points < 24 cycles; in "non-inflammatory" patients , the threshold C_t was ≥ 24 cycles. Three patients received Actemra^R infusions from their private physicians.

RT PCR assay of inflammatory gene mRNAs

The assay was done using a custom array of inflammatory genes (SABiosciences) on the Roche LightCycler using the ΔΔCt method [29]. A lower number of cycles (C_t) indicate a higher inflammation.

Th17 cell assay

PBMCs were isolated by the ficoll-hypaque gradient method, washed with 1XPBS and re-suspended in completed medium. 0.5-1.0×10^6 cells were incubated with medium, or medium with superoxide dismutase1 (SOD-1) or SOD-1 (2 μg/ml) plus tocilizumab (10 μg/ml) for 20 hours, and in presence of Brefeldin A (Golgiplug) (1 μl/ml) (BD Biosciences) for last 6 hours. The cells were harvested and washed with a FACS buffer (PBS in 0.02% NaN_3 (wt/vol) and 0.5% BSA), labeled with FITC-conjugated anti-CD3 antibodies, washed, fixed and permeabilized with cyofix/cytoperm (BD Biosciences) solution according to the manufacturer protocol and stained by PE or APC-conjugated anti-IL-17 (BD Biosciences, San Diego, CA). Flow cytometry was performed using a FACSCalibur instrument. Data were analyzed using FlowJo software (Ashland, OR) with lymphocyte gate, based on forward and side scatter.

Patient expression profiling

RNA-seq was performed on patient PBMCs using standard Illumina RNA-seq library construction protocols. RNA-seq libraries were sequenced on Illumina HiSeq 2000. Reads were aligned to the hg19 reference genome using TopHat [30]. Gene counts were quantified with HTSeq and normalized with DESeq, followed by adding a pseudo count of 1 for each gene [31]. To cluster the patient PBMC RNA-Seq data, gene expression signatures were identified from microarray intensity values of published works in which PBMC derived monocytes were stimulated by IL4 (IL4_06, IL4_24), IL10 (IL10_06, IL10_24), IL15 (IL15_06, IL15_24), NOD2L (NOD2L_06, NOD2L_24), TLR2/1L (TLR2/1L_06, TLR2/1L_24), Vitamin A (VIT_A), or Vitamin D (VIT_D) at different periods of exposure [18-20]. fRMA was used for normalization of the microarray intensity values [32]. To identify genes highly expressed in the stimulated monocytes, the ratio of the median intensity value of a perturbation set to the median intensity value of their respective control set was used to rank the genes. The top 50 ranking genes were chosen for each perturbation to serve as the signature for high expression. Additional signatures include the genes induced by interferon-β (*IFN-β*) and interferon-γ (*IFN-γ*) from IFN-treated PBMCs [33]. For each patient, a signature expression value was determined by calculating the arithmetic mean of the $\log_{10}$ gene count across the 50 genes within a signature for each signature. Patients were subsequently clustered on their set of mean signature values distanced by correlation.

Pathway enrichment analysis

Differentially expressed genes between the inflammatory patients (1, 5, and 11) and non-

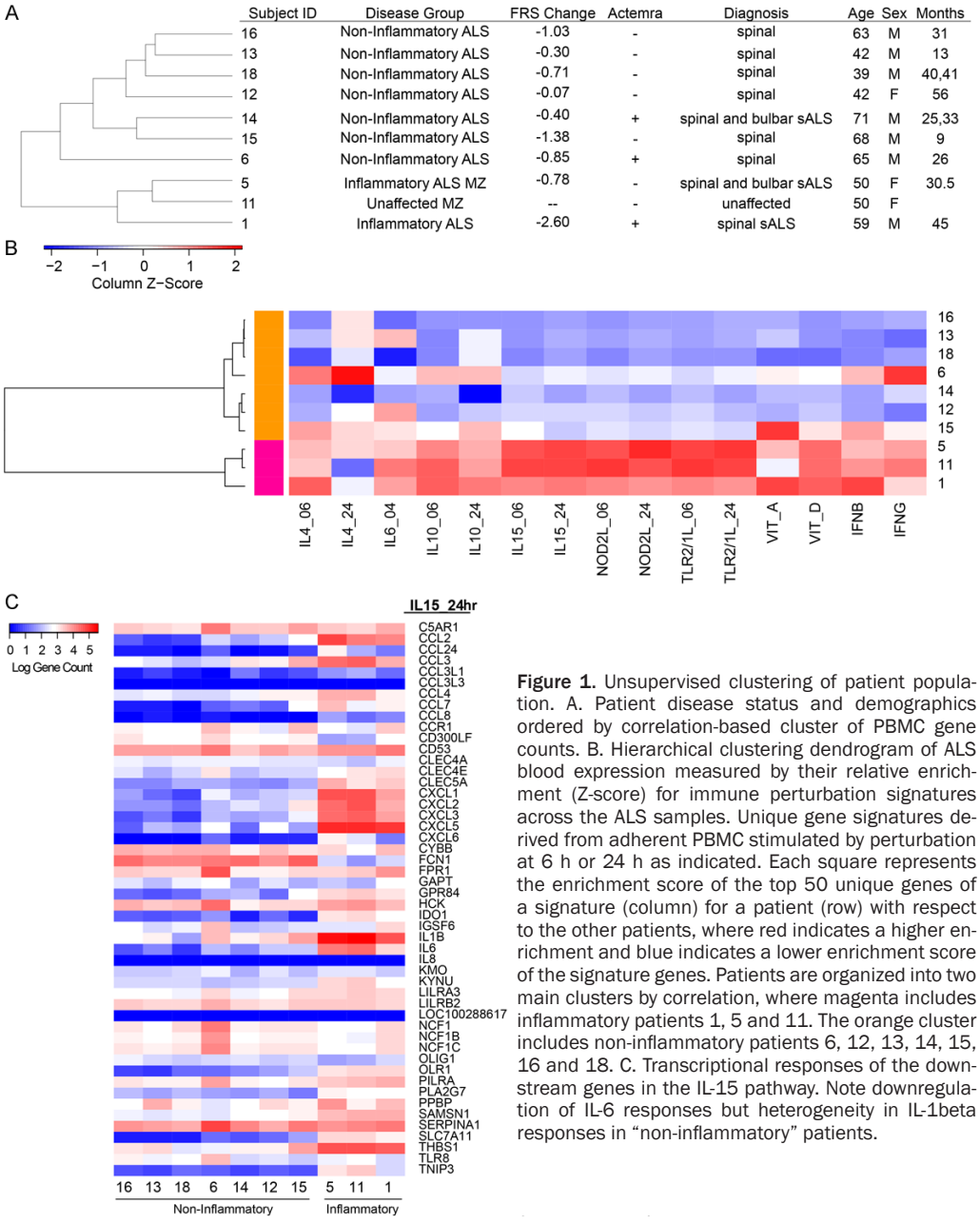


Figure 1. Unsupervised clustering of patient population. A. Patient disease status and demographics ordered by correlation-based cluster of PBMC gene counts. B. Hierarchical clustering dendrogram of ALS blood expression measured by their relative enrichment (Z-score) for immune perturbation signatures across the ALS samples. Unique gene signatures derived from adherent PBMC stimulated by perturbation at 6 h or 24 h as indicated. Each square represents the enrichment score of the top 50 unique genes of a signature (column) for a patient (row) with respect to the other patients, where red indicates a higher enrichment and blue indicates a lower enrichment score of the signature genes. Patients are organized into two main clusters by correlation, where magenta includes inflammatory patients 1, 5 and 11. The orange cluster includes non-inflammatory patients 6, 12, 13, 14, 15, 16 and 18. C. Transcriptional responses of the downstream genes in the IL-15 pathway. Note downregulation of IL-6 responses but heterogeneity in IL-1beta responses in "non-inflammatory" patients.

inflammatory patients (6, 12, 13, 14, 15, 16, and 18) were detected using DESeq (Anders et al. 2010). The genes were ranked by the fold change expression of the inflammatory patients to the non-inflammatory patients. To identify pathways enriched in the high expression genes among the inflammatory patients, the top 750 genes with a log2 fold change > 2 and an FDR < 0.05 were submitted to QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity). The bottom 539 genes with a log2 fold change < -2 and an FDR < 0.05 were submitted to Ingenuity's pathway analysis tool to identify en-

riched pathways for higher expression genes among the non-inflammatory patients.

Results

Clustering of sporadic ALS patients according to the signatures of Th1 and B-cell pathways

Clinical and demographic data of nine sALS patients and one identical non-affected twin of the patient #5 from a previous study [16] are displayed (**Figure 1A**). Through an unsupervised clustering method, the sALS patients #6, 12, 13, 14, 15, 16 and 18 were tightly clustered together based on the expression of 25,367 genes, while sALS patients #1, 5, and 11 formed a second cluster (Supplemental Figure 1).

To characterize the inflammatory state of the sALS patients, we intersected their PBMC mRNA expression profiles with gene expression signatures induced by activation of immune cells. These signatures are derived from previous studies involving the stimulation of adherent PBMCs with Th1-and Th2-like cytokines (IL-4, IL-10, IL-15, IFN-β, or IFN-γ), microbial ligands to innate immune receptors (NOD2L, TLR2/1L), and the vitamins A and D [18-20]. We generated a list of unique signature genes that were upregulated in response to each of these perturbations (see Methods). The relative enrichment score (Z-score) of each perturbation signature across the ALS blood samples was clustered and represented as a heat map of the enrichment score (**Figure 1B**), in which the ALS profiles are clustered based on their signature enrichment. Strikingly, we find that this analysis also generated two contrasting groups, much as the previous one in **Figure 1A**. This allows us to characterize the larger of the two groups (patients 16, 13, 18, 6, 14, 12, and 15) as B-cell "noninflammatory", as it has lower levels of *IL-15, NOD2L,* and *TLR2/1L*-induced gene signatures in comparison to the Th1 (patients 5,11, and 1) "inflammatory" group (**Figure 1B**). We find that the cytokine response signatures of the noninflammatory group are more heterogeneous. The patient #6 had elevated levels of *IL-4* and *IFNγ* signatures, the patient #15 showed higher levels of the vitamin A program, and the patients #6, 12, 16, and 18 had higher levels of both the ligands APRIL and BAFF and the receptors BCMA and TACI (Supplemental Figure 2). The transcriptional responses of the down-stream genes in the IL-15 pathway (**Figure 1C**) separate the patients in a greater detail and confirmed that some patients had a mixed pattern (e.g. patient #6 and #15).

Clustering of sporadic ALS patients with follow-up samples

A follow-up set of RNA-Seq samples were available for patients #18 and 14 and clustered on the expression of 25,367 genes along with other patient samples. The paired set of samples 18_R1, 18_R2, and 14_R1, 14_R2 reveal tight clustering within patients that is indicative of high correlation in gene expression profiles from the same individual at different time points (Supplemental Figure 1). This suggests that the gene expression signatures are quite robust over time, although they manifest some differences resulting from the treatment or disease progression.

Enriched pathways of high expression genes in inflammatory patients

We next asked whether the genes that were differentially expressed in the inflammatory and noninflammatory groups were enriched for certain functions and pathways using the Ingenuity Pathway Analysis tool (QIAGEN's Ingenuity Pathway Analysis IPA, QIAGEN Redwood City <http://www.qiagen.com/ingenuity>. Accessed 2015 Mar 5). We found that the pathways enriched among the inflammatory patients include granulocyte and agranulocyte adhesion and diapedesis, hepatic fibrosis, differential regulation of cytokine production in macrophages and T helper cells by IL-17A and IL-17F, inhibition of matrix metalloproteases, atherosclerosis signaling, and IL6-signaling (**Figure 2A**). Genes among the *IL-6* signaling pathway include *IL6* itself and many genes downstream of the *IL-6* responsive element (*TSG6, COL1A1*, and *IL8*) (**Figure 2B**). The differential regulation of cytokine production in macrophages and T helper cells by *IL-17A* and *IL-17F* pathways include pro-inflammatory genes that are induced in macrophages by either *IL-17A* or *IL-17F* (*IL-1β, IL-6, IL-10, IL-12, IL-13, CCL2, CCL3*, and G-CSF) (**Figure 2C**).

Enriched pathways of high expression genes in non-inflammatory patients

Pathways enriched among the genes that have higher expression in non-inflammatory patients

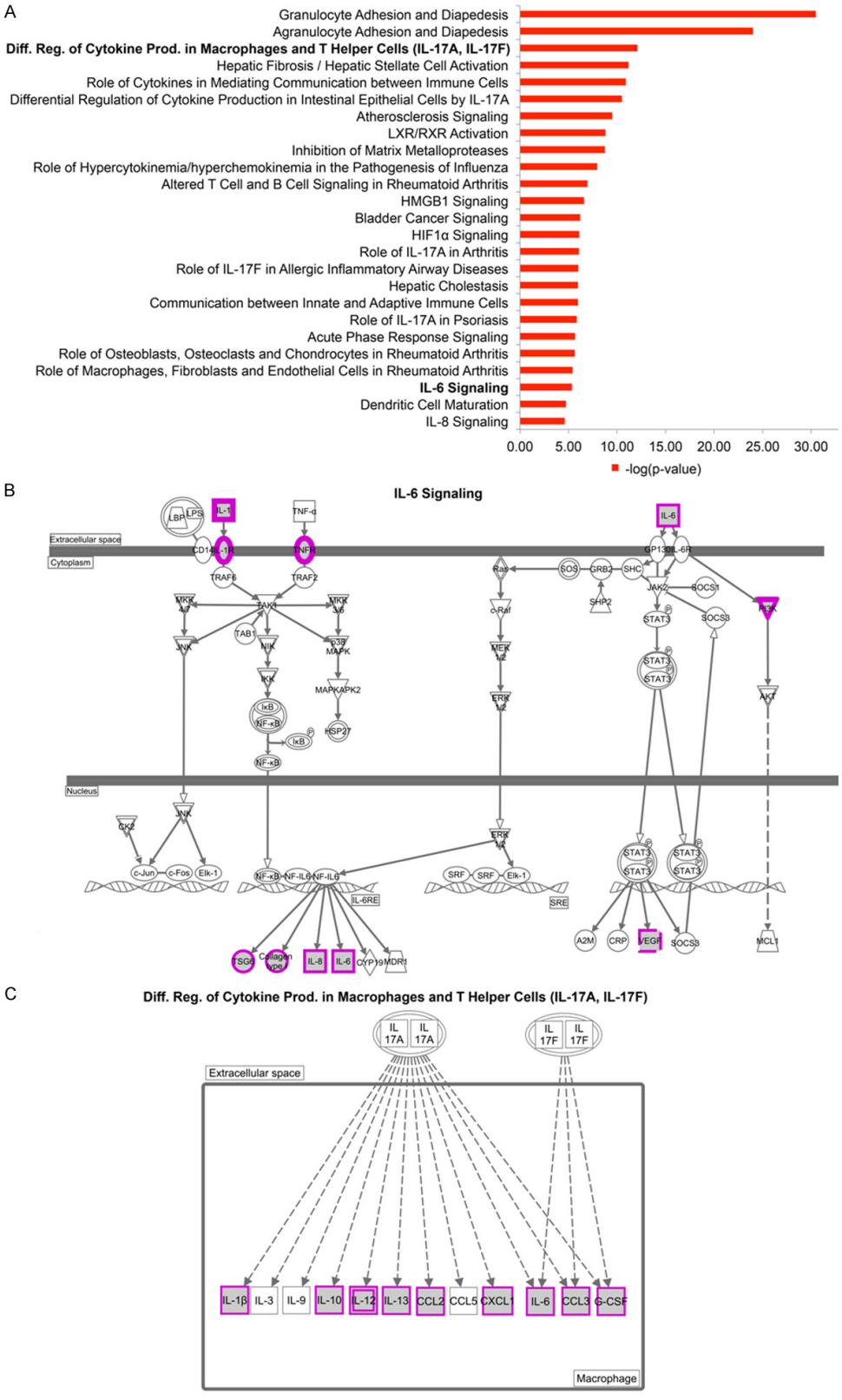

A
Granulocyte Adhesion and Diapedesis
Agranulocyte Adhesion and Diapedesis
Diff. Reg. of Cytokine Prod. in Macrophages and T Helper Cells (IL-17A, IL-17F)
Hepatic Fibrosis / Hepatic Stellate Cell Activation
Role of Cytokines in Mediating Communication between Immune Cells
Differential Regulation of Cytokine Production in Intestinal Epithelial Cells by IL-17A
Atherosclerosis Signaling
LXR/RXR Activation
Inhibition of Matrix Metalloproteases
Role of Hypercytokinemia/hyperchemokinemia in the Pathogenesis of Influenza
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis
HMGB1 Signaling
Bladder Cancer Signaling
HIF1α Signaling
Role of IL-17A in Arthritis
Role of IL-17F in Allergic Inflammatory Airway Diseases
Hepatic Cholestasis
Communication between Innate and Adaptive Immune Cells
Role of IL-17A in Psoriasis
Acute Phase Response Signaling
Role of Osteoblasts, Osteoclasts and Chondrocytes in Rheumatoid Arthritis
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis
IL-6 Signaling
Dendritic Cell Maturation
IL-8 Signaling
0.00 5.00 10.00 15.00 20.00 25.00 30.00
-log(p-value)
B
IL-6 Signaling
Extracellular space
Cytoplasm
Nucleus
C
Diff. Reg. of Cytokine Prod. in Macrophages and T Helper Cells (IL-17A, IL-17F)
Extracellular space
IL 17A
IL 17F
IL-1β
IL-3
IL-9
IL-10
IL-12
IL-13
CCL2
CCL5
CXCL1
IL-6
CCL3
G-CSF
Macrophage

Figure 2. Pathway Enrichment of Higher Expression Genes in Inflammatory Patients. A. The top 25 significant canonical pathways from the high expression genes in inflammatory patients. B. *IL-6* signaling pathway. Highly expressed genes in inflammatory patients are marked in purple. Omitted from the figure is the pathway involved in T helper cells. C. Differential regulation of cytokine production in macrophages and T helper cells by *IL-17A* and *IL-17F* pathways.

include alpha-tocopherol degradation, altered cytokine signaling, agranulocyte adhesion, serotonin degradation, retinoate biosynthesis, B cell development, atherosclerosis signaling, and other pathways (**Figure 3A**). Highly expressed genes in non-inflammatory patients include (a) the genes *CYP4F2* and *CYP4F3* involved in alpha-tocopherol degradation (**Figure 3B**); (b) the genes in the B cell development pathway including the members of MHCII (*HLA-DMA*, *HLA-DMB*, *HLA-DRB1*, *HLA-DRB5*) (**Figure 3C**); and (c) the genes *SULT1A1*, *SULT1B1*, *DHRS9*, *ALDH2*, and *ALDH1A1* involved in serotonin degradation (**Figure 3D**). In addition, the ligand for the B cell developmental pathway (*APRIL*) and the receptors (*BCMA/TNFRSF17* and *BCR)* were heterogeneously overexpressed in non-inflammatory patients (Supplemental Figure 2).

Th17 cells are increased in the inflammatory subgroup and are inhibited by tocilizumab

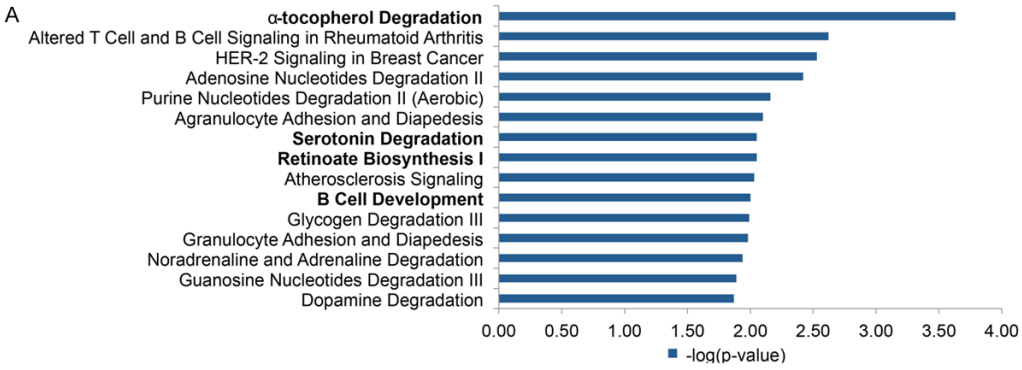
In freshly isolated PBMCs, Th17 cells were increased in the inflammatory subgroup patients in comparison to the non-inflammatory group patients and control subjects (**Figure 4**). After overnight stimulation by fibrillar SOD-1, Th17 increased 2- to 5-fold in the inflammatory subgroup subjects but did not change significantly in the non-inflammatory subgroup and control subjects. Overnight treatment with tocilizumab reduced (N.S.) Th17 induction by SOD-1 in the inflammatory subgroup, but did not have an effect in the non-inflammatory subgroup.

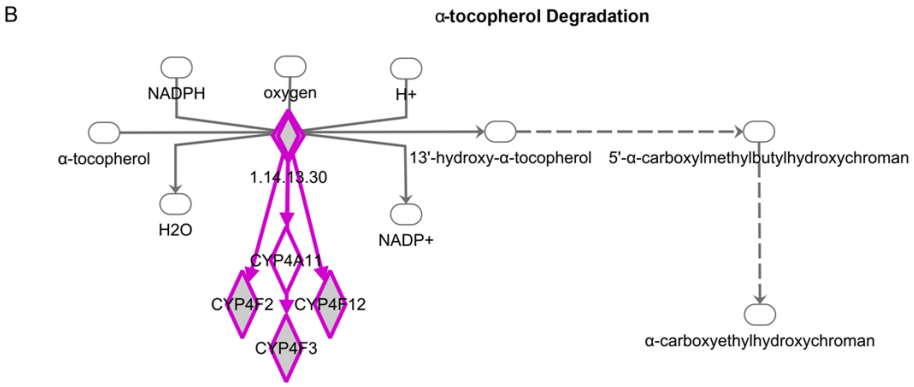
Inflammatory cytokines are down-regulated by Actemra^R therapy in the "inflammatory" subset but not in the "non-inflammatory" subset

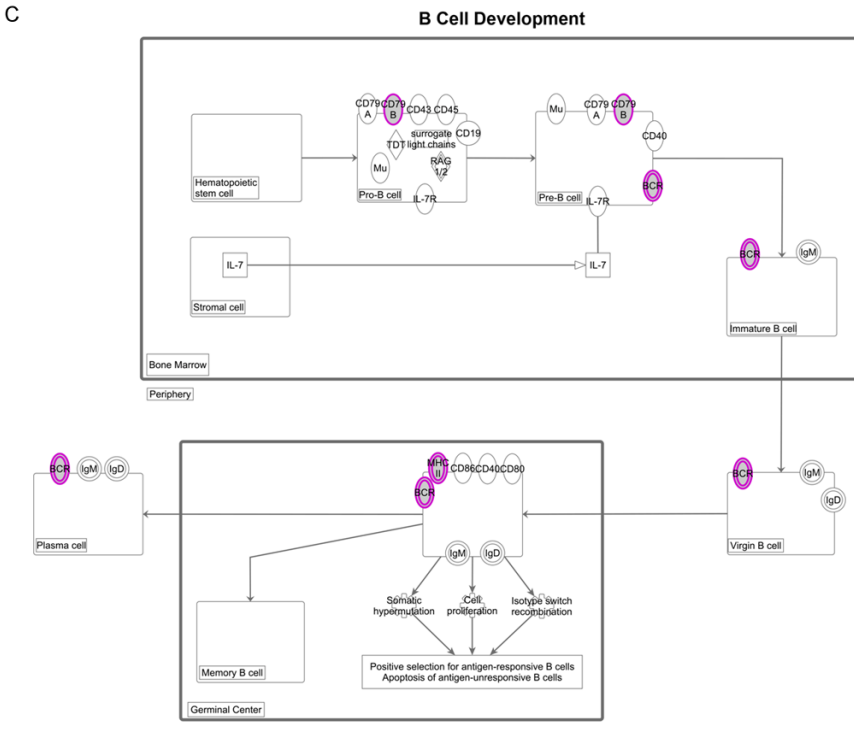
A brief follow-up of patients #1 and #6 was previously published [16]. Here we show that on a longitudinal follow-up of the inflammatory patient #1, and the non-inflammatory patients #6 and 14, inflammatory or non-inflammatory transcription, respectively, appeared to be a stable biomarker of these sALS patients in the course of the disease (600-1400, 100 to 600, and 750 to 950 days after onset, respectively) (**Figure 5**). Extended follow-up of the inflammatory patient 1 revealed up regulation of inflammatory cytokines on entry into the Actemra^R therapy in May 2012 (913 days after onset), but, following the start of therapy, down regulation of inflammatory cytokines for 4 months with immediate down regulation 2 hours after infusion of Actemra (see **Figure 5D**). The patient then developed a brief resistance to Actemra^R therapy for 2 months followed again by attenuation of inflammation. However, he stopped therapy in February 2013 (1200 days after onset) and afterwards became more inflammatory and expired 1930 days after onset. Two non-inflammatory sALS patients were given a short course of Actemra^R therapy. They had low transcription of inflammatory cytokines on entry into the therapy, but their inflammatory cytokine transcription actually increased during the therapy, which was then stopped (**Figure 5B** and **5C**). To determine the immediate response to Actemra, inflammatory cytokine transcription was measured before and 1-2 hr after treatment. For patient 1, the *IL1A* and *IL8* transcription decreased immediately after the infusion (**Figure 5D**). For patient 6, all inflammatory cytokine, with the exception of *IL1A,* increased following the infusion (**Figure 5E**).

Discussion

In this study, we confirmed the heterogeneity of inflammation in sALS patients, which was previously observed and believed to be important for selection of appropriate patients for tocilizumab (Actermra) therapy [16]. Only "inflammatory" patients are likely to respond to Actemra. Although inflammation has been demonstrated in a longitudinal study [16] and cross-sectional studies of sALS patients [10, 21], the progression of inflammation since the beginning to the end has not been completely analyzed. Here we observe that the inflammatory or non-inflammatory phenotype was maintained throughout the observed disease course, which however, started only 600 days after onset in the "inflammatory" patient. Thus, it is still plausible that a non-inflammatory phase preceded the inflammatory phase early after onset, as described by S. Appel's group [22]. It will be important to fol-


A
α-tocopherol Degradation
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis
HER-2 Signaling in Breast Cancer
Adenosine Nucleotides Degradation II
Purine Nucleotides Degradation II (Aerobic)
Agranulocyte Adhesion and Diapedesis
Serotonin Degradation
Retinoate Biosynthesis I
Atherosclerosis Signaling
B Cell Development
Glycogen Degradation III
Granulocyte Adhesion and Diapedesis
Noradrenaline and Adrenaline Degradation
Guanosine Nucleotides Degradation III
Dopamine Degradation
0.00 0.50 1.00 1.50 2.00 2.50 3.00 3.50 4.00
-log(p-value)


B
α-tocopherol Degradation
NADPH
oxygen
H+
α-tocopherol
1.14.13.30
13'-hydroxy-α-tocopherol
5'-α-carboxylmethylbutylhydroxychroman
H2O
NADP+
CYP4A11
CYP4F2
CYP4F12
CYP4F3
α-carboxyethylhydroxychroman


C
B Cell Development
Hematopoietic stem cell
Pro-B cell
Pre-B cell
Stromal cell
Immature B cell
Bone Marrow
Periphery
Plasma cell
Memory B cell
Virgin B cell
Somatic hypermutation
Cell proliferation
Isotype switch recombination
Positive selection for antigen-responsive B cells
Apoptosis of antigen-unresponsive B cells
Germinal Center

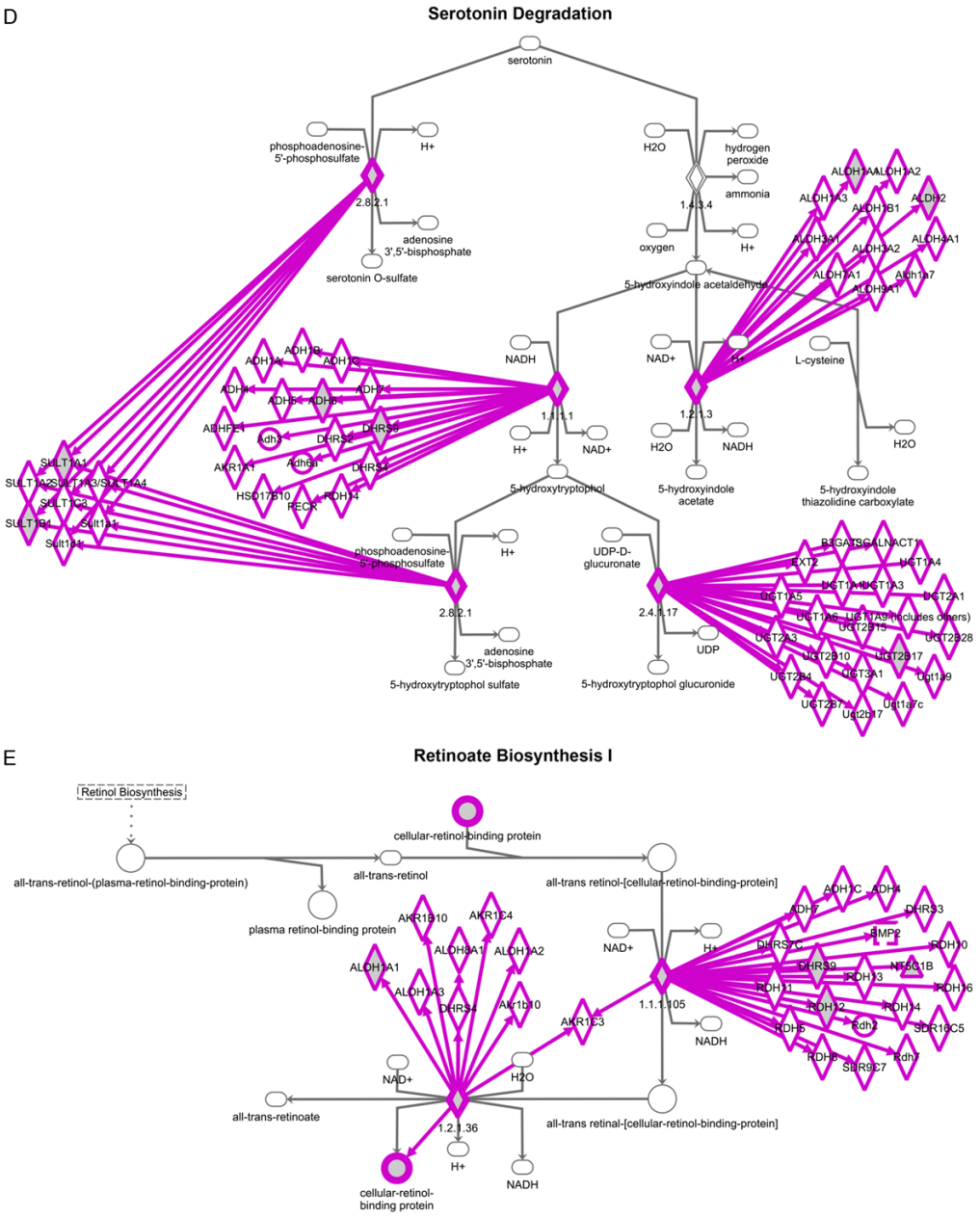


Figure 3. Pathway Enrichment of Higher Expression Genes in Non-Inflammatory Patients. A. The top 25 significant canonical pathways from the high expression genes in non-inflammatory patients. B. α-tocopherol (vitamin E) degradation. C. B-Cell development pathway with high expression genes in the non-inflammatory patients marked in purple. D. Serotonin degradation pathway. Highly expressed genes in inflammatory patients are marked in purple. E. Retinoate Biosynthesis I pathway. Highly expressed genes in inflammatory patients are marked in purple.

low ALS patients over time for establishing their "inflammatory" phenotype, as "non-inflammatory" ALS patients may become "inflammatory" and respond to Actemra therapy.

The "inflammatory" patient #1 had inflammatory phenotype before and after treatment with Actemra^{R}, but became less inflammatory during Actemra^{R} therapy, except for a transitory

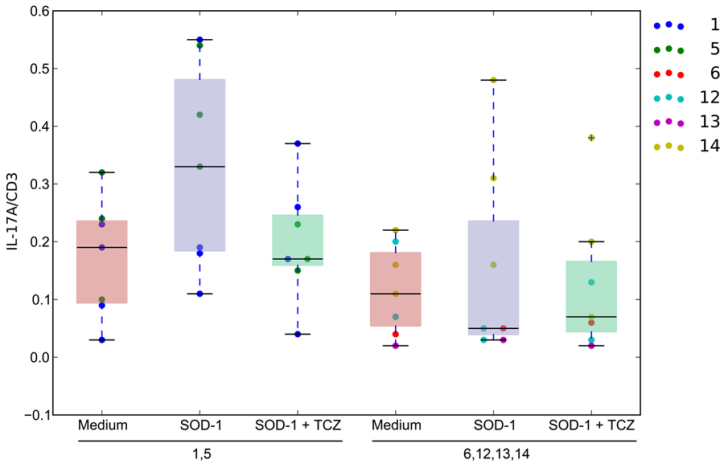


Figure 4. Th17 Cell Assay of Patient PBMCs. The distribution of the % of IL-17/CD3 cells in lymphocytes grouped by inflammatory patients (columns 1, 2, and 3) and non-inflammatory patients (columns 4, 5, and 6). The red distributions are the Th17 to CD3 cell ratios in medium. The blue distributions are the Th17 to CD3 cell ratios of PBMCs stimulated by SOD-1. The green distributions are the Th17 to CD3 cell ratios of PBMCs stimulated by SOD-1 in the presence of tocilizumab.

resistance to this therapy probably due to reduced uptake and degradation of IL-6 caused by tocilizumab [23]. The "inflammatory" patients had more Th17 cells in comparison to "non-inflammatory" patients. Tocilizumab therapy in vitro and in vivo attenuated Th17 cell induction as well as inflammatory cytokines. The patients in the "non-inflammatory" subgroup had low transcription and production of inflammatory cytokines and chemokines at baseline but, in response to the Actemra^{R} therapy, two patients actually increased the transcription of IL-6 and other inflammatory cytokines.

The up regulated Th1 and B-cell pathways clarify certain features of the pathophysiology of sALS patients. In both groups of patients, immune cell adhesion and diapedesis explain the infiltration by immune cells (granulocytes and agranulocytes) of the affected spinal cord segments [3]. In the inflammatory group, hepatic fibrosis and stellate cell activation is consistent with fibrosis and liver dysfunction in sALS patients [24]. Atherosclerosis signaling could be related to vascular changes with downregulation of tight junctions caused by SOD-1 in an animal model [25] and to inflammation caused by SOD-1 in human macrophages in vitro [3]. Increased α-tocopherol degradation may lead to deficiency of vitamin E. Patients with familial isolated vitamin E deficiency have neurological disorders characteristic of vitamin E deficiency such as cerebellar ataxia and dysarthria [26]. Importantly, in a large study of ALS incidence, long-term vitamin E supplement use was associated with lower ALS rates [27]. Increased serotonin degradation may be related to depression, which is common in ALS patients, as 37% patients are taking anti-depressant medication [28]. Increased retinoate may contribute to osteoporosis in ALS patients.

In conclusion, this study describes heterogeneity of signature pathways in sALS patients that may inform personalized therapy with drugs targeting various pathways, including inflammatory, vitamin, cytochrome P450 superfamily of enzymes, and B-cell developmental pathways. Currently, tocilizumab (Actemra) has shown positive effects in Th1 patients. However, tocilizumab would not be beneficial in the non-inflammatory subset of patients who might benefit from other approaches, such as a therapy targeted at B cells, since the B cell activation and survival genes were upregulated in these patients. These results provide a new rationale for a targeted molecular therapy of inflammatory sALS patients guided in a personalized fashion by the PBMC transcriptome.

Acknowledgements

We thank patients and families of patients in the study. LL was supported by 5P01GM099134-03, DM was supported by 1P50AR063020-01, and RRS was supported by R01 AI80778.

Disclosure of conflict of interest

M.F. received paid travel to meetings and honoraria for consultations from Smartfish AS, Oslo, Norway.

Authors' contribution

MF and MP designed research. LL and DM performed bioinformatics analysis. RH, LR, AR per-

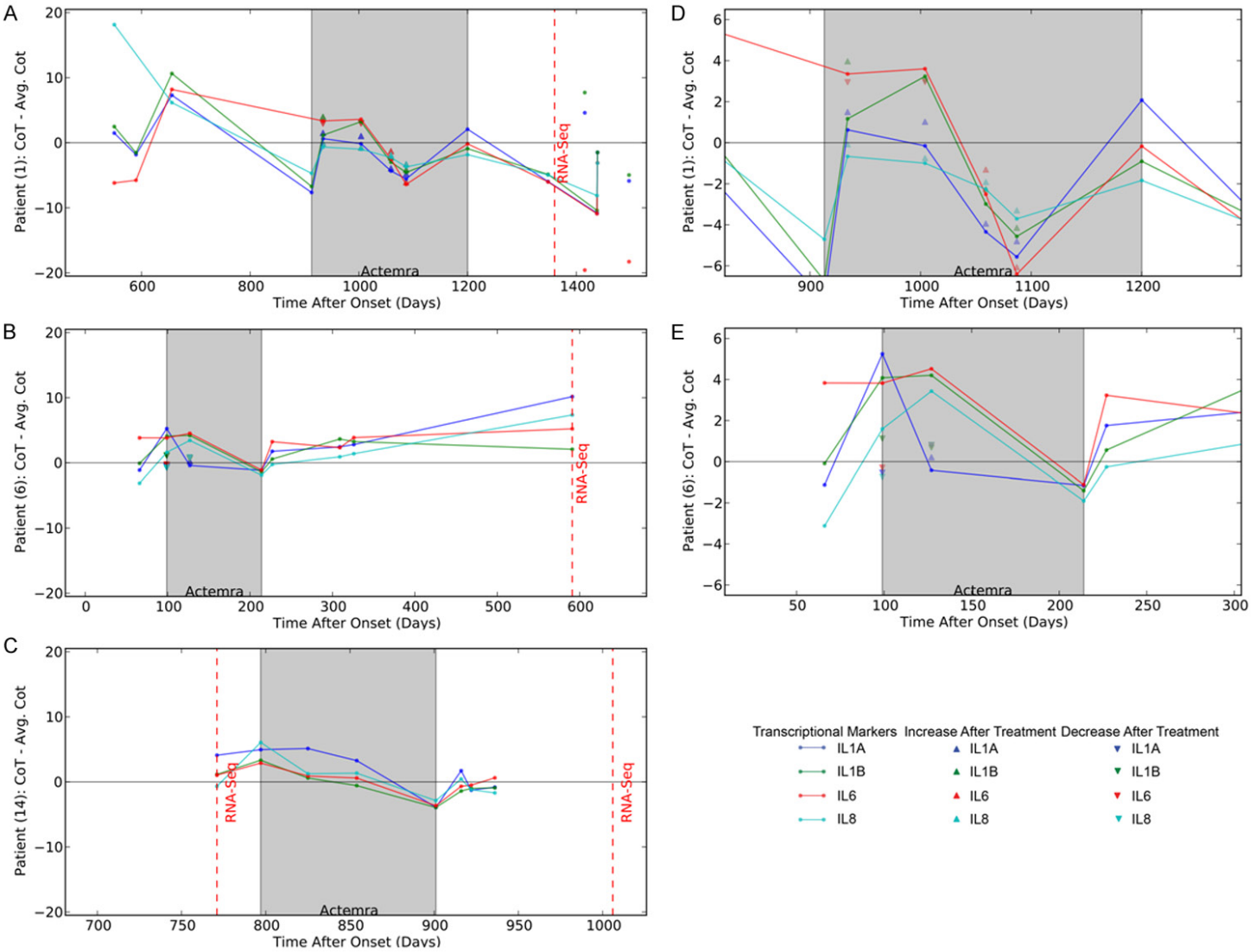

A
Patient (1): CoT - Avg. Cot
Actemra
RNA-Seq
Time After Onset (Days)
B
Patient (6): CoT - Avg. Cot
Actemra
RNA-Seq
Time After Onset (Days)
C
Patient (14): CoT - Avg. Cot
Actemra
RNA-Seq
RNA-Seq
Time After Onset (Days)
D
Patient (1): CoT - Avg. Cot
Actemra
Time After Onset (Days)
E
Patient (6): CoT - Avg. Cot
Actemra
Time After Onset (Days)
Transcriptional Markers
IL1A
IL1B
IL6
IL8
Increase After Treatment
IL1A
IL1B
IL6
IL8
Decrease After Treatment
IL1A
IL1B
IL6
IL8

Figure 5. Longitudinal Course of the Transcription of IL-1alpha, IL-1 beta, IL6 and IL-8 in three ALS patients. A. Longitudinal course of inflammatory patient #1; B. Longitudinal course of a non-inflammatory patient #6; C. Longitudinal course of a "non-inflammatory" patient #14; D. The rescaled transcriptional response of patient 1 during the Actemra treatment period. E. The rescaled transcriptional response of patient 14 during the Actemra treatment period. The blue dots indicate the difference of the observed *IL-1α* transcriptional level to the mean *IL-1α* level. The green dots indicate the difference of the observed *IL-1β* transcriptional level to the mean *IL-1β* level. The red dots indicate the difference of the observed *IL-6* transcriptional level from the mean *IL-6* level. The light blue dots indicate the difference of the observed *IL-8* transcriptional level to the mean *IL-8* level. All mean transcriptional level was calculated across all samples and days. The grey box indicates the days after onset of ALS in which transcriptional levels were measured after Actemra treatment. The dashed red line indicates the days after onset in which PBMCs were sampled for RNA-Seq analysis. Up arrows indicate the level of a positive (i.e. decreased transcription) immediate response to Actemra for the corresponding transcriptional marker, while down arrows indicate the level of a negative (increased transcription) response to Actemra.

formed molecular studies. MF, MP, LL, BS, SW, and RR analyzed the data. MF, MP, LL, RRS wrote the paper. All authors read and approved the final manuscript.

Address correspondence to: Dr. Milan Fiala, Department of Surgery, UCLA School of Medicine, Suite 220, UCLA Medical Plaza, Los Angeles, CA 90095-7022. E-mail: fiala@mednet.ucla.edu; Matteo Pellegrini, Department of Molecular, Cell and Developmental Biology UCLA, Box 951606, Los Angeles, CA 90095-1606. E-mail: matteope@gmail.com

References

[1] McGeer PL, McGeer EG. Inflammatory processes in amyotrophic lateral sclerosis. Muscle Nerve 2002; 26: 459-70.

[2] Henkel JS, Engelhardt JI, Siklós L, Simpson EP, Kim SH, Pan T, Goodman JC, Siddique T, Beers DR, Appel SH. Presence of dendritic cells, MCP-1, and activated microglia/macrophages in amyotrophic lateral sclerosis spinal cord tissue. Ann Neurol 2004; 55: 221-35.

[3] Fiala M, Chattopadhay M, La Cava A, Tse E, Liu G, Lourenco E, Eskin A, Liu PT, Magpantay L, Tse S, Mahanian M, Weitzman R, Tong J, Nguyen C, Cho T, Koo P, Sayre J, Martinez-Maza O, Rosenthal MJ, Wiedau-Pazos M. IL-17A is increased in the serum and in spinal cord CD8 and mast cells of ALS patients. J Neuroinflammation 2010; 7: 76-90.

[4] Liu G, Fiala M, Mizwicki MT, Sayre J, Magpantay L, Siani A, Mahanian M, Chattopadhyay M, La Cava A, Wiedau-Pazos M. Neuronal phagocytosis by inflammatory macrophages in ALS spinal cord: inhibition of inflammation by resolvin D1. Am J Neurodegener Dis 2012; 1: 60-74.

[5] Vaknin I, Kunis G, Miller O, Butovsky O, Bukshpan S, Beers DR, Henkel JS, Yoles E, Appel SH, Schwartz M. Excess circulating alternatively activated myeloid (M2) cells accelerate ALS progression while inhibiting experimental autoimmune encephalomyelitis. PLoS One 2011; 6: e26921.

[6] Butovsky O, Siddiqui S, Gabriely G, Lanser AJ, Dake B, Murugaiyan G, Doykan CE, Wu PM, Gali RR, Iyer LK, Lawson R, Berry J, Krichevsky AM, Cudkowicz ME, Weiner HL. Modulating inflammatory monocytes with a unique microRNA gene signature ameliorates murine ALS. J Clin Invest 2012; 122: 3063-87.

[7] Evans MC, Couch Y, Sibson N, Turner MR. Inflammation and neurovascular changes in amyotrophic lateral sclerosis. Mol Cell Neurosci 2013; 53: 34-41.

[8] Graves MC, Fiala M, Dinglasan LA, Liu NQ, Sayre J, Chiappelli F, van Kooten C, Vinters HV. Inflammation in amyotrophic lateral sclerosis spinal cord and brain is mediated by activated macrophages, mast cells and T cells. Amyotroph Lateral Scler Other Motor Neuron Disord 2004; 5: 213-9.

[9] Rentzos M, Rombos A, Nikolaou C, Zoga M, Zouvelou V, Dimitrakopoulos A, Alexakis T, Tsoutsou A, Samakovli A, Michalopoulou M, Evdokimidis J. Interleukin-17 and interleukin-23 are elevated in serum and cerebrospinal fluid of patients with ALS: a reflection of Th17 cells activation? Acta Neurol Scand 2010; 122: 425-9.

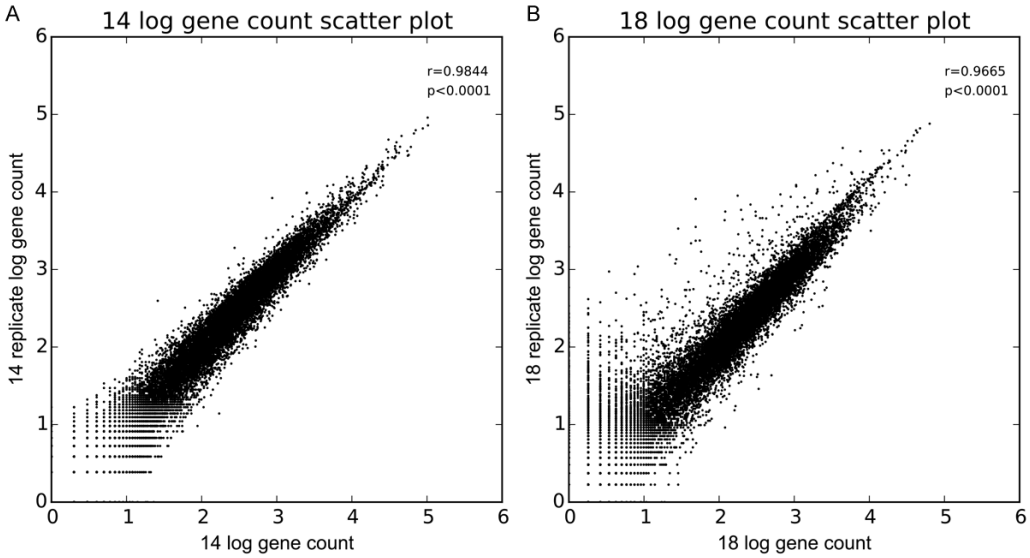
[10] Saresella M, Piancone F, Tortorella P, Marventano I, Gatti A, Caputo D, Lunetta C, Corbo M, Rovaris M, Clerici M. T helper-17 activation dominates the immunologic milieu of both amyotrophic lateral sclerosis and progressive multiple sclerosis. Clin Immuno 2013; 148: 79-88.

[11] Turner MR, Goldacre R, Ramagopalan S, Talbot K, Goldacre MJ. Autoimmune disease preceding amyotrophic lateral sclerosis: an epidemiologic study. Neurology 2013; 81: 1222-5.

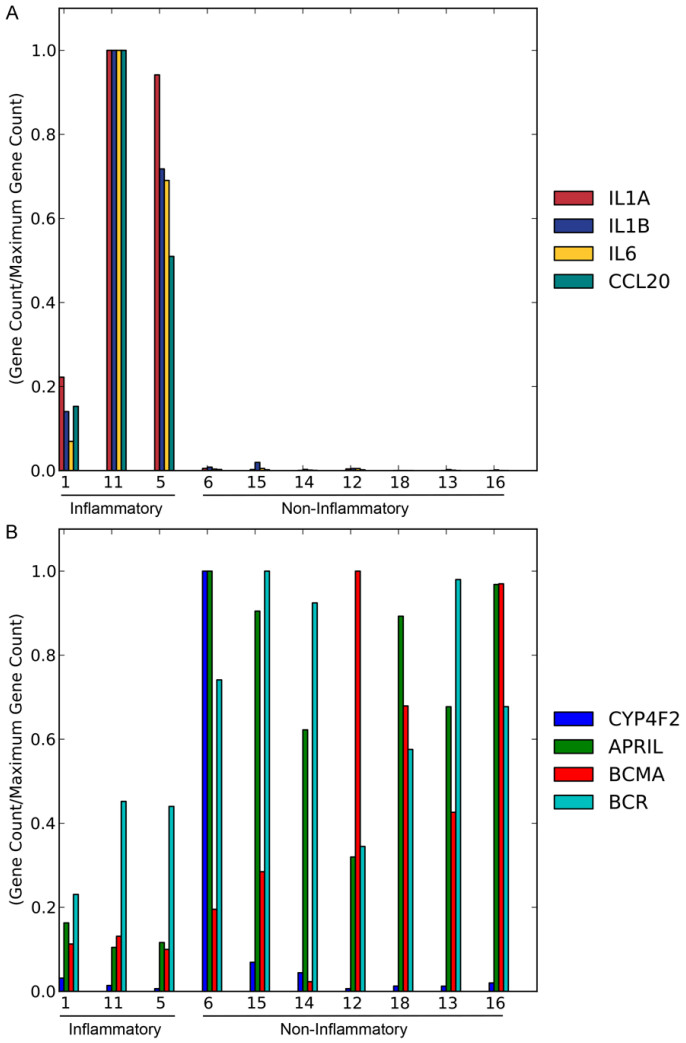
[12] Kebir H, Kreymborg K, Ifergan I, Dodelet-Devillers A, Cayrol R, Bernard M, Giuliani F, Arbour N, Becher B, Prat A. Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation. Nat Med 2007; 13: 1173-5.

[13] Campbell IL, Erta M, Lim SL, Frausto R, May U, Rose-John S, Scheller J, Hidalgo J. Transsignaling is a dominant mechanism for the

pathogenic actions of interleukin-6 in the brain. J Neurosci 2014; 34: 2503-13.
[14] Cudkowicz ME, Shefner JM, Schoenfeld DA, Zhang H, Andreasson KI, Rothstein JD, Drachman DB. Trial of celecoxib in amyotrophic lateral sclerosis. Ann Neurol 2006; 60: 22-31.
[15] Miller RG, Zhang R, Block G, Katz J, Barohn R, Kasarskis E, Forshew D, Gopalakrishnan V, McGrath MS. NP001 regulation of macrophage activation markers in ALS: A phase I clinical and biomarker study. Amyotroph Lateral Scler Frontotemporal Degener 2014; 15: 601-9.
[16] Fiala M, Mizwicki MT, Weitzman R, Magpantay L, Nishimoto N. Tocilizumab infusion therapy normalizes inflammation in sporadic ALS patients. Am J Neurodegener Dis 2013; 2: 129-39.
[17] Mizwicki MT, Fiala M, Magpantay L, Aziz N, Sayre J, Liu G, Siani A, Chan D, Martinez-Maza O, Chattopadhyay M, La Cava A. Tocilizumab attenuates inflammation in ALS patients through inhibition of IL6 receptor signaling. Am J Neurodegener Dis 2012; 1: 305-15.
[18] Schenk M, Fabri M, Krutzik SR, Lee DJ, Vu DM, Sieling PA, Montoya D, Liu PT, Modlin RL. Interleukin-1beta triggers the differentiation of macrophages with enhanced capacity to present mycobacterial antigen to T cells. Immunology 2014; 141: 174-80.
[19] Montoya D, Inkeles MS, Liu PT, Realegeno S, Teles RM, Vaidya P, Munoz MA, Schenk M, Swindell WR, Chun R, Zavala K, Hewison M, Adams JS, Horvath S, Pellegrini M, Bloom BR, Modlin RL. IL-32 is a molecular marker of a host defense network in human tuberculosis. Sci Transl Med 2014; 6: 250ra114.
[20] Wheelwright M, Kim EW, Inkeles MS, De Leon A, Pellegrini M, Krutzik SR, Liu PT. All-trans retinoic acid-triggered antimicrobial activity against Mycobacterium tuberculosis is dependent on NPC2. J Immunol 2014; 192: 2280-90.
[21] Italiani P, Carlesi C, Giungato P, Puxeddu I, Borroni B, Bossù P, Migliorini P, Siciliano G, Boraschi D. Evaluating the levels of interleukin-1 family cytokines in sporadic amyotrophic lateral sclerosis. J Neuroinflammation 2014; 11: 94.
[22] Hooten KG, Beers DR, Zhao W, Appel SH. Protective and Toxic Neuroinflammation in Amyotrophic Lateral Sclerosis. Neurotherapeutics 2015; 12: 364-75.
[23] Nishimoto N, Terao K, Mima T, Nakahara H, Takagi N, Kakehi T. Mechanisms and pathologic significances in increase in serum interleukin-6 (IL-6) and soluble IL-6 receptor after administration of an anti-IL-6 receptor antibody, tocilizumab, in patients with rheumatoid arthritis and Castleman disease. Blood 2008; 112: 3959-64.
[24] Nakano Y, Hirayama K, Terao K. Hepatic ultrastructural changes and liver dysfunction in amyotrophic lateral sclerosis. Arch Neurol 1987; 44: 103-6.
[25] Zhong Z, Deane R, Ali Z, Parisi M, Shapovalov Y, O'Banion MK, Stojanovic K, Sagare A, Boillee S, Cleveland DW, Zlokovic BV. ALS-causing SOD1 mutants generate vascular changes prior to motor neuron degeneration. Nat Neurosci 2008; 11: 420-2.
[26] Ouahchi K, Arita M, Kayden H, Hentati F, Ben Hamida M, Sokol R, Arai H, Inoue K, Mandel JL, Koenig M. Ataxia with isolated vitamin E deficiency is caused by mutations in the alpha-tocopherol transfer protein. Nat Genet 1995; 9: 141-5.
[27] Wang H, O'Reilly ÉJ, Weisskopf MG, Logroscino G, McCullough ML, Schatzkin A, Kolonel LN, Ascherio A. Vitamin E intake and risk of amyotrophic lateral sclerosis: a pooled analysis of data from 5 prospective cohort studies. Am J Epidemiol 2011; 173: 595-602.
[28] Pisa FE, Logroscino G, Casetta A, Cecotti L, Verriello L, Bratina A, Sartori A, Lazzarino de Lorenzo L, Eleopra R, Barbone F. The Use of Antidepressant Medication before and after the Diagnosis of Amyotrophic Lateral Sclerosis: A Population-Based Cohort Study. Neuroepidemiology 2015; 44: 91-8.
[29] Mizwicki MT, Fiala M, Magpantay L, Aziz N, Sayre J, Liu G, Siani A, Chan D, Martinez-Maza O, Chattopadhyay M, La Cava A. Tozilizumab attenuates inflammation in ALS patients through inhibition of IL-6 receptor signaling. Am J Neurodegener Dis 2012; 1: 305-15.
[30] Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 2009; 25: 1105-11.
[31] Anders S, Pyl PT, Huber W. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics 2015; 31: 166-9.
[32] McCall MN, Bolstad BM, Irizarry RA. Frozen robust multiarray analysis (fRMA). Biostatistics 2010; 11: 242-53.
[33] Teles RM, Graeber TG, Krutzik SR, Montoya D, Schenk M, Lee DJ, Komisopoulou E, Kelly-Scumpia K, Chun R, Iyer SS, Sarno EN, Rea TH, Hewison M, Adams JS, Popper SJ, Relman DA, Stenger S, Bloom BR, Cheng G, Modlin RL. Type I interferon suppresses type II interferon-triggered human anti-mycobacterial responses. Science 2013; 339: 1448-53.



Supplemental Figure 1. Clustering of patient population with follow-up samples. Correlation based cluster of PBMC gene counts of UCSC hg19 genes with the patient initial RNA-Seq samples along with follow-up results for (A) patient 14 (14_R2) and (B) patient 18 (18_R2).



Supplemental Figure 2. Transcriptional responses of Th1 and Th2 genes in individual ALS patients. Differential responses by the cytokines, the chemokine CCL20, the enzyme CYP4F2, and the genes in B cells developmental pathway. A. Transcriptional response high in the inflammatory patients (IL1A, IL1B, IL6, CCL2). B. Transcriptional responses high in the non-inflammatory patients (CYP4F2, APRIL, BCMA, BCR).

Chapter 5:

Epigenetic changes in T-cell and monocyte signatures and production of neurotoxic cytokines in ALS patients

THE FASEB JOURNAL • RESEARCH • www.fasebj.org

Epigenetic changes in T-cell and monocyte signatures and production of neurotoxic cytokines in ALS patients

Larry Lam,* Lydia Chin,† Ramesh C. Halder,† Bien Sagong,† Sam Famenini,† James Sayre,‡ Dennis Montoya,* Liudmilla Rubbi,* Matteo Pellegrini,* and Milan Fiala*,†,1

*Department of Molecular, Cell, and Developmental Biology and †Department of Surgery, UCLA School of Medicine, and ‡Department of Biostatistics, UCLA School of Public Health, University of California, Los Angeles, Los Angeles, California, USA

ABSTRACT: We have investigated transcriptional and epigenetic differences in peripheral blood mononuclear cells (PBMCs) of monozygotic female twins discordant in the diagnosis of amyotrophic lateral sclerosis (ALS). Exploring DNA methylation differences by reduced representation bisulfite sequencing (RRBS), we determined that, over time, the ALS twin developed higher abundances of the CD14 macrophages and lower abundances of T cells compared to the non-ALS twin. Higher macrophage signature in the ALS twin was also shown by RNA sequencing (RNA-seq). Moreover, the twins differed in the methylome at loci near several genes, including *EGFR* and *TNFRSF11A*, and in the pathways related to the tretinoin and H3K27me3 markers. We also tested cytokine production by PBMCs. The ALS twin's PBMCs spontaneously produced IL-6 and TNF-α, whereas PBMCs of the healthy twin produced these cytokines only when stimulated by superoxide dismutase (SOD)-1. These results and flow cytometric detection of CD45 and CD127 suggest the presence of memory T cells in both twins, but effector T cells only in the ALS twin. The ALS twin's PBMC supernatants, but not the healthy twin's, were toxic to rat cortical neurons, and this toxicity was strongly inhibited by an IL-6 receptor antibody (tocilizumab) and less well by TNF-α and IL-1β antibodies. The putative neurotoxicity of IL-6 and TNF-α is in agreement with a high expression of these cytokines on infiltrating macrophages in the ALS spinal cord. We hypothesize that higher macrophage abundance and increased neurotoxic cytokines have a fundamental role in the phenotype and treatment of certain individuals with ALS.—Lam, L., Chin, L., Halder, R. C., Sagong, B., Famenini, S., Sayre, J., Montoya, D., Rubbi L., Pellegrini, M., Fiala, M. Epigenetic changes in T-cell and monocyte signatures and production of neurotoxic cytokines in ALS patients. FASEB J. 30, 000–000 (2016). www.fasebj.org

KEY WORDS: inflammation · ALS · SOD-1

Neuronal death in the amyotrophic lateral sclerosis (ALS) spinal cord is related to intrinsic neuronal mechanisms related to pathogenic proteins that are aggregated, modified, and mislocalized into the cytoplasm, and immune mechanisms stimulated by these proteins. The pathogenic proteins include superoxide dismutase (SOD)-1, transactive response (TAR) DNA-binding protein (TDP)-43, and RNA-binding protein fused in sarcoma/translocated in liposarcoma (FUS) (1,2), as well as RNA-peptide aggregates (3). TDP-43 becomes ubiquitinated in sALS and frontotemporal dementia (4). RNA-binding proteins have prion-like domains important in assembly of stress granules, as well as in pathologic protein aggregation and intercellular propagation (5). TDP-43 proteinopathy is associated with antisense RNA foci in the motor neurons of patients with C9ORF72-ALS (6).

Since the 1990s, the pathophysiology of ALS has been thought to involve the participation of immune mechanisms (7–10), including autoimmunity (11). The roles of non-neuronal cells, including microglia (12), astrocytes (13, 14), and oligodendrocytes (15), have been described in animal models. However, in the spinal cord in human sporadic ALS (sALS), macrophages appear to assume the canonical role of microglia in animal models. The postmortem ALS spinal cord is densely infiltrated by large macrophages visibly exiting from the vessels and morphologically distinct from ramified microglia (8), thus suggesting their blood origin. The immigration of blood-borne macrophages into the CNS is well known in other human neuroinflammatory disorders, such as HIV-1 encephalitis (16) and Alzheimer disease (17).

ABBREVIATIONS: ALS, amyotrophic lateral sclerosis; APL, acute promyelocytic leukemia; ChIP-seq, chromatin immunoprecipitation sequencing; CSF, cerebrospinal fluid; EGF, epidermal growth factor; ExAC, exome aggregation consortium; FBS, fetal bovine serum; GREAT, Genomic Regions Enrichment of Annotations Tool; HCP, high-CpG-density promoters; IGV, integrative genomics viewer; IOD, integrated optical density; MAF, minor allele frequency; MEM, minimal essential medium; MIF, migration inhibitory factor; MRC1, mannose receptor, CD206; Neutro, neutrophil; PBMC, peripheral blood mononuclear cell; PGRC, polycomb group repressor complex; RNA-seq, RNA sequencing; RRBS, reduced representation bisulfite sequencing; sALS, sporadic ALS; SDExp, signature delta expression; SOD, superoxide dismutase; TAR, transactive response

[1] Correspondence: UCLA School of Medicine, 100 UCLA Medical Plaza, Suite 200, Los Angeles, CA 90095-7022, USA. E-mail: fiala@mednet.ucla.edu

doi: 10.1096/fj.201600259RR
This article includes supplemental data. Please visit *http://www.fasebj.org* to obtain this information.

SOD-1 has a central role in sALS immunopathology through the induction of inflammatory cytokines by the aggregated form of its wild-type (18). SOD-1 aggregation is a prominent feature of ALS, occurring both *in vitro* and in the spinal cord of transgenic mice (19). Denatured SOD-1 has been found in granular inclusions colocalizing with lysosomes in spinal motoneurons of patients with sALS (20). Disease progression in a mouse model is related to the abundance of mutant SOD-1 (3). Oxidized SOD-1 has a similar gain of toxic functions as mutant SOD-1 (21, 22). Both mutant and wild-type SOD-1 inhibit axonal transport (23).

Various epigenetic mechanisms, including DNA methylation (such as hypermethylation of CpG islands in C9orf72 expansion (24), histone remodeling, abnormal miRNA biogenesis, and other silencing mechanisms have been described in sALS (25). In the CNS, changes in the expression of *C9orf72*, *MATR*, and *VEGFA* are present in affected regions (26). Transcriptional alterations in peripheral blood mononuclear cells (PBMCs) involve the genes *B2M*, *ACTG1*, *DYNLT1*, *SKIV2L2*, *C12orf35*, *TARDBP*, and *ILKAP* (27). *TBK1* was identified as an ALS gene linking autophagy of ubiquitinated proteins with inflammation (28).

Despite the diversity of molecular mechanisms in sALS, a common finding in the disease is an infiltration of the gray matter in affected spinal cord segments by macrophages, CD4 and CD8 T cells, and mast cells (18), demonstrating that both innate and adaptive immune mechanisms are operative in the pathologic course of ALS. Immunopathologic mechanisms include phagocytosis of apoptotic and nonapoptotic neurons by inflammatory macrophages (29), toxicity induced by granzyme-positive CD8 T cells (30), disruption of the blood–brain barrier by Th17 cells (31), and IL-6 trans-signaling (demonstrated to be toxic in a dose-related fashion in the mouse brain) (32). Neuroprotective function declines through inhibition of microglia and T cells by TGF-β (33), decrease in regulatory T cells (34), and lack of trophic factors (35). Blocking accumulation of misfolded SOD-1 in mitochondria by elevating the cytokine macrophage migration inhibitory factor (MIF) enhances neuronal survival (36). In addition, proteomic analysis of cerebrospinal fluid (CSF) samples of patients with sALS, in comparison to control CSF samples, revealed enrichment of proteins related to inflammation (in particular complement components) and decreased levels of proteins related to synaptogenesis and extracellular matrix organization (37).

A study of 5 monozygotic twin pairs discordant in ALS phenotype did not reveal nucleotide differences (38). Another study of monozygotic ALS-discordant twins with the *C9orf72* repeat expansion did not find epigenetic modification of the genome (39). In the current study, we investigated by reduced representation bisulfite sequencing (RRBS) the methylome of a monozygotic twin pair that was discordant in the diagnosis of ALS and inferred differences in blood cell type abundances and pathways. Moreover, we hypothesized that a downstream cause of neuronal demise in the affected twin involves the production by macrophages of neurotoxic cytokines stimulated by effector T cells.

MATERIALS AND METHODS

Patients and controls

The immunologic and epigenetic investigation of patients and rat neurons was approved by the University of California, Los Angeles Institutional and Ethics Review Board. The twin pair in the study were monozygotic females 50 yr of age. The ALS twin had onset of ALS in the right arm in the spring of 2011 and subsequently progressed to bulbar involvement, whereas the non-ALS twin was not affected by 2015. Two other patients with sALS are included in the study of neuronal toxicity: a 72-yr-old man and a 56-yr-old woman, both with bulbar onset and upper extremity weakness.

RNA sequencing

RNA-sequencing (RNA-seq) was performed on PBMCs by using standard RNA-seq library construction protocols (Illumina, San Diego, CA, USA). RNA-seq libraries were sequenced on the Illumina HiSeq 2000. Reads were aligned to the hg19 reference genome by using TopHat Johns Hopkins University, Baltimore, MD, USA; *https://ccb.jhu.edu/software/tophat/index.shtml*) (40). Gene counts were quantified with HTSeq and normalized with DESeq (41, 42).

Expression cell-type signatures

To characterize expression differences between the twins, their gene counts were analyzed using multiple cell-type signatures derived from a human tissue atlas (43). Frozen robust multiarray analysis (fRMA) was used for normalization of the microarray intensity values of the atlas (44). To identify signature genes for each cell type, we calculated the ratio of the median intensity value of a cell type to the median intensity value. The top 50 ranking genes by this metric defined the cell-type–specific expression signature. For each RNA-seq sample, a signature δ expression (SD_{Exp}) value between the twins was determined by calculating the arithmetic mean difference of the $\log_{10}$ gene count (X) of the sample to the mean $\log_{10}$ gene count ($\bar{X}$) across all genes within a signature.

$$SD_{\mathrm{Exp}} = \frac{1}{50}\sum_{1}^{50}(X_{\mathrm{i}} - \bar{X}_{\mathrm{i}})$$

RRBS

Genomic DNA for making RRBS libraries was extracted from PBMCs according to the standard protocol (45). Genomic DNA was digested with a methylation-insensitive endonuclease, *Msp*I. Fragments of 40–220 bp were isolated, because they are enriched for CpG-rich regions, such as CpG islands, promoter regions, and enhancer elements. The *Msp*I-digested DNA was end repaired, A tailed, and ligated with Illumina adaptors. The double-stranded DNA was denatured, followed by bisulfite conversion and PCR amplification. These libraries were sequenced with Ilumina HiSeq 2000 sequencers (San Diego, CA, USA). The reads were aligned to the reference genome (human hg19) using the modified bisulfite aligner BS Seeker2 (46), and only the uniquely mapped reads were kept. We calculated the methylation level for each cytosine on the genome. Because bisulfite treatment converted unmethylated cytosines (Cs) to thymines (Ts), the methylation level at each cytosine were estimated by #C/(#C+#T), where #C is the number of methylated reads and #T is the number of

unmethylated reads. In this study, only cytosines that were covered by at least 4 reads in all samples were included in the subsequent analysis.

Epigenetic cell-type signatures

Cell-type methylation signatures were derived from the data generated by the Blueprint epigenome project (*http://www.blueprint-epigenome.eu/*) (47). For each cell type, we identified CpG sites that were uniquely hypomethylated compared to the other cell types. The criteria were that these CpG sites had methylation levels <0.3 in the specific cell types of interest, and the methylation was at least 0.5 higher in all other cell types. The Mann-Whitney *U* test with Bonferroni correction was used to compare the median signature values between the samples. Based on our selection criteria, we established B-cell, NK-cell, T-cell, neutrophil (Neutro), and $CD14^+$/monocyte (CD14) signatures that consist of 551, 463, 217, 324, and 184 CpG sites respectively (Supplemental Fig. 1*A*).

Differentially methylated RRBS fragments

Visualization of differentially methylated fragments was performed with the Integrative Genomics Viewer (IGV) (48) (Broad Institute, Cambridge, MA, USA; *http://www.broadinstitute.org/igv/*). RRBS fragment methylation levels were determined by calculating the mean CpG methylation levels within each fragment. Differentially methylated fragments between the twin samples were restricted to regions that contained at least a single CpG site, with a Fisher's exact test $P < 0.05$, and within 100 kb of a gene transcription start site. The fragments were ranked by the difference in fragment methylation between the twin samples.

Genomic Regions Enrichment of Annotations

We tested for enrichment of annotations of genes associated with the top and bottom 1000 differentially methylated fragments using Genomic Regions Enrichment of Annotations Tool (GREAT), and report terms with a false-discovery rate (FDR) < 0.05 from the MSigDB perturbation and MSigDB immunologic signatures (Broad Institute; *http://software.broadinstitute.org/gsea/msigdb/index.jsp/*) (49). Fragments were ranked by the difference in methylation level between samples. Fragment-to-gene associations were determined based on the single nearest gene within 100 kb, with the whole hg19 genome set as the background.

Genetic screening

DNA samples from the twins were screened using targeted sequencing of genes previously implicated in a series of neurodegenerative disorders, including the ALS genes *TARDBP*, *FUS*, *SQSTM1*, *VCP*, *OPTN*, *SOD1*, *CHCHD10*, *ALS2*, *MATR3*, *VAPB*, *PFN1*, *TAF15*, *HNRNPA1*, *HNRNPA2B1*, and *DCTN1*. Exonic regions for these genes were captured with a custom designed library and sequenced on an Illumina HiSeq 2500. Sequence reads were mapped to the GRCh37/hg19 reference genome, and variants were interactively joint-called with GATK Haplotype Caller according to GATK Best Practices recommendations (*https://www.broadinstitute.org/gatk/*) (50). The joint variant calling file was annotated with refGene, dbSNP138, functional effect prediction tools, as well as 1000 Genomes, Exome Variant Server and Exome Aggregation Consortium (ExAC) MAFs, according to Annovar (*http://annovar.openbioinformatics.org/*). In addition, all samples were screened for the hexanucleotide repeat of *C9orf72*, using both fluorescent and repeat-primed PCR (51).

Multiplex cytokine assays

The assays included multiplexed immunometric assay panels (Luminex platform; R&D Systems, Minneapolis, MN, USA) for the human cytokines IL1-β, IL-6, IL-8, IL-10, GM-CSF, and TNF-α and TGF-β1, -β2, and -β3 (R&D Systems). The Luminex xMAP system tests multiple analytes simultaneously, using spectrally addressed bead sets, each of which is conjugated with a different capture monoclonal antibody specific for a given target molecule. The antibody-conjugated beads are allowed to react with the sample and a secondary detection antibody, to form a capture sandwich immunoassay. After the assay is completed, the assay solution is drawn into the Bio-Plex 200 Luminex array reader (Bio-Rad, Hercules, CA, USA) to quantify the amount of each of the analytes.

Memory cells analysis

PBMCs were isolated from the healthy donors or patients with ALS by using a Ficoll-Hypaque (GE Healthcare, Pittsburgh, PA, USA) gradient method. PBMCs were surface-stained using anti-CD3 and anti-human CD127 PE, anti-human CD45RA PerCp-Cy5.5, or antihuman CD45RO APC and were analyzed with FlowJo software (TreeStar, Ashland, OR, USA) with lymphocyte gating, based on forward and side scatter.

Antibodies

Anti-human CD127 PE, anti-human CD45RA PerCp-Cy5.5, and anti-human CD45RO APC (Tonbo Biosciences, San Diego, CA, USA); anti-human Foxp3PerCp.Cy5.5 (eBioscience, San Diego, CA, USA); neutralizing IgG1 monoclonal antibody to TNF-α (InvivoGen, San Diego, CA, USA); neutralizing monoclonal human IL-1α antibody (R&D Systems); anti-IL6 receptor α antibody (tocilizumab; Actemra; Genentech, South San Francisco, CA, USA); anti-Neu-N and anti-MAP-2 (EMS-Millipore, Billerica, MA, USA); and goat anti-rabbit Alexa 488 and goat anti- mouse Alexa 555 (Thermo Fisher Scientific Life Sciences, Carlsbad, CA, USA).

Rat primary cortical neurons

All animal use protocols were approved by the University of California, Los Angeles, Chancellor's Animal Research Committee, and were in compliance with the NIH guidelines. Rat pups (postnatal d 0–1) were euthanized by decapitation, and cortices were dissected on ice in dissection medium [minimum essential medium (MEM)] supplemented with streptomycin (100 μg /ml), penicillin (100 U/ml) and 0.01 M HEPES (pH 7.5). The cortical tissues were digested for 15 min with 1 mg/ml papain, and cells were mechanically dissociated in MEM supplemented with 10% fetal bovine serum (FBS), 0.5% glucose, 2 mM Glutamax (Thermo Fisher Scientific Life Sciences). Cells were then centrifuged for 5 min at 145.2 *g*, resuspended in feeding medium (MEM supplemented with 5% FBS, 0.1 mg/ml transferrin, 2% B27, 0.5% glucose, 24 mg/L insulin, and 2 mM Glutamax) and plated on poly-D-lysine-coated coverslips (0.05 mg/ml) at a density of $4.5 \times 10^4/cm^2$. Neurons were incubated at 37°C and 5% CO^2 and, 72–96 h after plating, feeding medium was supplemented with 2 μM cytosine and β-D-arabinofuranoside to avoid astrocyte proliferation.

Testing of toxicity to rat neurons

Neurons were cultured on coverslips in 24-well plates. Eight-day *in vitro* neurons were treated with the PBMC supernatants at the

ratio of 3 μl of the supernatant in 300 μL of the media for 6 h. Anti-IL-1α, anti-TNF-α,, and anti-IL-6R were used in the inhibition studies at a concentration of 2 μg/ml.

Immunofluorescence

Neurons were washed with PBS, fixed with 4% paraformaldehyde, and, after permeabilization with 0.1% Triton X-100 (3 min) and blocking (30 min) with PBS containing 1% FBS and 1% BSA, were immunostained in blocking solution with antibodies against MAP-2 (green dendrites) and NeuN (red perinuclear cytoplasm). Three images were taken from the top, middle, and bottom of each coverslip using the Olympus research microscope with ×40/0.75 objective and a Hamamatsu camera (Hamamatsu City, Japan). The pictures were scanned with Image-Pro software (MediaCybernetics, Rockville, MD, USA).

Statistics

The differences in means and their 95% confidence limits (mean ± 2 SEM) of fluorescence intensities were analyzed in 2 experiments with 2 different supernatants by repeated-measures ANOVA pairwise comparison, with adjustment for multiple comparisons. Computations were performed with SPSS for Windows, ver. 20.0 (IBM, Armonk, NY, USA).

RESULTS

Genetic screening

To demonstrate lack of genetic differences between the twins discordant for ALS, the coding regions of 15 genes previously implicated in ALS were screened in both individuals. Both individuals shared common (ExAC MAF > 30%) coding variants within the *ALS* (2 synonymous and 1 nonsynonymous), *FUS* (2 synonymous), *OPTN* (1 nonsynonymous), *SQSTM1* (2 synonymous), and *CHCHD10* (2 synonymous) genes. They also shared one rare (ExAC MAF = 0.8%), nonsynonymous *DCTN1* variant (NM [lowen]001135041: c.A184G, p.I62V) predicted to be benign. These individuals carried repeats within the normal range for the *C9orf72* hexanucleotide repeat.

Analysis of cell types in monozygotic twins

We used the RRBS technique to profile the methylomes of the monozygotic twins. Blood specimens were obtained from the ALS twin in November 2013 and August 2014 and from the healthy twin in November 2013 and December 2014. The RRBS approach allowed us to measure ~2 million CpG sites in the genome, representing ~7% of

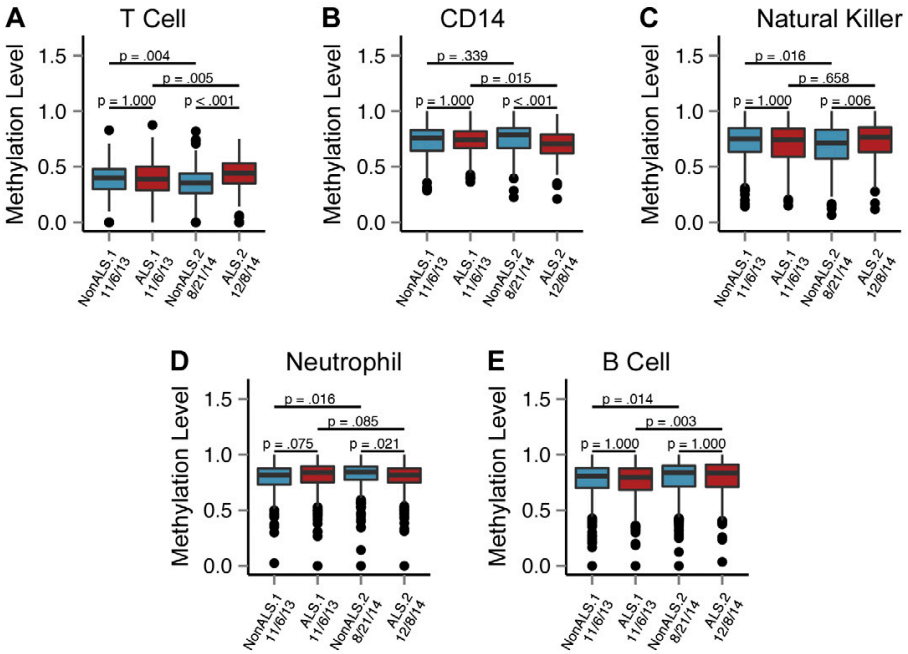


Figure 1. Methylation distribution at cell-type–specific CpG sites in the PBMC samples from ALS-discordant twins. Methylation-level box plots for the PBMC samples obtained in 2013 (box 1: November 6, 2013; box 2: November 6, 2013) and 2014 (box 3: August 21, 2014; box 4: December 8, 2014) at T-cell, $CD14^+$, NK, neutrophil, and B-cell-specific CpG sites. Blue: healthy twin; red: ALS-affected twin. The inner band of the box plot marks the median methylation value of the of the sample, with the lower and upper bounds of the box marking the first and third quartiles. The Bonferroni-adjusted *P* values were calculated to test for differences in the methylation level distribution between samples; the Mann-Whitney *U* test was used to analyze the differences between the twins in the 2013 and 2014 samples and in the batches from the unaffected and affected twins. Methylation distribution is shown at these CpG-specific sites: *A*) T-cell–specific, (*B*) CD14-specific, and (*C*) NK-specific CpG sites with adjusted *P* values. *D*) Neutrophil-specific and (*E*) B-cell–specific CpG sites.

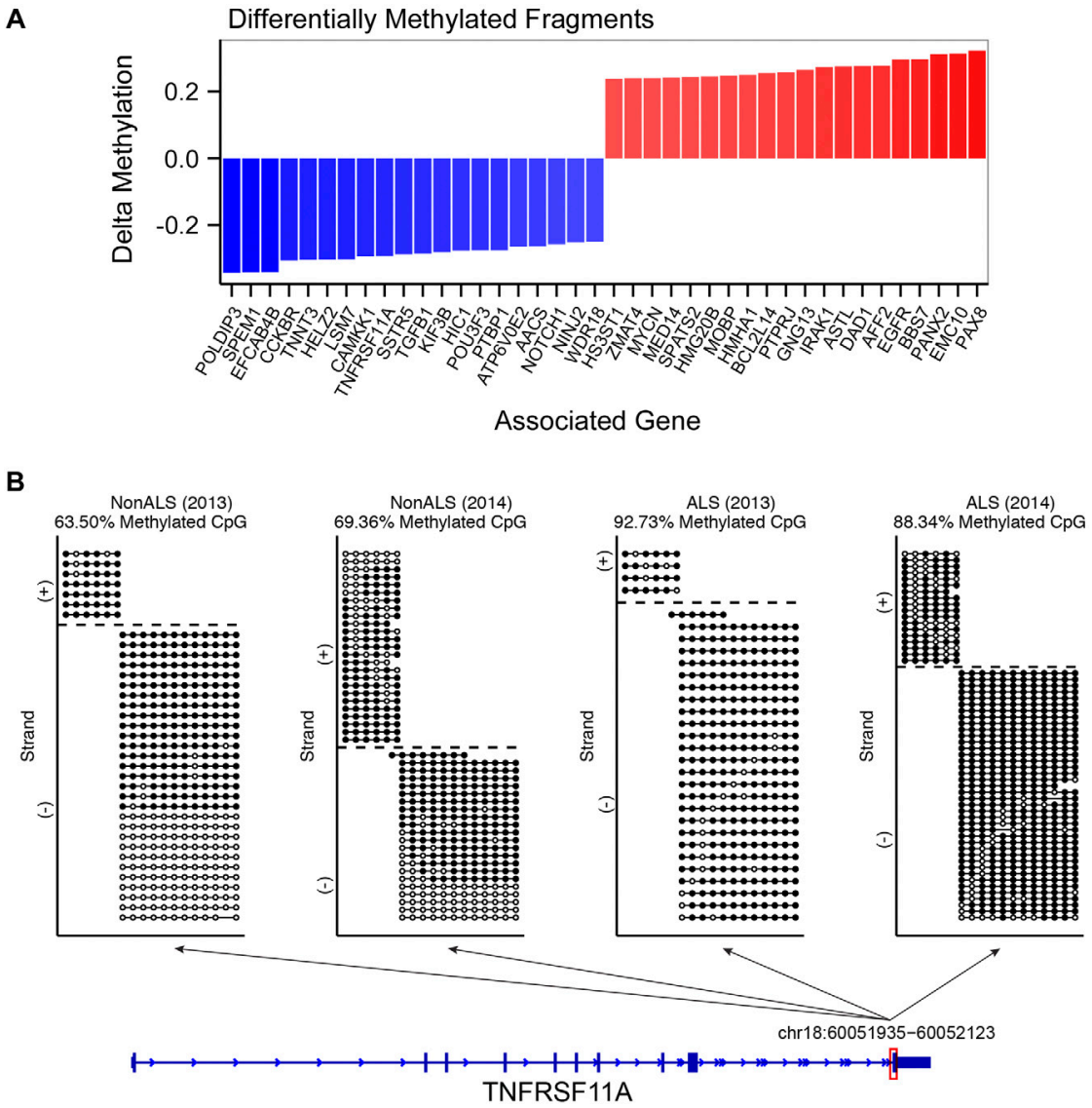


Figure 2. Differentially methylated RRBS fragments and associated genes in the PBMC samples from ALS-discordant twins. *A*) Delta methylation of the top 20 and bottom 20 fragments within 100 kb of a reference gene (RefGene) transcription start site. The top 20 and bottom 20 fragments were determined by pooling the 2013 and 2014 PBMC RRBS sample reads by individual. Individual CpG sites were filtered for a minimum coverage of 10 times common across the 4 samples (2,168,334 sites). The CpG methylation counts were pooled within replicates, and fragment-level methylation was determined by taking the mean CpG methylation within the fragment. The fragments were filtered for fragments with 1 or more CpGs that are significantly different in methylation level between the twins with a Fisher's exact test, $P < 0.05$. The fragments were ranked by the difference in fragment methylation (Δ methylation) of the affected twin to the unaffected twin and contained a minimum of a single CpG, which tested for differential methylation between the twins with a 2-sided Fisher's exact test, $P < 0.05$. Differentially methylated CpG sites were subsequently mapped to the mappable RRBS fragments. *B*) Visualization of the aligned RRBS reads to the fragment chr18:60,051,935–60,052,123 that spans exon 10 of the *TNFRSF11A* gene. The dotted lines in the non-ALS (2013) plot represent individual reads that aligned to the chr18 fragment within the *TNFRSF11A* gene body. The circles on each line represent individual CpG sites along the read. Filled circle: methylated CpG; open circle: unmethylated CpG. Within the fragment, the non-ALS (2013), non-ALS (2014), ALS (2013), and ALS (2014) samples had 63.5, 69.36, 92.73, and 88.34% of the aligned CpGs methylated, respectively.

the total. Because of the sequence preference of the MSPI enzyme used to digest the genome, we preferentially selected sites that are within CpG islands, promoters, and enhancers. We only kept sites with at least 10× the coverage across all samples, leaving us with a total of 2,168,334 sites.

Two effects can cause DNA methylation differences between the twins: *1*) changes in cell type composition

TABLE 1. *DNA Methylation differences between the twins related to the H3K27me3 marker*

Ontology	Enriched terms of differentially methylated fragments (2013)	Binomial FDR *q*	Observed gene hits (*n*)	Total genes (*n*)
Low methylation in ALS				
MSigDB perturbation	Genes up-regulated in NB4 cells (APL) in response to tretinoin [PubChem 444795]; based on ChIP-seq data	5.11E-30	111	781
MSigDB perturbation	Genes with HCP bearing a histone H3 trimethylation marker at K27 (H3K27me3) in MEF cells (embryonic fibroblast)	5.30E-23	87	573
High methylation in ALS				
MSigDB perturbation	Genes up-regulated in NB4 cells (APL) in response to tretinoin [PubChem 444795]; based on ChIP-seq data	2.61E-33	101	781
MSigDB perturbation	Genes with HCP bearing the trimethylation marker, at H3K27 (H3K27me3) in MCV6 cells (embryonic fibroblasts trapped in a differentiated state)	1.51E-27	70	418

Top and bottom 1000 differentially methylated fragments ranked by the difference in fragment methylation between the ALS twin and the unaffected twin 2013 samples were subjected to GREAT enrichment testing in the MSigDB perturbation and MSigDB immunologic signatures. The 5 most enriched terms for each ontology are reported, with a minimum FDR < 0.05.

in their PBMCs, and 2) changes in the methylation of specific cell types in activated *vs.* resting states. We first analyzed the changes that were related to differences in cell-type composition between the twins, by using the methylome data collected for different hematopoietic cell types by the Blueprint Consortium (Supplemental Fig. 1*A*). In brief, for each cell type, we identified the sites that fell within RRBS fragments that were uniquely demethylated in one cell type with respect to all the others. These signatures were then used as a proxy for the abundance of those cell types within the PBMCs as follows: lower average signature methylation implies that the cell type is more abundant, whereas higher levels indicate lower abundance.

We measured the differences in the cell-type signature in the twin PBMC samples collected in 2013 and 2014. We observed that the cell-type signatures were not significantly different between the twins in the 2013 samples, but the difference became significant in 2014 (**Fig. 1**). Specifically, we found that T- and NK-cell signatures were significantly higher in the ALS twin, indicating that these cell types were less abundant in the affected twin. In contrast, we found that the CD14 monocyte signature was significantly lower in the ALS twin ($P < 0.001$), suggesting that the affected twin had significantly higher levels of monocytes. We did not detect significant changes in the neutrophil and B-cell signatures, suggesting that these cell types did not vary between the twins (Fig. 1).

Differentially methylated RRBS fragments include immune-related regions

Considering that DNA methylation changes may be caused by activation of immune cells, we also directly compared the methylomes of the twins, to identify the regions with the most significant differential methylation. Using this approach, we identified 20 top and bottom fragments of the pooled samples ranked by the difference in fragment methylation (**Fig. 2*A***). The genes proximal to the differentially methylated fragments included immunity-related genes, such as *EGFR*, *TGFb1*, and *TNFRSF11A*. Specifically, we found that the fragment on chr18, at position 60,051,935–60,052,123, and 59,481 bp downstream of the *TNFRSF11A* transcription start site, had an increase in fully methylated reads in the ALS twin (Fig. 2*B*). On the contrary, the fragment on chr7 (55,073,097–55,073,297), which is 13,597 bp upstream of the *EGFR* transcription start site, had an increase in unmethylated reads in the ALS twin compared to the unaffected twin at both the 2013 and 2014 samples (Supplemental Fig. 3*A*). The fragment on chromosome 19 (41,831,930–41,832,049) had an increase in methylated reads in the ALS-affected twin compared with the unaffected twin at both the 2013 and 2014 samples (Supplemental Fig. 3*B*). An unpooled analysis intersecting the top 50 differentially methylated fragments high in ALS from 2013 and 2014 identified fragments associated with *GNG13*, *VPS53*, *EMC10*, *PAX8*, *TRPC3*, and *EGFR* as consistently differentially methylated in both the 2013 and 2014 samples (Supplemental Fig. 4*A*). Similarly, the intersection of the bottom 50 differentially methylated fragments low in ALS from 2013 and 2014 identified fragments associated with *CCKBR*, *TNFRSF11A*, *TGFB1*, *POFUT1*, and *GMEB2* to be consistently differentially methylated in both 2013 and 2014 samples (Supplemental Fig. 4*B*).

Specific pathways associated with epigenetic changes differentiate the twins

To determine whether the genes that are proximal to the differentially methylated regions are associated with

TABLE 2. *MSigDB immunologic signatures associated with differences between blood cell types*

Ontology	Enriched terms of differentially methylated fragments (2014)	Binomial FDR q	Observed gene hits (n)	Total genes (n)
Low methylation in ALS				
MSigDB perturbation	Genes up-regulated in NBA cells (APL) in response to tretinoin [PubChem 444795]; based on ChIP-seq data	3.39E-08	124	781
MSigDB perturbation	Genes up-regulated in the human mammary epithelial cells upon expression of TP53 [GeneID 7157] off adenoviral vector	3.30E-08	156	1065
MSigDB immunologic signatures	Genes down-regulated in comparison of monocytes treated with 1 ng/ml LPS (TLR4 agonist) *vs.* monocytes treated with vehicle.	2.25E-03	38	192
MSigDB immunologic signatures	Genes down-regulated in comparison of healthy CD4 [GeneID 920] T cells *vs.* healthy myeloid cells	1.44E-03	38	197
MSigDB immunologic signatures	Genes down-regulated in comparison of monocytes treated with anti-TREM1 [GeneID 54210] *vs.* monocytes treated with vehicle	1.28E-03	38	193
MSigDB immunologic signatures	Genes down-regulated in comparison to healthy B cells *vs.* healthy myeloid cells	1.88E-02	35	198
MSigDB immunologic signatures	Genes up-regulated in comparison of dendritic cells *vs.* central memory CD4 [GeneID 920] T cells	2.64E-02	34	197
High methylation in ALS				
MSigDB perturbation	Genes up-regulated in NB4 cells (APL) in response to tretinoin [PubChem 444795]; based on ChIP-seq data	1.59E-08	111	781
MSigDB perturbation	Genes with high-CpG-density promoters (HCP) bearing the trimethylation marker at H3K27 (H3K27me3) in MCV6 cells (embryonic fibroblasts trapped in a differentiated state)	2.58E-08	71	418
MSigDB perturbation	Genes within amplicon 16p13 identified in a study of 191 breast tumor samples.	5.91E-03	22	110
MSigDB perturbation	Genes up-regulated in comparison of naive CD4 [GeneID 920] T cells *vs.* d 0 monocytes	4.01E-02	30	187

Top and bottom 1000 differentially methylated fragments ranked by the difference in fragment mediation between the ALS twin and the unaffected twin 2014 samples were subjected to GREAT for enrichment testing in the MSigDB perturbation and MSigDB immunologic signatures. The 5 most enriched terms for each ontology are reported, with a minimum FDR < 0.05.

specific pathways or functional terms, we used the GREAT tool as previously described. The differentially methylated fragments in the 2013 ALS twin's samples were associated with the genes that are up-regulated in acute promyelocytic leukemia (APL) cells in response to tretinoin (retinoic acid or vitamin A) and with the genes associated with H3K27me3 markers (**Table 1**). The differentially methylated fragments in the 2014 samples also include both tretinoin terms and H3K27me3. However, the differentially methylated fragments from the 2014 samples also include gene sets from MSigDB immunologic signatures that are associated with differences between blood cell types (*e.g.*, genes down-regulated in comparison to monocytes treated with anti-TREM-1 *vs.* monocytes treated with a vehicle) (**Table 2**), reinforcing the hypothesis that a portion of the differences in methylation is because of the changes in cell-type abundances discussed above.

Differential mRNA expression supports cell-type abundance changes in PBMCs

Along with the extraction of DNA for RRBS analysis, we also extracted RNA from the PBMCs of both the twins. Using the rationale presented above, the differences in gene expression between the twins may be caused either by the

changes in cell-type abundances or by the changes in the activation state of the immune cells. We investigated these possibilities by gathering gene expression signatures that were specific to individual cell types as previously described. For each sample we computed the average log expression of the genes within the signatures: a higher score implied a higher abundance of that cell type and a lower score signified a lower abundance. We compiled signatures several of immune cells in both resting and activated states.

To interpret the differences in gene expression between the twins, we examined the difference in the average signature expression value between the affected and unaffected twins across multiple cell-type signatures (**Fig. 3*A***). We found that among the 13 expression signatures, only the dendritic cell signature reveals a significant difference between the twins.

We also examined the most differentially expressed genes between the twins (Fig. 3*B*). We observed that the genes with both significant fold changes and *P* values for those changes are strongly enriched for chemokines and metalloproteases (**Table 3**).

PBMCs of the ALS twin produce high concentrations of IL-6, TNF-α, and IL-1

The alterations in monocyte and T-cell abundances between the twins suggest that altered immune cell interactions in the ALS twin could explain the increased cytotoxicity by macrophages in the ALS spinal cord. To investigate this hypothesis, PBMCs of the twins were incubated overnight, and cytokines were measured in the supernatant. The PBMCs of the ALS twin, but not those of the non-ALS twin, spontaneously produced the inflammatory cytokines IL-1, IL-6, and TNF-α in the supernatant in nanogram per milliliter concentrations (Fig. 4*B*). In addition, IL-6, in nanogram per milliliter concentrations, was produced by macrophages of the ALS twin (Fig. 4C). Stimulation by SOD-1 increased cytokine production by non-ALS twin's PBMCs to 50% level in the ALS twin, and stimulation by SOD-1 in the presence of the mismatched ALS twin's serum strongly potentiated the cytokine production. Stimulation by SOD-1 and tocilizumab increased IL-6 production by PBMCs of both twins. Only picogram per milliliter concentrations of IL-10 and -17 were produced in PBMCs after overnight incubation.

Memory and effector T cells in the ALS twin

The strong spontaneous production of inflammatory cytokines by the ALS twin's, but not the healthy twin's, PBMCs suggested the presence of effector T cells only in the ALS twin. Flow cytometric testing showed a higher ratio of CD45RO/RA and a lower proportion of the CD127 memory T cells in the ALS twin (**Table 4**).

Neurotoxic cytokines in the ALS PBMC supernatants

Given the high concentrations of inflammatory cytokines produced by the ALS twin's PBMCs and macrophages, we investigated induction of neuronal toxicity by PBMC

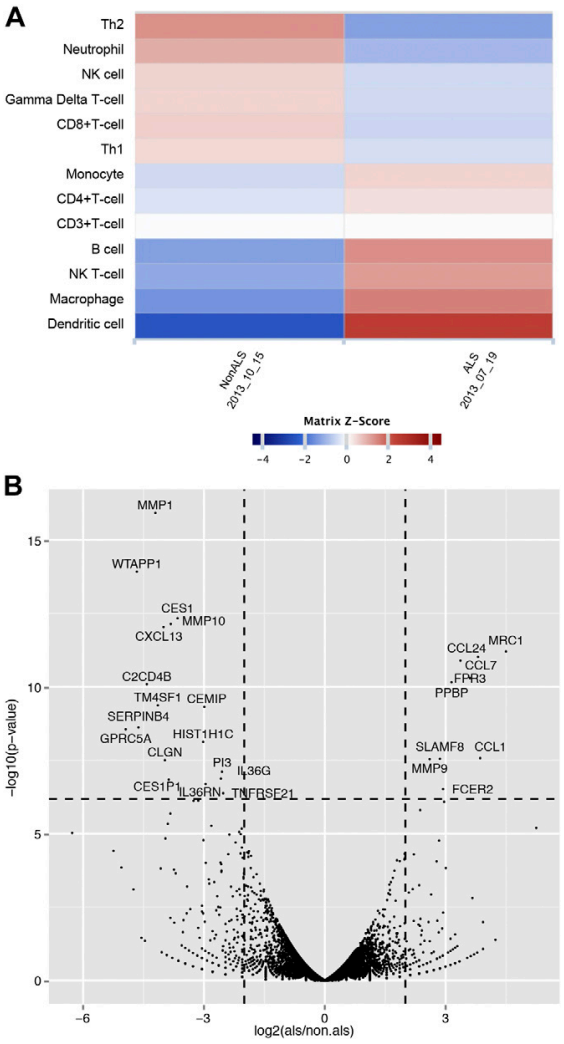


Figure 3. Differences in expression signatures between the twins. *A*) Heat map of the *z* scores for the signature Δ expression levels. Left column: signature Δ *z* score for the non-ALS PBMC sampled on October 15, 2013; right column: signature Δ *z* score for the ALS PBMCs sampled on July 19, 2013. Red: higher signature expression level *vs.* the mean Δ. Blue: lower signature expression level *vs.* the mean Δ. Among the 13 expression signatures, the dendritic cell signature *z* score reveals a significant difference from the matrix mean. *B*) A volcano plot of the 15,152 normalized gene count. Differences between the twins in gene expression are shown as $\log_2$ on the horizontal axis. The significance of differentially expressed genes is shown on the vertical axis, with a minimum Bonferroni corrected $-\log_{10}$, $P = 6.18$.

supernatants. In an *in vitro* assay of rat cortical neurons, the supernatants of the ALS twin's PBMCs, but not of the non-ALS twin's, caused attrition of neurites revealed by diminished staining with MAP-2 antibody and increasing neuronal destruction. The results of 2 experiments in 2 different rat neuronal cultures with 2 different ALS and non-ALS PBMC supernatants showed destruction of neurons exposed to the ALS twin's supernatants, but

TABLE 3. *Comparison between most differentially expressed genes in the twins*

Category	Terms associated with differentially expressed genes	Count	List total	FDR
UP_SEQ_FEATURE	Disulfide bond	16	21	2.73E-06
SP_PIR_KEYWORDS	Chemotaxis	6	21	9.80E-06
INTERPRO	Small chemokine, IL-8-like	5	21	1.42E-04
GOTERM_MF_FAT	Chemokine activity	5	16	1.70E-04

not the nonaffected twin's supernatants (**Fig. 5*A***). The difference in MAP-2 mean fluorescence intensities between the neurons exposed to the twin supernatants was analyzed by pairwise comparison by using repeated-measures ANOVA with adjustment for multiple comparisons with significant results ($P < 0.001$) (Fig. 5*B*). In another experiment, the effects of antibodies to cytokines on neurotoxicity by the ALS twin's supernatant were compared. Tocilizumab, the blocking antibody to the IL-6 α receptor, produced the strongest protection, and IL-1β

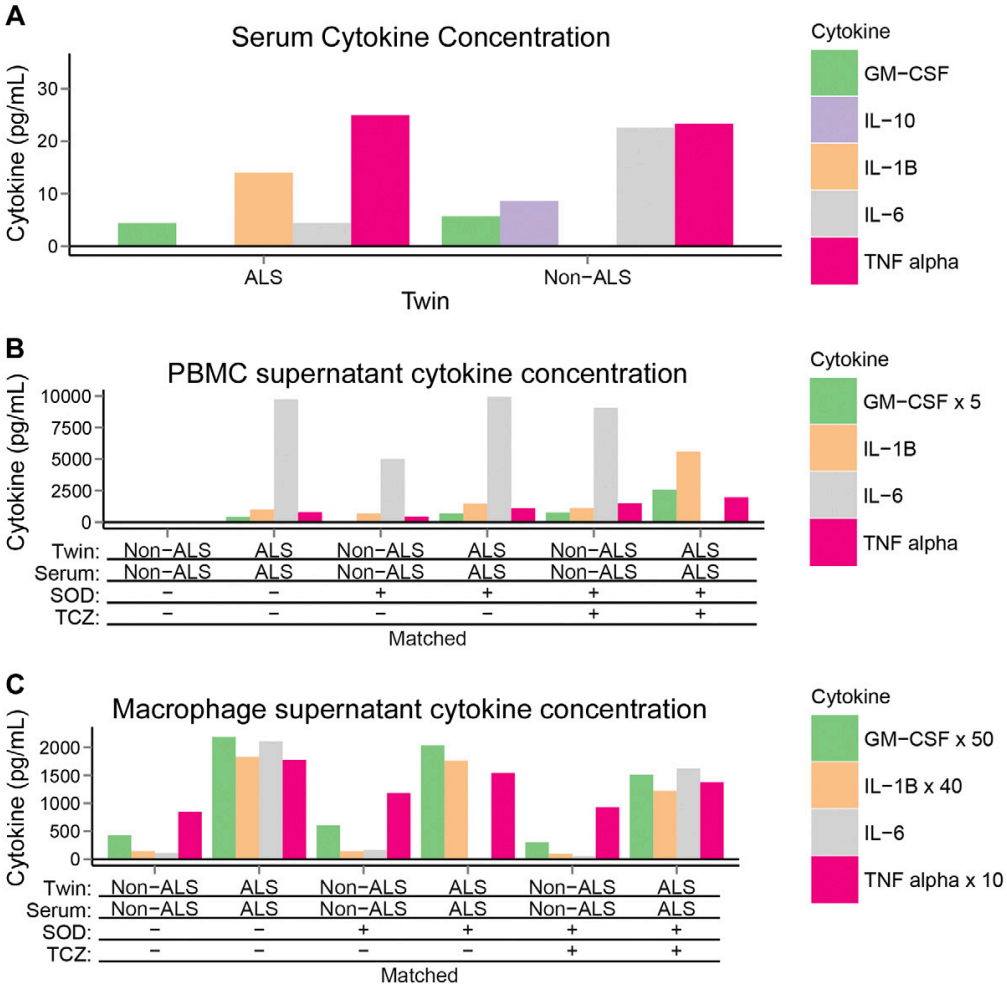


Figure 4. Cytokine concentrations in the twins. Cytokines in the 2014 twin samples. *A*) Serum cytokines: only picograms per milliliter concentrations of cytokine with equal concentrations of IL-6 in both twins. *B*) PBMC supernatant cytokines: up to 10 ng/ml concentrations of IL-6 in PBMC supernatants of the ALS twin in the absence or presence of SOD-1 or tocilizumab with matched or with mismatched serum; substantial IL-6 concentrations in the PBMC supernatants of the non-ALS twin were present only with SOD-1 stimulation; other cytokines, GM-CSF, IL10, IL-1β and TNF-α showed similar differences. *C*) Macrophage supernatant cytokines: up to 2 ng/ml of IL-6 in the supernatants of the ALS twin's macrophages, with or without SOD-1 or tocilizumab stimulation with matched or with mismatched serum; no IL-6 produced by non-ALS macrophages; and TNF-α, IL-10, and IL-1β were produced in negligible amounts.

TABLE 4. *Memory vs. effector T cells in the twins*

Patient	Ratio of CD45RO:CD45RA	CD127 (%)
ALS twin	36.5/20 = 1.8	32.4
Healthy twin	63.4/46.6 = 1.3	63.5

and TNF-α antibodies afforded lower protection [all pairwise comparisons by ANOVA repeated-measures after adjustment for multiple comparisons were significant ($P < 0.001$)]. Finally, neuronal toxicities of PBMC supernatants from 2 other ALS patients were tested and compared to the twins' supernatants. One patient had toxic supernatant with activity comparable to the ALS twin's. The other patient with ALS did not have neurotoxic supernatant, as the activity of his activity was equal to that of the non-ALS twin's in the same experiment.

DISCUSSION

Twins with the same genetic background, including common coding variants implicated in ALS, but different ALS phenotypes, differ in the abundance of blood cell types and regulation of inflammatory signaling resulting in increased production of neurotoxic cytokines from the ALS twin's PBMCs and macrophages. The neurotoxicity of ALS twin's supernatants was inhibited by IL-6 receptor and TNF-α antibodies.

We used a novel approach to estimate cell-type abundances using DNA methylation data. This method determines the methylation of specific cytosines that are found to be demethylated in only one hematopoietic cell type and hypermethylated in all others. Using this approach, we found that the ALS twin had comparable T-cell and monocyte composition when compared with the unaffected twin in 2013. The T-cell composition decreased as the monocyte composition increased over the course of the disease by 2014 (Fig. 1). These changes may account for altered interactions between T cells, macrophages, and dendritic cells that could lead to production of the inflammatory cytokines IL-1β, IL-6, and TNF-α (Fig. 4). The testing of transcriptional signature by RNA-seq largely supports the conclusion from the DNA methylation data that the affected twin loses T cells and gains monocytes during the course of the disease when compared to the unaffected twin (Fig. 3).

We also observed that the DNA methylation differences between the twins involved genes related to cell signaling, including hypermethylation of *TNFRSF11A* in the ALS twin and hypomethylation of *EGFR* in the ALS twin (Fig. 2). *TNFRSF11A* encodes a member of the TNF-receptor superfamily named tumor necrosis factor receptor superfamily member 11A, which interacts with various adaptor TRAF family proteins and induces activation of NF-κB and MAPK/JNK inflammatory pathways (52, 53). *EGFR* encodes a glycoprotein cell-surface receptor for members of the epidermal growth factor (EGF) family. The downstream signaling proteins from this receptor initiate several signal transduction cascades, principally the RAS-RAF-MEK-ERK, Akt, PLC-γ, and STATs pathways. EGFR inhibitors influence adaptive immune responses by altering immune gene expression (54), and EGFR participates in the down-regulation of immune responses by $Foxp3^+$ regulatory T cells (55). *EGFR* has an inhibitory function on neurite outgrowth, which could be important in the ALS twin (56).

We found that the pathways associated with genes that are proximal to DNA methylation differences between the twins were related to the gene sets "response to tretinoin (vitamin A)" and "H3K27me3 marks" (Table 1). This finding suggests that the activities of both vitamin A and the histone methyltransferase polycomb group repressor complex (PGRC)-2 could be increased in the ALS twin. PGRC2 is associated with the methylation of H3K27me3 and repression of transcription, typically during developmental processes or cellular differentiation (57). It has been shown that vitamin A induces *MMP9* in macrophages and increases macrophage phagocytosis (58). Indeed, *MMP9* was up-regulated in the ALS twin (Fig. 3*B*) who also displayed increased transcription of the genes related to immune cell migration and activation, including the metalloproteinase MMP9, the chemokines CCL1 and -7 that attract monocytes, the receptor FPR3 for monocytes, and the receptor on "proresolving" macrophages MRC1 (mannose receptor, CD206).

In animal models, the SOD-1-stimulated microglia-derived neurotoxic factors have been largely unidentified but are thought to involve excessive glutamate release (59). In the current study, we have found production of neurotoxic levels of inflammatory cytokines, in particular IL-6 and TNF-α, by the ALS twin's PBMCs and macrophages. The levels of IL-6 produced by macrophages were 5 times lower than those produced by PBMCs, suggesting that the interactions between effector T cells and macrophages are necessary for high IL-6 production. Tocilizumab increased IL-6 production, because it inhibits internalization of IL-6 (60). That finding notwithstanding, tocilizumab blockade of IL-6 signaling is anticipated to have therapeutic benefits in a current clinical trial of the drug.

The cells of the healthy twin required stimulation by SOD-1 for production of these cytokines, suggesting the presence of memory but not effector T cells. The PBMCs of the ALS twin had a higher CD45RO/RA ratio (Table 3), suggesting increased central and effector memory T-cell subset (61) and decreased CD127 proportion, indicative of an increased proportion of effector T cells (62) in the ALS twin. Macrophages of sALS patients (differentiated *in vitro* in autologous serum) showed increased inflammatory activation with progression of the disease (data not shown) compatible with stimulation by effector T cells.

The PBMC supernatants from the ALS twin containing IL-1, IL-6, and TNF-α caused demise of rat motor neurons, which was strongly inhibited by the antibody to the IL-6 receptor and less so by the antibodies to IL-1β and TNF-α. IL-6 signaling in a mouse model is known to cause neuropathology in a concentration-dependent fashion (32). In addition, IL-6 and TGF-β participate in the expression of IL-17A on CD8 T cells and mast cells in the affected spinal cord (18). Finally, we demonstrated the presence of infiltrating macrophages enclosing fragmented motor neurons in the ALS spinal cord. The epigenetic and

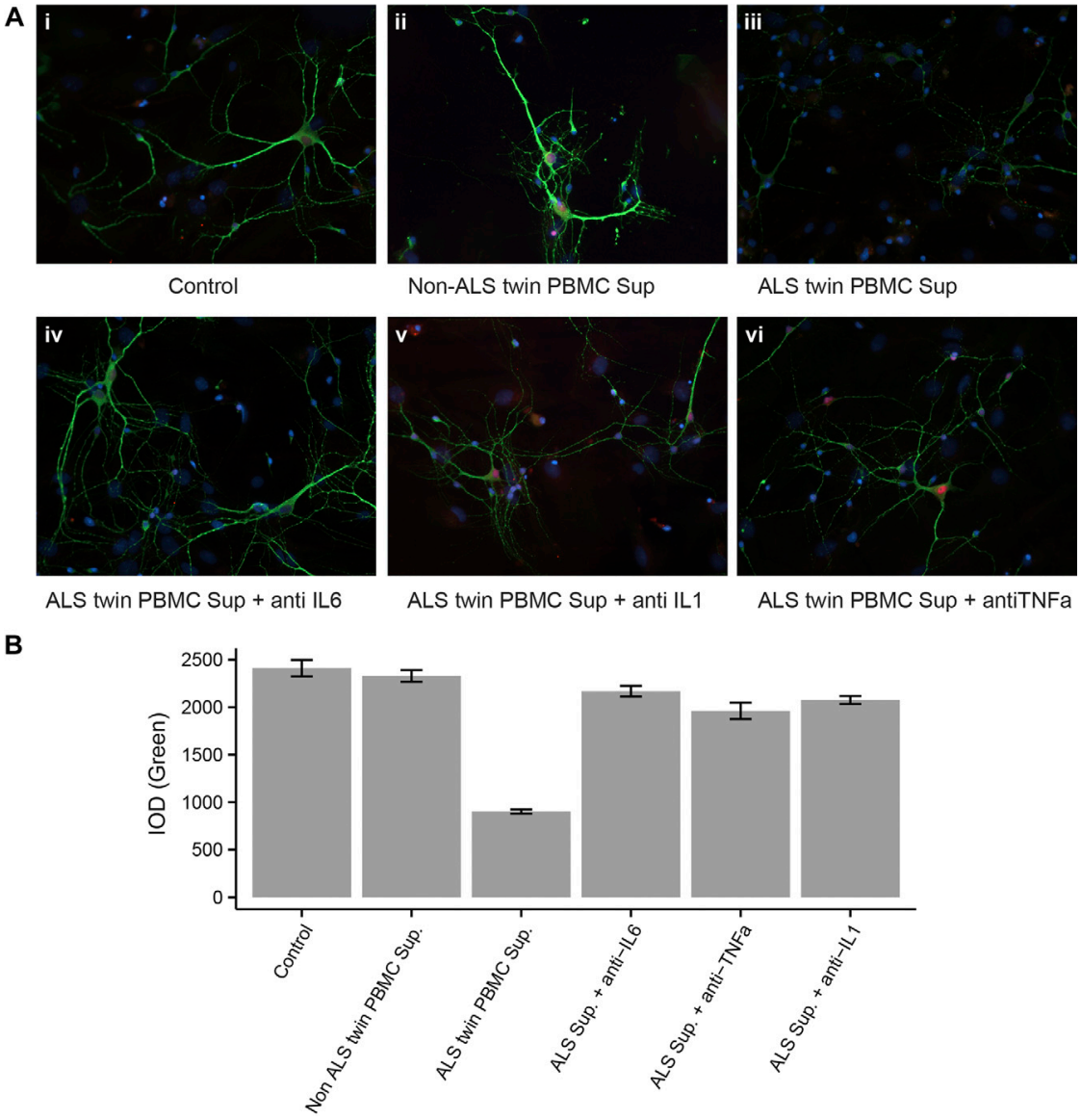


Figure 5. Toxicities of ALS PBMC supernatants to rat neurons. *A*) Fluorescence microscopy of rat neurons exposed to PBMC supernatants. Rat cortical neurons 8 d *in vitro* were exposed to the indicated supernatant (sup) for 6 h (5 μl in 300 μl/well). Neurons were then fixed and immunostained with anti-MAP2 (green dendrites) and anti-NeuN (red neuronal bodies). *B*) Integrated optical density (IOD) of the MAP-2 green fluorescence (determined by Image-Pro): left to right: a comparison of IOD in neurons exposed to the supernatant from each twin; comparison of IOD in neurons treated with the ALS twin's supernatant and antibodies to cytokines; comparison of IOD in neurons exposed to the supernatants from 2 other ALS patients and the twins. All comparisons are significant; $P < 0.001$.

transcriptional differences between the twins and higher abundance of IL-6- and TNF-α-producing macrophages in the ALS twin may account for their different phenotypes. The putative but heterogeneous role of neurotoxic cytokines was highlighted in 2 other sALS patients, one with and the other without neurotoxic cytokines. Our recent study of immune pathways in PBMCs of 9 sALS patients revealed that patients cluster into an 'inflammatory" subset with an inflammatory Th1/Th17 signature and a "noninflammatory" subset with a B-cell signature (63). We propose that RNA sequencing offers an approach to the selection of ALS patients for different anti-inflammatory therapies.

ACKNOWLEDGMENTS

The authors thank Giovanni Coppola for genetic screening of the twins, Anna Caputo for testing toxicities of PBMC supernatants in rat neurons, Julian Whitelegge for fibrillary

wild-type SOD-1, Larry Magpantay for testing cytokines, and Elizabeth Rigali (all from UCLA School of Medicine) for assistance with manuscript preparation. Funding was provided by U.S. National Institutes of Health (NIH) National Institute of Human Genome Research Grant T32 HG008553-01 (to L.M.) and NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases Grant 3P50AR063020-03S1 (to D.M.). M.F. was supported by and received paid travel to meetings from Genentech, Inc. (South San Francisco, CA, USA), and honoraria and paid travel from Smartfish AS (Oslo, Norway).

AUTHOR CONTRIBUTIONS

M. Fiala and M. Pellegrini designed the research; L. Lam, L. Chin, R. C. Halder, B. Sagong, S. Famenini, and L. Rubbi performed the research; L. Lam, J. Sayre, D. Montoya, M. Pellegrini, and M. Fiala analyzed the data; and M. Fiala, L. Lam, and M. Pellegrini wrote the paper.

REFERENCES

1. Andersen, P. M., and Al-Chalabi, A. (2011) Clinical genetics of amyotrophic lateral sclerosis: what do we really know? *Nat. Rev. Neurol.* **7**, 603–615
2. Vucic, S., Rothstein, J. D., and Kiernan, M. C. (2014) Advances in treating amyotrophic lateral sclerosis: insights from pathophysiological studies. *Trends Neurosci.* **37**, 433–442
3. Peters, O. M., Ghasemi, M., and Brown, R. H., Jr. (2015) Emerging mechanisms of molecular pathology in ALS. *J. Clin. Invest.* **125**, 1767–1779
4. Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M., Clark, C. M., McCluskey, L. F., Miller, B. L., Masliah, E., Mackenzie, I. R., Feldman, H., Feiden, W., Kretzschmar, H. A., Trojanowski, J. Q., and Lee, V. M. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133
5. Li, Y. R., King, O. D., Shorter, J., and Gitler, A. D. (2013) Stress granules as crucibles of ALS pathogenesis. *J. Cell Biol.* **201**, 361–372
6. Cooper-Knock, J., Higginbottom, A., Stopford, M. J., Highley, J. R., Ince, P. G., Wharton, S. B., Pickering-Brown, S., Kirby, J., Hautbergue, G. M., and Shaw, P. J. (2015) Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. *Acta Neuropathol.* **130**, 63–75
7. Kawamata, T., Akiyama, H., Yamada, T., and McGeer, P. L. (1992) Immunologic reactions in amyotrophic lateral sclerosis brain and spinal cord tissue. *Am. J. Pathol.* **140**, 691–707
8. Graves, M. C., Fiala, M., Dinglasan, L. A., Liu, N. Q., Sayre, J., Chiappelli, F., van Kooten, C., and Vinters, H. V. (2004) Inflammation in amyotrophic lateral sclerosis spinal cord and brain is mediated by activated macrophages, mast cells and T cells. *Amyotroph. Lateral Scler. Other Motor Neuron Disord.* **5**, 213–219
9. Henkel, J. S., Engelhardt, J. I., Siklós, L., Simpson, E. P., Kim, S. H., Pan, T., Goodman, J. C., Siddique, T., Beers, D. R., and Appel, S. H. (2004) Presence of dendritic cells, MCP-1, and activated microglia/macrophages in amyotrophic lateral sclerosis spinal cord tissue. *Ann. Neurol.* **55**, 221–235
10. McGeer, P. L., and McGeer, E. G. (2002) Inflammatory processes in amyotrophic lateral sclerosis. *Muscle Nerve* **26**, 459–470
11. Turner, M. R., Goldacre, R., Ramagopalan, S., Talbot, K., and Goldacre, M. J. (2013) Autoimmune disease preceding amyotrophic lateral sclerosis: an epidemiologic study. *Neurology* **81**, 1222–1225
12. Boillée, S., Vande Velde, C., and Cleveland, D. W. (2006) ALS: a disease of motor neurons and their nonneuronal neighbors. *Neuron* **52**, 39–59
13. Yamanaka, K., Chun, S. J., Boillee, S., Fujimori-Tonou, N., Yamashita, H., Gutmann, D. H., Takahashi, R., Misawa, H., and Cleveland, D. W. (2008) Astrocytes as determinants of disease progression in inherited amyotrophic lateral sclerosis. *Nat. Neurosci.* **11**, 251–253
14. Heath, P. R., and Shaw, P. J. (2002) Update on the glutamatergic neurotransmitter system and the role of excitotoxicity in amyotrophic lateral sclerosis. *Muscle Nerve* **26**, 438–458
15. Kang, S. H., Li, Y., Fukaya, M., Lorenzini, I., Cleveland, D. W., Ostrow, L. W., Rothstein, J. D., and Bergles, D. E. (2013) Degeneration and impaired regeneration of gray matter oligodendrocytes in amyotrophic lateral sclerosis. *Nat. Neurosci.* **16**, 571–579
16. Persidsky, Y., and Gendelman, H. E. (2003) Mononuclear phagocyte immunity and the neuropathogenesis of HIV-1 infection. *J. Leukoc. Biol.* **74**, 691–701
17. Zaghi, J., Goldenson, B., Inayathullah, M., Lossinsky, A. S., Masoumi, A., Avagyan, H., Mahanian, M., Bernas, M., Weinand, M., Rosenthal, M. J., Espinosa-Jeffrey, A., de Vellis, J., Teplow, D. B., and Fiala, M. (2009) Alzheimer disease macrophages shuttle amyloid-beta from neurons to vessels, contributing to amyloid angiopathy. *Acta Neuropathol.* **117**, 111–124
18. Fiala, M., Chattopadhay, M., La Cava, A., Tse, E., Liu, G., Lourenco, E., Eskin, A., Liu, P. T., Magpantay, L., Tse, S., Mahanian, M., Weitzman, R., Tong, J., Nguyen, C., Cho, T., Koo, P., Sayre, J., Martinez-Maza, O., Rosenthal, M. J., and Wiedau-Pazos, M. (2010) IL-17A is increased in the serum and in spinal cord CD8 and mast cells of ALS patients. *J. Neuroinflammation* **7**, 76–90
19. Lang, L., Zetterström, P., Brännström, T., Marklund, S. L., Danielsson, J., and Oliveberg, M. (2015) SOD1 aggregation in ALS mice shows simplistic test tube behavior. *Proc. Natl. Acad. Sci. USA* **112**, 9878–9883
20. Forsberg, K., Jonsson, P. A., Andersen, P. M., Bergemalm, D., Graffmo, K. S., Hultdin, M., Jacobsson, J., Rosquist, R., Marklund, S. L., and Brännström, T. (2010) Novel antibodies reveal inclusions containing non-native SOD1 in sporadic ALS patients. *PLoS One* **5**, e11552
21. Kabashi, E., Valdmanis, P. N., Dion, P., and Rouleau, G. A. (2007) Oxidized/misfolded superoxide dismutase-1: the cause of all amyotrophic lateral sclerosis? *Ann. Neurol.* **62**, 553–559
22. Guareschi, S., Cova, E., Cereda, C., Ceroni, M., Donetti, E., Bosco, D. A., Trotti, D., and Pasinelli, P. (2012) An over-oxidized form of superoxide dismutase found in sporadic amyotrophic lateral sclerosis with bulbar onset shares a toxic mechanism with mutant SOD1. *Proc. Natl. Acad. Sci. USA* **109**, 5074–5079
23. Bosco, D. A., Morfini, G., Karabacak, N. M., Song, Y., Gros-Louis, F., Pasinelli, P., Goolsby, H., Fontaine, B. A., Lemay, N., McKenna-Yasek, D., Frosch, M. P., Agar, J. N., Julien, J. P., Brady, S. T., and Brown, R. H., Jr. (2010) Wild-type and mutant SOD1 share an aberrant conformation and a common pathogenic pathway in ALS. *Nat. Neurosci.* **13**, 1396–1403
24. Xi, Z., Zinman, L., Moreno, D., Schymick, J., Liang, Y., Sato, C., Zheng, Y., Ghani, M., Dib, S., Keith, J., Robertson, J., and Rogaeva, E. (2013) Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. *Am. J. Hum. Genet.* **92**, 981–989
25. Paez-Colasante, X., Figueroa-Romero, C., Sakowski, S. A., Goutman, S. A., and Feldman, E. L. (2015) Amyotrophic lateral sclerosis: mechanisms and therapeutics in the epigenomic era. *Nat. Rev. Neurol.* **11**, 266–279
26. Jones, A. R., Troakes, C., King, A., Sahni, V., De Jong, S., Bossers, K., Papouli, E., Mirza, M., Al-Sarraj, S., Shaw, C. E., Shaw, P. J., Kirby, J., Veldink, J. H., Macklis, J. D., Powell, J. F., and Al-Chalabi, A. (2015) Stratified gene expression analysis identifies major amyotrophic lateral sclerosis genes. *Neurobiol. Aging* **36**, 2006.e1–2006.e9
27. Baciu, C., Thompson, K. J., Mougeot, J. L., Brooks, B. R., and Weller, J. W. (2012) The LO-BaFL method and ALS microarray expression analysis. *BMC Bioinformatics* **13**, 244
28. FALS Sequencing Consortium. (2015) Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347**, 1436–1441
29. Fiala, M., Mizwicki, M. T., Weitzman, R., Magpantay, L., and Nishimoto, N. (2013) Tocilizumab infusion therapy normalizes inflammation in sporadic ALS patients. *Am. J. Neurodegener. Dis.* **2**, 129–139
30. Holmoy, T. (2008) T cells in amyotrophic lateral sclerosis. *Eur. J. Neurol.* **15**, 360–366
31. Kebir, H., Kreymborg, K., Ifergan, I., Dodelet-Devillers, A., Cayrol, R., Bernard, M., Giuliani, F., Arbour, N., Becher, B., and Prat, A. (2007) Human TH17 lymphocytes promote blood-brain barrier disruption and central nervous system inflammation. *Nat. Med.* **13**, 1173–1175
32. Campbell, I. L., Erta, M., Lim, S. L., Frausto, R., May, U., Rose-John, S., Scheller, J., and Hidalgo, J. (2014) Trans-signaling is a dominant mechanism for the pathogenic actions of interleukin-6 in the brain. *J. Neurosci.* **34**, 2503–2513

33. Endo, F., Komine, O., Fujimori-Tonou, N., Katsuno, M., Jin, S., Watanabe, S., Sobue, G., Dezawa, M., Wyss-Coray, T., and Yamanaka, K. (2015) Astrocyte-derived TGF-β1 accelerates disease progression in ALS mice by interfering with the neuroprotective functions of microglia and T cells. *Cell Reports* **11**, 592–604
34. Hooten, K. G., Beers, D. R., Zhao, W., and Appel, S. H. (2015) Protective and toxic neuroinflammation in amyotrophic lateral sclerosis. *Neurotherapeutics* **12**, 364–375
35. Tovar-y-Romo, L. B., Ramírez-Jarquín, U. N., Lazo-Gómez, R., and Tapia, R. (2014) Trophic factors as modulators of motor neuron physiology and survival: implications for ALS therapy. *Front. Cell. Neurosci.* **8**, 61
36. Israelson, A., Ditsworth, D., Sun, S., Song, S., Liang, J., Hruska-Plochan, M., McAlonis-Downes, M., Abu-Hamad, S., Zoltsman, G., Shani, T., Maldonado, M., Bui, A., Navarro, M., Zhou, H., Marsala, M., Kaspar, B. K., Da Cruz, S., and Cleveland, D. W. (2015) Macrophage migration inhibitory factor as a chaperone inhibiting accumulation of misfolded SOD1. *Neuron* **86**, 218–232
37. Collins, M. A., An, J., Hood, B. L., Conrads, T. P., and Bowser, R. P. (2015) Label-free LC-MS/MS proteomic analysis of cerebrospinal fluid identifies protein/pathway alterations and candidate biomarkers for amyotrophic lateral sclerosis. *J. Proteome Res.* **14**, 4486–4501
38. Meltz Steinberg, K., Nicholas, T. J., Koboldt, D. C., Yu, B., Mardis, E., and Pamphlett, R. (2015) Whole genome analyses reveal no pathogenetic single nucleotide or structural differences between monozygotic twins discordant for amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Frontotemporal Degener.* **16**, 385–392
39. Xi, Z., Yunusova, Y., van Blitterswijk, M., Dib, S., Ghani, M., Moreno, D., Sato, C., Liang, Y., Singleton, A., Robertson, J., Rademakers, R., Zinman, L., and Rogaeva, E. (2014) Identical twins with the C9orf72 repeat expansion are discordant for ALS. *Neurology* **83**, 1476–1478
40. Trapnell, C., Pachter, L., and Salzberg, S. L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111
41. Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106
42. Anders, S., Pyl, P. T., and Huber, W. (2015) HTSeq: a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169
43. Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M. P., Walker, J. R., and Hogenesch, J. B. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **101**, 6062–6067
44. McCall, M. N., Bolstad, B. M., and Irizarry, R. A. (2010) Frozen robust multiarray analysis (fRMA). *Biostatistics* **11**, 242–253
45. Smith, Z. D., Gu, H., Bock, C., Gnirke, A., and Meissner, A. (2009) High-throughput bisulfite sequencing in mammalian genomes. *Methods* **48**, 226–232
46. Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M. Q., Chen, P. Y., and Pellegrini, M. (2013) BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics* **14**, 774
47. Martens, J. H., and Stunnenberg, H. G. (2013) BLUEPRINT: mapping human blood cell epigenomes. *Haematologica* **98**, 1487–1489
48. Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011) Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26
49. McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., Wenger, A. M., and Bejerano, G. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* **28**, 495–501
50. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303
51. ITALSGEN Consortium. (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268
52. Feng, X. (2005) Regulatory roles and molecular signaling of TNF family members in osteoclasts. *Gene* **350**, 1–13
53. Locksley, R. M., Killeen, N., and Lenardo, M. J. (2001) The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell* **104**, 487–501
54. Pollack, B. P. (2012) EGFR inhibitors, MHC expression and immune responses: can EGFR inhibitors be used as immune response modifiers? *OncoImmunology* **1**, 71–74
55. Zaiss, D. M., Gause, W. C., Osborne, L. C., and Artis, D. (2015) Emerging functions of amphiregulin in orchestrating immunity, inflammation, and tissue repair. *Immunity* **42**, 216–226
56. Povlsen, G. K., Berezin, V., and Bock, E. (2008) Neural cell adhesion molecule-180-mediated homophilic binding induces epidermal growth factor receptor (EGFR) down-regulation and uncouples the inhibitory function of EGFR in neurite outgrowth. *J. Neurochem.* **104**, 624–639
57. Young, M. D., Willson, T. A., Wakefield, M. J., Trounson, E., Hilton, D. J., Blewitt, M. E., Oshlack, A., and Majewski, I. J. (2011) ChIP-seq analysis reveals distinct H3K27me3 profiles that correlate with transcriptional activity. *Nucleic Acids Res.* **39**, 7415–7427
58. Lo, H. M., Wang, S. W., Chen, C. L., Wu, P. H., and Wu, W. B. (2014) Effects of all-trans retinoic acid, retinol, and β-carotene on murine macrophage activity. *Food Funct.* **5**, 140–148
59. Mesci, P., Zaïdi, S., Lobsiger, C. S., Millecamps, S., Escartin, C., Seilhean, D., Sato, H., Mallat, M., and Boillée, S. (2015) System xC− is a mediator of microglial function and its deletion slows symptoms in amyotrophic lateral sclerosis mice. *Brain* **138**, 53–68
60. Nishimoto, N., Terao, K., Mima, T., Nakahara, H., Takagi, N., and Kakehi, T. (2008) Mechanisms and pathologic significances in increase in serum interleukin-6 (IL-6) and soluble IL-6 receptor after administration of an anti-IL-6 receptor antibody, tocilizumab, in patients with rheumatoid arthritis and Castleman disease. *Blood* **112**, 3959–3964
61. Sallusto, F., Lenig, D., Förster, R., Lipp, M., and Lanzavecchia, A. (1999) Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* **401**, 708–712
62. Huster, K. M., Busch, V., Schiemann, M., Linkemann, K., Kerksiek, K. M., Wagner, H., and Busch, D. H. (2004) Selective expression of IL-7 receptor on memory T cells identifies early CD40L-dependent generation of distinct CD8+ memory T cell subsets. *Proc. Natl. Acad. Sci. USA* **101**, 5610–5615
63. Lam, L., Halder, R. C., Montoya, D. J., Rubbi, L., Rinaldi, A., Sagong, B., Weitzman, S., Rubattino, R., Singh, R. R., Pellegrini, M., and Fiala, M. (2015) Anti-inflammatory therapies of amyotrophic lateral sclerosis guided by immune pathways. *Am. J. Neurodegener. Dis.* **4**, 28–39

Received for publication February 10, 2016.
Accepted for publication June 21, 2016.

Chapter 6:

CEllFi: cellular epigenetic fingerprinter -- a bisulfite sequencing based method for cellular deconvolution of heterogeneous samples

ABSTRACT

Variation of immune cells across patient blood and biopsy samples provides insights to the immunobiology of auto-immune disorders and infectious diseases. Research has demonstrated the importance of DNA methylation of CpG dinucleotides in defining cellular identity. Here we present a method that utilizes bisulfite sequencing to measure millions of CpG sites to yield a robust platform for interpreting blood profiles into their constituent cell types. Through the selection of cell type-specific DNA methylation signatures and non-negative least squares regression, we demonstrate the estimation of cell type proportions from *in vitro* cell mixture experiments. Our method provides a translational computational approach to quantify immune cellular quantities with potential for the diagnosis or prognosis of disease.

BACKGROUND

Over the past few years there has been a greater availability of immunotherapies for cancer. These newer classes of cancer treatments such as, ipilimumab, target proteins produced by tumors that suppress the immune system. This drug unmasks the cancer cells from the immune system, leading to the proliferation of the T-cells and improved survival rates (Olson & McNeel, 2013). A recent clinical study has shown that patients with delayed increases in CD4 and CD8 T cells after ipilmumab treatment proceeding an early increase in absolute lymphocyte counts are associated with positive clinical outcomes of melanoma, which highlights the need for immune response monitoring for personalized treatments (A. Martens et al., 2016). One proposed method of monitoring immune response is to quantify the cell type composition of blood in treated individuals. Works by Houseman and Accomando have shown that estimating the leukocyte composition of peripheral blood can be achieved through DNA methylation

profiling of bulk samples using DNA methylation microarrays (Houseman et al., 2012). Their approach provides researchers a means of monitoring both inflammatory and suppressive immune activity based on the methylation profile of multiple purified references. In addition, their efforts provide a high throughput means of sub cell type distribution for heterogeneous samples.

As the gold standard for studying DNA methylation has shifted to NGS based methods, a growing set of whole genome methylomes from purified cell types have been made accessible to researchers. The BLUEPRINT project generated several reference epigenomes of healthy individuals to identify key epigenetic features involved in hematopoietic differentiation programs (J. H. A. Martens & Stunnenberg, 2013). This reference set provides methylation calls at millions of CpG loci as opposed to the hundreds of thousands available to the microarrays utilized in previous DNA methylation deconvolution methods. Here we present a method that utilizes bisulfite sequencing to yield a robust platform for decomposing blood profiles into their constituent cell types. Following the selection of cell type specific DNA methylation signatures and non-negative least squares regression, we demonstrate our method on *in vitro* mixtures of hematopoietic cell types. Our method provides a comprehensive pipeline to detect unique CpG markers and estimate cell type composition in blood that utilizes a greater number of markers available in DNA methylation array based approaches. with potential applications beyond immune-biology.

IMPLEMENTATION

Linear Model

CEllFi is a computational pipeline for the estimation of cell mixture composition in heterogeneous samples using DNA methylation profiles from BS-Seq methylation calls. Using a reference set of purified cell type specific methylation profiles; CEllFi estimates the proportion of each referenced cell type within a sample's methylation profile. To estimate the cell type proportions, CEllFi utilizes a non-negative least squares algorithm where A is a m by n matrix of m CpG methylation calls for n reference cell types to minimize x (cell type proportions), where b is the vector of corresponding CpG methylation calls for a heterogeneous sample and x is constrained to be non-negative (Lawson & Hanson, 1974).

$$\min ||Ax - b||_2 , \ x \geq 0$$

The CEllFi pipeline incorporates multiple pre-processing steps to organize the methylation calls of replicate reference methylation profiles as well as select common CpGs across a multi-sample analysis. The preprocessing step identifies regions with optimal differences in methylation between groups.

Processing Overview

The input data for the CEllFi pipeline consists of methylation call files in UCSC bedgraph or BSSeeker2 CGmap file formats for the reference cell types. The bedgraph files are joined and low variance CpG sites may be removed from additional processing. The sample methylation call files are joined and filtered to identify common CpG sites with a user defined minimum coverage level across all samples.

The CEllFi workflow begins with the "meth_matrix" process to join multiple reference methylation calls in bedgraph format into a single matrix. The "matrix_metrics_std" process

selects for CpG sites with large variance across the different reference files. The default minimum methylation standard deviation of 0.20 across all references files per CpG site. The "cgmap_data_frame" process is used to merge the sample methylations, that have been generated by BS_Seeker_2, into a single matrix. The "join_ref_sample" process joins the reference matrix and the sample matrix into a merged matrix file. The "bind_dmr_col" process will perform the fragment detection step by performing the region detection algorithm previously described. Single site CpG methylation values are subsequently aggregated into fragment methylation values by calculating the arithmetic mean for both the reference and sample input. In the "matrix_metrics_delta" step, the minimum fragment methylation differences is calculated between cell type groups and recorded. The "filter_ref_sample," step removes fragments with a threshold delta of -0.20 and a minimum unadjusted anova p-value of 0.05. "decon_int_est" is the final process in which the non negative least squares regression estimates the cell type proportions for each sample using the cell type group as the reference set (Figure 1).

Region Detection

A sliding CpG method is used to identify regions with similar levels of methylation between neighboring CpG sites. Starting at a single end of a chromosome, a seed region is set as the starting CpG. The methylation difference between the seed region methylation call and a neighboring CpG is calculated for all reference values. When the maximum methylation difference between the seed region and bp distance is below a maximum threshold, the CpG is merged into the region. In the merge step, the region methylation values are updated as the mean methylation for the CpG members and the region boundaries are expanded to the most distal CpGs. When neighboring CpGs exceed the maximum methylation difference or distance

threshold, the CpG is initialized as the starting point for the new region. Once the reference regions are defined, the region methylation values are calculated for the samples by calculating the arithmetic mean of the CpG methylation values of sites within the defined region.

Differential Methylation

To identify regions of differential methylation, replicate methylation calls are aggregated into an average cell type methylation value. Pairwise differences in the cell type methylation are calculated between cell types. For each cell type, cell specific regions are ranked based on the minimum difference to the remaining cell types. Regions can be filtered by an anova and t-test p-value between cell types. A minimum difference threshold may also be set to choose regions with cell specific methylation levels.

MATERIALS/METHODS

In Vitro Cell Mixture

Six *in vitro* mixtures were prepared composed of CD4 positive T cells, CD8 positive T cells, Neutrophils, Natural Killer, B cells, and Monocytes. The cell types were isolated from healthy human whole blood using Ficoll-Paque density centrifugation. Isolated cell types were subsequently mixed into six different test tubes with varying percentages of each cell type (Table 1).

RRBS Alignment

Genomic DNA for making RRBS libraries was extracted from the six *in vitro* mixtures according to the standard protocol (Smith, Z.D., et al.). Genomic DNA was digested with a

methylation-insensitive endonuclease, MspI. Fragments of 40–220 bp were isolated, because they are enriched for CpG-rich regions, such as CpG islands, promoter regions, and enhancer elements. The MspI digested DNA was end repaired, A tailed, and ligated with Illumina adaptors. The double-stranded DNA was denatured, followed by bisulfite conversion and PCR amplification. These libraries were sequenced with Ilumina HiSeq 2000 sequencers (San Diego, CA, USA). The reads were aligned to the reference genome (human hg19) using the modified bisulfite aligner BS Seeker2, and selecting for the uniquely mapped reads (Guo et al., 2013). We calculated the methylation level for each cytosine on the genome. Because bisulfite treatment converted unmethylated cytosines (Cs) to thymines (Ts), the methylation level at each cytosine were estimated by #C/(#C+#T), where #C is the number of methylated reads and #T is the number of unmethylated reads. In this study, only cytosines that were covered by at least 10 reads in all samples were included in the subsequent analysis.

In Vitro Cell Mixture Estimation

Six cell types were selected for deconvolution, B Cell, CD14, CD4, CD8, Natural Killer, and Neutrophils. The reference methylation calls for the six cell types were download from the Blueprint Epigenome project (J. H. A. Martens & Stunnenberg, 2013). Together with the six cell mixture methylation calls, the inputs were submitted to the CEllFi processing pipeline to generate % composition of each cell type for all samples. The CEllFi pipeline utilized the default parameters for detected fragments with two or more fragments per cell type, use hypo methylated fragments with a methylation distance of at least 20% from the reference cell type.

RESULTS/DISCUSSION

The preprocessing steps in selecting CpG signatures hypomethylated in the different cell types reveal that 1154, 372, 443, 13, 30, and 235 CpG sites are uniquely low in methylation for the Neutrophil, CD14, Natural Killer, CD4 T Cell, CD8 T Cell, and B Cells may be incorporated into a fragment for estimation (Figure 2). Our method is able to identify unique markers, specific to each cell type. Unfortunately, identifying a large number of CpG sites to distinguish between similar cell types like CD4 and CD8 cells remain a challenge. To overcome this limitation we may incorporate hyper methylated CpG sites rather than focus solely on hypomethylated features.

Our method was able to estimate the % cell type composition for each of the cell mixtures with mean error levels of 4.30% in tube 1, 8.21% in tube 2, 16.06% in tube 3, 3.06% in tube 4, 3.25% in tube 5, and 4.09% in tube 6 (Table 2). A correlation analysis between the estimated and expected values for each cell type reveal that a positive correlation is consistent between the observed and estimated cell type mixture levels with a Pearson's correlation coefficient of 0.716, 0.996, 0.894, 0.977, 0.991 and 0.996 for the B Cell, CD14, CD4, CD8, Natural Killer, and Neutrophil cell types respectively (Figure 3). The weakest performing estimation occurred in the tube 2 and tube 3 mixtures, which consist of the highest percentage of CD4 or CD8 T cells. By identifying additional features for CD4 and CD8, we may be able to improve on their estimation. When we assess the error of the mixtures with moderate to low levels of CD4 or CD8 T cells, the average sample estimation error remains below 10% and maintains strong correlation between estimated and expected values within cell types (Figure 4).

CEllFi provides a means of both identifying cell type specific epigenetic marks as well as estimate the cell type composition of heterogeneous samples using WGBS or RRBS. As new immune therapies for cancers undergo development, it will be informative to utilize a high

throughput method for both the monitoring of immune response as well as identify key epigenetic biomarkers.

TABLES

Table 1. *In Vitro* Blood Cell Mixtures

	Tube 1	Tube 2	Tube 3	Tube 4	Tube 5	Tube 6
CD4 T cell	19.40%	82.80%	0.50%	1.10%	8.00%	5.00%
CD8 T cell	19.40%	10.20%	84.60%	0.50%	3.20%	2.50%
Neut	3.00%	0.40%	0.80%	1.70%	39.50%	0.10%
NK	19.40%	5.10%	10.40%	88.50%	1.60%	1.00%
B cells	19.40%	0.50%	1.00%	2.70%	15.90%	10.00%
Mono	19.40%	1.00%	2.60%	5.50%	31.90%	81.30%

Table 2. Cell Composition

Tube No.	Tube Label	Cell Type	Expected %	Estimated %	Error
1	uniform	B Cell	19.40%	14.29%	5.11%
1	uniform	CD14	19.40%	22.46%	3.06%
1	uniform	CD4	19.40%	21.41%	2.01%
1	uniform	CD8	19.40%	10.57%	8.83%
1	uniform	Natural Killer	19.40%	22.97%	3.57%
1	uniform	Neutrophil	3.00%	6.20%	3.20%
2	hi_t_cd4	B Cell	0.50%	8.96%	8.46%
2	hi_t_cd4	CD14	1.00%	1.06%	0.06%
2	hi_t_cd4	CD4	82.80%	57.28%	25.52%
2	hi_t_cd4	CD8	10.20%	13.45%	3.25%
2	hi_t_cd4	Natural Killer	5.10%	11.83%	6.73%
2	hi_t_cd4	Neutrophil	0.40%	5.66%	5.26%
3	hi_t_cd8	B Cell	1.00%	8.10%	7.10%
3	hi_t_cd8	CD14	2.60%	0.50%	2.10%
3	hi_t_cd8	CD4	0.50%	23.04%	22.54%
3	hi_t_cd8	CD8	84.60%	37.35%	47.25%
3	hi_t_cd8	Natural Killer	10.40%	21.84%	11.44%
3	hi_t_cd8	Neutrophil	0.80%	6.72%	5.92%
4	hi_nk	B Cell	2.70%	2.70%	0.00%
4	hi_nk	CD14	5.50%	4.96%	0.54%
4	hi_nk	CD4	1.10%	0.87%	0.23%
4	hi_nk	CD8	0.50%	1.95%	1.45%
4	hi_nk	Natural Killer	88.50%	79.80%	8.70%
4	hi_nk	Neutrophil	1.70%	9.10%	7.40%
5	hi_neut	B Cell	15.90%	12.59%	3.31%
5	hi_neut	CD14	31.90%	33.80%	1.90%
5	hi_neut	CD4	8.00%	0.24%	7.76%
5	hi_neut	CD8	3.20%	4.09%	0.89%
5	hi_neut	Natural Killer	1.60%	4.19%	2.59%
5	hi_neut	Neutrophil	39.50%	42.53%	3.03%
6	hi_mono	B Cell	10.00%	6.06%	3.94%
6	hi_mono	CD14	81.30%	75.21%	6.09%
6	hi_mono	CD4	5.00%	1.42%	3.58%
6	hi_mono	CD8	2.50%	6.29%	3.79%
6	hi_mono	Natural Killer	1.00%	3.57%	2.57%
6	hi_mono	Neutrophil	0.10%	4.65%	4.55%

FIGURES

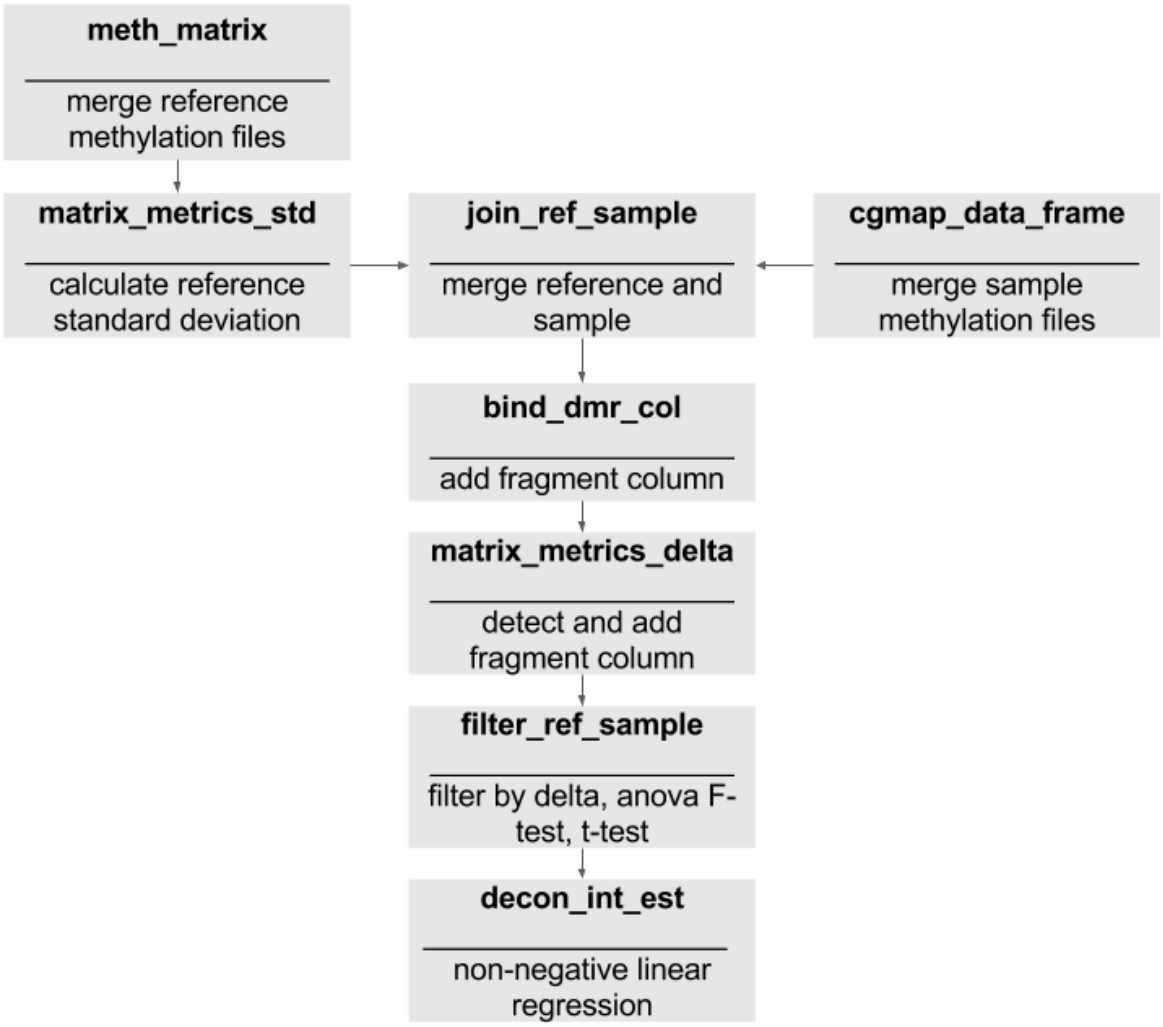


Figure 1. CEllFi Workflow

The CEllFi workflow consists of 8 processes. The "meth_matrix" combines reference bedgraph files into a single matrix. The "matrix_metrics_std" selects for CpG sites with high variance. The "cgmap_data_frame" process is used to merge the sample methylations. The "join_ref_sample" process joins the reference matrix and the sample matrix. The "bind_dmr_col" process will perform the fragment detection and aggregate the methylation calls into fragments. In the "matrix_metrics_delta" step, the fragment delta is calculated. The "filter_ref_sample," step

removes fragments with a minimum threshold delta. “decon_int_est” is the final process to estimate the cell type proportions for each sample.

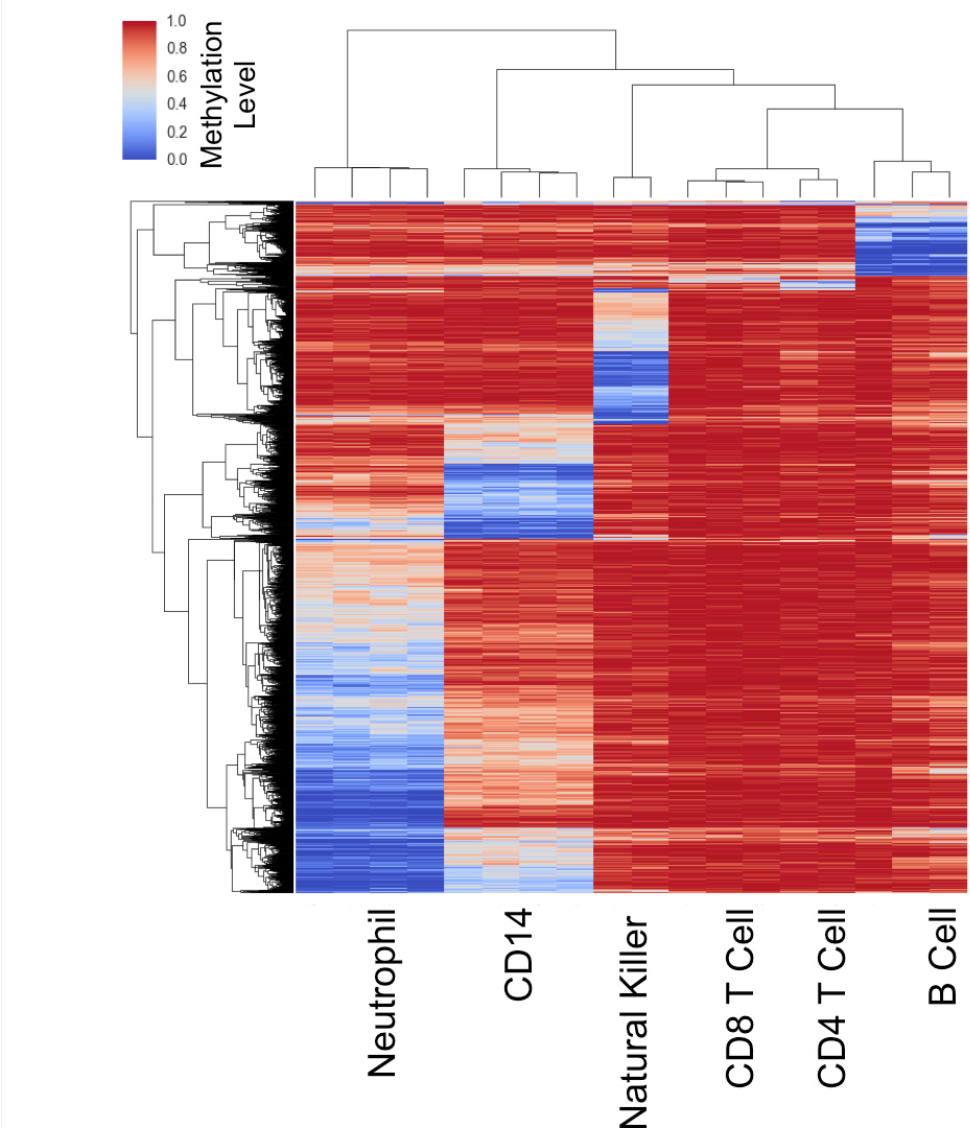


Figure 2. Heatmap of Cell Type Specific CpG Sites

The heatmap is a visualization of the methylation level across the reference methylation data set of sites uniquely low in methylation with high levels of methylation in red and low methylation levels in blue. 1154, 372, 443, 13, 30, and 235 CpG sites were identified as uniquely hypo methylated for the Neutrophil, CD14, Natural Killer, CD4 T Cell, CD8 T Cell, and B Cells respectively.

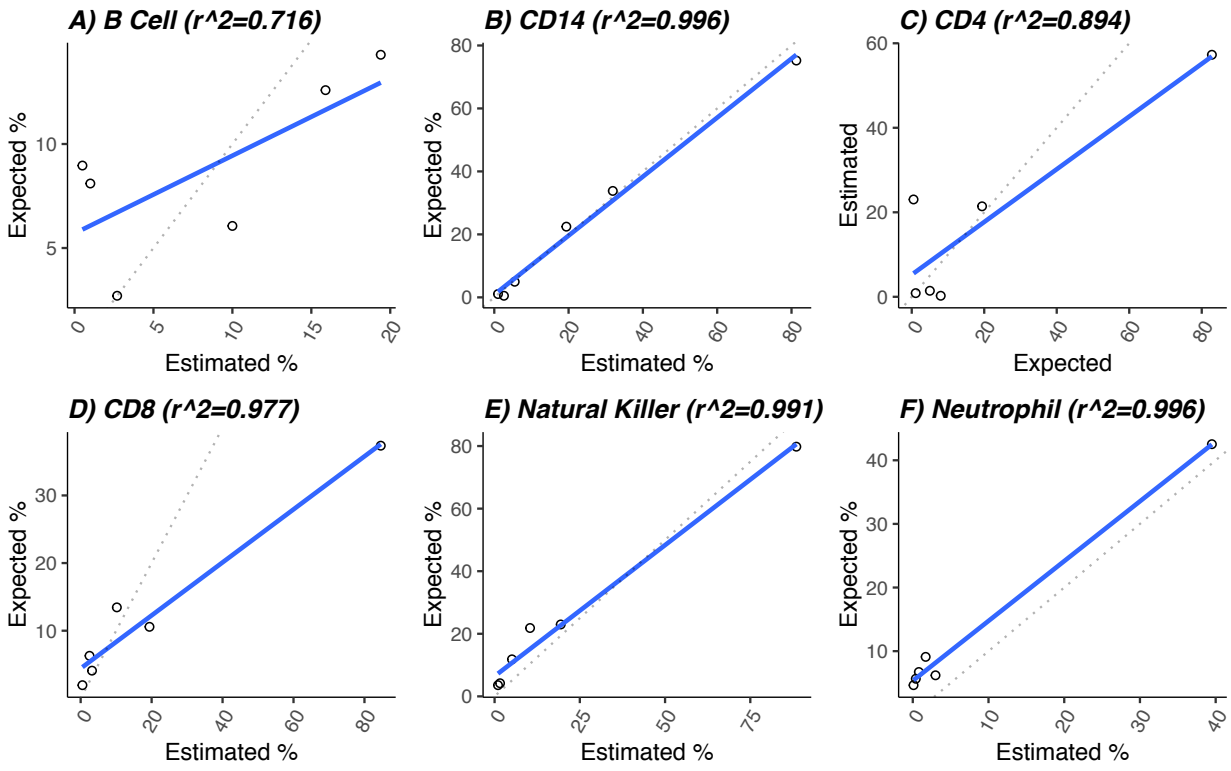


Figure 3. Correlation of Estimated and Expected Cell Type Composition

Scatter plot of the expected cell type percentage within the tube mixture against the estimated percentage across all tube mixtures. A) Scatter plot of B Cell expected and estimated cell mixture contribution with a Pearson's correlation coefficient of 0.716. B) Scatter plot of CD14 cell expected and estimated cell mixture contribution with a correlation coefficient of 0.996. C) CD4 T Cell expected and estimated cell mixture contribution, with a Pearson's correlation coefficient of 0.894. D) CD8 T Cell expected and estimated cell mixture contribution, with a Pearson's correlation coefficient of 0.977. Natural Killer expected and estimated cell mixture contribution, with a Pearson's correlation coefficient of 0.991. F) Neutrophil expected and estimated cell mixture contribution with a Pearson's correlation coefficient of 0.996.

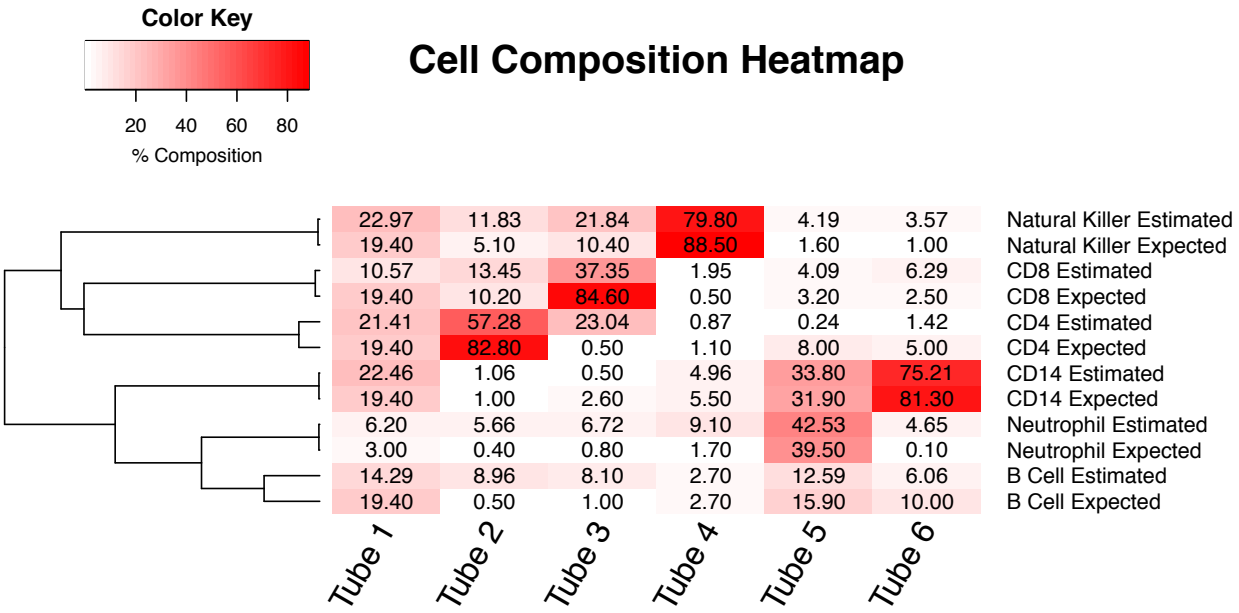


Figure 4. Cell Composition Heatmap

Heatmap of both the expected and estimated % composition of B Cell, CD14, CD4, CD8, Natural Killer, or Neutrophils within a sample mixture. A white cell indicates an undetected or low % of a cell type within a sample. A red cell indicates a high % composition of a cell type within a sample. The % composition values are clustered based on correlation across the % values across the *in vitro* mixtures (tubes 1 – 6).

REFERENCES

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., & Zhang, M. Q. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*, *14*. http://doi.org/10.1186/1471-2164-14-774

Houseman, E. A., Accomando, W. P., Koestler, D. C., Christensen, B. C., Marsit, C. J., Nelson, H. H., … Kelsey, K. T. (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, *13*(1), 86. http://doi.org/10.1186/1471-2105-13-86

Lawson, C. L., & Hanson, R. J. (1974). *Solving Least Squares Problems*. (R. E. J. (University of W. O'Malley, Ed.) (SIAM). Philidelphia: Society of Industrial and Applied Mathematics.

Martens, A., Wistuba-Hamprecht, K., Yuan, J., Postow, M. A., Wong, P., Capone, M., … Weide, B. (2016). Increases in absolute lymphocytes and circulating CD4+ and CD8+ T cells are associated with positive clinical outcome of melanoma patients treated with ipilimumab. *Clinical Cancer Research*, 1–11. http://doi.org/10.1158/1078-0432.CCR-16-0249

Martens, J. H. A., & Stunnenberg, H. G. (2013). BLUEPRINT: Mapping human blood cell epigenomes. *Haematologica*, *98*(10), 1487–1489. http://doi.org/10.3324/haematol.2013.094243

Olson, B. M., & McNeel, D. G. (2013). Monitoring regulatory immune responses in tumor immunotherapy clinical trials. *Frontiers in Oncology*, *3*(May), 109. http://doi.org/10.3389/fonc.2013.00109

Chapter 7:

Conclusion

The biological interpretation of DNA methylation or RNA-Seq data has been aided by leveraging biomedical publications and publicly available datasets, like MSigDB. By performing an enrichment analysis, it is possible to associate differentially expressed genes or differentially methylated regions to curated pathways or sets of genes. Analysis approaches are greatly simplified through the use of online tools, like DAVID or GREAT, with access to numerous gene set ontologies. Unfortunately, as newer expression and epigenetic data becomes available, the online annotation tools may fail to include the latest pathways or gene sets. Throughout this work, tools and analysis strategies employed underused data resources to provide insight into a series of biomedical experiments.

The work in the second chapter was to build a collection of DNA methylation analysis tools. Features of the bisulfite sequencing tool, MethGO, includes several methylation characterization analyses to examine methylation coverage, sample methylation levels, methylation level distribution, methylation characterization of specific genomic features, single nucleotide polymorphism calling, and copy number variation calling. Among the methylation characterization tools, MethGO is able to visualize the methylation levels across transcription factor targets that may correspond to changes in expression. MethGO is coded in Python and is publicly available. One opportunity for expansion of MethGO, is to provide a web or user interface for biologists. Automation and simplification of methylation analysis can reduce time and effort involved in established a computational pipeline. It may also be informative to incorporate methylation profiling at additional genomic features, such as chromatin states, CpG islands, or chromatin states.

The analysis approach in the third chapter incorporated multiple external data resources to interpret the expression and epigenetic changes that occurred in the increasingly

chemoresistant mouse Burkitt's lymphoma cell lines. A PCA analysis of the expression data of the lymphoma cell lines along with control B-cells undergoing maturation reveals that the transcriptional changes of the increasingly resistant B-cells corresponds with the dedifferentiation of the control B-cells. In working with the DNA methylation data, multiple transcription factor targets were identified from external CHiP-Seq experiments. An enrichment analysis of the genes associated with differential methylation and transcription targets indicate an association with transcription factors involved in B-cell development. The analysis also identified transcription factors involved in mediating epigenetic changes. Through the integration of the expression and epigenetic analyses, an epigenetic mechanism of chemoresistance was proposed. A potential application of the study is to identify epigenetic markers of chemoresistance or B-cell dedifferentiation in B-cells. By utilizing healthy WGBS data from healthy B-cells at different stages in differentiation, like Blueprint Epigenome Project, it should be possible to identify DMRs to be used as a marker for chemoresistance in Burkitt's lymphoma.

In the fourth and fifth chapters, transcriptional and epigenetic profiles of PBMCs were examined in patients affected with ALS. An unsupervised cluster analysis of the expression profiles identified two main groups. By defining a series of inflammatory expression signatures from microarray expression data from GEO, it was possible to annotate the group as having an inflammatory transcriptional profile or non-inflammatory profile. A subsequent differential expression analysis followed by pathway enrichment analysis reveals differences in mechanistic activity between the inflammatory and non-inflammatory samples. One challenge to analyzing PBMCs is gauging the effect of differences in blood cell type composition. In the fifth chapter we download the methylation calls from multiple haematopoietic cell lines to build a cell type specific set of hypomethylated CpG sites. By examining the methylation level distribution of the

cell type specific CpG sites across the ALS samples it was possible to identify qualitative differences in cell type composition between the samples.

The sixth chapter describes the methylation based cellular deconvolution tool CEllFi. The methylation analysis scripts used in the fifth chapter were modified into a quantitative tool to estimate the cell type composition of heterogeneous samples. By using a non-negative least squares regression function, CEllFi is able to quantify the cellular composition of blood samples mixed *in vitro*. The cell type estimates across our *in vitro* mixtures reveal strong correlation between the expected and estimated values.

Interpretation of differentially methylated or differentially expressed genes has been greatly simplified by web-based tools like DAVID or GREAT. By aggregating gene sets and gene families to a single access point, biologists are able to annotate results on a large set of reference data. However, when datasets have not been fully disseminated to the databases of popular annotation tools, it becomes necessary to develop an analysis pipeline able to profile the methylation and expression activity in our studies. By working with expression or methylation reference data directly, biologists may exploit information from new or underutilized data.