THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

Functions of ARID3a in multiple hematopoietic cell types

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSPHY

BY

JOSHUA WADE GARTON

Norman, Oklahoma

2020

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF CHEMISTRY AND BIOCHEMISTRY

BY THE COMMITTEE CONSISTING OF

Carol F. Webb, Ph.D., Chair

Christina Bourne, Ph.D., Chair

Valentin Rybenkov, Ph.D.

Si Wu, Ph.D.

Rakhi Rajan, Ph.D.

David McCauley, Ph.D.

© Copyright by Joshua Wade Garton 2020

All Rights Reserved.

ACKNOWLEDGEMENTS

I would like to thank my family, lab mates, and my mentor, Dr. Carol Webb, for their continued support.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF ILLUSTRATIONS

Figure Page

ABSTRACT

Since the completion of the Human Genome Project in 2003, scientists have had a map of all the base pairs that make up human DNA. Beginning in 2008, RNA-seq was developed that allowed scientists to directly sequence messenger RNA that is transcribed by genes in cells. In 2011, computing power and cost of sequencing (1) was just beginning to reach a point to allow for high-throughput sequencing experiments that could define a person's genome and transcriptome and identify potential disease-causing mutations, having important implications for personalized medicine. One year later, the capability to interrogate the transcriptome of a single cell became possible with the introduction of single-cell RNA-seq (scRNA-seq) (2). Previous bulk studies represent an ensemble of transcriptomes from many cells and therefore could mask important differences. Studies have shown that individual cells have remarkable heterogeneity, even when isolated from a seemingly homogeneous populations (3). Although single-cell technology exists, the need for experts that can analyze and interpret the large and high-dimensional data that is produced from next-generation sequencing experiments is high. For example, the number of scRNA-seq datasets uploaded to the Gene Expression Omnibus (GEO) has increased from 15 in 2012 to over 15,000 in 2020 (4). Therefore, my primary objective for graduate school was to develop a solid background in bioinformatics and to apply next-generation sequencing and microfluidic technology to questions in human disease.

To meet this objective, I have used multiple biologic systems to garner expertise in bioinformatics by exploring functions of the transcription factor, ARID3a. ARID3a is an understudied protein, with unknown functions, that is expressed in hematopoietic cells during development, and in various types of adult stem cells (5). It is a member of a large family of epigenetic regulators and dimerization of ARID3a is necessary for binding to DNA in a sequence specific fashion (6,7). The few studies that investigate the function of ARID3a indicate that it has functions in both activating and repressing gene expression. The first role of ARID3a in activating expression came from studies in B cells isolated from mice. These studies revealed that ARID3a is required for proper immunoglobulin heavy chain expression through binding the intronic heavy chain enhancer (8-10). Later studies would reveal that ARID3a expression is developmentally restricted and that it binds directly to promoter/enhancer regions of the pluripotency master regulators OCT4, SOX2, and NANOG to contribute to their repression in mouse embryonic fibroblasts and stem cells (11,12). However, the mechanism by which ARID3a contributes to activation or repression of gene expression remains elusive. Therefore, my goals were to use sequencing technology to investigate 1) the importance of ARID3a in biologically important systems (i.e. lupus, hematopoietic cells, erythropoiesis) and 2) provide insights into the functions of ARID3a using different techniques (i.e. RNA-seq, ATAC-seq, single-cell RNA-seq).

To meet these goals, I have used RNA-seq to show that ARID3a expression is associated with disease activity in two cell types, plasmacytoid dendritic cells (pDCs) and low density neutrophils (LDNs) which are key players in inflammatory pathways observed in patients with systemic lupus erythematosus (SLE) (13). The RNA-seq data I generated also show that ARID3a is lowly expressed in pDCs and LDNs isolated from both SLE patients and healthy controls.

Therefore, it was not possible to perform typical differential expression analysis based on ARID3a^{+} and ARID3a^{-} samples. This finding also explains why other investigators using traditional RNA-seq analysis have not identified ARID3a to be important to disease activity in these cell types. Instead, I performed unsupervised hierarchical clustering to show that the RNA-seq data cluster based on the levels of ARID3a protein and that SLE disease activity scores strongly correlate with ARID3a protein (13). I also show through correlation analysis that ARID3a functions epigenetically through repressing and activating many genes.

Previous work revealed that ARID3a deletion in mice resulted in embryonic lethal phenotype but the rare survivors had a significant reduction in erythrocytes and B cells (14). Using the early hematopoietic cell line, K562, which can be stimulated with hemin to induce erythrocyte differentiation, I show with RNA-seq that ARID3a protein is necessary for fetal globin expression and erythrocyte development. My data identify genes affected by ARID3a expression and indicate that ARID3a functions in a cell type specific fashion. The Assay for Transposase-Accessible Chromatin (ATAC)-seq data I generated using K562 cells show for the first time that ARID3a alters chromatin accessibility of enhancer regions essential for the induction of erythroid-specific genes. These findings allowed me to learn bulk RNA-seq and ATAC-seq analyses and led to the investigation of ARID3a in primary human cells using single cell RNA-seq.

Single cell RNA-seq was performed on hematopoietic stem cells (HSCs) from aged and young donors and reveal that ARID3a levels have an impact on B cell fate decisions. This work show that single cell technology can be used to identify changes based on ARID3a transcript in HSCs.

Finally, to accomplish my goal of applying the technologies described above to human disease, I have employed microfluidic technology to capture single naïve B lymphocytes from SLE patients. Healthy naïve B cells do not express ARID3a. However, ARID3a is present in ~50% of the naïve B cells of SLE patients. Additionally, since ARID3a is an intracellular protein, it is not possible to isolate ARID3a^{+} B lymphocytes without affecting RNA integrity. Therefore, I have employed microfluidic technology to capture single ARID3a-expressing and ARID3a negative naive B lymphocytes from SLE patients. Briefly, B lymphocytes were isolated using known surface markers representing naïve B cell subsets and were then captured on a Fluidigm C1 Single-Cell Auto Prep system. At present, there are significant knowledge gaps, which will be detailed in Chapter 2, regarding how the expression of ARID3a contributes to increased disease activity and what causes the increased inflammatory responses in SLE patients. We hypothesize that ARID3a expression will identify autoreactive naïve B cells that have broken tolerance and/or are part of the inflammatory responses observed in SLE.

CHAPTER 1

NEW FRONTIERS: ARID3a in SLE

Joshua Garton 1,† , M. David Barron 2,† , Michelle L. Ratliff 3 and Carol F. Webb 4*

1 Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK 73072, USA; joshgarton@ou.edu

2 Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA; matthew-barron@ouhsc.edu

3 Department of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA; michelle-ratliff@ouhsc.edu

4 Departments of Medicine, Microbiology and Immunology, Cell Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA

* Correspondence: carol-webb@ouhsc.edu; Tel.: +405-271-4188 † These authors contributed equally to this work

Acknowledgements

I would like to acknowledge David Barron, a co-first author on this paper for his contributions in conducting conceptualization, literature review and writing. I would like to also acknowledge other co-authors, Dr. Michelle Ratliff and Dr. Carol Webb for their contributions to writing and generating figures. This work was supported by funding from the National Institutes of Health (AI118836 and AI123951 (CFW), T32 AI007633 (JG), and K99AG055717 (MLR)).

Copyright Information

The original version of this work was published in Cells on September 24th, 2019. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (**http://creativecommons.org/licenses/by/4.0/**). This work is reproduced directly with changes to merge referencing with the rest of this work.

Allocation of Contribution

This publication was created from a group effort and has major contributions from the co-first author, Mr. David Barron. I have included the entirety of the published work in this chapter of my dissertation. My contribution to this work was literature review and writing for all sections, except for the section on hematopoiesis, and generating figure 3. I also isolated, prepared sequencing libraries, sequenced, preprocessed, analyzed the data and performed pathway analysis for **Figure 3**. My contributions to **Figure 4** were made through RNA-seq analyses of pDCs and neutrophils. I aided in the generation of all other figures through helpful commentary.

ABSTRACT

Systemic lupus erythematosus (SLE) is a devastating and heterogeneous autoimmune disease that affects multiple organs, and for which the underlying causes are unknown. The majority of SLE patients produce autoantibodies, have increased levels of type-I inflammatory cytokines, and can develop glomerulonephritis. Recent studies indicate an unexpected but strong association between increased disease activity in SLE patients and the expression of the DNA-binding protein ARID3a (A + T rich interaction domain protein 3a) in a number of peripheral blood cell types. ARID3a expression was first associated with autoantibody production in B cells; however, more recent findings also indicate associations with expression of the inflammatory cytokine interferon alpha in SLE plasmacytoid dendritic cells and low-density neutrophils. In addition, ARID3a is expressed in hematopoietic stem cells and some adult kidney progenitor cells. SLE cells expressing enhanced ARID3a levels show differential gene expression patterns compared with homologous healthy control cells, identifying new pathways potentially regulated by ARID3a. The associations of ARID3a expression with increased disease severity in SLE, suggest that it, or its downstream targets, may provide new therapeutic targets for SLE.

INTRODUCTION

The autoimmune disease systemic lupus erythematosus (SLE) affects approximately one million Americans (15), with symptoms ranging from rash and fatigue, to severe organ dysfunction (16,17). Disease severity varies according to the degree of organ involvement, and the level of inflammation and systemic deposition of autoantibody-containing immune complexes, and is quantified using scores that combine measurements of these criteria as SLE

disease activity indices (SLEDAI scores) (16,17). The underlying mechanisms that lead to SLE are unknown, but several key features of the disease are present in the majority of patients. These include breaks in humoral tolerance that result in the production of autoimmune antibodies; inflammation typically characterized by increased levels of Type I cytokines, such as interferon alpha (IFNα); and, lupus-induced kidney damage, known as lupus nephritis. Renal involvement occurs in nearly 60% of SLE patients (18), and results in high healthcare costs (15) and mortality (19). Very few effective therapies for SLE exist. Current treatments include Rituximab, an antibody that targets the majority of B lymphocytes, and Belimumab, an antibody that also targets B cells through the B cell survival factor, BLyS (B Lymphocyte Stimulator) or BAFF (B cell Activating Factor) (20). Both drugs have broadly immunosuppressive effects through general B-cell inhibition, and fail to target the specific sources of autoimmune antibodies directly. Therefore, new therapeutic targets are badly needed. Our data now suggest that the DNA-binding protein ARID3a (A + T rich interaction domain protein 3a) is overexpressed in a number of hematopoietic cell types in SLE peripheral blood compared with healthy controls, and that expression of ARID3a is associated with increased SLEDAI scores. Although a number of contributing factors are likely to influence SLE pathology, in this review, we will highlight the molecular and cellular associations of ARID3a with the common lupus pathologies of autoantibody production, interferon induction, and the promotion of clinical nephritis.

ARID3a, or Bright (B cell regulator of immunoglobulin (Ig) heavy chain transcription) as it was first called in the mouse, was originally discovered as a transcription factor that increases Ig-heavy chain transcription and binds to the intronic heavy chain enhancer, where it is associated with epigenetic effects and the organization of chromatin into transcriptionally active domains by interactions with nuclear matrix associated regions (8-10,21,22). ARID3a is not

required for the initiation of Ig production, but requires dimerization and association with both Bruton's tyrosine kinase (BTK) and TFII-I (Transcription Factor II-I) for activity in B lymphocytes (9,23,24)\. Binding sites for ARID3a were first identified 5′ of the promoter of the V1 S107 gene, a gene necessary for responses to phosphorylcholine (25). In humans, there are 15 ARID family proteins, most of which have epigenetic functions, and are members of larger chromatin-modifying complexes that act in part through protein domains associated with discrete epigenetic functions (6,7,26). The three members of the ARID3 family are distinguished as smaller proteins that have an extended DNA-binding domain that confers increased sequence specificity compared to other ARID family members but lack obvious protein domains associated with epigenetic functions (Figure 1A). Because ARID3a can act as both an activator and suppressor of transcription (8,9,12), ARID3a may regulate gene expression as a DNA-specific tether that recruits other epigenetic modifiers to those sites (Figure 1B,C).

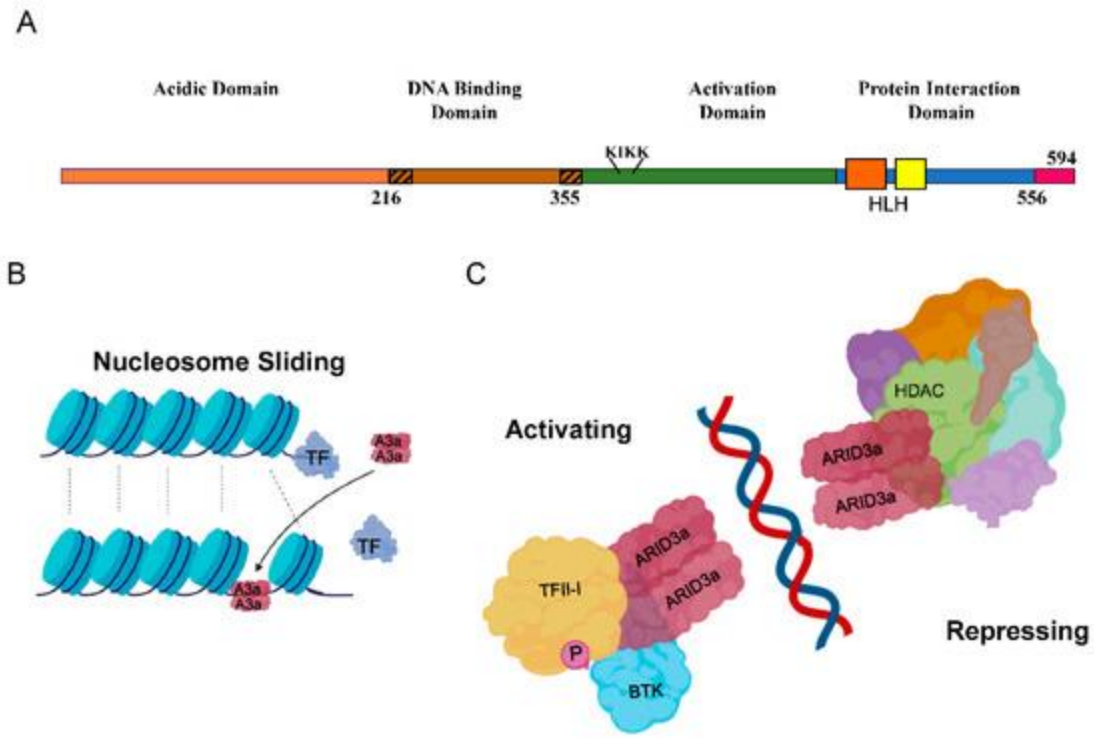


Figure 1. Schematic diagrams of ARID3a (A + T rich interaction domain protein 3a) domains and functions. (A) The four protein domains of ARID3a are shown, including the extended DNA-binding domain (hashed ends), the nuclear localization motif (KIKK), the helix-loop-helix regions (orange and yellow boxes), and amino acid numbers. The short carboxyl terminus has not been given a name or function. (B). ARID3a is proposed to induce nucleosome sliding, potentially disrupting the binding of other transcription factors, and potentially recruiting new transcription factors. (C). ARID3a DNA-interacting protein complexes can have either activating or repressing functions. Some of the known interacting proteins are shown. A3a—ARID3a; TF—transcription factor; HDAC—histone deacetylase; TFII-I—transcription factor II-I; P—phosphorylation; BTK—Bruton's tyrosine kinase.

Although most ARID family members are expressed ubiquitously, ARID3a expression is tissue and developmentally restricted to hematopoietic and other adult stem cells, and to particular types of mature cells in the hematopoietic lineage (27). The expression of ARID3a is tightly regulated during B cell development in both mice and humans, and is the highest in the bone marrow pre-B and germinal center-activated B cells, while other B cell subsets, including the majority of mature splenic B cells, lack both ARID3a mRNA and protein (28,29). During B lymphocyte development in mice, ARID3a expression is tightly regulated, such that it is

expressed in transitional T1 B cells, down-regulated at the T2 cell stage where tolerance checkpoints have been identified (30,31), and turned off at the level of transcription in naïve follicular B cells (32,33). Others have shown recently that ARID3a is required for fetal lineage B1 B cell development (29,34). ARID3a is also expressed in mature B1 and marginal-zone (MZ) B cells, two cell types associated with autoimmunity in several systems (32,35-39). Figure 2 summarizes where ARID3a is expressed during healthy B lymphocyte development.

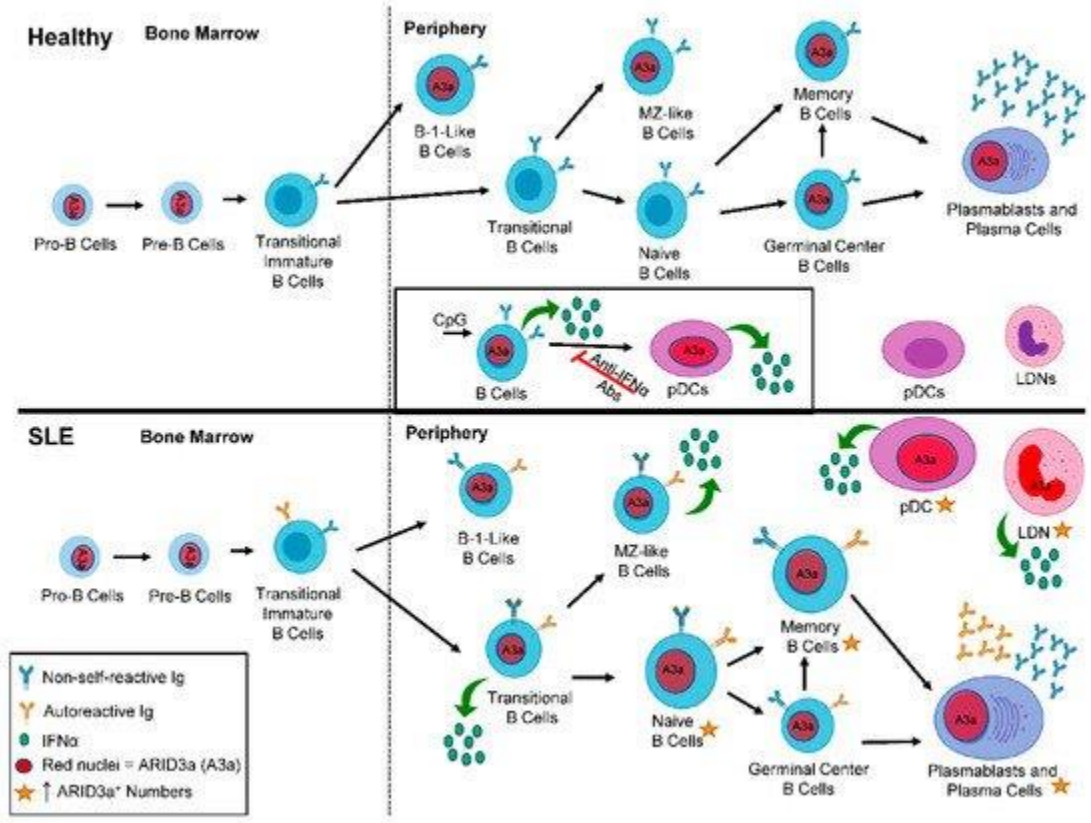


Figure 2. ARID3a expression is enhanced in systemic lupus erythematosus (SLE) B cell subsets compared to healthy controls. A diagram depicts B lineage development in healthy (top panel) and SLE patients (bottom panel) from the bone marrow to the periphery. Red nuclei denote cell subsets that express ARID3a. The inset shows the induction of ARID3a and interferon alpha (IFNα; green circles), and the effect of those ARID3a-expressing B cells on healthy pDCs. In SLE patients, larger cells and gold stars indicate those cell subsets with increased numbers of ARID3a^{+} cells and associated IFNα production. Blue immunoglobulin (Ig) indicates non-self-reactive Ig, while gold Ig indicates the subsets of cells that can also produce autoreactive Ig. pDCs—plasmacytoid dendritic cells; LDN—low density neutrophils; A3a—ARID3a.

2. ARID3a and Autoantibodies in SLE

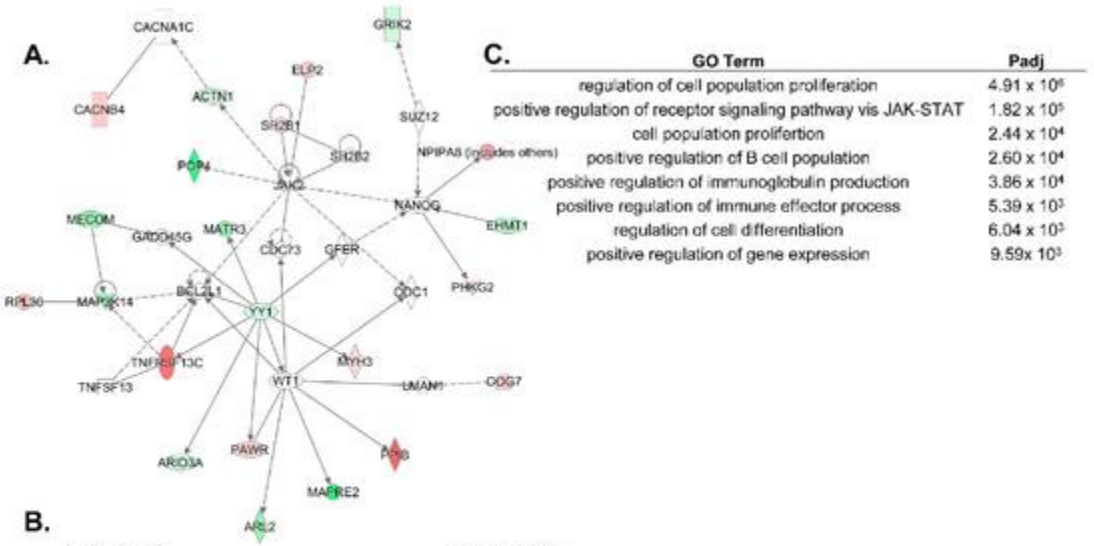
Several years ago, we discovered that the constitutive expression of murine ARID3a/Bright in all B lymphocyte lineage cells in transgenic mice resulted in antinuclear antibody (ANA) production in young mice and immune complex deposits in the kidneys of older mice (35,36). ANAs and anti-DNA antibodies typically increase during flares of SLE disease activity and are hallmarks of lupus in 50%–70% of SLE patients (35,40). ANA production in ARID3a-transgenic mice was observed on two genetic backgrounds and was associated with increases in transitional T1 B cells and MZ B cells (36), subsets previously associated with autoimmune disease activity in mice and humans (39,41-43). In addition, chimeric mice generated from hematopoietic stem cells taken from the transgenic mice also developed ANAs and increased MZ B cell subset numbers, indicating that the constitutive ARID3a expression was directly linked to the production of ANAs and increases in MZ cells (35).

To determine if ARID3a was over-expressed in lupus B lymphocytes, we assessed the numbers of peripheral blood B cells for ARID3a expression in healthy controls, patients with rheumatoid arthritis, and SLE patients (44). Surprisingly, SLE circulating blood peripheral B lymphocytes had dramatically increased ARID3a expression (44). Examination of a random cross section of 115 lupus patients revealed abnormally high numbers of circulating ARID3a$^+$ B cells (up to 40-fold increases) compared with healthy controls and patients with rheumatoid arthritis (44). ARID3a was expressed in all of the nine SLE B cell subsets we examined, including naïve B cells that typically do not express ARID3a transcripts in healthy controls (**Figure 2**) (28,44). Longitudinal analyses of 37 SLE patients revealed that numbers of ARID3a-expressing B lymphocytes varied over time in each patient and within individual B cell subsets, but the total numbers of ARID3a-expressing B cells, irrespective of the subset, were associated

with increased SLEDAI scores (p = 0.0039) (44). ARID3a expression in each B cell subset was bimodal, with only a fraction expressing ARID3a, and there was no direct relationship with specific organ involvement or any autoantibody specificity (44,45). ARID3a was expressed in both healthy and SLE MZ-like B cells, suggesting it may have innate immune functions in those cells (44,46,47). Intriguingly, Epstein Barr virus (EBV) exposure has been associated with increased lupus susceptibility (47), and some anti-DNA antibodies cross react with the EBNA1 protein (47,48). Others showed that EBV requires and recruits ARID3a for expression of the EBNA C promoter that maintains viral latency, associating this virus with ARID3a expression (49). Thus, ARID3a likely plays important roles in innate immunity in healthy B cells and may be over-expressed in SLE in a fashion similar to our ARID3a transgenic mice that developed autoantibodies.

These data led us to hypothesize that ARID3a-expressing naïve B cells might be predisposed to produce autoantibodies. We sorted naïve B cells from both healthy controls and SLE patients, and generated 37 monoclonal antibodies from those cells, but failed to observe skewing toward specific Igs associated with autoimmunity in the SLE naïve B cells, perhaps because of the small number of Igs examined (45). However, when we isolated total B cells from SLE patients with high versus low ARID3a-expression, and examined them for differential gene expression (46,50), we found associations with known mediators of disease activity. Specifically, many genes associated with IFNα expression (*IRF3, IRF5, IFI44,* and *OAS1*), toll-like receptors (*TLR9* and *TLR7*), and anti-apoptotic genes (*BCL2* and *BCL2L1*) were upregulated in the samples with increased ARID3a expression (50). Additionally, our unpublished scRNA-seq data from naïve SLE B cells also show an upregulation of *OAS1*, *IFI27,* and *IRF3*, and confirm that *BCL2L1* is linked to pathways associated with ARID3a (**Figure 3**A). YY1, an important

suppressor and epigenetic regulator, binds to sites in the IgH enhancer that overlap the ARID3a binding sites, and was co-expressed with ARID3a in naïve B cells from SLE patients (**Figure 3**A) (51). Transcripts from 13 IFN signature genes were significantly upregulated in SLE B cells, and five of those, including *EPSTI1*, *IFI27*, and *OAS1*, were increased over 20-fold (50). The gene signatures for the ARID3a-expressing B cells indicate a strong correlation to IFN, suggesting a potential connection between ARID3a and the lupus-associated cytokine, IFNα.



C.

GO Term	Padj
regulation of cell population proliferation	4.91 x 10^{6}
positive regulation of receptor signaling pathway vis JAK-STAT	1.82 x 10^{5}
cell population proliferation	2.44 x 10^{4}
positive regulation of B cell population	2.60 x 10^{4}
positive regulation of immunoglobulin production	3.86 x 10^{4}
positive regulation of immune effector process	5.39 x 10^{3}
regulation of cell differentiation	6.04 x 10^{3}
positive regulation of gene expression	9.59x 10^{3}

B.

Gene Name	Gene Function
AMPH	Promotes endocytosis in synapses
CCN1	Promotes endothelial cell adhesion, promotes neutrophil efferocytosis
IFNG	Antiviral response, activates plasma cells
IL4	Participates in B cell activation and induces MHCII expression, enhances IgE and IgG1
IL5	B cell and Eosinophil Development, induces B1 cell autoantibody production
IL9	Leukocyte survival, Autoantibody production
IL13	B cell development, Allergic responses
NANOG	Pluripotency Factor
SOX2	Pluripotency Factor, Cytosolic sensing of microbial DNA

Figure 3. ARID3a expression is associated with diverse genes in distinct cell types. (**A**). An ingenuity pathway analysis (IPA) network was generated using genes differentially expressed in single cell RNAseq of naïve SLE B cells with a differential ARID3a expression. Green indicates upregulation and red indicates downregulation, while the genes associated with these pathways, but not differentially regulated in the B cells, are colorless. (**B**). SLE LDNs and pDCs shared nine differentially expressed genes associated with the ARID3a expression and listed here. (**C**). GO analyses of the nine ARID3a-associated genes differentially regulated in both pDCs and LDNs indicate the most significant pathways associated with their expression.

A number of labs demonstrated that SLE patients display altered epigenetic marks in peripheral blood cells (50,52). Elegant studies by Scharer et al. indicate that SLE resting naïve B

cells are already epigenetically distinct from healthy control B cells (52,53), suggesting that epigenetic changes might define unique transcriptomes in SLE. However, no epigenetic regulator was specifically associated with the signatures observed in these studies, and the data were obtained from all SLE patients, regardless of disease activities (53). ARID3a has several characteristics of an epigenetic regulator. It alters chromatin accessibility of the Ig enhancer in mice ((21), our unpublished data), and the mouse homologue co-precipitated with at least one histone modifying enzyme (12). In addition, our unpublished data suggest that ARID3a recruits epigenetic machinery. We found that whole blood from SLE patients with increased numbers of ARID3a-expressing B cells showed decreased methylation of promoters for a number of Type I interferon genes, including *IFNα2* and *IFNα6*, compared to SLE samples with normal levels of ARID3a-expressing B cells (46,50). While these data do not directly associate ARID3a with open chromatin at those IFNα loci, these data suggested that ARID3a could be specifically associated with alterations in IFNα expression at the epigenetic level.

3. ARID3a and IFNα in SLE

Consistent with the data from other labs indicating that human B lymphocytes can secrete IFNα (54), the intracellular staining of SLE B lymphocytes revealed that IFNα is produced by a subset of B cells, most of which co-express ARID3a (46). Indeed, we found that a subset of MZ-like healthy B cells can be stimulated with the TLR9 agonist CpG to increase ARID3a and IFNα expression (46). EBV-transformed B cell lines express both ARID3a and IFNα, and were used to demonstrate that ARID3a inhibition reduced the production of both proteins (46). These data, along with time course data indicating ARID3a expression could be induced in healthy cells prior to IFNα expression (46), link ARID3a with IFNα expression in B cells, and suggest that ARID3a may contribute to IFNα regulation.

The major sources of type-I IFN production in SLE are thought to be plasmacytoid dendritic cells (pDCs) and granulocytes, rather than B lymphocytes (55,56). Therefore, we investigated if ARID3a-expressing B cells might have a helper cell activity and induce IFNα production in healthy pDCs. The addition of previously CpG-stimulated healthy donor B cells expressing ARID3a and IFNα to autologous pDCs upregulated IFNα production in the pDCs (**Figure 2** inset) (46). To our surprise, the pDCs that expressed IFNα also co-expressed ARID3a, emphasizing the association between ARID3a and IFNα expression, and demonstrating that mature cells other than B lymphocytes express ARID3a (46). We next queried whether SLE patient pDCs and low-density neutrophils (LDNs) express ARID3a in association with IFNα. Indeed, we observed a very strong correlation between ARID3a and IFNα expression in lupus pDCs (**Figure 2**) (13). However, the ARID3a and IFNα expression in the pDCs did not correlate highly with the SLEDAI scores, suggesting that other factors or cell types more strongly influenced disease activity (13).

LDNs are much more abundant than pDCs in SLE, comprising up to 54% of patient peripheral blood mononuclear cells, and they have been associated with both IFNα production and increases in autoimmune activity (57). Neutrophils exposed to circulating chromatin produce significant amounts of IFNα in healthy individuals and in SLE patients (58). We found that SLE LDNs also showed increased levels of ARID3a protein expression compared with LDNs from healthy controls (**Figure 2**) (13). ARID3a was also strongly associated with IFNα expression in SLE LDNs, and was, surprisingly, more closely associated with increases in SLEDAI scores than IFNα levels (13). These data support a role for ARID3a in IFNα-associated pathologies in SLE and suggest that ARID3a may contribute to other disease-associated activities not directly

correlated with IFNα levels. Indeed, others showed that IFNα levels do not associate well with SLE disease activity (59).

LDNs undergo spontaneous NETosis, a process whereby inflammatory extracellular nucleic acids decorated with histones, enzymes, and antimicrobial peptides capable of entrapping bacteria are extruded from cells, and this occurs at higher rates in LDNs than in normal density neutrophils (57,60,61). This raises the possibility that LDNs may serve as a source for DNA and chromatin that could trigger TLR9 responses and reactivate self-reactive memory B cells (55,60,62-64). Indeed, others showed that NETs that contact neighboring B cells can stimulate nucleotide-sensing TLRs and promote T-independent B cell activation (65). In a lupus mouse model, the depletion of neutrophils reduced the splenic B cell numbers, serum BAFF, and anti-dsDNA IgG levels, and immune complex deposits in the kidneys (66). In addition to NET production, LDNs and other neutrophils possess immunomodulatory properties capable of directly and indirectly stimulating autoimmune B cells (67-70). B-helper neutrophil (N_{BH}) subsets identified in the spleen by Puga et al. (65), produce the B-cell stimulating factors BAFF, APRIL (A Proliferation-Inducing Ligand), and IL-21, and parallel functions of T cells. N_{BH} cells also exhibit high rates of spontaneous NETosis, making them oddly similar to LDNs (65); however, these cells have not been examined for ARID3a expression. Because ARID3a is over-expressed in lupus pDCs, LDNs, and B cells (13,44,46), we hypothesize that ARID3a contributes to interactions between these three cell types in lupus. Others found that bone marrow-derived neutrophils generate IFNα and stimulate early B cell precursors (71). Therefore, inflammatory events in LDNs associated with ARID3a expression may precede the activation of autoimmune anti-self B cells in lupus. Indeed, recent studies indicate that transitional bone marrow B cells,

the precursors of mature B cells in patients with SLE, already have an IFN and autoimmune signature (41).

Gene expression data from pDCs and LDNs with high levels of ARID3a protein provide new insights into genes associated with autoimmunity and increases in SLEDAI scores. Transcriptomic analyses of pDCs stratified by ARID3a protein expression levels identified 189 genes associated with both ARID3a and IFNα expression in pDCs, and only 122 differentially expressed genes with significant R^2 values in LDNs (13). These genes were tightly co-regulated in both cell types suggesting that ARID3a is either tightly associated with a master regulator of gene transcription, or that it is a master regulator itself (13). Only nine differentially expressed genes were shared between LDNs and pDCs, in keeping with the cell-type specific effects of ARID3a on gene pathway regulation (**Figure 3**B,C) (13). In addition, because ARID3a protein expression was more tightly associated with increases in the SLEDAI scores in LDNs than in pDCs, we postulate that the relatively small number of genes associated with ARID3a expression, but not with IFNα, will identify new mediators of disease activity in SLE. The ARID3a-associated LDN genes include eight transcription factors and two epigenetic factors that may function downstream of ARID3a, as well as a number of noncoding RNAs with potential regulatory functions.

4. ARID3a and Nephritis

While there is no direct evidence that ARID3a is associated with the development of nephritis in SLE patients, there are a number of interesting observations that suggest it could be. *NANOG* and *SOX2* are two genes frequently associated with ARID3a expression (**Figure 3**B), and with cell fate commitment (72-74). These two genes play important functions in stem cells, and along with *OCT4*, may be regulated by ARID3a (11,12). In *X. laevis,* ARID3a is

expressed in kidney stem cells, where it binds to enhancers of the genes required for the regeneration of nephric tubules and changes histone 3 lysine 9 (H3K9me3) levels, allowing the expression of the *LHX1* gene critical for tubule formation(72,75). Our unpublished data also reveal ARID3a expression in human adult kidney progenitor cells, suggesting it may play a role in human nephrogenesis as well. Additionally, resident renal cells secrete IFNα in a lupus nephritis mouse model (75), but it is not known whether IFNα secretion in these cells is associated with ARID3a expression. One might envision that the over-expression of ARID3a within kidney cells could alter gene expression patterns, contributing to inflammation and the autoimmune complexes observed in SLE that ultimately result in renal dysfunction.

5. ARID3a and Hematopoiesis

ARID3a is also expressed in a number of hematopoietic progenitors (27,76,77), and is required for B lineage development in both mouse and man (14,76). The knockdown of ARID3a in human cord blood leads to increases in myeloid lineage development, with associated reductions in the B lymphoid lineage (76). Although the precise mechanisms of ARID3a function in stem cells have not been fully elucidated, knockout mice die between days 12 and 14 of gestation when hematopoiesis moves from the yolk sac to the fetal liver (14). Homozygous knockout embryos exhibit 90% depletion of hematopoietic stem cells (HSCs), suggesting ARID3a is critical for normal numbers of HSC development, and embryos without ARID3a are deficient in erythrocyte development, perhaps explaining lethality (14). HSCs are included in the hematopoietic stem and progenitor cells (HSPCs), a heterogeneous population of cells that consists of both primitive progenitor cells (HSCs, and multipotent progenitors (MPPs)), and committed progenitor cells (multi-lymphoid progenitors or MLPs and multi-myeloid progenitors or MMPs), and these cells are ultimately responsible for generating all mature hematopoietic

cells, including B cells, pDCs, and LDNs (Figure 4). HSPCs from SLE patients have been proposed to be dysfunctional and exhibit defects in proliferation (78,79), as well as deficiencies during hematopoietic transplantation for SLE, which result in the re-emergence of disease and/or engraftment failure (80,81).

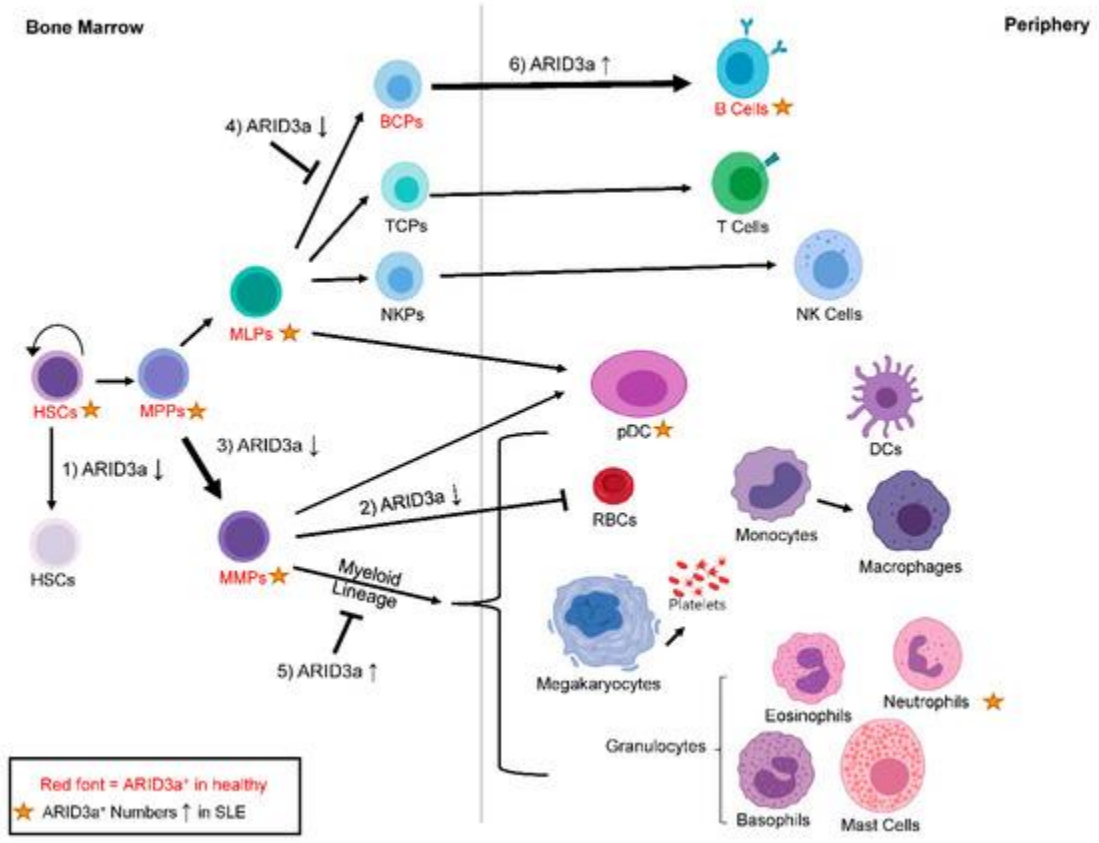


Figure 4. ARID3a is expressed in healthy and SLE hematopoietic progenitors, and the levels of ARID3a affect lineage decisions. Cell subsets with ARID3a expression in healthy individuals are indicated with the red font, while gold stars indicate populations with increased numbers of ARID3a-expressing cells in SLE patients. The effects of increased or inhibited levels of ARID3a (arrows up or down) on HSCs (1), RBCs (2) early myeloid lineage development (3), early B lymphoid development (4), late myeloid lineage development (**5**), and late B lineage development (6) are shown by thicker arrows for increases, or with a symbol showing that development is blocked. HSCs—hematopoietic stem cells; MPPs—multipotent progenitors; MMPs—multi-myeloid progenitors; MLPs—multi-lymphoid progenitors; TCPs—T cell progenitors; NKPs—natural killer cell progenitors; pDCs—plasmacytoid dendritic cells; RBCs—red blood cells.

Our studies suggest that SLE HSPCs are also associated with the autoimmune dysregulation of ARID3a. ARID3a is expressed at varying frequencies in a subset of cells in all defined progenitor populations of healthy controls (76,77) (Figure 4). In SLE patients, increased numbers of ARID3a^{+} cells existed in all progenitor subpopulations compared to healthy controls, although the total number of circulating hematopoietic progenitors did not differ between SLE patients and healthy individuals (77). This suggests that ARID3a expression is abnormal, even in early hematopoietic precursors in SLE patients. Other studies found slightly reduced numbers of circulating HSPCs in SLE patients (78,79), but they did not assess ARID3a expression. When HSPCs from SLE patients were transplanted into immunodeficient mice, the engraftment and development were similar between the samples with low numbers of ARID3a-expressing cells and those with high numbers of ARID3a-expressing cells (77), suggesting that ARID3a over-expression does not affect engraftment potential. However, all of the mice that received HSPCs with high numbers of ARID3a-expressing cells generated human ANAs, while only 16% of the mice that received HSPCs with low numbers of ARID3a-expressing cells generated ANAs (77), again associating ARID3a levels with autoantibody production. As mentioned above, SLE patient HSPCs that have increased numbers of ARID3a-expressing cells show increased frequencies of ARID3a expression in all progenitor populations, suggesting that if ARID3a is induced in the early populations of cells, it is either maintained throughout development, or that it is continually induced in maturing subsets.

There are few mechanistic studies defining the differences between healthy and SLE HSPCs. Moonen et al. reported that circulating hematopoietic progenitors, which also have the capacity to repair vascular damage associated with atherosclerosis, exhibit reduced functional capacity in SLE patient-derived samples, but they did not assess hematopoietic development

(78). We found that SLE HSPCs with low numbers of ARID3a$^+$ cells proliferated less well compared with both healthy control and SLE samples with increased numbers of ARID3a$^+$ HSPCs, possibly because of the lower expression levels of IL7R required for cell proliferation (77). Importantly, HSPCs can also respond to TLR9 ligands in vitro, with increased expression of both ARID3a and IFNα ((46), unpublished data). IFNα signaling in HSCs induces proliferation and exit from quiescence, and under chronic conditions, reduces the numbers of quiescent HSCs in the bone marrow pool (82,83). However, others reported that HSCs under chronic IFNα exposure return to quiescence rapidly after initial exposure (84). Because IFNα can act autonomously on the cells that produce it (85), the associated outcomes of exposure are likely to be diverse.

The importance of ARID3a for B lineage development in mice was demonstrated in multiple systems, and has increased our understanding of ARID3a function (14,29,32,34-36). Several studies revealed that ARID3a expression in B lineage progenitors drives development toward B1 lineage over B2 lineage (29,34). Interestingly, when ARID3a was overexpressed in murine bone marrow progenitor B cells, the B1a cells that do not require the expression of a surrogate light chain for generation were increased, with those cells showing autoreactive BCR repertoires (29). Pre-B and immature B cells also exhibited increased expression of *MYC* and *BHLHE41*, and showed decreased expression of *SIGLEC-G* and *CD72* compared with wild type progenitor cells (29). The reduction of *SIGLEC-G* and *CD72* was hypothesized to contribute to the autoreactive nature of the B1 cells generated, as both are negative regulators of BCR signaling, while the increases in *MYC* and *BHLHE41* expression likely promote the survival of autoreactive B cells (29). These data provide mechanistic insights for the

autoantibody production in ARID3a transgenic mice and imply that ARID3a expression in B cells from SLE patients may also contribute to the autoantibody phenotype.

ARID3a inhibition in human HSCs also resulted in an increased expression of the myeloid lineage-associated transcription factors *CEBPB* and *CSF1*, as well as a reduction in *TCF3*, the gene for E2A (76). Furthermore, when ARID3a was overexpressed in cord blood HSPCs, myeloid lineage cell development was impaired (76). The HSPCs from SLE patients with fewer ARID3a-expressing cells produced fewer B lineage cells in culture compared to both healthy donor cells and SLE HSPCs with high numbers of ARID3a-expressing cells (77). Conversely, SLE samples with increased numbers of ARID3a-expressing HSPCs showed in vitro expansion equivalent to healthy controls, with increased B lineage cell development (77). Together, these data suggest that alterations in ARID3a levels during early hematopoiesis in SLE patients could have profound effects on mature cell phenotypes and lineage pathways, as shown in Figure 4, even before known B cell tolerance checkpoints occur.

Elegant studies by a number of laboratories reveal that multiple hematopoietic cell types, which we have not discussed here, are likely to play important roles in SLE pathogenesis, including T cells, macrophages, basophils, and innate lymphoid cells (86,87). Hematopoiesis is an inter-regulatory process, so these cells may also be affected differentially through interactions with cells that express ARID3a. Indeed, ARID3a expression modifies gene expression pathways in a cell-type specific fashion in the cells in which it is expressed, based on our transcript analyses (13,46,50,76). To date, we have not observed ARID3a expression in any T cell or monocyte subset; however, our unpublished data show ARID3a transcripts in a subset of NK cells. Our data suggest that ARID3a can contribute to hematopoietic lineage decisions, perhaps resulting in skewing of cell frequencies. Skewing of mature hematopoietic subsets is a common

characteristic of SLE (44,57,88-91), and although we have not definitively shown that ARID3a contributes to that skewing, the data certainly suggest that possibility.

6. Regulation of ARID3a Expression

Although ARID3a expression levels clearly change in both healthy inflammatory responses and in multiple SLE cell types, the presumed extracellular triggers that induce ARID3a expression have not been clearly elucidated. Many differentially expressed genes in SLE have been identified through genome-wide associated studies (GWAS), and are associated with single nucleotide polymorphisms (SNPs) that provide a genetic component to the risk of developing the disease (92). Although SNPs can be found in the coding regions of ARID3a, they do not affect the amino acid sequence, likely because defects in ARID3a function are embryonic lethal in multiple organisms, including mice (14,93,94). Nearly 80% of the SNPs associated with lupus are located in non-coding DNA (92). Therefore, it is possible that ARID3a may interact differentially with the regulatory regions affected by these SNPs. This is an area of current interest. A number of studies suggest that epigenetic alterations, including changes in histone acetylation, are associated with SLE (95). ARID3a has been shown to co-precipitate with histone deacetylases in mice (73), but it is unclear whether it is associated with deacetylases in human SLE cells. Interestingly, in mice, the inhibition of histone deacetylases was effective in reducing autoantibody responses (96). Clearly, additional studies are required in order to determine how ARID3a regulates other genes and how it is regulated.

Chromatin and other nuclear antigens exposed by NETs are highly opsonized, can activate IFNα responses through TLR9, and have been postulated to trigger autoimmune responses in multiple inflammatory diseases (62,63,97). Our data indicate that healthy B lineage cells and hematopoietic progenitors respond to TLR9 engagement by increasing ARID3a

expression ((46), unpublished data). In early B lineage cells, ARID3a is down-regulated by miRNAs of the 125 family (98), and it is likely to be similarly regulated in pDCs and LDNs, because transcripts for ARID3a are present in healthy control pDCs and LDNs in the absence of detectable ARID3a protein (13,29). Let-7 regulates ARID3a levels in mouse hematopoiesis, and controls the switch from fetal to adult B lineage development (29,34), but the role of Let-7 in human B lymphopoiesis is not known. The dysregulation of miRNAs has also been implicated as a contributing factor for disease pathogenesis in SLE (99), and it is possible that ARID3a levels are upregulated in multiple hematopoietic cells in SLE as a consequence of deficiencies in the appropriate regulatory RNAs. Additional studies will be required in order to determine whether ARID3a overexpression in SLE is the result of increased activation, decreased regulation, or a combination of both processes.

7. Clinical Implications

The heterogeneous nature of SLE pathogenesis and the lack of understanding of the underlying mechanisms that cause the disease limit therapeutics to those that treat symptoms. Treatments often include glucocorticoids and broad immunosuppressive therapies that act to diminish immune responses as a whole and perturb normal immune functions against disease. Our data show increased expression of ARID3a in B cells, pDCs, and LDNs from patients with SLE, and associate that increased expression with increases in disease activity. Nothing is known regarding the function or expression of ARID3a in other autoimmune diseases. In a limited patient sample, we did not observe ARID3a over-expression in the B lymphocytes of patients with rheumatoid arthritis (44). However, ARID3a-expressing cells have been observed in colon cancer (100) and B cell lymphomas (101). The miRNA Let-7 may regulate ARID3a expression in some cell types, and the dysregulation of Let-7 has been associated with a number of different

cancers, where ARID3a expression might also occur (102). It will be critical to determine in those cases if ARID3a expression is intrinsic to the cancer tissue, or if it is the result of infiltrating immune cells. Many diseases are expected to yield activated immune cells that may express ARID3a in a higher abundance than in healthy tissues.

Our data show strong associations between ARID3a expression and production of IFNα. IFNα is an inflammatory cytokine proposed to play a major role in SLE, and it is currently the target of several clinical trials that indicate that targeting IFNα responses may be beneficial (103-106). Our data in B lymphocytes suggest that ARID3a may function upstream of IFNα (46). Therefore, we propose that new therapies targeting ARID3a directly may be beneficial for the treatment of SLE. Because ARID3a over-expression in SLE is limited to a percentage of cells within any given cell type (i.e., B cells, LDNs, and pDCs), one might imagine that the reversible inhibition of this protein could provide directed therapies that would inhibit the ARID3a-expressing subsets of cells without impairing the normal immune responses in cells that do not express ARID3a. It should be noted that there is no evidence that increases in ARID3a are causally related to SLE disease pathogenesis. However, the plethora of associations of increased ARID3a expression with disease activity in multiple cell types provides strong circumstantial evidence that ARID3a has a major role in SLE pathogenesis, and that ARID3a inhibition may have significant therapeutic benefits in this difficult to treat disease.

Author Contributions

J.G. and M.D.B. contributed to the conceptualization and writing of the manuscript. M.L.R. contributed to the writing of the manuscript. C.F.W. supervised the writing and contributed to the conceptualization and writing of the manuscript.

Funding

This work was supported by funding from the National Institutes of Health (AI118836 and AI123951 (CFW), T32 AI007633 (JG), and K99AG055717 (MLR)).

CHAPTER 2

ARID3a GENE PROFILES ARE STRONGLY ASSOCIATED WITH HUMAN INTERFERON ALPHA PRODUCTION

Michelle L. Ratliff[a1], JoshuaGarton[b1], Lori Garman[cd], M. David Barron[e], Constantin Georgescu[c], Kathryn A. White[c], Eliza Chakravarty[c], Jonathan D. Wren[cf], Courtney G. Montgomery[cd], Judith A. James[cg], Carol F. Webb[aeh]

a Department of Medicine, Oklahoma City, OK, USA

b Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK, USA

c Arthritis and Clinical Immunology Program, Oklahoma City, OK, USA

d Division of Genomics and Data Sciences, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA

e Department of Microbiology and Immunology, Oklahoma City, OK, USA

f Department of Biochemistry, Oklahoma City, OK, USA

g Department of Pathology, and Cell Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA

h Department of Cell Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA

[1] These authors contributed equally to the manuscript

Acknowledgements

I would like to acknowledge Dr. Michelle Ratliff, a co-first author on this paper for her contributions in conducting experimental design, data interpretation, and writing. I would like to also acknowledge other co-authors, Dr. Lori Garman, Dr. Constantin Georgescu, and Dr. Kathryn White for their contributions to writing R scripts and correlation analysis. Dr. Eliza Chakravarty and Dr. Judith James were gracious enough to characterize and provide human samples. Finally, I would like to acknowledge Dr. Carol Webb for her contributions to conceptualization, resources, data interpretation, and writing. Studies were supported by AI118836 (CFW), K99AG055717 (MLR), GM110766 (CGM), GM103636 (JDW) and US4GM104938 and P30AR053483 (JAJ).

Copyright Information

The original version of this work was published in the Journal of Autoimmunity on October 5th, 2018. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). The published work is reproduced directly with changes throughout the chapter to merge formatting and referencing with the rest of this work.

Allocation of Contribution

This chapter is reproduced directly from a published article (13) that was created from a collaboration between multiple labs and has many co-authors. The major contributors were myself and Dr. Michelle Ratliff. The RNA-seq data presented in **Figure 7, Figure 8,** and **Figure 9** was preprocessed and analyzed by me. Dr. Constantin Georgescu aided in writing R scripts to combine expression values from multiple samples into a single gene by sample matrix. Dr. Lori Garman and Dr. Kathryn White aided in correlation analysis of the RNA-seq data.

ABSTRACT

Type I interferons (IFN) causes inflammatory responses to pathogens, and can be elevated in autoimmune diseases such as systemic lupus erythematosus (SLE). We previously reported unexpected associations of increased numbers of B lymphocytes expressing the DNA-binding protein ARID3a with both IFN alpha (IFNα) expression and increased disease activity in SLE. Here, we determined that IFNα producing low density neutrophils (LDNs) and plasmacytoid dendritic cells (pDCs) from SLE patients exhibit strong associations between ARID3a protein expression and IFNα production. Moreover, SLE disease activity indices correlate most strongly with percentages of ARID3a+ LDNs, but were also associated, less significantly, with IFNα expression in LDNs and pDCs. Hierarchical clustering and transcriptome analyses of LDNs and pDCs revealed SLE patients with low ARID3a expression cluster with healthy controls and identified gene profiles associated with increased proportions of ARID3a- and IFNα-expressing cells of each type. These data identify ARID3a as a potential transcription regulator of IFNα-related inflammatory responses and other pathways important for SLE disease activity.

INTRODUCTION

ARID3a (A-T rich interacting domain 3a) is a DNA-binding protein that modulates gene expression, increasing immunoglobulin gene expression in B lymphocytes and repressing expression of other genes in neonatal fibroblasts (11,12,23). We previously described increases in the number of circulating ARID3a+ B lymphocytes in SLE patients, and found that these cells are associated with increases in disease activity indices as defined by SLEDAI scores (SLE disease activity indices) (44). Others found that B lymphocytes produce Type I IFNs during development and in response to infections (107-109). We have also shown that induction of ARID3a protein in healthy B lymphocytes results in Type I IFN production (46), suggesting an unexpected link between ARID3a and IFN. Furthermore, SLE patient B lymphocytes with high levels of ARID3a expression also exhibited increases in IFN signature gene expression compared to healthy controls and SLE B cells with low ARID3a expression levels (46,50). Together, these data suggest that ARID3a expression in both SLE and healthy B cells is associated with IFN production.

In healthy individuals, Type I IFNs are essential for immune responses against intracellular pathogens, including viruses that trigger anti-DNA and anti-RNA responses (110). Virtually all cells can produce and respond to IFNs, with differing cell type-specific responses that can be detrimental (111). Increased plasma Type I IFN, mainly IFNα, occurs in approximately half of adult SLE patients, and is associated with increased disease activity (112-116). IFNα expression in autoimmune diseases results in different gene signatures than those produced by viral infections (117), and dysregulated IFNα levels in SLE activate a panel of IFN-regulated genes, commonly referred to as an IFN signature (118). Transcriptomic analyses of whole blood from pediatric SLE patients further defined this gene signature (119). Although

increases in IFNα accelerated disease onset (120-122), and depletion of major subsets of IFNα-producing cells ameliorated autoimmunity in several mouse models (123,124), the reasons for elevated IFNα expression in SLE patients is unknown. Several clinical trials using treatments that inhibit IFN production as therapies for SLE are ongoing, [reviewed in ref. (125). Thus, understanding underlying mechanisms associated with increased IFNα production is important.

Plasmacytoid dendritic cells (pDCs) are thought to be a major source of IFNα in SLE patients, [reviewed in ref. (126). Other cell types, including neutrophils (55,127,128), also contribute to IFNα production in SLE. A population of low density neutrophils (LDNs) present in the peripheral blood mononuclear fraction of pediatric SLE patient samples have enhanced capacity to synthesize IFNα (55), contributed to tissue damage in adult SLE patients (57,129), and were also associated with increased disease activity in SLE (57,61,130). We demonstrated that healthy, ARID3a+ B lymphocytes induced both IFN production and ARID3a expression in autologous pDCs (46). These data led us to hypothesize that ARID3a might be a biomarker for IFN production in pDCs and other cell types, including LDNs.

Nothing is known regarding ARID3a expression in pDCs and LDNs, including whether its expression is correlated in these cells with increased IFN production. Therefore, we assessed levels of intracellular ARID3a and IFNα in pDCs and LDNs from healthy controls and SLE patients with a wide range of disease activity. ARID3a protein levels correlated with IFNα levels in both pDCs and LDNs. To identify genes associated with ARID3a expression in LDNs and pDCs, RNA-seq was performed and expressed genes were correlated with ARID3a protein levels. Our data define gene profiles associated with ARID3a and IFN expression in both pDCs and LDNs and implicate ARID3a as a regulator of innate immune responses in these cells.

MATERIALS AND METHODS

Patients and healthy controls

Healthy age and sex-matched controls and patients who met a minimum of four American College of Rheumatology Classification Criteria for SLE (131) were recruited after informed consent from the Oklahoma Medical Research Foundation (OMRF) Oklahoma Lupus Center in accordance with OMRF (IRB compliance #06-19) and OUHSC (IRB compliance 5946) Institutional Review Board approvals and in accordance with the Declaration of Helsinki. SLE patients (127) ranging in age from 21 to 73 (96% female, SLEDAI 0 to 8, Table 1) and 11 healthy controls were recruited for this study. All patients were under treatment regimens at blood draw (Table 1).

Table 1. Study subject demographics.

Subject	Race	Gender	Age	SLEDAI	Treatment
SLE 1	C	F	71	4	HCQ
SLE 2	C	F	73	3	HCQ
SLE 3	C	F	62	0	HCQ, MTX
SLE4	C	F	71	0	AZA
SLE 5	AA	F	51	6	HCQ, RTX
SLE 6	C	F	43	8	PDN,Q
SLE 7	AA	F	34	5	HCQ
SLE 8	NA/H	F	29	4	AZA, HCQ
SLE 9	AA	F	32	4	HCQ
SLE 10	AA	M	56	6	PDN, AZA, HCQ
SLE 11	H	F	39	0	MTX, HCQ, PDN
SLE 12	A	F	52	0	AZA
SLE 13	C	F	36	1	AZA, HCQ
SLE 14	C	F	34	2	MTX, HCQ, PDN
SLE 15	AA	F	18	0	PDN, AZA, HCQ
SLE 16	A	M	35	6	HCQ
SLE 17	AA	F	21	2	MMF, PDN, HCQ
SLE 18	C	F	28	2	PDN, HCQ
SLE 19	C	F	45	6	HCQ, RTX
SLE 20	AA	F	42	2	PDN, HCQ
SLE 21	C	F	64	2	PDN, HCQ, HC
SLE 22	AA	F	59	6	MTX, HCQ
SLE 23	AA	F	50	4	MMF, ADA, HCQ
SLE 24	AA	F	36	4	PDN, HCQ
SLE 25	AA	F	53	4	HCQ, RTX, PDN
SLE 26	C	F	55	2	MTX, HCQ, B
SLE 27	C	F	62	0	PDN, AZA, HCQ
C1	C	F	48	N/A	
C2	C	F	55	N/A	
C3	C	F	35	N/A	
C4	C	F	46	N/A	
C5	C	M	57	N/A	
C6	C	F	50	N/A	
C7	C	F	34	N/A	
C8	AA	F	63	N/A	
C9	AA	F	40	N/A	
C10	C	F	44	N/A	
C11	C	F	33	N/A	

SLE patients (SLE) and healthy controls (C) are listed. Races are: A- Asian, AA- African American, C- Caucasian, H-Hispanic, NA - Native American. Age at time of blood draw and gender (F-female, M-male) are indicated. Disease activity indices (SLEDAI scores) and drug treatments are indicated for patients. Drugs: HCQ-hydroxychloroquine, MTX-methotrexate, AZA- Azathioprine, RTX- Rituximab, PDN-Prednisone, HC- Hydrocortisone, ADA- Humira, B- Benlysta, MMF-Cellcept, Q-Quinacrine.

Sample preparation

Total peripheral blood mononuclear cells (PBMCs) from patients and healthy controls were isolated from heparinized peripheral blood with Ficoll Paque Plus (GE Healthcare, Fisher Scientific cat# 45-001-749). Two million PBMCs were stained for flow cytometric analysis and the remaining cells were either cryopreserved for later use or were immediately enriched as described below for either LDNs or pDCs. Patient plasma samples were stored at −80 °C until assay. LDNs were isolated for RNA analyses from either cryopreserved or freshly isolated PBMCs using the EasySep Human Neutrophil Enrichment Kit (cat# 19257, Stemcell Technologies) according to manufacturer instructions. Isolation of pDCs was performed using the human Plasmacytoid Dendritic Cell Isolation Kit II (cat# 130-097-415, Miltenyi Biotec) according to manufacturer instructions using LD columns (cat# 130-042-901, Miltenyi Biotec). Patient samples chosen for RNA analyses were based on population cell numbers and ARID3a expression levels identified by flow cytometric evaluation of total PBMCs.

Flow cytometry analyses

pDCs were defined as CD3−CD20−CD56−CD11c−CD123+CD304+ [35,36]. LDNs were defined as CD3−CD20−CD56−CD14−CD15+CD16+CD16b+ [32]. The use of CD16b was used as a marker to ensure exclusion of NK cells (132) that may express ARID3a (our unpublished data). Appropriate matching isotype controls from Biolegend were used for gating. Following surface marker staining, cells were fixed with fixation buffer (BD Biosciences),

permeabilized with Transcription Factor Fixation/Permeabilization Buffer kit (eBioscience) and stained for IFNα and ARID3a. IFNα-Phycoerythrin (Clone LT27:295, cat# 130-092-601) recognizes the majority of the IFN-α subtypes, but not IFN-α2b. Human-specific anti-ARID3a antibodies were generated in goats by our laboratory against a peptide sequence from the amino terminal portion of ARID3a (G-R-G-R-E-G-P-G-E-E-H-F-E), and were purified over a peptide column, verified by western blot and mobility shift against in vitro translated human ARID3a and B cell nuclear extracts containing ARID3a (23,28). A rabbit anti-goat Fluorescein (Cat# 6160-02) was used as the secondary antibody. Doublet exclusion was used to ensure analyses of single cells prior to forward/side scatter gating. Data were collected using an LSRII (BD Biogenics) and FACSDiva (BD Biosciences) software version 4.1 or Stratedigm S1200Ex and CellCapTure acquisition software and were analyzed using FlowJo (Tree Star) software version 10. Specific antibodies used were: human lineage markers CD3 Pacific Blue (clone UCHT1, cat# 300434), CD20 Pacific Blue (Clone 2H7, cat# 302330), and CD56 Pacific Blue (Clone 5.1H11, cat# 362552), CD11c-Brillian Violet 605 (Clone 3.9, 301636), CD16-Allophycocyanin-Cyanin7 (Clone 3G6, cat#302018), CD14-Allophycocyanin (Clone 63D3, cat#367118), CD304-Brilliant Violet 510 (Clone 12C2, cat# 354515), CD15-Brilliant Violet 605 (Clone W6D3, cat# 323032), and CD19-Phycoerythrin-Cyanin5.5 (Clone HIB19, cat# 302210) from Biolegend, CD123-Allophycocyanin-Vio770 (Clone AC145, cat#130-104-196) from Miltenyi Biotec, and CD16b-Alexa fluor700 (Clone 245514, cat# FAB1597N) from R&D Systems.

RNA-seq

RNA from enriched LDNs and pDCs was isolated using the MagMAX mirVana Total RNA Isolation Kit (Applied Biosystems, cat# A27828) according to manufacturer instructions. RNA concentrations were measured with an Impen Nanophotometer. RNA samples were

prepared for sequencing using the Ovation RNA-Seq v2 (NuGEN Technologies) kit and libraries were constructed using Ovation Ultralow Library System V2. Library concentration and fragment size distribution was determined using Agilent High Sensitivity D1000 kit on an Agilent 2200 TapeStation (Agilent Technologies). Paired-end (2 × 75bp) sequencing was performed on a NextSeq platform using SBS v2 chemistry.

Fastq files were analyzed using FastQC and low quality reads and sequencing adapters were trimmed using Trimmomatic (133). A Bowtie index of raw fastq files was created based on the UCSC knownGene (hg38) transcriptome, and paired-end reads were aligned directly to this index using default parameters (134). RSEM (135) was run using default parameters on the aligned reads to estimate gene expression levels. Library quality metrics, such as genomic mapping rates, library and the fraction of ribosomal RNA in each library were calculated. Two SLE pDC, one healthy control pDC, and one SLE neutrophil samples were excluded due to low library size and poor alignment. Differential gene expression was analyzed using limma (136). Unsupervised hierarchical clustering was performed on differentially expressed genes (FDR < 0.05). Pathway analyses were performed using Ingenuity Pathway Analysis (Qiagen). Heatmaps in Fig. 8 were constructed of all genes that significantly correlated with % IFNα+ or % ARID3a+ cells by Spearman's correlation (unadjusted $p < 0.05$). All correlation coefficients of all top genes with respect to each other and % protein-expressing cells were then plotted using the unsupervised clustering heatmap R package corrplot.

Plasma IFNα activity

The WISH endothelial cell line (ATCC, CCL-2; gift from S. Kovats) that expresses IFNα receptors, but cannot use endogenous IFN pathways, was used to measure IFN-responsive gene activation [43,44]. WISH cells (50,000 cells/well) were cultured 1:2 with SLE patient or control

plasma, or recombinant human IFNα2a (Invitrogen, cat# 111001) at 75U per well, in DMEM supplemented with 10% FBS for 6 h at 37 °C prior to lysis for RNA isolation with Tri-reagent (Sigma, cat# T9424). RNA concentration was determined by Implen Nanophotometer. cDNA synthesis was performed using iScript™ gDNA Clear cDNA Synthesis Kit (Bio-Rad, cat# 1725034). qPCR was performed on a Bio-Rad CFX96 real time PCR machine with the following program: 1: 1 min at 95 °C, 2: 10 s at 95 °C, 3: 30 s at 57 °C, repeat 2 & 3 40 times, hold at 4 °C. Primers for IFI44 and HPRT were purchased from IDT (137). IFI44 forward primer: 5′CTC GGT GGT TAG CAA TTA TTC CTC 3′, reverse primer: 5′AGC CCA TAG CAT TCG TCT CAG 3'. HPRT forward primer: 5′ TTG GTC AGG CAG TAT AAT CC 3′, reverse primer: 5′GGG CAT ATC CTS CAA CAA AC 3'. Data were normalized to HPRT1. Fold-increases were calculated relative to the values of IFI44 for unstimulated cells.

Data analyses

Simple linear regression analyses of SLEDAI scores and proportions of cells producing ARID3a and IFNα were performed by best fit linear regression using Prism (Graphpad) version 7 ($p < 0.05$ indicates the slope of the best-fit line differs from 0). Data was log-transformed where indicated to satisfy assumptions of normality. For RNA-seq data, genes with FDR-adjusted p values of less than 0.05 from Limma were considered differentially expressed. Genes that were expressed at a transcript per million (TPM) of greater than one in at least one sample were kept for downstream analysis. TPM values were log transformed. For correlations with protein expression, TPM values were correlated with % ARID3a+ or % IFNα+ cells using Spearman's Correlation (R statistical software, p values were not adjusted). A subset of those genes was also analyzed by linear regression. Stepwise additive and subtractive multiple linear

regression models of SLEDAI were performed independently for pDCs and LDNs using the R function stepAIC within the MASS package.

Data sharing statement

RNA-seq data are publicly available through the GEO NCBI database under the accession number GSE117836. For original data, contact carol-webb@ouhsc.edu.

RESULTS

Association of ARID3a and IFNα protein expression in LDNs and pDCs

Healthy controls and SLE patients with a range of SLEDAI scores from 0 to 8 were recruited (Table 1). Total PBMCs were subjected to flow cytometry using surface markers to identify LDNs and pDCs, and intracellular staining identified IFNα and ARID3a protein expression (Fig. 5A, B). T lymphocytes typically do not express ARID3a or IFNα and served as internal controls for intracellular staining integrity for each PBMC sample (Fig. 5C). Healthy controls showed fewer than 20% ARID3a and IFNα-expressing LDNs (Fig. 5D). However, ARID3a protein expression correlated strongly with IFNα expression ($R2 = 0.64$, $p < 0.0001$) in LDNs of SLE patients (Fig. 5D). Similarly, in pDCs from SLE patients, ARID3a and IFNα protein expression were strongly correlated ($R2 = 0.81$, $p < 0.0001$), while healthy controls had few cells expressing either protein (Fig. 5E). Interestingly, increased percentages of ARID3a-expressing LDNs were found in SLE patients with increased disease activity indices (SLEDAI scores) by univariate linear regression (Fig. 5F), while disease activity was less strongly associated with percentages of ARID3a+ pDCs (Fig. 5G). Surprisingly, while ARID3a and IFNα protein levels are correlated in both pDCs and LDNs, ARID3a expression in LDNs is more significantly associated with disease activity scores.

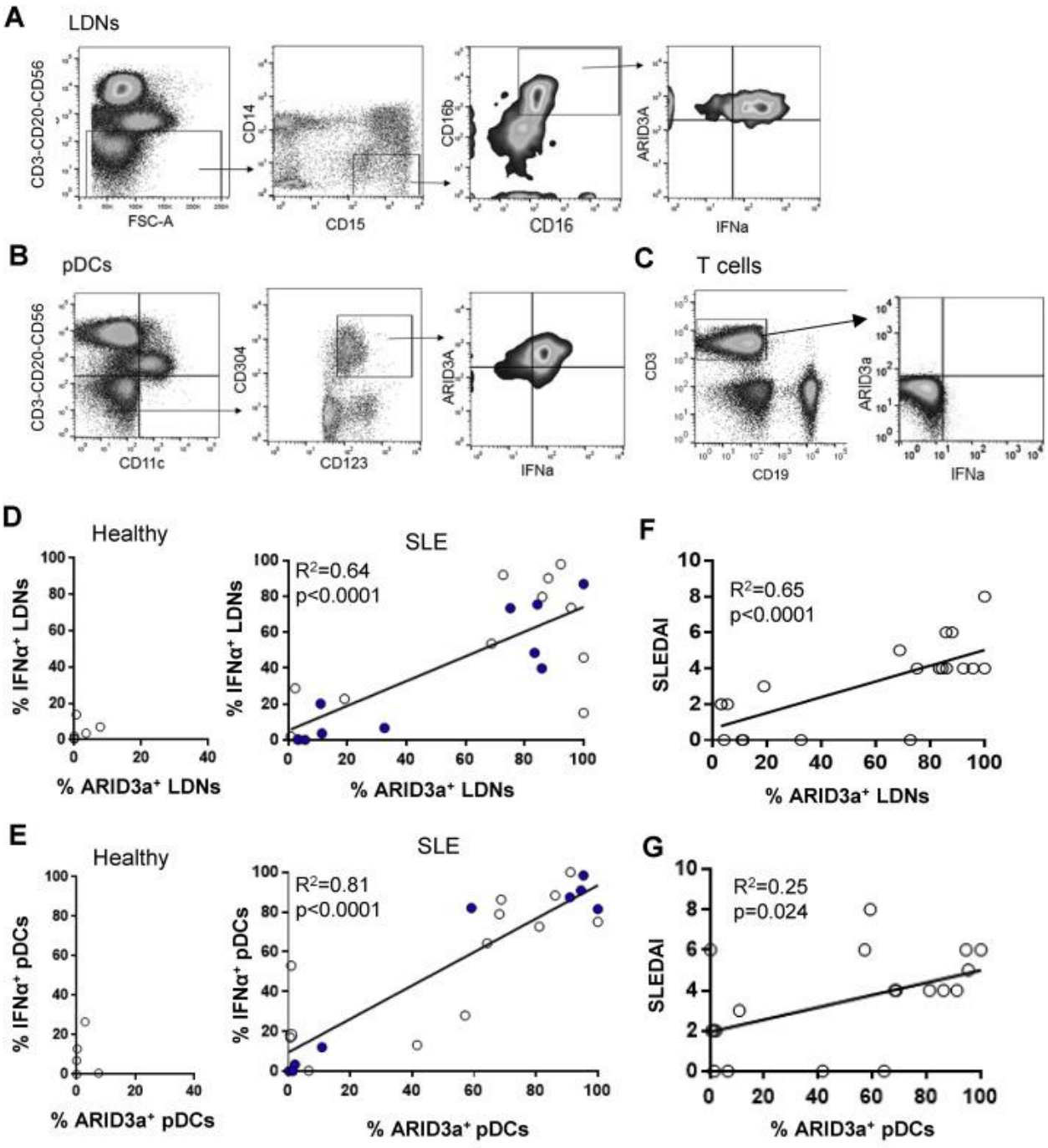


Figure 5. ARID3a expression in SLE LDNs and pDCs is correlated with IFNα production, and with disease activity in LDNs. Flow cytometry of healthy control (n = 9) and SLE patient LDNs (CD3−CD20−CD56−CD14−CD15+CD16+CD16b+) and pDCs (CD3−CD20−CD56−CD11c−CD123+CD304+) (n = 23) were evaluated for intracellular ARID3a and IFNα protein expression. Representative gating strategies for LDNs (A), pDCs (B) and internal control T cells (C) are shown. Associations between %ARID3a+ and %IFNα+

LDNs (D) and pDCs (F) were analyzed by linear regression. Associations between patient SLEDAI scores at the time of blood draw and % ARID3a+ LDNs (E) and pDCs (G) were determined by linear regression analyses. R^2 and p values are presented. Each point represents an individual patient sample. Solid points are samples used for later transcriptome analyses.

IFNα expression is weakly associated with disease activity

We next assessed associations of IFNα protein expression levels in LDNs and pDCs with increased disease activity scores. Only poor correlations with SLEDAI scores were observed for IFNα expression in the LDNs (Fig. 6A), unlike what we observed for ARID3a (Fig. 5F). Furthermore, IFNα expression was only weakly associated with SLEDAI scores in pDCs (Fig. 6B). Evaluation of IFN signatures and IFNα plasma activity (represented by IFI44 gene expression) revealed weak associations with numbers of IFNα-expressing LDNs or pDCs (Fig. 6C and D). Likewise, IFNα plasma activity was only weakly associated with numbers of ARID3a-expressing LDNs or pDCs (Fig. 6E and F). Similar results were found using the IFIT1 gene (not shown). To further examine the potential influence of numbers of ARID3a- or IFNα-producing cells on SLEDAI, we performed additive and subtractive step-wise multiple linear regression on all SLE samples using R. Multiple regression was performed independently for LDNs ($n = 19$) and pDCs ($n = 20$). SLEDAI was associated with % ARID3a+ LDNs ($R^2 = 0.65$, $p = 0.00003$), while % IFNα+ LDNs did not contribute significantly to the predictive effect of the model. In contrast, while % IFNα+ pDCs contribute significantly to the model predicting SLEDAI ($R^2 = 0.29$, $p = 0.015$), % ARID3a+ pDCs does not. Therefore, ARID3a was more strongly associated with disease activity in LDNs than in pDCs.

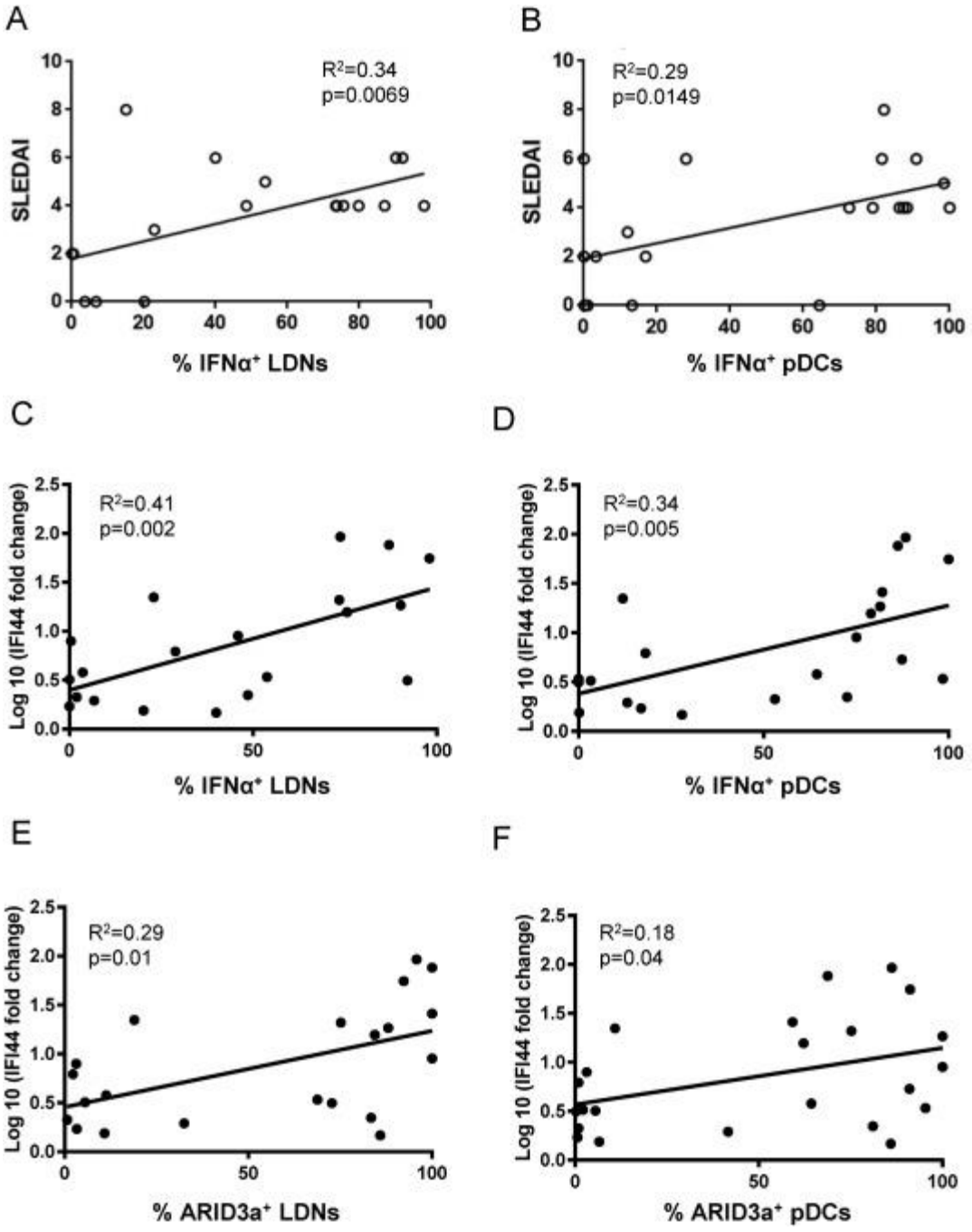


Figure 6. IFNα only weakly associates with disease activity and ARID3a expression. SLE samples were analyzed for associations between SLEDAI scores and % IFNα+ LDNs (A) and pDCs (B) via linear regression. Plasma IFNα activity, via induction of the IFNα signature gene IFI44 was measured by qRT-PCR, and log-transformed values were tested for association with % IFNα+ (C,D) and %ARID3a+ (E,F) LDNs (C,E) and pDCs (D,F) via linear regression. R^2 and p values are given.

High ARID3a expression is associated with distinct gene profiles in LDNs and pDCs

To better understand how ARID3a, a transcription regulator, contributes to disease activity, RNA-seq analyses were performed on isolated samples of LDNs and pDCs from healthy controls and SLE patient samples with a wide range of ARID3a protein expression. ARID3a is an intracellular protein, so it is not possible to sort for ARID3a+ cells without damaging RNA integrity. In addition, pDCs and LDNs are non-abundant subsets that are notoriously short-lived. To assess the feasibility of performing transcriptome analyses with these cell types, we undertook a pilot study using cryopreserved PBMCs from 3 healthy samples, 3 SLE samples with low numbers of ARID3a+ cells (<40%, Fig. 5D, solid dots), and 3 SLE samples with high numbers of ARID3a+ cells (>70%). Samples that failed to meet library quality control criteria were eliminated from analyses. In both LDNs and pDCs, high-ARID3a SLE samples clustered together by gene expression, while low-ARID3a SLE samples clustered with healthy controls and were indistinguishable by gene expression profiles (Fig. 7A). Unlike our findings in B lymphocytes where ARID3a transcripts correlate closely with ARID3a protein levels (46,50), all LDN and pDC samples expressed ARID3A transcripts regardless of SLE or control status. These data were confirmed with freshly isolated cells from 2 healthy controls, and 2 each of high and low ARID3a-expressing SLE samples. As in cryopreserved samples, fresh SLE samples with low numbers of ARID3a-expressing cells clustered with healthy controls using unsupervised hierarchical clustering of gene expression patterns (Fig. 7B). These data suggest that ARID3a, as a transcription regulator, could directly or indirectly contribute to differences observed in gene expression patterns in SLE patients.

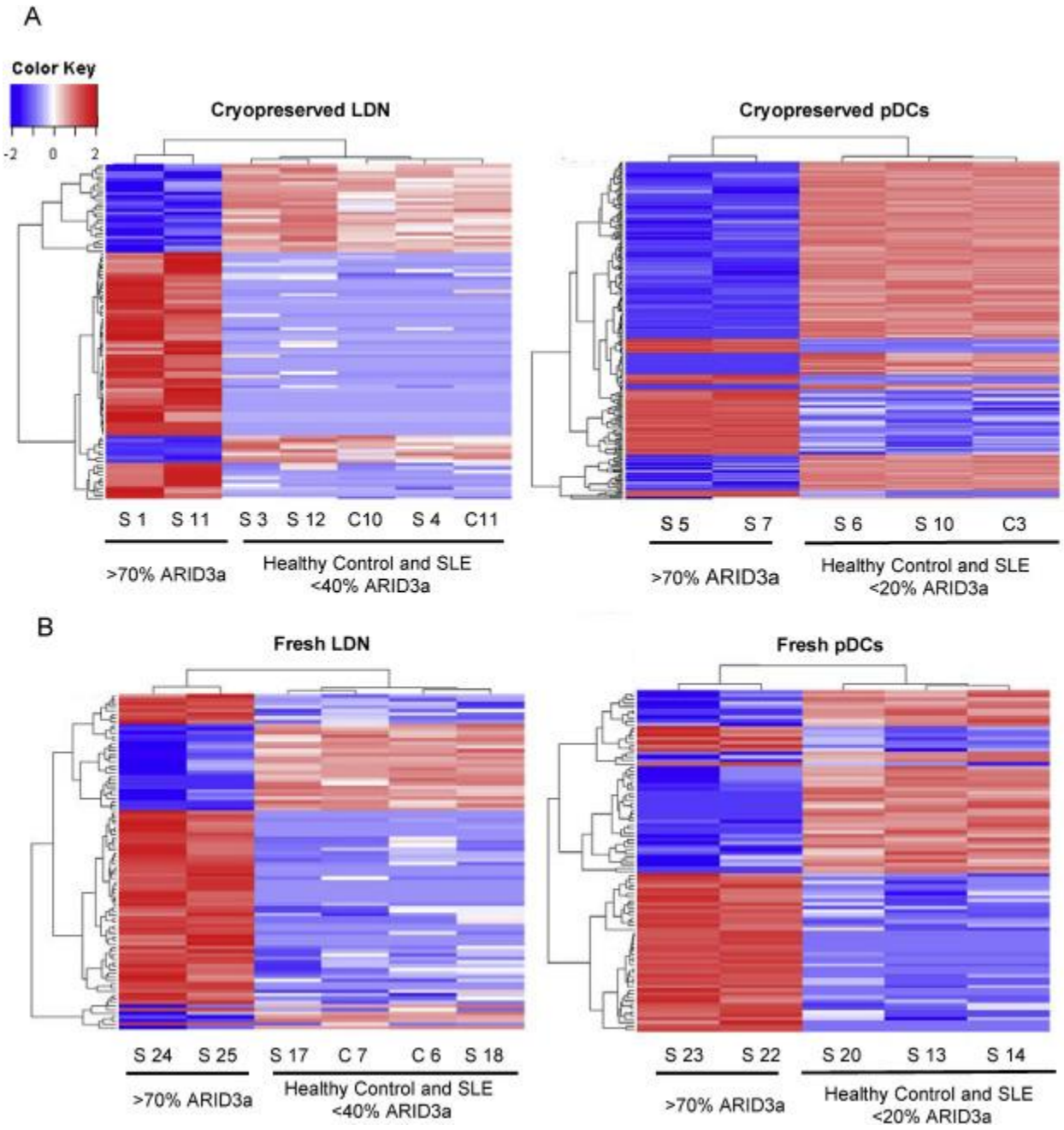


Figure 7. Gene expression patterns of SLE LDN and pDCs samples with low ARID3a protein levels resemble healthy controls. Gene expression patterns obtained from RNA-seq of healthy control and SLE patient LDN and pDC samples with a wide range of ARID3a protein expression (solid points in Fig. 1D, E) were subjected to unsupervised hierarchical clustering of genes with FDR adjusted p values < 0.05. Heat maps of top differentially expressed genes from cryopreserved (A) and freshly isolated (B) LDNs and pDCs are shown. Patient (S) and control (C) samples are indicated with relative levels of ARID3a protein expression.

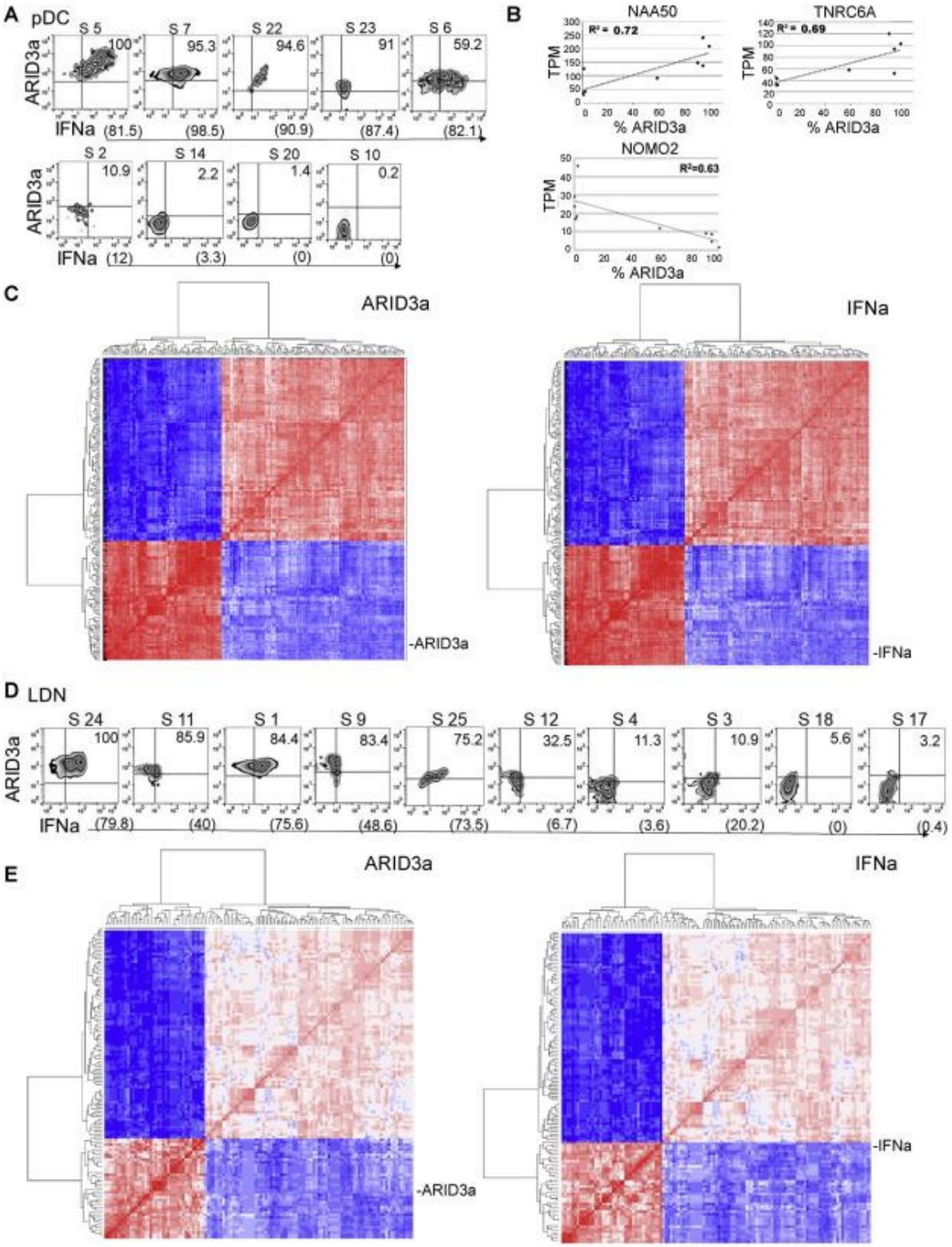


Figure 8. Cells expressing both ARID3a and IFNα protein exhibit co-regulated gene profiles in pDCs and LDNs. Transcriptomes from 9 SLE pDC samples with varying levels of ARID3a and IFNα protein expression as shown by flow cytometry (A) were evaluated for associations between TPM and % ARID3a+ or % IFNα+ cells by Spearman's correlation. (B) Examples of linear regression analyses of 3 representative genes are shown with R^2. (C) Genes

significantly ($p < 0.05$) correlated with expression of both ARID3a and IFNα in pDCs were clustered via heatmap with % ARID3a+ pDCs by Spearman's correlation coefficient value (left panel); similarly, those genes were clustered with numbers of IFNα+ pDCs (right panel). The analysis was repeated in LDNs (D, E). Color of cells indicates positive (red) or negative (blue) correlation coefficient of gene expression when compared to % ARID3a+ or % IFNα+ cells; intensity indicates strength of coefficient.

ARID3a and IFNα protein levels are associated with distinct gene profiles

Due to constitutive ARID3a transcript expression in LDNs and pDCs, determining how ARID3a is associated with changes in gene expression in Fig. 7 required analysis of data in association with ARID3a protein levels. Therefore, Spearman correlation analyses were performed to identify genes associated directly or indirectly with ARID3a or IFNα protein levels, as determined by flow cytometry (Fig. 8A). Variable total numbers of pDCs and LDNs are common in SLE blood samples [30,32,45], consistent with our findings. Hierarchical clustering analyses were used to cluster the samples according to protein expression with transcriptome data from nine samples of SLE patient pDCs with ARID3a protein frequencies ranging from 100% to virtually undetectable levels. In pDCs, 592 genes were significantly correlated with ARID3a protein expression levels (Spearman's correlation, unadjusted, $p < 0.05$), while 680 genes were found to be correlated with IFNα protein levels in the same samples (unadjusted $p < 0.05$). Three examples of genes associated with % ARID3a+ pDCs (Fig. 8B) have potential roles in gene regulation. In the pDCs, 189 genes were significantly associated with both ARID3a and IFNα protein expression, and heat maps representing the pairwise correlation coefficients (Spearman's correlation) among genes correlated with either ARID3a or IFNα expression are shown in Fig. 8C. These data reveal two tightly correlated clusters of genes that appear to be up- or down-regulated in tandem, suggesting the presence of genetic profiles tightly correlated with ARID3a protein expression. Similar analyses were performed using transcriptome data obtained

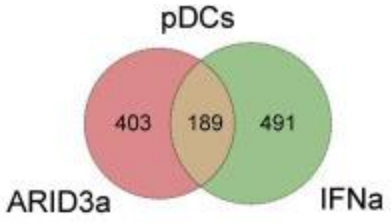
from 10 samples of SLE patient LDNs that ranged in ARID3a protein expression levels from 100 to 3.2% (Fig. 8D). These data show 223 genes correlated with ARID3a protein expression in LDNs, while 1552 genes are correlated with IFNα protein expression in these same cells (unadjusted $p < 0.05$). Genes correlated with both ARID3a and IFNα protein expression (122 genes) are shown in association with ARID3a or IFNα in hierarchical clustered heat maps (Fig. 8E). Two major groups of up- or down-regulated genes are visible when plotted by association with ARID3a (left panel) or IFNa (right panel) for both pDCs and LDNs (Fig. 8C and E). Therefore, ARID3a may be linked to other master regulators that control expression of large groups of genes, or it may itself function as a master regulator.

A subset of genes are associated with both ARID3a and IFNα expression

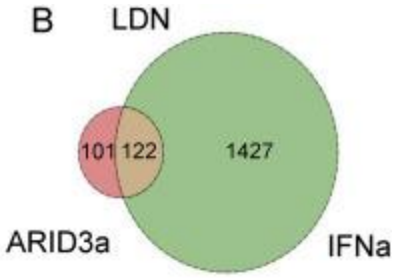
Numbers of genes significantly associated with ARID3a and/or IFNα protein levels ($p < 0.05$) are presented as Venn diagrams for both pDCs (Fig. 9A) and LDNs (Fig. 9B). Overlap in the Venn diagrams showing genes associated with both ARID3a and IFNα protein expression are the genes depicted in the heat maps in Fig. 9C and E. Genes most strongly associated with both ARID3a and IFNα protein expression in pDCs with absolute r values of 0.7 or greater are listed (Fig. 9C). Similarly, genes with absolute r values of 0.6 or greater that correlated with both ARID3a and IFNα protein levels in LDNs are indicated (Fig. 9D). Gene Ontology enrichment analyses of gene sets correlated with both ARID3a and IFNα in both cell types are significantly enriched for sulfite oxidation and metabolic degradation pathways. The interferome webtool (http:interferome.org) also validated associations of these genes with interferon pathways. Genes associated with ARID3a expression alone in pDCs are enriched for IL17 antiviral responses. The genes associated only with ARID3a expression in LDNs are enriched for other transcription factors (31%), long non-coding RNAs and miRNAs, RNA-binding proteins and enzymes that

modify chromatin. These data are consistent with ARID3a functions as a master transcription regulator.

A



B



C

pDCs

Gene	ARID3a r	Gene	ARID3a r	Gene	ARID3a r
VIPR1-AS1	-0.92	*GIN1*	-0.76	*LINC00702*	0.72
IL17RB	-0.91	*E2F1*	-0.76	*NF1*	0.72
THNSL1	-0.90	*FAM47E-STBD1*	-0.76	*TFDP2*	0.72
PCDHGB3	-0.89	*CACNB3*	-0.75	*RASAL2-AS1*	0.73
GDI2P2	-0.87	*FAH*	-0.75	*ATP6V1E1*	0.74
EID3	-0.87	*AVEN*	-0.75	*KIAA0100*	0.74
KIAA1614	-0.87	*IGKV6-21*	-0.75	*TOR1AIP1*	0.74
NCLP1	-0.87	*CTLA4*	-0.75	*TTC17*	0.74
FAM21FP	-0.87	*CHST15*	-0.74	*MGAT4EP*	0.74
NOMO2	-0.86	*TNS4*	-0.74	*FOXE1*	0.74
SUOX	-0.85	*CHCHD4*	-0.74	*AKT3*	0.74
H2AFY2	-0.85	*ST6GALNAC2*	-0.74	*PPP3CB-AS1*	0.74
WNT4	-0.85	*FAT2*	-0.73	*SRR*	0.74
C17orf77	-0.85	*ANK1*	-0.73	*ASB7*	0.75
AMN	-0.84	*CAGE1*	-0.73	*CAPZA2*	0.75
LCA5	-0.84	*EXOC3-AS1*	-0.73	*NEMP1*	0.75
SNORD97	-0.83	*OSGEPL1-AS1*	-0.72	*CCDC14*	0.76
WRB	-0.83	*PTPN13*	-0.72	*PWWP2A*	0.76
PREX2	-0.82	*POTEC*	-0.72	*TSC22D2*	0.76
TRIB1	-0.82	*ATRN*	-0.72	*PIP5KL1*	0.76
NUP62CL	-0.82	*CAD*	-0.72	*ZNF99*	0.77
ZMYND15	-0.82	*TLDC1*	-0.72	*GPR141*	0.77
SYNGAP1	-0.82	*MARCKSL1P1*	-0.72	*HIVEP3*	0.77
OXCT2	-0.81	*ENAM*	-0.71	*PREPL*	0.77
HIPK1-AS1	-0.81	*TTC30B*	-0.71	*NTRK2*	0.78
GGT7	-0.81	*C16orf86*	-0.71	*DDHD2*	0.79
TMEM132A	-0.80	*IQCE*	-0.71	*CCL2*	0.80
PCDHGB5	-0.80	*KLRG1*	-0.71	*CCDC114*	0.81
CDC20	-0.79	*NAB2*	-0.71	*TC2N*	0.81
RPS17P13	-0.79	*CYP2D6*	-0.71	*GTF2E2*	0.82
STARD7-AS1	-0.79	*SDC1*	-0.70	*TMEM189*	0.82
PIPSL	-0.78	*LINC00680*	-0.70	*EIF4ENIF1*	0.82
AGFG2	-0.78			*TRAF6*	0.82
MELK	-0.78	*CRYZL1*	0.70	*CCNE2*	0.83
GATA6	-0.78	*FAM91A1*	0.70	*HDX*	0.83
TROAP	-0.77	*TNRC6A*	0.70	*IPO9*	0.83
C10orf128	-0.77	*PRLR*	0.71	*OR51E2*	0.84
MIR600HG	-0.77	*ANAPC13*	0.71	*DGKH*	0.85
NBPF13P	-0.76	*ATP6V1D*	0.71	*NHSL2*	0.85
CAPN10-AS1	-0.76	*WDR45B*	0.71	*NAA50*	0.88
C4A-AS1	-0.76	*RBMS3-AS2*	0.71	*SFXN1*	0.89
C4B-AS1	-0.76	*AC005562.1*	0.71	*ZC3H8*	0.89
CDK8	-0.76	*RPS4XP14*	0.72		

D

LDN

Gene	ARID3a r
CLLU1	-0.87
DDAH1	-0.75
MAP3K7CL	-0.74
BMS1P2	-0.74
XCL2	-0.70
RBM43	-0.70
TJP2	-0.68
SLAMF9	-0.67
MRPL35	-0.66
CORO7-PAM16	-0.66
FAM81A	-0.65
HCG25	-0.65
CYP1B1	-0.64
RPS10-NUDT3	-0.63
LPCAT2	-0.63
DGKZ	-0.63
GCSAML	-0.62
CXCL17	-0.61
FOLH1	0.60
CFAP57	0.60
TMEM191B	0.61
CPLX3	0.61
ZFP91-CNTF	0.61
GAP43	0.61
GOLGA8F	0.62
CAPN11	0.63
CYP4B1	0.63
FAM167A	0.64
CLDN4	0.66
NPM1P12	0.66
NPTX1	0.66
RPL36AP33	0.68
LAMC2	0.68
ITGA11	0.69
TGIF2P1	0.71
C6orf10	0.72
AC105342.1	0.72
AC012066.1	0.72
PRRX1	0.73
LINC00487	0.75
ERICH3	0.81

Figure 9. Distinct gene subsets are associated with both ARID3a and IFNα protein in pDCs and LDNs. Venn diagrams depict genes associated with ARID3a and IFNα protein levels in pDCs (A) and LDNs (B). Genes that correlated with both ARID3a and IFNa protein expression in the Venn diagrams above with absolute ARID3a associated r values > 0.7 in pDCs (C) and with absolute ARID3a-associated r values > 0.6 in LDNs (D) are listed.

DISCUSSION

The ability to segregate SLE patients with high and low disease activity would be clinically beneficial. We found that the transcription regulator ARID3a is expressed in LDNs and pDCs, and that expression of ARID3a is associated with IFNα production in those cells. Surprisingly, high numbers of ARID3a-expressing LDNs are also associated with increased disease activity indices in SLE. Transcriptome analyses of patient samples with broad ranges of ARID3a expression revealed that samples with the highest percentage of ARID3a-expressing cells showed distinct gene profiles compared to SLE samples with low numbers of ARID3a-expressing cells which clustered with the healthy control samples for both pDCs and LDNs. Finally, gene expression profiles suggest that ARID3a is a regulator of inflammatory pathways involved in SLE.

While increased plasma IFNα has been associated with increased disease activity (112-116), others noted that IFNα gene signatures were not associated with longitudinal changes in disease activity (59). Similarly, we found weak correlations between IFNα-expression and increased disease activity in both LDNs and pDCs, and were surprised that ARID3a expression is more highly associated with disease activity in LDNs. Our data suggest that IFNα may also contribute to disease activity, but are consistent with data suggesting that factors other than IFNα production contribute to SLEDAI scores in SLE (57,129,138,139).

Our data identify genes associated with both ARID3a and IFNα protein expression. Interestingly, the large majority of genes that are correlated with ARID3a and/or IFNa production differ between pDCs and LDNs, indicating that ARID3a expression identifies gene profiles that are cell type specific. The large numbers of genes that are coordinately up- and/or down-regulated as a group in association with ARID3a expression in both pDCs and LDNs suggest that ARID3a may act as a master gene regulator, or that it is closely associated with other master gene regulators. In LDNs, thirty-one percent of the genes associated with variation in ARID3a protein expression are transcription factors, including SOX2 and NANOG. Other ARID family proteins affect large numbers of genes epigenetically, [reviewed in ref. (7,140). In single cell analyses of the K562 cell line, ARID3a is associated with distinct regulatory states and chromatin configurations (141). More recently, ARID3a was shown to bind near the edges of enhancer regions in many cell types (142), suggesting it may also function epigenetically.

We did not observe upregulation of any of the 13 IFNA subtypes in our analyses, despite the presence of IFNα protein in both LDNs and pDCs. The levels of individual IFNA transcripts may be below the level of detection using RNA-seq technology. IFNA detection typically requires quantitative PCR analyses. Therefore, RNA-seq data have limitations that may preclude detection of all transcripts differentially expressed in association with ARID3a protein expression. Another limitation of RNA-seq data highlighted by our study is that RNA levels do not always correlate with protein levels. Others demonstrated that ARID3a is regulated by miRNAs in early hematopoiesis (34,98). These studies suggest that regulation of ARID3a by miRNAs is cell type-specific. We speculate that miRNA regulation of ARID3a also occurs in pDCs and LDNs, and that miRNA control of protein expression would allow rapid responses to extracellular signals potentially explaining the discrepancy between ARID3A transcripts and

protein levels in these cells. Our RNA-seq data did not detect the miR125b and lin28b miRNAs associated with ARID3a regulation in early B lymphocyte progenitors. Further experiments will be required to determine if other differentially expressed miRNAs with potential ARID3a target sites may function in neutrophils and pDCs.

Our data implicate ARID3a as an important regulator of inflammatory responses and IFNα production in several cell types, and identify gene profiles associated with ARID3a expression in pDCs and LDNs from SLE patients. Most surprising was the finding that increased disease activity in SLE is more strongly associated with frequencies of ARID3a+ LDNs than with IFNα+ cells. While a major function of pDCs is cytokine secretion, LDNs also undergo NETosis (57,129). Therefore, ARID3a may regulate genes important for disease activity, but that are distinct from IFNα regulatory pathways.

We do not know whether ARID3a regulates IFN production, or if IFNα production stimulates ARID3a expression in pDCs and LDNs. However, in B lymphocytes, developmentally less mature cells expressed ARID3a without IFNα, and induction of ARID3a transcripts occurred prior to detection of IFNα transcripts (46). Furthermore, inhibition of ARID3a in a B cell line also inhibited IFNα expression (46). Regardless of which protein is upstream in pDCs and LDNs, a large number of genes appear to be co-regulated in correlation with ARID3a and IFNα. These data strongly suggest that ARID3a is associated with other master regulators of gene expression, and that it could be mechanistically involved in innate immune responses in SLE. We speculate that ARID3a and its gene targets may offer new therapeutic approaches.

Author contributions

CFW – conceptualized studies; MLR, CFW- designed experiments; MLR, JG, CFW – wrote manuscript; MLR, JG, MDB – performed research; MLR, JG, LG, CM, CG, KW, CFW – analyzed data; MLR, JG, LG, CM, JW, JAJ, CFW – interpreted data; LG, CM, MDB, EC, JAJ – edited manuscript; EC, JAJ-provided human samples; EC- characterized patients.

Acknowledgements

We thank the Clinical Genomics Core Facility and Quantitative Analysis Core at Oklahoma Medical Research Foundation and the Flow Cytometry Core Facility at OUHSC; Drs. P. Gaffney, G. Wiley and R. Pelikan for helpful discussions; M. Shankar for manuscript preparation; Drs. M. B. Humphrey and J. Metcalf for manuscript review. Studies were supported by AI118836 (CFW), K99AG055717 (MLR), GM110766 (CGM), GM103636 (JDW) and US4GM104938 and P30AR053483 (JAJ).

Chapter 2 Addendum

The following manuscript (13), Ratliff ML, Garton J, Garman L, et al. ARID3a gene profiles are strongly associated with human interferon alpha production. *J Autoimmun*. 2019;96:158-167. doi:10.1016/j.jaut.2018.09.013, provided an important opportunity to learn how to analyze bulk RNA-seq data by performing differential expression and correlation analyses on SLE patient samples. My contributions were to process and perform differential expression analysis on plasmacytoid dendritic cells (pDCs) and low density neutrophils (LDNs) from SLE patients with varying levels of both disease activity and ARID3a protein. Both pDCs and LDNs are important cell types where ARID3a is overexpressed in SLE patients and is associated with interferon expression.

Bulk RNA-seq analyses were performed on two separate batches of pDCs and LDNs from SLE patients. The first batch was processed from frozen samples while the second batch was processed the day the samples were collected. Briefly, the raw sequencing files were processed by removing sequencing adapters and low quality reads from each sample using Cutadapt with the following code:

cutadapt -a ADAPTER_FWD -A ADAPTER_REV -o out.1.fastq -p out.2.fastq reads.1.fastq reads.2.fastq

Where ADAPTER_FWD represents 5' Nextera XT sequencing adapters:

GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCATCTCGTATGCCG TCTTCTGCTTG, CTGTCTCTTATACACATCT, AGATGTGTATAAGAGACAG, TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG, GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

And ADAPTER_REV represents the 3' Nextera XT sequencing adapters:

AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGT ATCATT, CTGTCTCTTATACACATCT, AGATGTGTATAAGAGACAG, CTGTCTCTTATACACATCTGACGCTGCCGACGA, CTGTCTCTTATACACATCTCCGAGCCCACGAGAC

Reads.1.fastq and reads.2.fastq represent the forward and reverse raw sequencing reads.

This code results in the generation of two fastq files, now designated out.1.fastq and out.2.fastq, for both the forward and reverse reads that no longer contain Illumina Nextera XT sequencing adapters. This allows for more accurate mapping of raw reads to the reference human genome.

Next, a Bowtie2 index was generated based on the UCSC knownGene transcriptome and paired-end reads were aligned directly to this index using Bowtie2. All alignments were performed on OUs Supercomputing Center for Education and Research (OSCER) supercomputer.

bowtie2-build genomes.fna *Homo_sapiens*

This code creates a database of indexed .bt2 files that will be used to align the paired end reads using RSEM (135). Next, RSEM, which uses the Bowtie2 index described above, was employed on the OSCER supercomputer using default parameters to obtain transcripts per million (TPM) values for each gene:

```
rsem-calculate-expression -p 8 \
--paired-end \
--bowtie2 –bowtie2-path /opt/oscer/software/Bowtie2/bin \
--estimate-rspd \
--append-names \
--output-genome-bam \
/scratch/jgart10/sample1_R1.fastq \
/scratch/jgart10/sample1_R2.fastq \
/home/jgart10/Homo_sapiens/RSEM_GRCh38ref_topleveldna \
/scratch/jgart10/<Sample name>/<Sample name>_trim_RSEM_expression
```

The code above generates a gene x sample matrix and contains all estimated expression values (i.e expected counts, fragments per kilobase of transcript per million mapped reads (FFPKM), and transcripts per million (TPM), for each gene that had reads mapping to their respective gene. However, this code generates the expression matrix for each sample individually. Since it is unreasonable to combine gene expression values for each sample

manually, the following code was written in R to automate this process and generate a single file containing TPM values all samples:

```
dr="C:/Users/jgart10/Desktop/SLE/"
fls=list.files(dr,"results")
fOneSpl=read.table(paste(dr,fls[1],sep=""),header=T,sep="\t")
bigTable=data.frame(fOneSpl[,1:4])
flsN=sapply(fls,function(x) strsplit(x,"_")[[1]][1])
for(j in 1:length(fls)){
fOneSpl=read.table(paste(dr,fls[j],sep=""),header=T,sep="\t")

jj=match(as.character(bigTable[,"transcript_id.s."]),as.character(fOneSpl[,"transcript_id.s."]))
 cat(j,sum(is.na(jj)),"\n")
mat=fOneSpl[jj,c("expected_count","TPM","FPKM")]
colnames(mat)=paste(flsN[j],colnames(mat),sep="__")
bigTable=data.frame(bigTable,mat)
}
bigTable=data.frame(rowId=paste("r",1:nrow(bigTable),sep=""),bigTable)
write.csv(bigTable,file=paste(dr,"allSamples_RSEM_toplevel_expression.genes.results",sep=""),row.names=F,quote=T)
```

The above code generates a single file containing the expected counts, FPKM, and TPM values for all samples. To obtain only the TPM values for all of the samples into a single file, the following code was used:

```
bigMat=bigTable[,c(2,grep("_TPM",colnames(bigTable)))]
write.csv(bigMat,file=paste(dr,"allSamples_RSEM_TPM",sep=""),row.names=F,quote=T)
```

Next, genes that had expression values of 0 for all samples were removed to reduce noise in downstream analyses and then TPM values were log transformed:

```
jjj=which(apply(bigMat[,-1],1,function(x) sum(x==0))==(ncol(bigMat)-1))
```

```
bigMat=bigMat[-jjj,]
bigMat[,-1]=apply(bigMat[,-1],2,function(x) log(x+1))
```

Genes with expression values of TPM >1 in half of the samples were retained for downstream analyses. Differential expression analysis was performed using limma (136) within R using the following code:

```
library(limma)
mat=bigMat[,grep("TPM",colnames(bigMat))]
prtQF=factor(c("Ctr","Prot")[1*(prtQ>0)+1],levels=c("Ctr","Prot"))
jjc=match(tolower(sapply(colnames(mat),function(x)
strsplit(gsub("X","",x),"__")[[1]][1])),tolower(sapply(names(prtQ),function(x) gsub("
","",x))))
prtQF=prtQF[jjc]

design <- model.matrix(~prtQF)
fit <- lmFit(mat, design)
fit <- eBayes(fit); names(fit)
range(p.adjust(fit$p.value[,2],"fdr"))
fitM=data.frame(bigMat,tScore=fit$t[,2],pVal=fit$p.value[,2],qVal=p.adjust(fit$p.value[,
2],"fdr"))
colnames(fitM)[match(c("tScore","pVal","qVal"),colnames(fitM))]=paste(c("tScore","pV
al","qVal"),"over70",sep="__")
write.csv(fitM[fitM$pVal__num<0.05,],file=paste(dr,"statResultsSelected.csv",sep=""),r
ow.names=F,quote=T)
```

The above code analyzes differential expression of samples that had ARID3a protein levels greater than 70% and generates a csv file containing the p values, fold change values, and expression values for genes identified to be significant for each sample and was used to make

heatmaps in Figure 7 of this chapter. Hierarchical clustering was performed using R on differentially expressed genes (DEGs) identified by limma.

```
library("ggplot2")
fitM=read.csv(file=paste(dr,"statResultsSelected.csv",sep=""),as.is=T)

jj=sort(which(fitM$pVal__num<0.01));length(jj)

mat=log(as.matrix(fitM[jj,grep("TPM",colnames(fitM))])+1)

colnames(mat)=sapply(colnames(mat),function(x) strsplit(x,"__")[[1]][1])

row.names(mat)=sapply(as.character(fitM$gene_id)[jj],function(x)

heatmap.2(mat,scale="row",trace="none",dendrogram="none",Colv=F,Rowv=F,col=blue
red(41),density.info="none",key.xlab=NA,cexCol=1,cexRow=1,keysize=0.8,main="Sele
cted genes, pval ordered")
```

An unexpected finding from these analyses was that protein levels of ARID3a did not correlate with RNA transcript levels in either neutrophils or pDCs. All samples had near equal expression values of ~2-3 TPM for ARID3a. Even healthy controls without ARID3a protein produced transcripts. Therefore, it was not possible for perform traditional differential expression analysis based on samples expressing ARID3a and those that did not express ARID3a. This then required analyses of gene transcript data in association with ARID3a or IFNa protein levels determined by flow cytometry. Finally, with the help of Lori Garman and Kathryn White, the expression files generated were used to find genes that correlated with ARID3a protein levels in pDCs and LDNs from SLE patients.

CHAPTER 3

ARID3a INHIBITION AFFECTS CHROMATIN ACCESSIBILITY REQUIRED FOR HUMAN FETAL GLOBIN EXPRESSION

Joshua Garton [1], Malini Shankar[2], Brittany Chapman[2], Kira Rose[3], Patrick M. Gaffney[4] and Carol F. Webb[2,3,5*]

[1]Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK, 73019, USA

[2]Departments of Medicine, [3]Microbiology and Immunology, and [5]Cell Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, 73104, USA
[4]Genes and Human Disease Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, 73104, USA

Present address-Kira Rose - Department of Chemistry, University of Central Oklahoma, Oklahoma City, OK, 73034, USA

* To whom correspondence should be addressed: Carol F. Webb
ORCID: 0000 0003-4164-9634
Tel: 405-271-4188
Email: carol-webb@ouhsc.edu
Mailing Address: Department of Medicine, OUHSC
800 Research Park, Ste. 419
Oklahoma City, OK 73104

Acknowledgements

I would like to acknowledge Dr. Malini Shankar and Mrs. Brittany Chapman for their contributions to cell culturing, flow cytometry, and conducting time course assays. I would also like to acknowledge Dr. Patrick Gaffney for helpful discussions about RNA-seq and ATAC-seq data analysis. Finally, I would like to acknowledge Dr. Carol Webb for her contributions to conceptualization, resources, data interpretation, and writing. This work was supported by the National Institutes of Health [AI123951 and AI118836 to C.F.W., T32 AI007633 to J.W.G.].

Copyright Information

The original version of this work was submitted to Nucleic Acids Research on July 22nd, 2020.

Allocation of Contribution

This chapter is reproduced directly from the submitted article. The major contributors were myself and Dr. Malini Shankar. In this dissertation chapter, all figures were generated by myself, with the exception of **Figure 10**, which was produced by Dr. Malini Shankar and Mrs. Brittany Chapman. However, I repeated the experiment in **Figure 10** using the CRISPR-Cas9 ARID3a knockout clones that were used for ATAC-seq. The sequencing experiments, including RNA and nuclei extraction, purification, library preparation was performed by me. Data processing and analysis of both RNA-seq and ATAC-seq data was also performed by myself with helpful discussions with Dr. Patrick Gaffney. CRISPR-Cas9 ARID3a$^{-/-}$ knockout clones were ordered from Synthego. Single cell sorting and isolation of homozygous ARID3a$^{-/-}$ clones and analysis of flow cytometry data was performed by Dr. Malini Shankar. In addition, I performed western blot, PCR, and flow cytometry experiments presented in **Figure 14** to confirm homozygous deletion of ARID3a.

ABSTRACT

ARID3a is an understudied member of a large family of epigenetic regulators associated with both increases and decreases in gene expression in hematopoietic cells. Recent molecular analyses in human erythropoietic systems revealed increases in ARID3a transcript levels implicating potential roles for ARID3a in human erythrocyte development. However, ARID3a transcript levels do not faithfully reflect protein levels in many cell types, and potential functions and requirements for ARID3a during erythropoiesis have not been explored. Therefore, we used the human erythroleukemic cell line K562 as an *in vitro* model system to elucidate functions of ARID3a protein in human erythropoiesis. Stimulation of K562 cells with hemin induced visible globin production and accompanying modifications in gene expression by three days post induction. Knockdown of ARID3a in hemin-stimulated cells inhibited fetal globin production and reduced levels of surface proteins associated with erythroid differentiation. Temporal RNA-seq data link ARID3a expression with the important erythroid regulators GATA1, GATA2 and KLF1. ARID3a inhibition resulted in down regulation of 227 genes that were upregulated in control hemin-stimulated cells. Ablation of ARID3a using CRISPR-Cas9 in K562 cells

confirmed the requirement for ARID3a in globin gene expression in this model system and allowed ATAC-seq analyses of epigenetic structures affected by ARID3a, identifying roles for ARID3a in intergenic regions that may function as enhancers. These data define biologically relevant roles for ARID3a in maintenance of chromatin structure and show the protein is required for human fetal globin gene expression.

INTRODUCTION

Multipotent hematopoietic stem cells (HSCs) and progenitor cells are essential for maintaining appropriate levels of red blood cells and platelets during erythropoiesis (143,144). Although studied extensively, the gene programs that drive erythropoiesis have not been fully elucidated. Decreased levels of hemoglobin gene transcripts and maturational arrest of erythroid lineages in thalassemia's are associated with defects in expression of the transcription factors GATA1, GATA2 and KLF1, as well as with changes in chromatin accessibility of enhancer regions for the globin genes (145). Temporal studies using defined hematopoietic systems reveal the possibility that precise manipulation of these important regulatory events could lead to modulation of specific erythropoietic diseases and regeneration of blood cells (146-150). The genes within the β-globin gene cluster are expressed in a developmentally controlled manner with embryonic globin (HBZ) being expressed first in the yolk sack, followed by fetal globin (HBG1/HBG2) during primitive erythropoiesis, and finally adult (HBB/HBD) globin genes during definitive erythropoiesis (149,151). The erythroid-specific globin gene cluster is regulated by an upstream cis-regulatory region, the locus control region (LCR) (152,153), that binds to specific transcription factors allowing accessibility to the embryonic and adult globin in a developmentally controlled fashion (154-156). However, gene expression studies alone provide limited information about the causative regulators involved.

The A+T rich binding protein ARID3a is a member of a large family of proteins, many of which have important roles as epigenetic regulators (157). Modulation of ARID3a levels can both suppress and enhance gene expression in a cell type-specific fashion (reviewed in (5,158)). ARID3a was identified as a highly differentially expressed gene in primitive erythropoiesis in the mouse (144) and as a potential regulator of hemogenic reprogramming in conjunction with GATA2 through motif associations (150). In addition, single cell ATAC-seq analyses in K562 cells implicated ARID3a as a potential transacting factor marginally associated with epigenomic variability along with GATA1 and GATA2 (141). However, recent data emphasize potential discrepancies in protein versus transcript levels during human erythropoiesis, particularly for GATA1 (159). Our recent data in human granulocytic cells further emphasize the fact that ARID3a transcripts do not always correlate with protein levels (13). Although we previously found that ARID3a-deficient mice die in utero at E12.5 due to failed erythropoiesis and 90% reduction of hematopoietic stem cells (14), suggesting that ARID3a could be important for erythropoiesis in mice, it was unclear if failed erythropoiesis resulted directly from requirements for ARID3a during erythropoiesis or from earlier hematopoietic progenitor defects. Thus, the role of ARID3a protein in human erythropoiesis and epigenetic regulation has not been defined.

Our earlier data indicated that the human monomyelocytic cell line K562 constitutively expresses ARID3a protein (28). This human cell line has long been used as a model for erythrocyte and myeloid lineage development and can be induced with several different stimuli to differentiate into erythroid cells that express high levels of embryonic and fetal globin genes (159-163). Therefore, we used this model system to determine if ARID3a protein is required for human globin gene expression. Furthermore, we used temporal transcriptome analyses and integrated chromatin accessibility data from ATAC-seq to determine how ARID3a affects gene

expression patterns during induction of erythroid lineage differentiation. These results identify new epigenetic functions for ARID3a in human erythropoiesis.

MATERIALS AND METHODS

Cell Culture and Transfection

K562 (ATCC® CCL-243™) cells were plated in triplicate at $1x10^5$ per well in 6-well plates with RPMI 1640 + 7.5% fetal calf serum overnight at 37°C prior to treating cells with 0.04 mM hemin (Sigma), as previously described (164). Cells were harvested at 24, 48 and 72 hours and viabilities were assessed using trypan blue exclusion. To evaluate erythroid lineage differentiation, cells were stained with benzidine to detect globin expression, as reported previously (165). Briefly, cells were resuspended in 25μL phosphate-buffered saline (PBS) and stained at a 1:1 ratio with benzidine solution made with 30% fresh hydrogen peroxide. At least 200 cells were evaluated per replicate. Lentivirus expressing shRNA specific for ARID3a, or an unrelated control shRNA, both of which co-express green fluorescent protein (GFP) allowing visualization of infected cells, were purchased from Genecopoeia, Inc,Rockville,MD and used at a multiplicity of infection of 0.6 to 1.0, as previously described (11). The ARID3a sequence targeted was GCAGTTTAAGCAGCTCTA from exon 2, and does not react with other ARID family members (11). K562 cells were infected with virus 30 minutes to 3 hours prior to stimulation with hemin in the presence of 8 μg/ml polybrene as we have done previously (11). Lentivirus transfection efficiency was assessed via GFP expression using a Zoe Fluorescent Imager, BioRad, on day two and was typically >70%.

Flow Cytometry

Antibodies to the transferrin receptor, an erythroid precursor marker, CD71 APC-Cy7 (Biolegend Cat # 33410) and the glycophorin A erythrocyte marker CD235a PE-Cy7 (Biolegend Cat # 349112) were used for surface staining to evaluate erythroid lineage differentiation. Appropriate isotype controls from Bio Legend were used for gating. Myeloid lineage detection was evaluated using surface markers CD24 APC (Biolegend Cat # 311118) and CD33 PE-Cy5 (Biolegend Cat # 303406). Following surface marker staining, cells were fixed with fixation buffer (Bio Legend Cat # 420801), permeabilized with Foxp3/Transcription Factor Staining Buffer Set (Invitrogen eBioscience Cat # 00552300) and stained for ARID3a with goat anti-human ARID3a peptide-specific antibody, as we described previously (76). Donkey anti-goat IgG PE (Invitrogen Cat# Pl31860) was used as the secondary antibody. Data were collected on a Stratedigm S1200Ex and data were analyzed using FlowJo (Tree Star) software version 10.

RNA-seq and Analyses

Total RNA from triplicate samples treated with and without hemin, ARID3a shRNA, and/or scrambled shRNA was isolated using NucleoSpin RNA XS kits (Macherey-Nagel, Cat # 740902.50). RNA concentrations were measured with an Impen Nanophotometer. RNA integrity numbers were obtained using an Agilent 2200 TapeStation. Library construction was performed as described previously (13). Briefly, the Ovation RNA-Seq v2 (NuGEN Technologies) kit was used to generate sequencing libraries. Paired-end (2 x 50bp) sequencing was performed on a NovaSeq platform. Fastq files were demultiplexed and sequencing adapters were removed using Cutadapt (166). Briefly, we created a Bowtie (134) index based on the UCSC knownGene (167) transcriptome, and aligned paired-end reads directly to this index using Bowtie2. The average sequence depth was 21M reads with an average alignment of 83% mapping to the hg19 genome assembly. Next, we ran RSEM v1.3.0 (135) using default parameters to obtain transcript per

million (TPM) values for each gene. Genes with expression values of TPM > 1 in half of the samples were retained, leaving 11,869 transcripts for downstream analyses. Differential gene expression was analyzed using DESeq2 v3.5 (168). Differentially expressed genes (FDR < 0.05) with fold changes ≥ 2 were used for Ingenuity Pathway Analysis (Qiagen) (IPA). Hierarchical clustering (Euclidean) was performed on differentially expressed genes (FDR value adjusted < 0.05) and heatmaps were generated with the pHeatmap package in R. Principal component analysis (PCA) was performed in R using the prcomp function.

ARID3a knockout

Genome editing of ARID3a was performed via CRISPR/ Cas9 mutation of the K562 cell line contracted through Synthego (Redwood City, CA). Briefly, modified guide RNA ARID3a-932711 (5'-CCTCGTAAGTCCAGTCGCCG-3' [TGG]-PAM) targeting exon 3 was chosen to be specific for ARID3a. A bulk knockout sample of greater than 70% knockout was then single cell sorted via flow cytometry for isolation of homozygous ARID3a knockout clones. Sixty-six clones visually confirmed to have only one cell per well after sorting were allowed to grow, and 19 clones were screened by flow cytometry for ARID3a protein expression. Eight of the 19 clones were then selected as being wild type, or potentially homozygous knockout, and levels of ARID3a expression were confirmed by Western blotting using a commercial ARID3a antibody (mouse monoclonal IgG Catalog# sc-398367, Santa Cruz Biotech). Homozygous colonies and wild type colonies were used for ATAC-seq analyses.

Western Blotting

For protein detection, total cell extracts from $1x10^6$ cells were resuspended in 50µL of Laemmli sample buffer containing 5% 2-mercaptoethanol. Following 5 min boiling at 90 °C, 10µL of extract was loaded onto pre-cast Mini-PROTEAN TGX (BioRad Cat# 456-1093) gel and transfer was done for 1 h on nitrocellulose 0.2µM (BioRad). Membranes were blocked in 1% gelatin in TBST for 1 h at room temperature, as previously described (169). Blots were probed overnight for ARID3a and actin with mouse anti-ARID3a and rabbit anti-β-actin, respectively. Following incubation with primary antibody, blots were washed three times for 10 mins with TBST and probed with secondary antibody for 1 h at room temperature. The secondary antibody for ARID3a was goat anti-mouse IgG and rabbit anti-goat IgG for actin. Blots were then washed three times for 10 min with TBST. Proteins were detected using the AP conjugate substrate kit (BioRad Cat # 170-6432).

ATAC-seq and analyses

ATAC-seq libraries were generated from wild type and ARID3a$^{-/-}$ K562 clones treated with or without hemin. Duplicate samples of 30,000 cells were washed in cold PBS, pelleted by centrifugation and lysed using cold lysis buffer (10mM Tris-HCl, pH 7.4, 10mM NaCl, 3mM $MgCl_2$, 0.1% IGEPAL CA-630). Nuclei were collected by centrifugation and the pellet was resuspended in 50µL transposase reaction mix (25µL 2x TD buffer, 2.5µL transposase (Illumina Nextera FC121-1030 TDE1 and TD buffer) and 22.5µL nuclease-free water). The transposition reaction was incubated at 37°C for 30 min. Samples were cleaned using a MiniElute kit (Qiagen) following manufacturer's protocol and eluted in 10µL buffer EB. Library construction was done by PCR in a reaction mix containing 25µL 2x NEBNext PCR master mix (New England Biolabs), 10µL transposed sample, and 5µL primer mix (1.25µM each of Nextera XT adapter1

and adapter2 primer mix). PCR conditions were 72°C for 5 min, 98°C for 30s, and 11 cycles (98°C for 10s, 68°C for 30s, 72°C for 1 min), ending with 72°C for 5 min. PCR amplified sequencing libraries were cleaned with AMPure Beads (Beckman Coulter). Library quality was determined by analysis on an Agilent TapeStation.

For each sample, 25-99 million 50bp paired-end reads were obtained on an Illumina NextSeq sequencer. All data processing steps were performed within the Partek Flow Genomics Analysis software. Fastq files were processed and both sequencing primers and Nextera transposase adapters were removed using Cutadapt. Trimmed reads were aligned to the hg19 GRCh37 reference genome using Bowtie2v2.2.5 with parameters –very-sensitive -X 2000 (134). Low quality (-Q 30) and duplicate reads, and reads mapping to the ENCODE project blacklist, mtDNA and rDNA genes were removed. MACS2 (170) was used to call peaks for duplicate samples using the parameters -q 0.05 --nolambda --slocal 1000 --llocal 10000 -m 5 50 --shift 0 --extsize 200 --fe-cutoff 1.0.The ATAC peaks of pooled replicate samples were annotated to genomic regions such as transcription start sites(TSS), introns, and exons using RefSeq version 89. Each peak was annotated in relation to these genomic elements and may have multiple gene annotations. DESeq2 v3.5 was run on the ATAC peaks to identify differential chromatin accessibility in wild type vs ARID3a$^{-/-}$ samples with an FDR cutoff of 0.05 (168).

Statistics

Data for viability, benzidine stain and time course were plotted and all statistical analyses were performed using Prism (Graphpad) version 7. A one-way ANOVA was used for comparisons of multiple groups, followed by Tukey posttest for multiple comparison corrections. All statistical tests and corresponding *P* values are stated in the figure legends. *P* values <0.05 were considered significant.

RESULTS

ARID3a knock down inhibits globin production and expression of erythrocyte markers.

The human K562 erythroleukemia cell line can be used as an *in vitro* model of erythroid and megakaryocyte lineage differentiation when stimulated with hemin (171). K562 cells treated with hemin for five days showed visible production of red, hemoglobin-producing cells *in vitro* (Figure 10A). While cells treated with control shRNA (Figure 10A, right) resembled those treated with hemin only, cells that received ARID3a shRNA showed no obvious red cells (Figure 10A, middle). This effect was robustly evident in pelleted cells, and in cells treated with benzidine to enhance globin visibility (bottom, Figure 10A). Time course analyses indicated that near maximal globin production (77.5 %) was achieved in hemin treated cells by day three of treatment, and reduced numbers of globin-producing cells were apparent as early as day one after inhibition of ARID3a (Figure 10B). Inhibition of globin expression in ARID3a-inhibited samples varied slightly but averaged 24% percent on day three (Figure 10C). Viabilities and cell numbers were equivalent on day three in all cultures (Figure 10D), suggesting ARID3a did not function by causing cell death or by inhibiting cell division. These data suggested that most responses necessary for induction of globin occur within the first three days of culture.

Flow cytometric analyses of cells on day three of culture confirmed that ARID3a inhibition resulted in less ARID3a protein (Figure 10E). Furthermore, erythrocyte lineage markers CD71 (TFRC) and CD235a (GYPA) were reduced and enhanced as expected by hemin treatment (Figure 10F). However, cells treated with ARID3a-specific shRNA, with or without hemin stimulation, more closely resembled unstimulated cells with respect to expression of these

surface markers (Figure 10F). Samples treated with ARID3a shRNA alone did not show significant reductions in either CD235a or CD71or CD33 (Figure 10F, G), suggesting that the reduction in these surface markers is due to a block in differentiation. Similarly, while hemin stimulation resulted in increased expression of monocyte marker CD33 and CD24, cells treated with ARID3a-specific shRNA more closely resembled untreated cells (Figure 10G). The control shRNA stained cells expressed surface markers similar to those of hemin treated cells. Together, these data suggest that ARID3a protein is necessary for globin production and differentiation down the erythrocyte lineage in hemin-stimulated K562 cells.

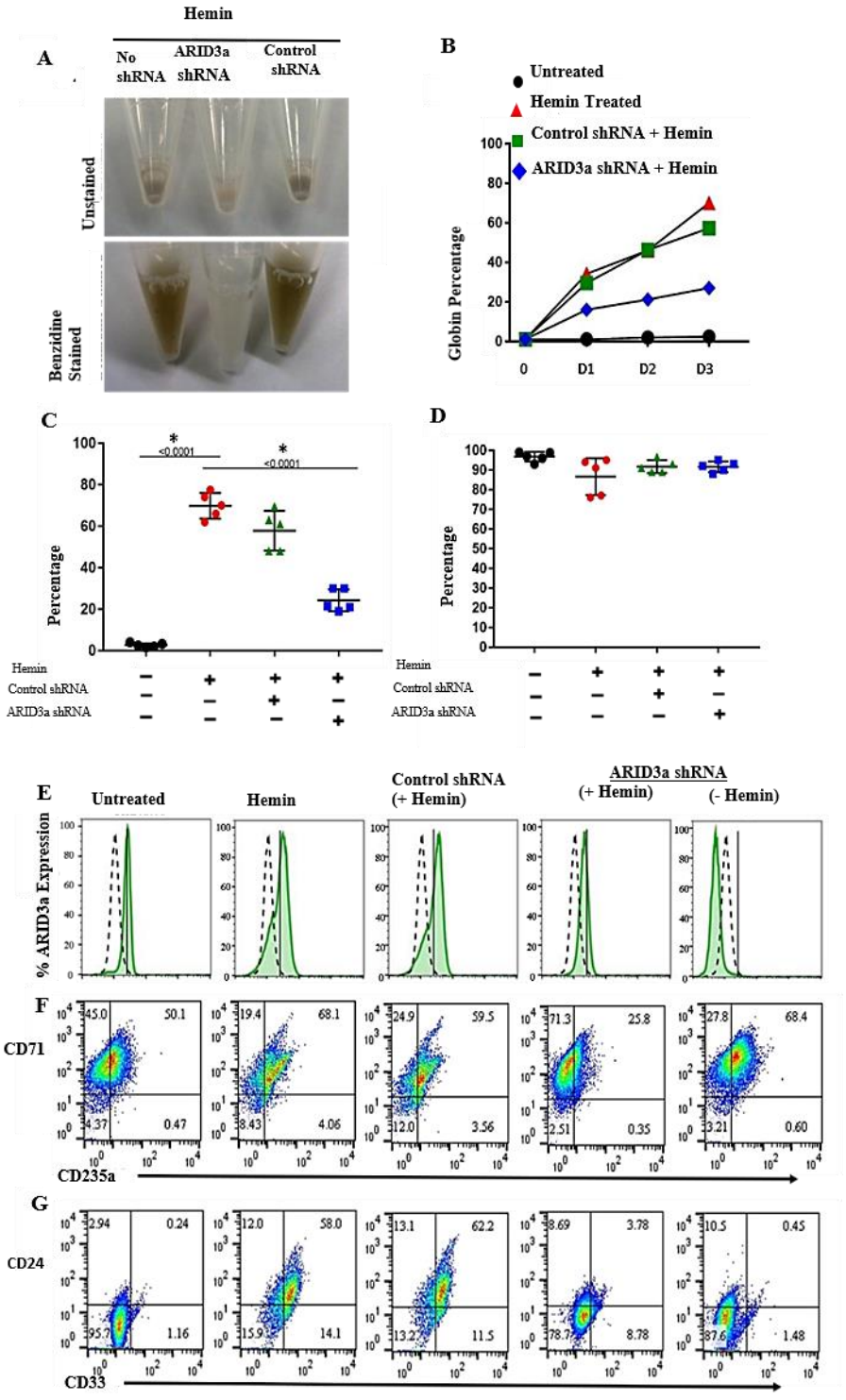


Figure 10. ARID3a is required for hemin-induced fetal globin production and erythroid maturation. K562 cells were treated with hemin with and without prior transduction of cells with lentivirus expressing ARID3a-specific shRNA or unrelated shRNA control virus for six days. (A) Untreated and benzidine stained cell pellets of hemin-stimulated K562 cells reveal

brown colored globin production. (B) Time course data from three individual experiments show percentages of globin producing cells elucidated microscopically over 6 days with and without ARID3a inhibition. (C) Cumulative data show percentages of benzidine positive cells from 5 individual experiments analyzed on day three ($P < 0.0001$, one-way ANOVA). (D) Percentages of viable K562 cells counted via trypan blue exclusion on day three are shown. (E) Flow cytometric histograms indicate numbers of cells expressing intracellular ARID3a (shaded peaks) compared to isotype controls (dotted lines). Solid vertical lines depict peak intensities of control unstimulated cells. Graphs are presented in normalized mode. Flow cytometry indicates surface staining of proteins associated with erythrocyte (F) and monocyte lineage differentiation (G) are shown for each treatment condition. Data are representative of 5 experiments.

ARID3a inhibition of hemin stimulated cells results in down regulation of genes associated with erythroid differentiation.

To further explore the block in erythroid differentiation in ARID3a inhibited samples, we performed a time course RNA-seq experiment over three days in K562 cells treated with or without hemin and with or without ARID3a inhibition. Triplicate samples from each treatment condition were sequenced and differential expression analyses were performed to identify genes affected by ARID3a inhibition. Principal component analysis revealed treatment-dependent patterns that ordered as expected during erythropoiesis (Figure 11). Untreated samples clustered away from hemin-stimulated samples and ARID3a-inhibited samples clustered more closely to untreated samples (Figure 11). Furthermore, samples on days two and three clustered closely together but away from day one samples, suggesting that most cells had terminally differentiated by day three. Samples treated with shRNA targeting ARID3a clustered between untreated and hemin treated samples (Figure 11), indicating perturbed differentiation.

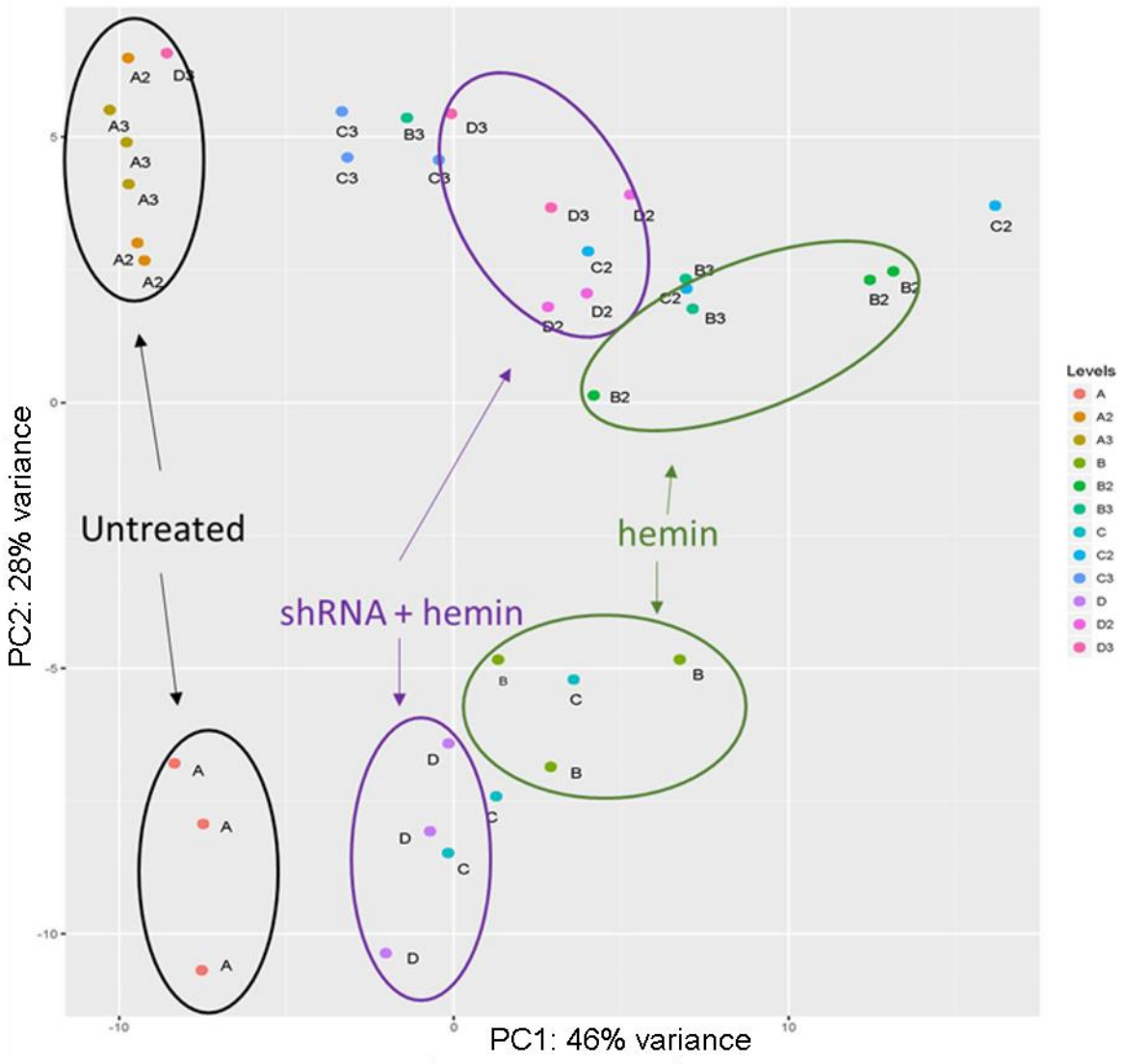


Figure 11. ARID3a inhibition alters transcription profiles of hemin-induced K562 cells. Principal component analysis of K562 cells treated with hemin both with and without ARID3a inhibition every day for three days is shown. Individual dots represent triplicate cultures sequenced on days 1-3. Numbers indicate the day of isolation and letters (A-D) represent the different treatment conditions (A=untreated, B=hemin treated, C=hemin-treated with control shRNA and D= hemin treated with ARID3a-shRNA). Cell treatments are grouped and labeled accordingly. Variance along PC2 is due to differentiation and variance along PC1 is due to treatment.

Analyses of the transcriptomes within and between samples treated with or without hemin and ARID3a shRNA revealed tight clustering of transcriptomes from different days. There are numerous significant (n=3 per group, FDR < 0.05) temporal changes in gene expression on each

day, with each treatment (hemin, shRNA) exhibiting unique transcriptomes (Figure 12). To identify the genes required for erythroid lineage differentiation in this model system, we first performed hierarchical clustering on differentially expressed genes in hemin versus untreated cultures at each of days one through three (Figure 12A). Differential expression analyses identified 846, 2,228, and 1,055 differentially expressed genes (DEGs) on days one, two, and three, respectively, in K562 cells stimulated with hemin as compared to untreated controls. Select genes, including those previously reported to be important for globin expression and erythroid differentiation are indicated in Figure 12A. As expected, we confirmed expression of key genes involved in the differentiation toward the erythroid lineage, such as *GATA1, GATA2, ALAS2,* and *NFE2,* although they were not all differentially expressed on all three days. Alpha- (*HBA1, HBA2)* and β-globin (*HBG1, HBG2*) genes were significantly induced on all three days upon treatment with hemin. Erythroid genes *TAL1, KLF1, MYB*, and *GFI1B* were differentially expressed on day 2. These data agree with previous transcriptome analyses of hemin-stimulated K562 cells using microarrays (164,172,173). As expected, pathway analyses identified erythrocyte development among the top pathways. GO pathway analysis of upregulated genes on day one revealed enrichment of pathways involved in heme synthesis, iron hemostasis, thioredoxin and translation initiation (data not shown). Transcription, mRNA splicing, nuclear export, and autoimmune pathways were enriched in hemin-treated samples by day two. Histone and nucleosome processes were also enriched in the DEGs from day two. These pathways were also enriched at day three indicating that most of the changes at the level of transcription were evident by day two.

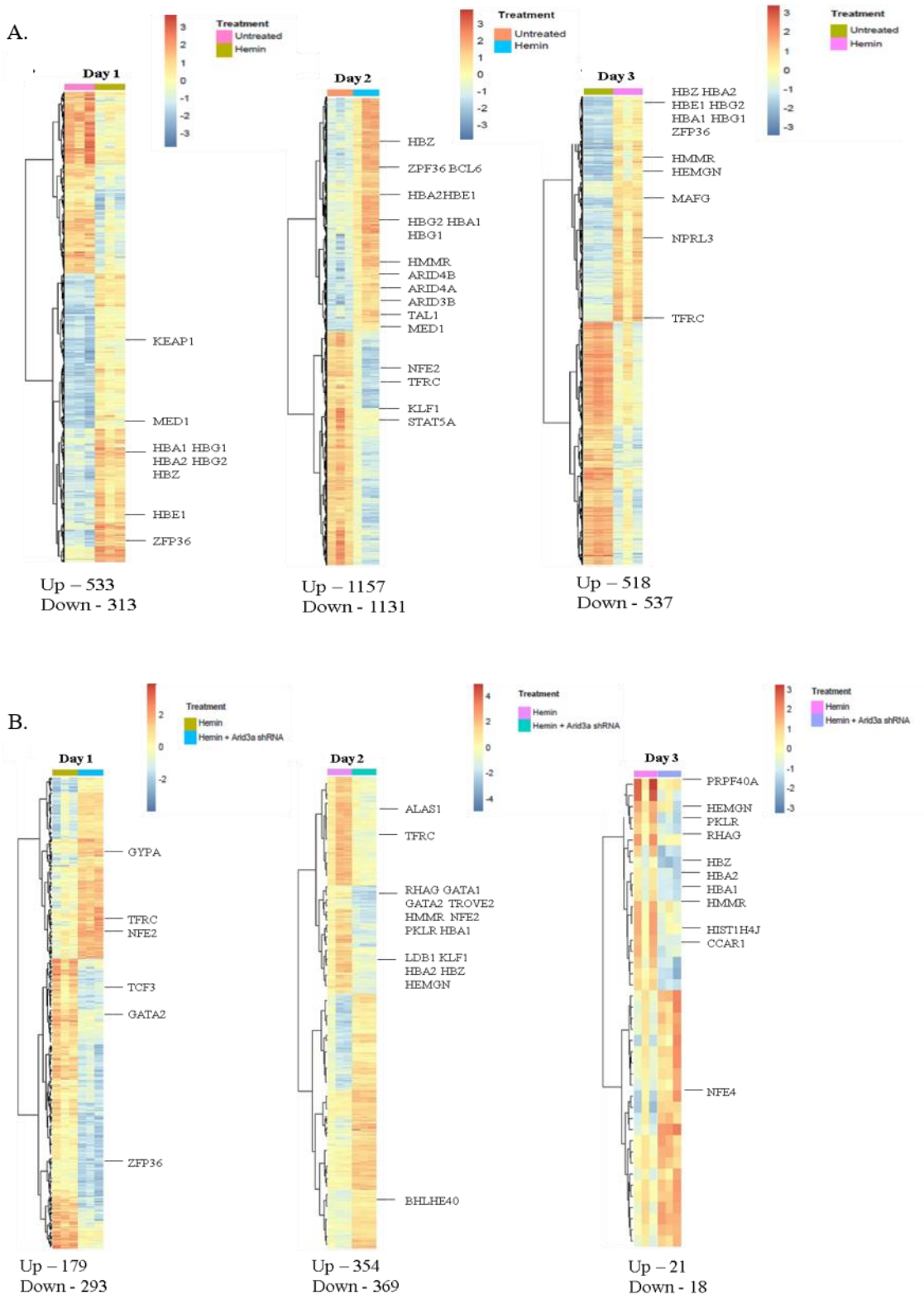


Figure 12. Both hemin induction induces globin-associated genes within three days, while ARID3a inhibition alters those gene expression profiles. (A) Heatmaps of differentially expressed genes in triplicate cultures of untreated vs hemin-treated cells for days one to three.

Numbers of differentially expressed genes (FDR < 0.05, FC ≥ 1.5, n = 3) are indicated below the heatmaps. Select genes previously associated with erythrocyte differentiation are indicated. (B) Hierarchical clustered heatmaps of genes affected by ARID3a inhibition during three days of hemin stimulation of triplicate cultures are shown as in (A). Total numbers of differentially expressed genes are given at the panel bottoms and select genes are indicated.

To identify genes affected by ARID3a inhibition that perturbed hemin-induced erythroid lineage differentiation, we performed similar analyses of DEGs from hemin-treated cultures with or without ARID3a shRNAs and performed hierarchical clustering of those genes on days one through three (Figure 12B). Numbers of DEGs ranged from 39 to 723, with day two having the highest number of DEGs (Figure 12B). Indeed, when ARID3a expression was suppressed by RNA interference, hemin-induced transcriptional activation of GATA1-targeted genes encoding α-globin, NFE2, PRG2, HMBS, PPOX, and ANK1 was strongly attenuated two days after hemin treatment. The expression of erythroid differentiation markers CD71 (TFRC) and GYPB were also significantly reduced upon ARID3a inhibition. Additionally, both α- and β-globin genes were significantly reduced on the first two days while only α-globin genes were downregulated on day 3. This is consistent with the reduction in the number of erythroid-differentiated benzidine positive cells by three days of hemin treatment (Figure 10).

Principal component analyses (Figure 11) revealed that hemin stimulated cells were differentiated by day two, and that most of the genes affected by ARID3a inhibition were apparent by that time as well. Therefore, we focused on genes differentially expressed on day two. A Venn diagram indicates overlapping DEGs affected by each treatment condition (Figure 13A). Hemin induction affected more genes than were affected by ARID3a inhibition. These analyses identified 227 overlapping DEGs upregulated by hemin treatment and downregulated by ARID3a inhibition (Figure 13A). Pathway analyses of this gene list identified SLE signaling, mRNA processing (GO:0006396), RNA splicing (GO:0008380) and chromatin binding (GO:0003682) pathways (Figure 13B). Identification of over-represented transcription factor binding sites in the

227 genes induced by hemin and repressed by ARID3a shRNA showed significant enrichment in genes with binding sites for *GATA1 (163/227), GATA2 (62/227), TAL1, PAX5 (127/227), SP1(66/227),* and *ARID3a*, among others (Figure 13C). Essential TFs for erythropoiesis (*GATA1, GATA2, KLF1, NFE2*) and their cofactors/mediators (*MED1* and *LDB1*) were all inhibited on day two in samples treated with hemin and shRNA. Additionally, members of the polycomb repressive complex, such as *EZH2*, are essential regulators of erythropoiesis and were also inhibited by ARID3a shRNA on day two. A list of the top 35 most significantly differentially expressed genes affected by hemin stimulation and those repressed by ARID3a inhibition is given in Table 2. Among the top DEGs repressed by ARID3a, the majority were either transcription factors, micro RNAs, or other small nuclear RNAs involved in splicing.

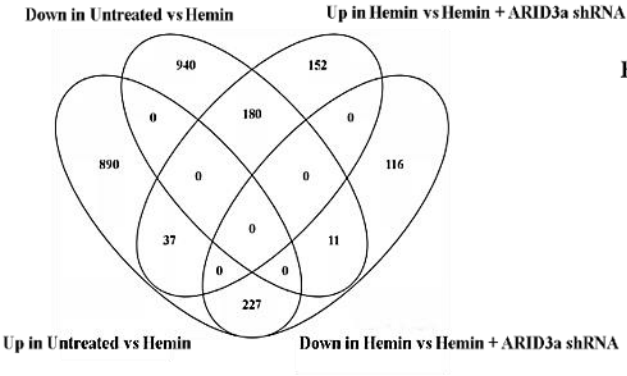
Table 2. Most significantly differentially expressed genes on Day 2

Untreated Vs Hemin			Hemin Vs Hemin + ARID3a shRNA		
Gene	Log_2 Fold Change	Adj P Value	Gene	Log_2 Fold Change	Adj P Value
TXNIP	5.24	6.25E-15	*SH3BGR*	2.29	4.02E-07
OSGIN1	3.68	3.63E-16	*HSPA5*	2.05	7.51E-10
HBZ	3.52	6.64E-16	*HERPUD1*	1.83	4.88E-11
AKR1C1	3.01	1.89E-14	*ARG2*	1.75	2.36E-09
SQSTM1	2.99	1.22E-24	*ALDH1A1*	1.61	7.07E-07
MCM5	2.74	1.07E-18	*NFE4*	1.43	3.92E-12
HBA2	2.54	2.68E-15	*SEC24D*	1.40	3.58E-07
HBE1	2.47	3.58E-21	*BEX2*	1.29	4.11E-06
NQO1	2.41	2.22E-32	*CREG1*	1.28	2.85E-06
HBG2	2.12	2.33E-15	*LCP1*	1.22	4.33E-07
GCLM	2.08	5.83E-19	*SERPINH1*	1.20	5.91E-06
PPP1R15A	2.05	2.15E-13	*RTN4*	1.16	3.14E-08
FTL	2.02	3.63E-16	*ACOT13*	1.15	6.60E-07
HBA1	1.98	3.73E-13	*TDP2*	1.11	1.49E-07
HBG1	1.94	1.02E-11	*TPM4*	1.11	4.53E-08
TXNRD1	1.52	9.58E-12	*CTSD*	1.11	1.78E-08
FTH1	1.47	4.70E-12	*PGD*	1.05	3.26E-06
CREM	1.41	1.28E-14	*ACTB*	1.02	2.38E-06
TXN	1.20	6.53E-14	*CTSB*	1.02	1.39E-06
PRKCSH	-1.18	3.70E-16	*SND1*	0.96	8.32E-07
FKBP2	-1.44	1.47E-11	*COPA*	0.89	2.56E-07
UCA1	-1.52	6.71E-13	*GSR*	0.88	2.85E-06
SERPINH1	-1.60	1.47E-11	*ANXA5*	0.88	1.47E-10
PDIA3	-1.63	2.29E-19	*SEC61A1*	0.80	1.54E-06
DNAJB11	-1.69	4.11E-14	*CALU*	0.80	1.71E-09
AC068631.2	-1.95	9.98E-12	*SSX1*	0.77	1.23E-07
HSP90B1	-1.97	3.99E-19	*NQO2*	0.73	5.36E-06
HERPUD1	-1.99	2.64E-14	*PSMC1*	0.68	1.35E-06
NMU	-2.01	3.65E-11	*SEM1*	0.61	4.33E-07
HYOU1	-2.07	7.01E-19	*TNNI3*	-1.04	8.32E-07
PDIA6	-2.19	2.54E-17	*VARS*	-1.06	2.82E-07
MANF	-2.24	8.93E-16	*LINC01029*	-1.10	1.26E-06
PDIA4	-2.35	1.43E-14	*MARCKSL1*	-1.14	4.84E-07
HSPA5	-2.35	7.91E-14	*HEMGN*	-1.95	7.51E-10
CALR	-2.57	9.34E-17	*TXNIP*	-4.15	9.81E-11

Ingenuity Pathway Analysis (IPA) of differentially expressed genes (FDR < 0.05, FC > 1.5) on day one indicated that transcription factors such as *MYC, STK11,* and *TP53* were inhibited

while let7 gene targets, *ATF4*, *FN1*, and *ELK1* were activated. In addition, when the 227 genes affected both by hemin and ARID3a were interrogated, networks with functions related to cell cycle, cell death and survival, and cell morphology were identified. Interestingly, *ARID3a* was included in the network and was associated with *FOS* and *YY1* (Figure 13D-E).

A.



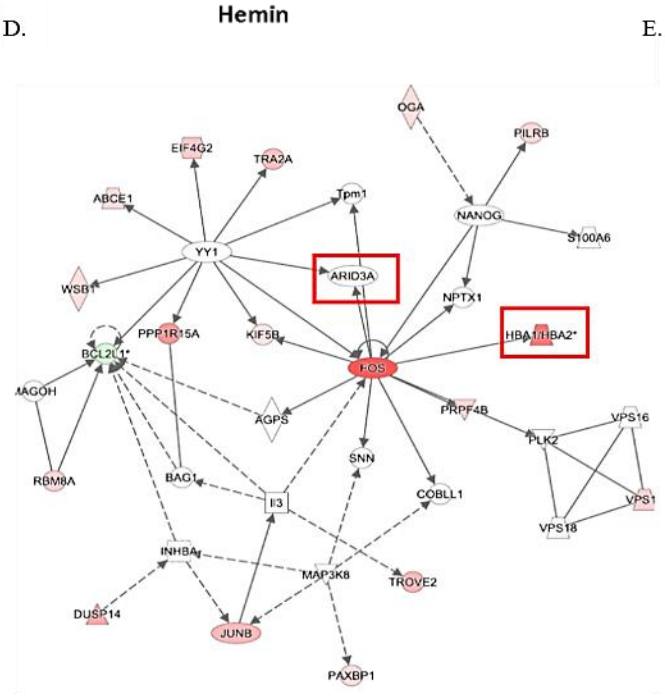
B.

GO Terms	Term ID	Adjusted P Value	Number of DEGs
RNA binding	GO:0003723	3.99E-13	59
Nucleic acid binding	GO:0003676	1.43E-10	89
Heterocyclic compound binding	GO:1901363	8.32E-08	105
Organic cyclic compound binding	GO:0097159	1.44E-07	105
Protein binding	GO:0005515	5.39E-05	150
RNA methyltransferase activity	GO:0008173	2.72E-04	7
Catalytic activity, acting on RNA	GO:0140098	1.18E-03	14
Chromatin binding	GO:0003682	2.64E-03	17

C.

Name	p-value	Motif
PAX5	6.14E-06	
GATA1	2.40E-04	
NRF2	2.11E-03	
YY1	1.04E-02	
GATA2	1.65E-02	

D.



E.

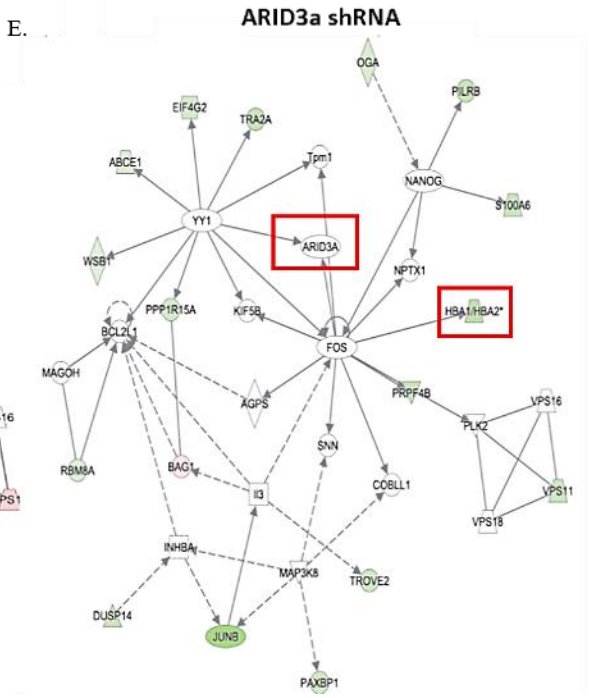


Figure 13. ARID3a expression is linked to key genes and transcription factors important for erythropoiesis. (A) A Venn diagram indicates numbers of differentially expressed genes on day two of treatment with and without hemin and ARID3a (FDR < 0.05, FC ≥ 1.5, n=3). (B) GO analyses indicates pathways important for the 227 genes that are affected by ARID3a inhibition and hemin induction. (C) The most highly represented transcription factor binding motifs of the

227 overlapping genes are shown. (D, E) Network analyses of the 227 differentially expressed genes in (A) reveal related genes. Log2 fold-change values from the untreated vs hemin comparison (left) were overlaid onto the top network identified by IPA using the 227 overlapping genes. Red color indicates upregulated genes and green color indicates downregulated genes. (E) Log2 fold-change values from the hemin vs hemin and ARID3a shRNA were overlaid onto the top network.

Chromatin accessibility is altered in ARID3a KO K562 cells

Alpha-globin genes (*HBA1*, *HBA2*, and *HBZ*), components of mediator complexes important for erythropoiesis (*MED1*), and histone subunits were among the 227 genes induced by hemin and down-regulated by ARID3a, suggesting ARID3a has a role in globin transcription or mediates cofactor binding to transcription start sites (TSS) of erythroid-specific genes through epigenetic mechanisms. To explore the hypothesis that ARID3a regulates lineage-specific genes by altering chromatin accessibility, we used CRISPR-Cas9 gene editing in K562 cells to generate clones with biallelic inactivation of *ARID3a* for ATAC-seq analyses (Figure 14). Single guide RNAs (sgRNAs) targeting exon 3, which codes for the extended DNA-binding domain specific to ARID3 family members, were used to generate genomic deletions of *ARID3a* (Figure 14A). Bulk deleted clones were then single cell sorted and individual clones were screened for deletion by PCR (not shown), flow cytometry and western blotting (Figure 14B-C). Two clones (BH and AL) selected via reduced intracellular staining were confirmed to exhibit no detectable protein via western blotting and were used with wild type clones for ATAC-seq analyses.

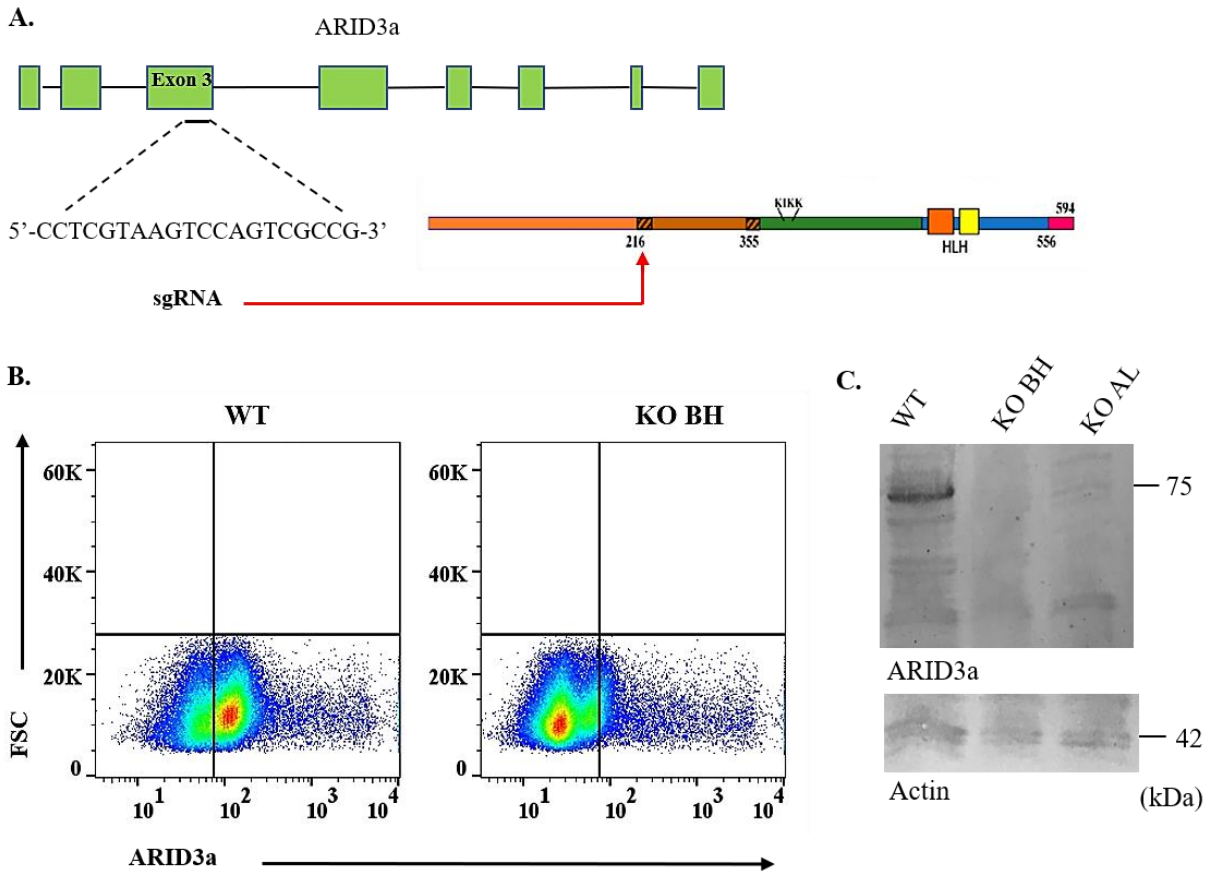
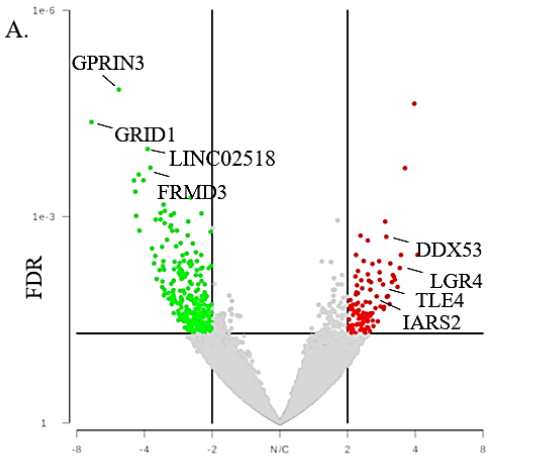
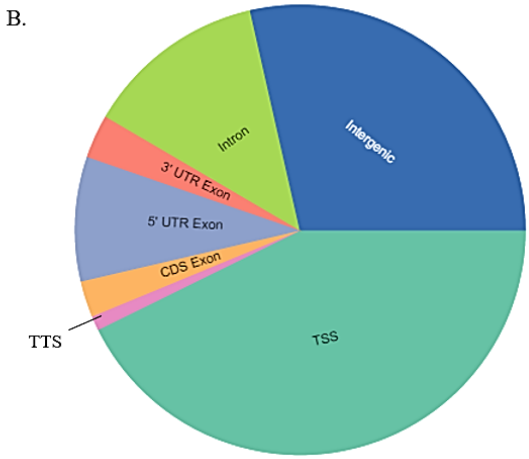


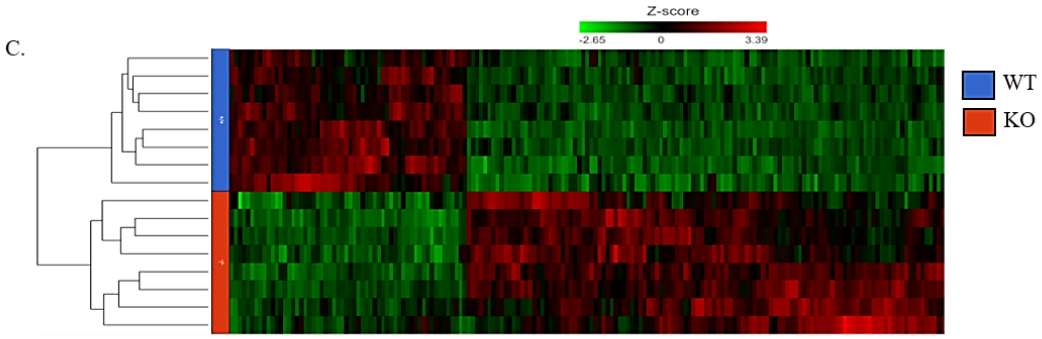
Figure 14. Clones of homozygous CRISPR-Cas9 ARID3a knockouts in the K562 cell line were identified and isolated. (A). Schematic diagram of the sgRNA cut-site that causes a deletion of part of the extended-ARID DNA-binding domain. (B) Flow cytometry was used to identify clones with reduced staining for intracellular ARID3a protein, as shown for clone BH versus the wild type clone. Quadrant gates were set according to isotype controls. (C) ARID3a protein levels were measured by western blotting of wild type and CRISPR-Cas9 KO clones BH and AL for 100,000 cells per lane. Actin was used to confirm protein loading in each lane (lower panel).

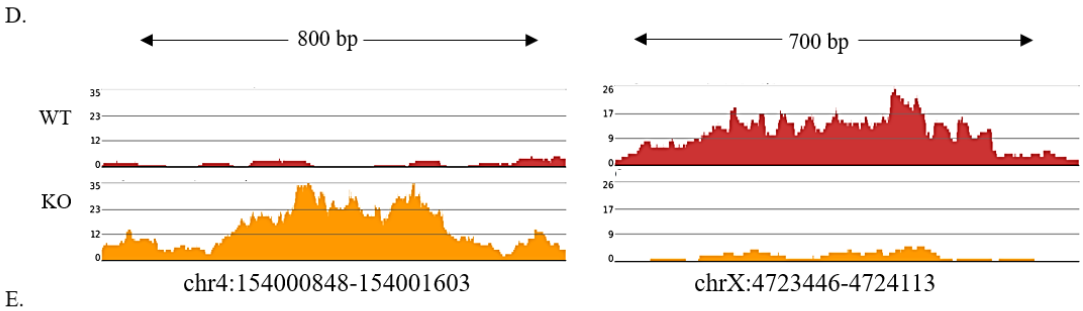
DESeq2 identified 690 genomic regions with differential chromatin accessibility in ARID3a$^{+/+}$ vs ARID3a$^{-/-}$ K562 cells, and 385 were mapped near genes (Figure 15). Among the 385 genes was MAPK14 (P38), which acts as a break for erythropoiesis through JNK-mediated inhibition (174). Overlap analysis of the 385 differentially expressed genes identified by ATAC-seq with the 227 genes from the RNA-seq data revealed 5 genes (*SNHG12*, *SERTAD1*, *VPS37B*,

PPP1R15A, *SLC25A25*) that are associated with ARID3a expression in hemin-induced K562 cells. *SERTAD1*, *VPS37B*, *PPP1R15A*, and *SLC25A25* all contain TFBSs for GATA1, GATA2, and TAL1. SERTAD1 has implications in blood cell development in mice (175). These genes have no known functions in erythropoiesis; however, SHNG12 and SERTAD1 have been implicated to inhibit cell proliferation (176). PPP1R15A and VPS37B are involved in translation initiation and protein trafficking, respectively. These data suggest ARID3a may function in regulating cell differentiation through inhibiting certain histone methyltransferases or by activating genes involved in translation initiation. A large percentage of differentially accessible sites were also intergenic regions (Figure 15B). Unsupervised hierarchical clustering of all 690 regions indicated that wild type and knockout regions grouped together and showed both increased and decreased regions of accessibility associated with ARID3a deficiency (Figure 15C). Alterations in chromatin accessibility were particularly evident in large intragenic regions (Figure 15D) that may represent enhancers. A list of the top 10 regions that were most significantly up- or downregulated between wild type and ARID3a KO clones are listed in Figure 15E, and eight of those are intragenic regions of unknown significance.









E.

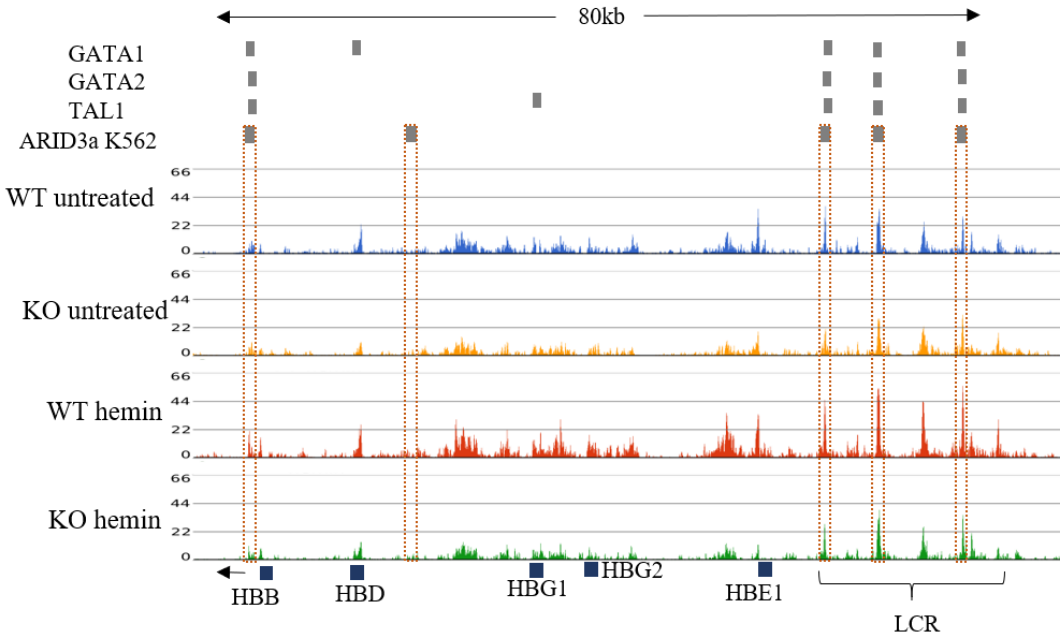
ARID3a WT vs ARID3a KO			
Region	**Gene**	**FDR**	**Log$_2$ Fold Change**
chr4:154000848-154001603		4.22E-05	-6.877
chr4:90220055-90221377	GPRIN3	1.42E-05	-5.197
chr6:113747366-113750109	LINC02518	2.98E-04	-4.453
chr1:120523623-120525510	NOTCH2NLC	4.34E-04	-4.389
chr5:148439793-148441189	SH3TC2	9.75E-04	-4.351
chr9:85886229-85887102	FRMD3	2.45E-04	-4.240
chr2:109440940-109442012	CCDC138	1.60E-03	-4.211
chr2:181569274-181570630	SCHLAP1	2.96E-04	-4.041
chr10:87661221-87662893	GRID1	1.04E-04	-3.870
chr3:14274766-14277305		1.95E-04	-3.765
chr9:82288731-82289428	TLE4	8.971E-03	3.149
chr1:220297417-220298613	IARS2	7.076E-03	3.178
chr6:104861292-104862095		7.489E-03	3.213
chr4:70411492-70412684		8.248E-03	3.246
chr2:58054814-58055701		1.052E-02	3.334
chr11:27485604-27486309	LGR4	5.590E-03	3.411
chrX:23017878-23018299	DDX53	3.612E-03	3.453
chr4:93102904-93104970		1.973E-04	3.599
chr8:33579010-33580960		2.276E-05	3.960
chrX:4723446-4724113		3.584E-03	4.092

Figure 15. Ablation of ARID3a in K562 cells results in changes in chromatin accessibility. (A) Differential expression analyses of wild type and ARID3a$^{-/-}$ KO K562 cells (n = 2 clones per group, FDR < 0.05, log FC > 2). Dark red dots represent individual peaks with FDR < 0.05 (y-axis) and log FC > 2 (x-axis). Several genes with high differential accessibility are labeled. (B) The genomic distribution of differentially expressed ATAC peaks in K562 wild type versus ARID3a knockout cells is shown. The human genome was portioned into seven bins relative to RefSeq genes. TSS, transcriptional start site; TTS, transcriptional termination site. (C) An unsupervised hierarchical clustered heatmap of differentially expressed ATAC regions between wild type and ARID3a$^{-/-}$ KO K562 clones reveals clustering of knockout versus wildtype clones and revealed two clusters of genomic regions with reciprocal expression. (D) Intergenic regions show differential chromatin accessibility of two representative regions highly affected by deletion of ARID3a. Gray bar indicates ARID3a ChIP sites from ENCODE. (E) List of the top 10 upregulated and downregulated differentially accessible regions between wild type and ARID3a KO K562 cells.

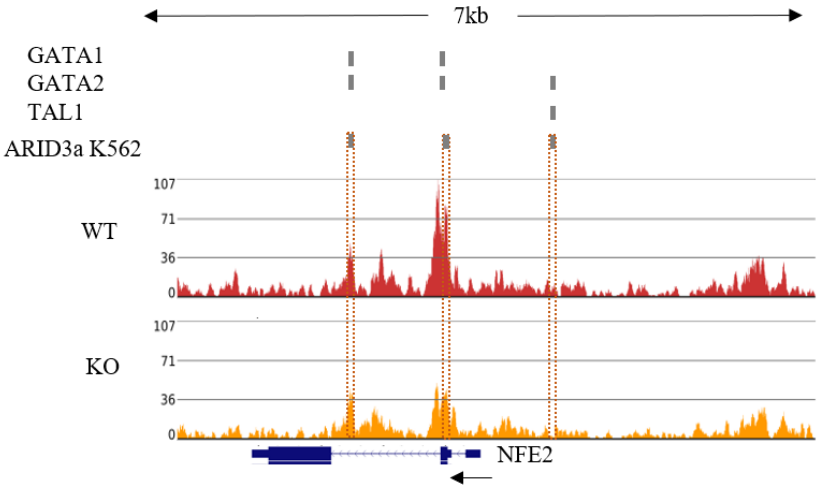
ARID3a deficiency directly affects chromatin regions associated with globin gene regulation

The globin locus control (LCR) region upstream of the fetal globin genes is critical for developmental regulation of those genes. This LCR revealed reduced accessibility in ARID3a$^{-/-}$ clones compared to wild type clones (Figure 16A). Data from ENCODE identified ARID3a binding sites via ChIP-seq (Figure 16A), in many cases with overlapping GATA1, GATA2, and TAL1 TF binding sites. The erythroid-specific TF loci for NFE2 and TAL1 were also significantly less accessible in ARID3a$^{-/-}$ clones than in wild type cells (Figure 16B-C). The surface marker CD71 that was down-regulated at both the protein and transcript level (Figures 10 and 12), also exhibited reduced accessibility in ARID3a deficient cells (not shown). Together, these data suggest that ARID3a is required to maintain appropriate chromatin configurations for globin gene expression and erythropoietic differentiation in K562 cells.

A.


80kb
GATA1
GATA2
TAL1
ARID3a K562
WT untreated
KO untreated
WT hemin
KO hemin
66
44
22
0
HBB
HBD
HBG1
HBG2
HBE1
LCR

B.


7kb
GATA1
GATA2
TAL1
ARID3a K562
WT
KO
107
71
36
0
NFE2

C.

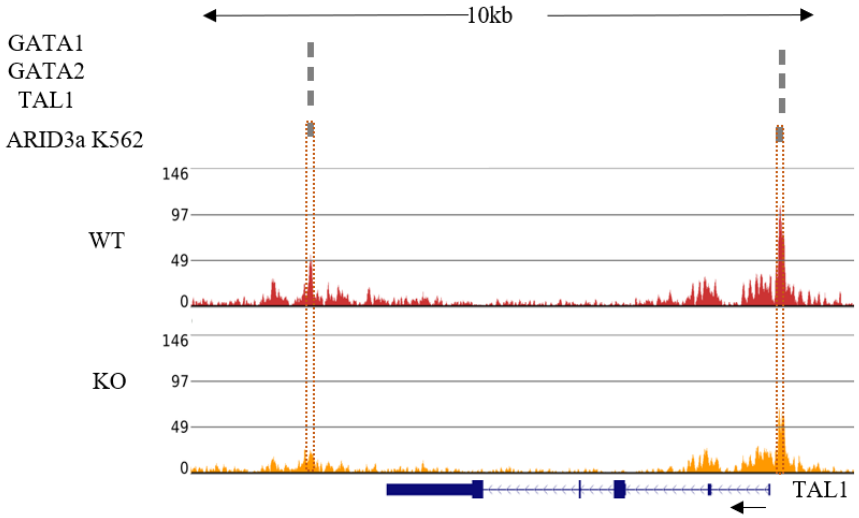

10kb
GATA1
GATA2
TAL1
ARID3a K562
WT
KO
146
97
49
0
TAL1

Figure 16. ARID3a affects chromatin accessibility of transcription factors and regulatory regions important for globin expression. (A). Chromatin accessibility of the β-globin genes and locus control region in wild type and ARID3a$^{-/-}$ KO K562 cells (n = 4 clones per group, FDR < 0.05, logFC > 2) stimulated with and without hemin. Gray bars and dotted rectangles indicate transcription factor binding sites for ARID3a, GATA1, GATA2, and TAL1 identified in K562 cells by ENCODE. Chromatin accessibility in wild type and ARID3a$^{-/-}$ KO clones of the erythroid-specific TFs NFE2 (B) and TAL1 (C) are shown.

DISCUSSION

In this study, we demonstrated that ARID3a protein is required for hemin-induced, erythrocyte lineage differentiation of the human K562 model cell line. Knockdown of ARID3a with shRNA resulted in a visible reduction in globin production and downregulation of erythroid-lineage associated surface markers CD235a (GYPA) and CD71 (TFRC) without affecting viability. RNA-seq analyses of ARID3a-inhibited samples identified 227 genes that required ARID3a for hemin induced differentiation, confirmed downregulation of globin genes (*HBA1*, *HBA2*, and *HBZ*) and revealed that ARID3a deficiency led to blocks in differentiation and repression of key erythroid-specific TFs (*GATA1*, *GATA2*, *KLF1*, *NFE2*) and genes encoding critical cofactors (*MED1*, *LDB1*, *CCAR1*) for hemoglobin expression. Furthermore, our data reveal that ARID3a acts as an epigenetic regulator that is required for appropriate chromatin accessibility of genes and enhancer regions essential for erythroid maturation. These data indicate a previously unappreciated role for ARID3a in human erythropoiesis and are the first to document genomic sites altered by depletion of ARID3a.

We previously found that *ARID3a* knockout embryos exhibited profound defects in erythropoiesis at day 12.5 of gestation (14). Consistent with our observations, Kingsley et al., found ARID3a transcripts were enriched 22-fold in primitive erythropoiesis in the mouse (144). Additional studies in human erythroid progenitors indicated that the *ARID3a* gene locus was differentially methylated during early erythropoiesis with higher expression in fetal erythroblasts

(177). More recently, the *ARID3a* locus was identified to be differentially methylated in human primary basophilic erythroblasts (178). Further, others linked ARID3a with the erythroid master regulators, GATA1 and TAL1, in human erythropoietic studies that determined transcription factor landscapes of enhancers in erythroid progenitors (156,179). Our studies extend those data, and definitively demonstrate a requirement for ARID3a protein in human erythroid development.

We identified 227 differentially regulated genes associated with ARID3a and erythropoiesis in this system. These genes showed significant enrichment of the transcription factor binding sites, GATA1 and GATA2. Both GATA1 and GATA2 are critical regulators of erythropoiesis (180-183). GATA2 is expressed in erythroid precursors (184), and as GATA1 levels increase, GATA2 is replaced by GATA1 at many sites throughout the genome, a process called GATA switching (185,186). Studies of enhancer turnover in $CD34^{+}$ cells suggested that ARID3a could be associated with the GATA2-to-GATA1 switch, raising the possibility that ARID3a could be involved in epigenetic modifications in those cells (144). ENCODE data of K562 cells showed considerable overlap between GATA1, GATA2, and ARID3a binding sites in many genes important for erythropoiesis, suggesting ARID3a may function with those factors, either as a transcription factor or as an epigenetic regulator mediating opening/closing of chromatin in enhancer/promoter regions. Indeed, four of the five genes identified by overlap analysis of the RNA-seq and ATAC-seq data contained ARID3a, GATA1, GATA2, and TAL1 binding sites. Moreover, the GATA switch is mediated by positioning of polycomb subunits EZH2 (187). *EZH2* was significantly downregulated upon inhibition of ARID3a in our studies and our unpublished data suggest it may interact directly with ARID3a in K562 cells. Further studies will be needed to explore if ARID3a is important for the GATA switch through regulation of, or interactions with EZH2.

Our ATAC data reveal decreases in chromatin accessibility in erythroid-specific enhancer regions in ARID3a$^{-/-}$ clones, including the LCR region that is essential for the expression of embryonic, fetal, and adult globin genes in a developmentally controlled manner. EHMT1 adds repressive H3K9me2 marks to the LCR region (188), and showed 2-fold increased accessibility in ARID3a KO clones. EHMT1 also adds repressive histone marks to H3K9me2 at the γ-globin locus in human adult erythroid cells, thereby reducing expression of both γ-globin and fetal globin (62). Future studies will be required to how ARID3a contributes to these effects. However, our data suggest that ARID3a may participate in mediation of multiple epigenetic events necessary for erythropoiesis, and that these events may require context-dependent transcriptional activities.

ARID3a was originally discovered as a DNA-binding protein in the mouse (Bright) that upregulated transcription of the immunoglobulin heavy chain locus in B lymphocytes in response to cytokine and antigen stimulation (189,190), and was later discovered to suppress the *Oct4* gene, participating in pluripotency regulation (11,12). In addition, our data here and in other human blood cells suggest that ARID3a over- and under-expression is associated with differential gene expression patterns in a cell type-specific fashion (13). Therefore, it is likely that ARID3a functions in coordination with other epigenetic factors to mediate its effects. Indeed, these data indicate that ARID3a knockdown contributes to alterations in chromatin accessibility in a wide range of genomic sites that are both associated with coding and intragenic regions (Figure 6). We postulate that many of the intragenic regions may function as enhancers; although it is likely that that function will also exhibit cell type specificity and the requirement for transcription factors that may be absent from the K562 cell line.

Out of the 690 differentially accessible regions identified by ATAC-seq, 305 of these regions were located in intergenic regions with unknown function. Moreover, there were 8 intergenic regions among the top 20 most differentially accessible regions (Figure 15). The importance of intergenic regions is emphasized by a study on the super-enhancer-derived RNA, alncRNA-EC7/Bloodlinc, which is required for terminal erythropoiesis and red blood cell production (191). Our ATAC data show reduced accessibility of this enhancer region (not shown). It is not currently possible to distinguish which of these intergenic regions with altered chromatin accessibility directly contribute to erythropoietic functions, versus other hematopoietic events. For example, some of these regions may be important in other hematopoietic cells where ARID3a is linked to disease activity, such as lupus (13,192). Histone subunits (*HIST1H2BN*, *HIST1H3B*, *HIST1H3H*, and *HIST1H4J*), chromatin remodelers (*SAT2B*), heme biosynthesis enzymes (*ALAS1*), and genes implicated in Systemic Lupus Erythematosus (SLE) signaling pathways (*PPP1R15A* and *TROVE2*) were repressed by ARID3a shRNA on day two. Thus, it is not possible from these data alone to elucidate the functions of ARID3a-regulated regions in other cell types.

Understanding how hemoglobin expression and erythropoiesis are regulated is critical for the development of new therapeutics for diseases such as sickle cell disease and thalassemia's. Further elucidation of how ARID3a functions in erythropoiesis could lead to development of therapeutic agents for blood disorders and for SLE, for which few drugs are currently available. Together, these data expand our knowledge of the importance of ARID3a in hematopoiesis, and particularly in erythroid lineage development and define new regulatory roles for ARID3a.

DATA AVAILABILITY

RNA-seq data are publicly available through the GEO NCBI database under the accession number GSE131649. For original data, contact carol-webb@ouhsc.edu**.**

ACKNOWLEDGEMENTS

We thank Dr. Gaffney for carefully reviewing the manuscript, the Clinical Genomics Core Facility at Oklahoma Medical Research Foundation and the Flow Cytometry Core Facility at OUHSC.

FUNDING

This work was supported by the National Institutes of Health [AI123951 and AI118836 to C.F.W., T32 AI007633 to J.W.G.].

CONFLICT OF INTEREST

The authors declare no competing financial interests.

Chapter 3 Addendum

Although, chapter 3 elucidated new functions for ARID3a in lupus, it did not provide opportunities for traditional analyses (i.e. differential expression analysis by analyzing ARID3a^{+} vs ARID3a^{-} samples) of RNA-seq data because ARID3a was post-transcriptionally regulated in both LDNs and pDCs (13)(Chapter 3). ARID3a was expressed at low levels in all samples while protein levels varied greatly. Therefore, we took advantage of an additional project ongoing in the lab that suggested ARID3a is a critical for erythropoiesis. ARID3a knockout mice exhibited low numbers of erythrocytes and died between days 9 and 12 of development due to failed

erythropoiesis. The role of ARID3a in human erythropoiesis has not been studied. Stimulation of the human K562 cell line with hemin has been shown to induce fetal globin synthesis and erythrocyte differentiation. Because K562 cells constitutively express ARID3a, we used this cell line to determine if ARID3a is important for human erythrocyte differentiation. Knockdown of ARID3a with shRNA in hemin stimulated K562 cells resulted in inhibition of globin production by >75%. To determine what genes were affected by loss of ARID3a, we performed a series of RNA-seq experiments at various times after hemin stimulation with and without ARID3a. The processing of raw fastq sequencing files were performed as described in Chapter 3. However, instead of using limma, DESeq2 was used to perform differential expression analysis.

```
pAnno <- read.csv(file="colDataB2.csv")
dds <- DESeqDataSetFromMatrix(countData = round(counts), colData = pAnno, design = ~ Levels)
#remove genes that are expressed less than an average of 10 TPMs across all samples
keep <- rowSums(counts(dds)) >= 10
dds <- dds[keep,]
```

The above code loads in the gene expression file, removes lowly expressed genes from the gene expression matrix, and the pAnno represents an annotation file that lets the program know which sample is in a specific condition (i.e. wild type, hemin treated, etc).

```
resA2vB2<- results(dds, contrast=c("Levels", "A2","B2"))
resOrderedA2vB2 <- resA2vB2[order(resA2vB2$pvalue),]
summary(resA2vB2)
sum(resA2vB2$padj < 0.1, na.rm=TRUE)
resdataA2vB2 <- merge(as.data.frame(resA2vB2), as.data.frame(counts(dds,normalized =TRUE)), by = 'row.names', sort = FALSE)
#write.csv(resdataA2vB2, file = paste0("A2vB2_A3AK562-results-with-normalized.csv"))
```

This is an example code for differential expression analysis between untreated K562 samples processed on day 2 (“A2”) and hemin treated samples processed on day 2 (“B2”), creates a matrix that is ordered by p values, concatenates gene names to the DESeq2 results file, and then saves the results as a csv file. Next, a heatmap of the top significantly differentially expressed genes were generated with the following code:

```
library(pheatmap)
rld <- rlog(dds, blind = FALSE)
topgenes <- head(rownames(resOrderedA2vB2))
mat <- assay(rld)[topgenes,]
mat <- mat - rowMeans(mat)
df <- as.data.frame(colData(dds)[,c("Condition","Levels")])
pheatmap(mat, annotation_col=df)
```

The code above was performed for each conditions analyzed in Figure12 of this chapter. The PCA plot in Figure 11 was generated in R using the code:

```
plotPCA(rld, intergroup = “Levels”)
```

ARID3a shRNA-inhibited cells confirmed inhibition of globin gene products. Erythroid specific transcription factors were also inhibited in samples treated with ARID3a shRNA while myeloid surface markers were induced. Several critical upstream regulators of ARID3a were also identified in these analyses. These data are the first to suggest that ARID3a is important for cell fate decisions in erythropoiesis in human cells and identify potential important targets of ARID3a in this process. In addition, ATAC-seq was performed on *ARID3a*$^{-/-}$ K562 cells to look for differences in chromatin accessibility.

CHAPTER 4

ARID3A EXPRESSION IN HUMAN HEMATOPOIETIC STEM CELLS IS ASSOCIATED WITH DISTINCT GENE PATTERNS IN AGED INDIVIDUALS

Michelle L Ratliff[1], Joshua Garton[2], Judith A. James[1,3,4,5], and Carol F. Webb[1,5,6]

[1]Department of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104

[2]Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK 73019

[3]Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104

[4]Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104

[5]Department of Microbiology and Immunology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104

[6]Department of Cell Biology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104

Address correspondence to Michelle L Ratliff, East Carolina University, 132 Biotechnology Building MS 629, 600 Moye Blvd, Greenville, NC 27834. Email: ratliffm19@ecu.edu or Carol F Webb, University of Oklahoma Health Sciences Center, 800 Research Parkway Rm 418, Oklahoma City, OK 73104. Email: carol-webb@ouhsc.edu

Copyright Information

The original version of this work was accepted with revisions to Immunity and Ageing on June 16th, 2020.

Allocation of Contribution

I contributed to this work by processing fastq files and performing dimensional reduction (tSNE plots) and differential expression and GO analysis used in Figure 18.

ABSTRACT

Immunologic aging leads to immune dysfunction, significantly reducing the quality of life of the elderly. Aged-related defects in early hematopoiesis result in reduced lymphoid cell development, functionally defective mature immune cells, and poor protective responses to vaccines and pathogens. Despite considerable progress understanding the underlying causes of decreased immunity in the elderly, the mechanisms by which these occur are still poorly understood. The DNA-binding protein ARID3a is expressed in a subset of human hematopoietic progenitors. Inhibition of ARID3a in bulk human cord blood $CD34^+$ hematopoietic progenitors led to developmental skewing toward myeloid lineage at the expense of lymphoid lineage cells *in vitro*. Effects of ARID3a expression in adult-derived hematopoietic stem cells (HSCs) have not been analyzed, nor has ARID3a expression been assessed in relationship to age. We hypothesized that decreases in ARID3a could explain some of the defects observed in aging.

INTRODUCTION

The US Census Bureau estimates that nearly a quarter of the US population will be over the age of 65 by the year 2060 (193). A major consequence of aging is a decline in immune function. Both murine and human studies revealed age-related defects in early hematopoietic development, and functional defects in mature immune cell populations, that result in decreased potentials to mount protective immune responses in aged individuals (reviewed in 194), as exemplified by increased susceptibility to influenza and pneumonia in the elderly. Human hematopoiesis is a dynamic process requiring complex regulation of multiple gene expression pathways for lineage commitment and resulting in development of diverse numbers of blood cell types (195). Several studies indicate that aged hematopoietic stem cell (HSC) frequencies are increased in both human and mouse (196,197). Bulk CD34-expressing hematopoietic stem and

progenitor cells (HSPCs) from aged donors exhibit epigenetic and transcriptional changes that promote self-renewal over differentiation (198,199). In old age, cells in the hematopoietic progenitor pool accumulate decreased telomere lengths and DNA damage markers, and their developmental potential becomes increasingly skewed toward myeloid versus lymphoid lineage development (198-201). Identification of the changes in old age that alter the development of mature immune cells, and possibly contribute to their dysfunction, will require mechanistic studies that better define potential differences in gene regulatory mechanisms critical for lineage choices.

The transcription factor ARID3a is an understudied member of a large family of proteins with epigenetic functions (27,202,203). Previous studies indicated that ARID3a can contribute to repression and enhancement of transcription in a cell type-specific fashion (169,204,205). Regulation of transcription through ARID3a may be associated with epigenetic functions that affect large subsets of genes and lineage decisions (21,192, our unpublished data,204,206). ARID3a deficient mice die in utero between days 12 and 14 of gestation, but exhibited a 90% reduction in HSCs numbers in the fetal liver associated with defective erythropoiesis and B lymphopoiesis (14). Functional loss of ARID3a in B lineage cells, either through ARID3a dominant negative transgenic mice or rare adult ARID3a$^{-/-}$ mice, revealed important roles for ARID3a in B1 B cell lineage development and function (14,207). These data were recently confirmed using conditional knockout mice, showing definitively that ARID3a is required for B1 B lineage development in mice. Loss of functional ARID3a in B lineage cells in mice directly impaired normal protective immune responses to infection with *S. pneumoniae* (207), an organism associated with pneumonia in aged individuals (24, 25). Forced expression of ARID3a in mouse B lineage cells resulted in enhanced development of B1 and MZ B cells versus

conventional follicular B cells (208), suggesting ARID3a levels can modulate B lineage responses in mice. Mechanisms responsible for generating B1 lineage B cells in man remain controversial (209,210) . Together, these data identify ARID3a as an important regulator of B lymphopoiesis.

Roles for ARID3a in human hematopoiesis are less clear. We found that ARID3a is variably expressed in healthy human HSPCs, including total $CD34^+$ HSPCs, HSCs, multipotent progenitor (MPP), multi-lymphoid progenitors (MLP), and multi-myeloid progenitors (MMP) derived from adult peripheral blood (211), but the functional significance of expression in those progenitors is not clear. In functional studies with human cord blood HSPCs, where ARID3a expression dominates the majority of those cells, manipulation of ARID3a resulted in skewing of lineage development with promotion of myeloid over lymphoid lineage differentiation upon loss of ARID3a expression and increased B lymphopoiesis upon over-expression of ARID3a (76). ARID3a expression in circulating peripheral blood HSPCs from lupus erythematosus patients is upregulated compared to similar cells from healthy individuals, although the role of ARID3a in those cells is unknown. These data suggest the need for further experiments to determine how ARID3a levels affect adult human hematopoiesis.

We hypothesized that one explanation for reduced B lymphopoiesis and increased numbers of myeloid cells in aged versus young individuals is that ARID3a expression is reduced in HSCs from healthy aged individuals compared to healthy young individuals, or that its function in those cells is impaired. Our results indicate that peripheral blood HSCs from aged donors exhibit reduced frequencies of ARID3a-expressing cells compared with young donors. Furthermore, modulation of ARID3a levels in aged and young donor-derived HSCs altered B lymphopoiesis in vitro. Finally, single cell RNA-seq analyses of revealed differences in gene

expression patterns in ARID3a-expressing progenitors from aged versus young individuals that are linked to age-associated hematopoietic changes.

METHODS

Single-cell RNA-seq

HSCs from 4 young (ages 19-40) and 4 aged (ages 61-70) individuals (2 males and 2 females each) were FACS-sorted and immediately used for single cell analyses. Smart-seq/C1 libraries were prepared on the OUHSC Consolidated Core Laboratory Fluidigm C1 system using the SMARTer Ultra Low RNA Kit (Clontech) according to the manufacturer's protocol. Cells were loaded on a 5-10 μm RNA-seq microfluidic IFC at a concentration of 200,000/ml. Capture site occupancy was surveyed using a standard light microscope and recorded to verify cell capture. Sequencing library amplification was performed using Nextera XT Index primers (Illumina) according to manufacturer's protocol. Barcoded library concentration and fragment size distribution was determined using Agilent High Sensitivity D1000 kit on an Agilent 2200 TapeStation (Agilent Technologies) at the OMRF Genomics Core Facility. Paired-end (2 x 50bp) sequencing was performed on a NovaSeq 6000 platform by the OMRF Genomics Core Facility.

Paired-end reads were aligned to the hg38 genome assembly with STAR using default parameters. The Partek Genomics Suite was run on the aligned reads to estimate gene expression levels. Cells that had greater than 25% of counts mapping to mitochondrial genes indicate stressed or dying cells and were excluded from analyses. The number of detected genes was 44,025 across 301 cells for aged samples and 39,417 genes across 264 cells for young samples. Log normalized counts per millions (CPM) were $\log_2$-transformed and used for tSNE visualization and differential expression analyses. Data from aged and young samples were

interrogated separately based on ARID3a expression, with >0.5 CPM denoting *ARID3A*$^{+}$ cells and <0.5 CPM denoting *ARID3A*$^{-}$ cells for further analysis. ANOVA analysis was performed within the Partek Flow software to identify DEGs with 2-fold or greater changes in expression and a False Discovery Rate (FDR) < 0.05. IPA analyses were performed on significant DEGs. RNA-seq data are publicly available through the GEO NCBI database under the accession number GSE138544.

Data Analyses

Data were statistically evaluated using the non-parametric Mann Whitney U test or the non-parametric Wilcoxon paired test to compare distribution of variables between groups. Correlations were evaluated using Pearson's Correlations. Statistical analysis was performed with Prism (Graphpad) software version 8.2. Differential gene expression was analyzed using ANOVA within the Partek Genomics Suite. Gene ontology analyses on differentially expressed genes were analyzed within the Partek Flow Genomics Suite using the gene set enrichment tool. P values of less than 0.05 were considered significant.

RESULTS

To identify genes associated with ARID3a expression in HSCs from aged versus young donors, we performed single cell RNA-seq analyses and analyzed each cell for the presence and level of ARID3a transcription as shown by the scatter plot (Fig. 18A). Analyses of *ARID3A* transcript by qPCR from bulk HSCs of known ARID3a protein expression suggest that transcript and protein expression in bulk HSCs correlate (data not shown). There were 153 ARID3a^{+} and 148 ARID3a^{-} cells from aged donors and 172 ARID3a^{+} and 92 ARID3a^{-} cells from the young

donors. Three-dimensional t distributed stochastic neighbor embedding plots (tSNE) of 301 aged and 264 young HSCs from 8 donors revealed considerable spread in dimensionality in the aged (circles) versus young (squares), shown as overlays (Fig. 18B and C). This suggests that isolation of HSCs using standard surface markers (Fig. 17A) results in cells that are heterogeneous with respect to their transcriptomes in both aged and young donors. Identification of ARID3a-expressing cells, as indicated by blue symbols, reveals widespread ARID3a expression in both aged and young HSCs, with increased clustering of ARID3a-expressing cells toward the right-hand side in the aged donor cells (Fig. 18B). Ingenuity Pathway Analyses (IPA) analyses of ARID3a-associated genes and non-*ARID3a* associated genes from aged donors revealed enrichment in pathways associated with cell cycle, regulation of B cell apoptosis, negative regulation of B cell activation, and positive regulation of histone methylation in the *ARID3A* cells (Fig 18D). Similar analyses of young donor cells indicated enrichment of pathways associated with lymphocyte homeostasis, JAK-STAT signaling and nucleic acid binding in the *ARID3A* cells (Fig 18E).

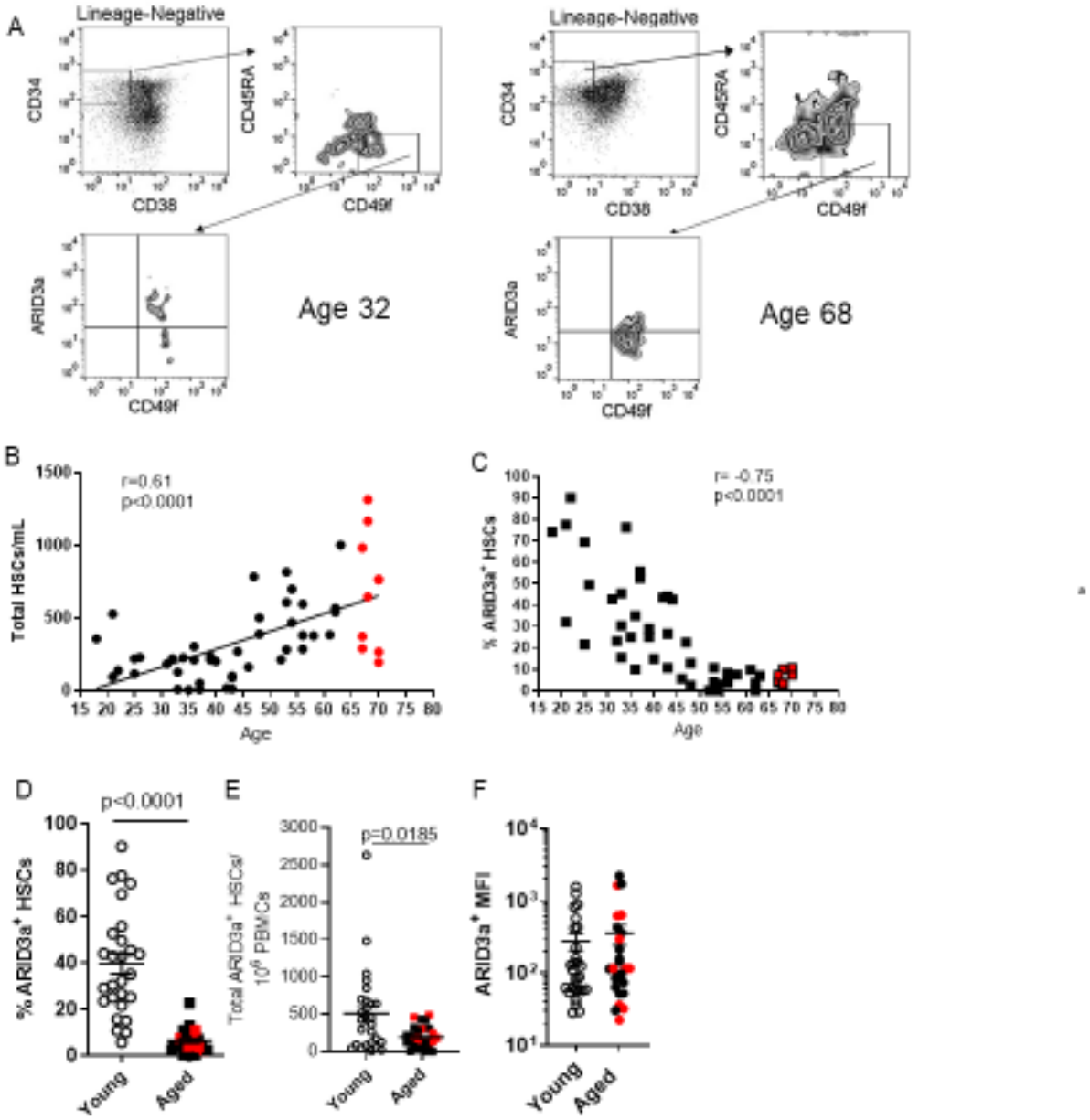
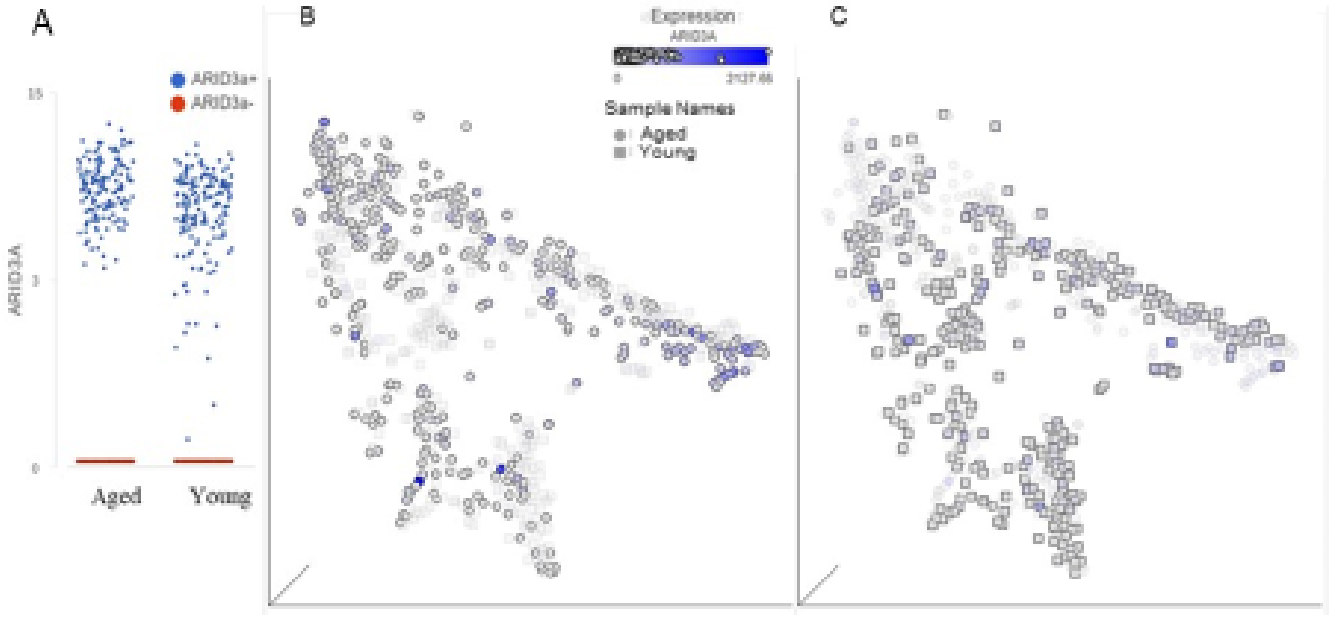


Figure 17. PBMCs from aged individuals show reduced frequencies of ARID3a$^+$ HSCs. Human HSCs expressing ARID3a were identified by flow cytometry from total PBMCs of 54 healthy adults (ages 18-70). (A) Representative gating from young (left) and aged (right) donors are shown. (B) Numbers of HSCs/ml of blood are plotted against age with a line of best fit. Red samples indicate samples from donors ≥ 65 years of age. (C) Frequencies of ARID3a$^+$ HSCs are plotted against age. Frequencies (D) and numbers (E) of ARID3a$^+$ HSCs for young (open circles) and aged (closed circles) donors are shown. (F) Mean Fluorescence Intensities (MFI) of ARID3a$^+$ HSCs from young and aged donors are presented. Averages and standard error bars are shown. Statistical significance was determined by Mann Whitney U test and Pearson's Correlation.

Direct comparison of *ARID3a*-expressing cells from aged versus young donors revealed 253 genes down-regulated >2-fold and 195 up-regulated genes. IPA of the ARID3a-associated genes from the young donors and the *ARID3a*-associated genes in the aged donors revealed enrichment in pathways associated with B1 B cell differentiation lymphocyte differentiation, type 1 interferon signaling, and regulation of gene expression in the aged donor cells (Fig 18F). The most highly up-regulated gene in the aged ARID3a^{+} HSCs versus young donor ARID3a^{+} HSCS is *AIF1*, a gene associated with macrophage activation (Fig. 18G), while the most highly down-regulated gene encodes a chemokine, *PPBP*. In addition, at least four transcription factors *HES1, MAFB, ATF4* and *JAK1* are differentially regulated in aged versus young ARID3a-expressing HSCs. Thus, alterations in gene expression profiles in young versus aged ARID3a-expressing cells may be the result of changes in these regulatory proteins. Together, *ARID3A*-expressing cells in aged donors differ in gene expression patterns from *ARID3A*-expressing young donor HSCs.



D

ARID3a+ Aged vs ARID3a- Aged
Top GO Terms Enriched in DEGs (1893genes)
Lymphocyte activation (GO 0046649) Padj <1.27 x 10-4
Negative regulation of glycogen biosynthetic process (GO 0045719) Padj <2.86 x 10-3
Cell cycle process (GO 0022402)) Padj <9.38 x 10-3
Regulation of B cell apoptotic process (GO 002902) Padj < 0.01
Negative regulation of B cell activation (GO 0050869) Padj <0.03
Positive regulation of histone methylation (GO 0045787) Padj <0.04

E

ARID3a+ Young vs ARID3a- Young
Top GO Terms Enriched in DEGs (548 genes)
translation (GO 0006412) Padj <1.11 x 10-8
translational initiation (GO 00064163) Padj <1.38 x 10-8
RNA processing (GO 0006396)) Padj <2.77 x 10-5
Nucleic acid binding (GO 003676) Padj < 2.01 x 10-4
JAK-Stat signaling (GO 0060397) Padj < 3.10 x 10-3
Lymphocyte homeostatsis (GO 0002260) Padj < 0.01

F

ARID3a+ Aged vs ARID3a+ Young
Top GO Terms Enriched in DEGs (448 genes)
translational initiation (GO 00064163) Padj <4.40 x 10-82
Protein transport (GO 0015833) Padj <4.29 x 10-27
Regulation of gene expression (GO 0010468)) Padj <2.88 x 10-15
Type I interferon signaling pathway (GO 0060337) Padj < 3.08 x 10-6
Lymphocyte differentiation (GO 0030098) Padj < 2.99 x 10-4
B1a B cell differentiation (GO 00002902) Padj < 8.59 x 10-3

G

ARID3a+ Aged vs Young	
Gene	FC
PPBP	-255.4
RPS4YI	-119.88
RPL36AL	-110.69
TRBC1	-108.6
TRAC	-88.53
DDX3Y	-75.47
CDC123	-66.54
TRBC2	-60.42
OST4	-53.57
PARL	-44.55
SHC3	14.42
AC139749.1	14.98
OR52M1	16.56
DUT	17.75
HES1	27.06
MRPL47	54.61
DAD1	60.82
S100A8	66.07
OR2B3	144.85
AIF1	299.23

Figure 18. ARID3a$^+$ HSCs from aged donors express altered transcriptomes compared to ARID3a$^+$ HSCs from young donors. Single-cell RNA-seq expression profiles from 4 young (ages 19, 21, 37, and 40) and 4 aged (ages 61, 66, 68, and 70) donors were obtained and analyzed based on *ARID3A* transcript levels (*ARID3A*$^+$ >0.5 CPM, *ARID3A*$^-$ <0.5 CPM). Data from 264 young donor cells (172 *ARID3A*$^+$ and 92 *ARID3A*$^-$) and 301 aged donor cells (153 *ARID3A*$^+$ and 148 *ARID3A*$^-$) were assessed. (A) ARID3A transcript levels for young and aged donor cells are shown. TSNE plots of aged donor HSCs (circles) are overlaid with shaded young donor HSCs (squares) (B) and ARID3a expression levels are indicated by intensities of blue dots. In (C), TSNE plots are overlaid with young donor HSCs over the shaded aged donor HSCs for better visibility of the ARID3a-expressing cells in each group. The top GO terms enriched in DEGs from ARID3a$^+$ vs ARID3a$^-$ Aged HSCs are given with p values (D) and for young ARID3a$^+$ versus ARID3a$^-$ cells in (E). (F) Top GO terms directly comparing ARID3a$^+$ cells in aged versus young donors are shown. (G) The most differentially expressed genes in ARID3a+ HSCs in aged versus young donors are presented with negative or positive fold change (FC).

DISCUSSION

The data presented here suggest that HSCs from both aged and young donors contain equivalent numbers of ARID3a-expressing cells, although total frequencies of ARID3a+ cells were reduced in aged individuals. Analyses of hematopoietic lineage potential revealed that aged donor HSCs yielded reduced numbers of B cells compared to those derived from young donor HSCs, despite the presence of equivalent total numbers of ARID3a-expressing progenitors. These data suggest ARID3a expression alone is insufficient to cause age-associated shifts in B cell generation. Surprisingly, the aged donor-derived B cells retained CD34 expression to a greater degree than B lymphoid lineage cells derived from young donors, implying the aged donor B lineage cells could be less mature than young donor derived B cells. Artificial over-expression of ARID3a in aged donor-derived HSCs suggested that increased ARID3a expression resulted in better B lineage development without retention of surface CD34. ARID3a expression was critical for B lineage development from young donor peripheral blood-derived HSCs, consistent with our previous findings using HSCs from cord blood (76). Importantly, single cell RNA-seq of both aged and young donor HSCs expressing ARID3a revealed a number of

transcriptome differences between these cells, suggesting that the ARID3a-expressing cells in aged individuals differ at the molecular level from ARID3a-expressing cells from young individuals. These differences may contribute to the functional differences observed in hematopoiesis in vitro. Together, these data suggest that ARID3a-expressing progenitor cells in aged individuals differ at the transcription and functional level from those present in young individuals.

In support of our surface expression data suggesting that the ARID3a-expressing cells from aged individuals differ from ARID3a^{+} HSCs in the young, single cell transcriptome analyses revealed major differences in gene expression as well. GO analyses of young *ARID3A*-expressing HSCs and aged *ARID3A*-expressing HSCs reveal an enrichment in the pathways associated with B1 B cells differentiation in the aged cells (Fig 18D). Several elegant studies from the Hardy group in the last few years have elucidated the role of ARID3a in B1 B cell development in fetal/neonatal murine development (212-214). While human B1 cell identification remains controversial, others reported that cord blood HSPCs and mouse fetal liver HSPCs are very similar phenotypically and functionally (215,216). We previously found that ~75% of human umbilical cord blood HSCs express ARID3a where it is important for B lineage development (76). These data suggest the interesting possibility that the ARID3a-expressing HSCs from aged donors could be enriched for B1 lineage-like development and might have fewer precursors that would ultimately result in conventional B lymphocytes. Murine bone marrow derived B1 B cells are suggested to be continually generated into old age (217), a population that has been reported to have separate progenitors than conventional B cells (218). Alternatively, age-associated B cells (ABCs) are a unique heterogeneous subset of B cells that have been the focus of study for several laboratories in the last decade. First described separately

by both the Cancro and Marrack research groups, these cells expand in old age and have been attributed to origins from multiple types of B lineage cells, including B1 B cells (219-221). However, no one has demonstrated the origins of those cells from early hematopoietic progenitors. Our data suggest that alterations in humoral immunity attributed to alterations in B cell maturation in aged individuals are already evident at the gene level in early hematopoietic stem cells. The differences in gene expression that we observed between young and aged donor HSCs are generally consistent with published reports from others using young and aged HSPCs (222,223). However, as shown in Figure 18, considerable heterogeneity remains at the transcriptome level even in purified HSCs. The ability to directly compare populations with the B cell-associated protein ARID3a underscores additional differences in aged versus young donor-derived HSCs, and present a new caveat that even within the ARID3a-expressing cells, some progenitors may be predisposed to develop into B1 versus conventional B cells as suggested by the GO analyses. More studies are clearly needed to better define the discrete progenitors within HSCs and the associated transcriptome differences between young and aged individuals.

CONCLUSIONS

The data presented here demonstrate that ARID3a-expressing HSCs total numbers are equivalent between old and young donors, yet the aged donor HSCs produced fewer B cells in vitro. Functional defects in aged donor ARID3a-expressing HSCs were reflected both in expression of surface protein markers and at the transcriptome level, suggesting either that the ARID3a-expressing HSCs from aged donors are functionally defective in maturation responses, or that the HSCs expressing ARID3a in aged donors reflect skewing of a precursor subset of B

lineage cells with different capacities to expand in these in vitro cultures. In addition, these data highlight the expansion of HSCs in aged individuals that do not express ARID3a, raising the possibility that these cells contribute to age-associated defects in hematopoiesis.

Chapter 4 Addendum

ARID3a is expressed in hematopoietic stem cells (HSCs) and a colleague in the lab discovered that ARID3a expression changed in HSCs from aged individuals. My goal was to help analyze differences in ARID3a expression at the single cell level to prepare for analyzing naïve B cells using scRNA-seq. Therefore, my contribution to this manuscript (Ratliff ML, Garton J, James JA,Webb CF. ARID3a Expression in Human Hematopoietic Stem Cells is Associated with Age-Related Reduction in B Lineage Development. *Immunity & Aging*. 2020; In Press) was to analyze the scRNA-seq data from 4 aged and 4 young samples where we found reduced ARID3a expression in stem cells from individuals over the age of 55 when compared to samples from young individuals. Briefly, quality control was performed on the raw fastq files whereby sequencing adapters were trimmed and low quality paired-end reads were removed before aligning them to the hg38 assembly of the human genome using the STAR program (224). The Partek Genomics Suite was run on the aligned reads to estimate gene expression levels. Cells that had greater than 25% of counts mapping to mitochondrial genes indicate stressed or dying cells and were excluded from analyses. The number of detected genes was 44,025 across 301 cells for aged samples and 39,417 genes across 264 cells for young samples. Log normalized counts per millions (CPM) were $\log_2$-transformed and used for t-Distributed Stochastic Neighbor Embedding (t-SNE) visualization and differential expression analyses. Data from aged and young samples were interrogated separately based on ARID3a expression, with >0.5 CPM denoting ARID3a$^+$ cells and <0.5 CPM denoting ARID3a$^-$ cells for differential expression

analysis. ANOVA was performed to identify DEGs with 2-fold or greater changes in expression and a FDR < 0.05. Significant DEGs were then subjected to IPA analysis to identify enriched pathways and gene networks.

CHAPTER 5

ROLE OF ARID3A IN SLE NAÏVE B CELLS

The entirety of the work in this chapter was processed and analyzed by me. Colleagues in the Webb lab aided in collection of B cell subsets from PBMCs. I was tasked with staining and sorting using flow cytometry, capture of single cells, extraction of RNA, cDNA synthesis, quality control (i.e. obtaining RNA integrity numbers, fragment size distribution, concentration of RNA/cDNA, preprocessing of sequencing reads), ligation of sequencing adapters, design of sequencing library indices, fragment size purification of libraries, pooling/normalization of libraries for sequencing, preprocessing of sequencing files and data analysis (as described in Chapter 2-4).

Copyright Information

N/A

Allocation of Contribution

The entirety of this chapter, including sample processing and data analysis, was written by me.

ABSTRACT

Systemic autoimmune diseases, which include systemic lupus erythematosus (SLE), systemic sclerosis, and rheumatoid arthritis, among others are a complex group of disorders that effect millions of people. Systemic lupus erythematosus is an autoimmune disorder that primarily effects young women and is characterized by the loss of tolerance to self with the destruction of host tissue. The cause of SLE is currently unknown and new treatment options have been scarce. Previous data also linked ARID3a expression to interferon alpha (IFNa) expression. IFNa is a cytokine previously associated with inflammatory responses in nearly half

of all lupus patients. Recently, the DNA-binding protein ARID3a was found to be highly correlated with disease activity in SLE patients (44,46). Furthermore, ARID3a is not normally expressed in naïve B cells from healthy individuals but is highly expressed in naïve B cells from SLE patients. We hypothesize that ARID3a-expressing naïve B cells are precursors to marginal zone (MZ)- like B cells that secrete interferon and can induce interferon production in other cell types, leading to the associated disease phenotypes in SLE (46). The long-term goal of this project is to develop a better understanding of how ARID3a expression contributes to disease activity in SLE, and to identify therapeutic targets and surface biomarkers specific to ARID3a^{+} B cell subsets. Next-generation sequencing (NGS) and microfluidic technology will be used to capture single ARID3a-expressing and ARID3a negative B lymphocytes and to examine the transcriptome within individual B cell subsets in SLE patients. At present, there are significant knowledge gaps regarding how the expression of ARID3a contributes to increased disease activity and what causes the increased inflammatory responses in SLE patients. Because ARID3a is associated with alterations in chromatin accessibility, we expect there to be differentially expressed genes under the control of ARID3a regulation. Therefore, our current studies have taken advantage of the bimodal expression of ARID3a within individual SLE samples and single-cell technology to perform RNA-seq analyses. We predict that these studies will be important for the understanding of how ARID3a expression contributes to disease activity in SLE patients and will provide new information on its role in development.

INTRODUCTION

B cells are essential for adaptive immunity and their development requires strict regulation (225,226). They are important in modulating immune responses to self and foreign antigens, and dysregulation of complex signaling networks are implicated in increased

autoimmune disease activity, such as in systemic lupus erythematosus (SLE) (227). The cause of SLE is unknown and only one new treatment has been developed to treat SLE in more than 60 years (228); however, previous studies demonstrated that AT-rich interacting domain 3a (ARID3a), a DNA-binding protein, is more abundant in SLE B lymphocytes than in healthy control B cells (44). Increased levels of ARID3a-expressing B cells correlate with increased disease activity (44,46). Interestingly, recent studies indicated that ARID3a is important for the expression of the key inflammatory cytokine, interferon-alpha (IFNa), in B lymphocytes and other cell types (44-46). Others have observed a correlation between IFN transcriptional signatures in peripheral blood cells and SLE disease activity (SLEDAI) (112). Indeed, ARID3a-expressing B lymphocytes appear to denote a new type of B effector cell that can induce IFNa expression in other cell types (46). Single cell analysis on PBMCs from SLE patients identified clusters of autoreactive naive B cell subsets that have high levels of TLR7 expression, which may be precursors double negative (DN2) (229,230) and activated naïve B cells (230). It could be possible that these ARID3a-expressing cells have broken the tolerance checkpoint, which allows autoreactive B cells to survive and expand (227). Therefore, it is important to better define the characteristics of these cells. ARID3a is bimodally expressed in both SLE and healthy controls, such that only a fraction of the cells within a given B cell subset express ARID3a at any given time. We hypothesize that ARID3a is an important contributor to increased disease activity in SLE patients. Further, since deletion of ARID3a in K562 resulted in changes to chromatin accessibility (Chapter 3, manuscript submitted) we suppose that it will influence gene expression patters, either directly or indirectly, in B lymphocytes. The proposed research objective will use new single-cell technology to perform RNA-seq analyses of ARID3a-expressing and ARID3a

negative SLE B cells to better define differentially expressed genes. This manuscript aims will shed light on how ARID3a expression contributes to disease pathogenesis in patients with SLE.

METHODS

Flow cytometry and B cell isolation

SLE B cells from PBMCs were enriched for B lymphocytes via negative selection using magnetic beads (Stem Cell Technologies), and were then fluorescence activated cell sort (FACS) purified using doublet exclusion to isolate single cells of > 99% purity. PBMCs were isolated from heparinized peripheral blood (~15 ml) with Ficoll-Paque Plus (GE Healthcare), and stained with the following fluorochrome-labeled antibodies: CD19 PE-Cy5, CD10 Pacific Blue (BioLegend), IgD PerCP-Cy5.5, CD27 PE-Cy7, and IgM APC (Southern Biotech). When possible, B cells were fixed (3% paraformaldehyde) and permeabilized (0.1% Tween-20) prior to staining with goat anti-human ARID3a antibody and a rabbit anti-goat IgG FITC secondary (Invitrogen). Gating for individual B cell subsets was described previously (44) and used with the following B ($CD19^+$) cell subset markers: naïve ($IgD^+CD27^-CD10^+$) B cells. Isotype controls (Caltag, BD Pharmingen, and eBioscience) were used for gating. Data (500,000 events per sample) were collected using an LSRII (BD Biogenics) and FACSDiva (BD Biosciences) software version 4.1 and were analyzed using FlowJo (Tree Star) software version 9.5.2.

Single-cell RNA-seq

Total ($CD19^+$) B cells from one SLE patient and naïve B cells from three SLE patients were FACS-sorted and immediately used for single cell analyses. Smart-seq/C1 libraries were prepared on the Fluidigm C1 system using the SMARTer Ultra Low RNA Kit (Clontech)

according to the manufacturer's protocol. Cells were loaded on a 5-10 μm RNA-seq microfluidic IFC at a concentration of 200,000/ml. Capture site occupancy was surveyed using a standard light microscope and recorded to verify cell capture. Sequencing library amplification was performed using Nextera XT Index primers (Illumina) according to manufacturer's protocol. Barcoded library concentration and fragment size distribution was determined using Agilent High Sensitivity D1000 kit on an Agilent 2200 TapeStation (Agilent Technologies) at the OMRF Genomics Core Facility. Paired-end (2 x 50bp) sequencing was performed on a NovaSeq 6000 platform by the OMRF Genomics Core Facility.

Paired-end reads were aligned to the hg38 genome assembly with Bowtie2 using default parameters. Gene expression estimates (TPM) were generated using RSEM. Sequencing adapters and low quality reads were removed within the Parteks Genomics Suite using Cutadapt. Cells that had greater than 25% of counts mapping to mitochondrial genes indicate stressed or dying cells and were excluded from analyses. Gene expression, in transcript per millions (TPM), were $\log_2$-transformed and used for PCA visualization and differential expression analyses. Data from naïve ($CD19^+$ IgM^+ IgD^+) B cell SLE samples were batch corrected and interrogated together based on ARID3a expression, with >0.5 TPM denoting $ARID3A^+$ cells and <0.5 TPM denoting $ARID3A^-$ cells for further analysis. GSA analysis was performed within the Partek Flow software to identify DEGs with 2-fold or greater changes in expression and a False Discovery Rate (FDR) < 0.05. DEGs were then subjected to pathway and network analysis using Ingenuity Pathway Analysis (Qiagen).

RESULTS

We processed naïve ($CD19^+$ IgM^+ IgD^+) B cells from three SLE patients (Figure 18). We previously identified ARID3a protein, but not transcript, levels to be associated with large

changes in gene expression in LDNs and pDCs of SLE patients (13). Flow cytometry was used to determine investigate if ARID3a protein levels correlate with transcript in B cells, when possible (Figure 19). Indeed, one representative sample had 93% of naïve B cells expressing ARID3a protein and 84% of cells expressing ARID3a transcript. Although it was not possible to measure ARID3a protein levels for every sample, data from the several that we looked at suggest that transcript correlates with protein in these cell types. A representative diagram of our single cell isolation, sequencing, and analysis is shown in Figure 20.

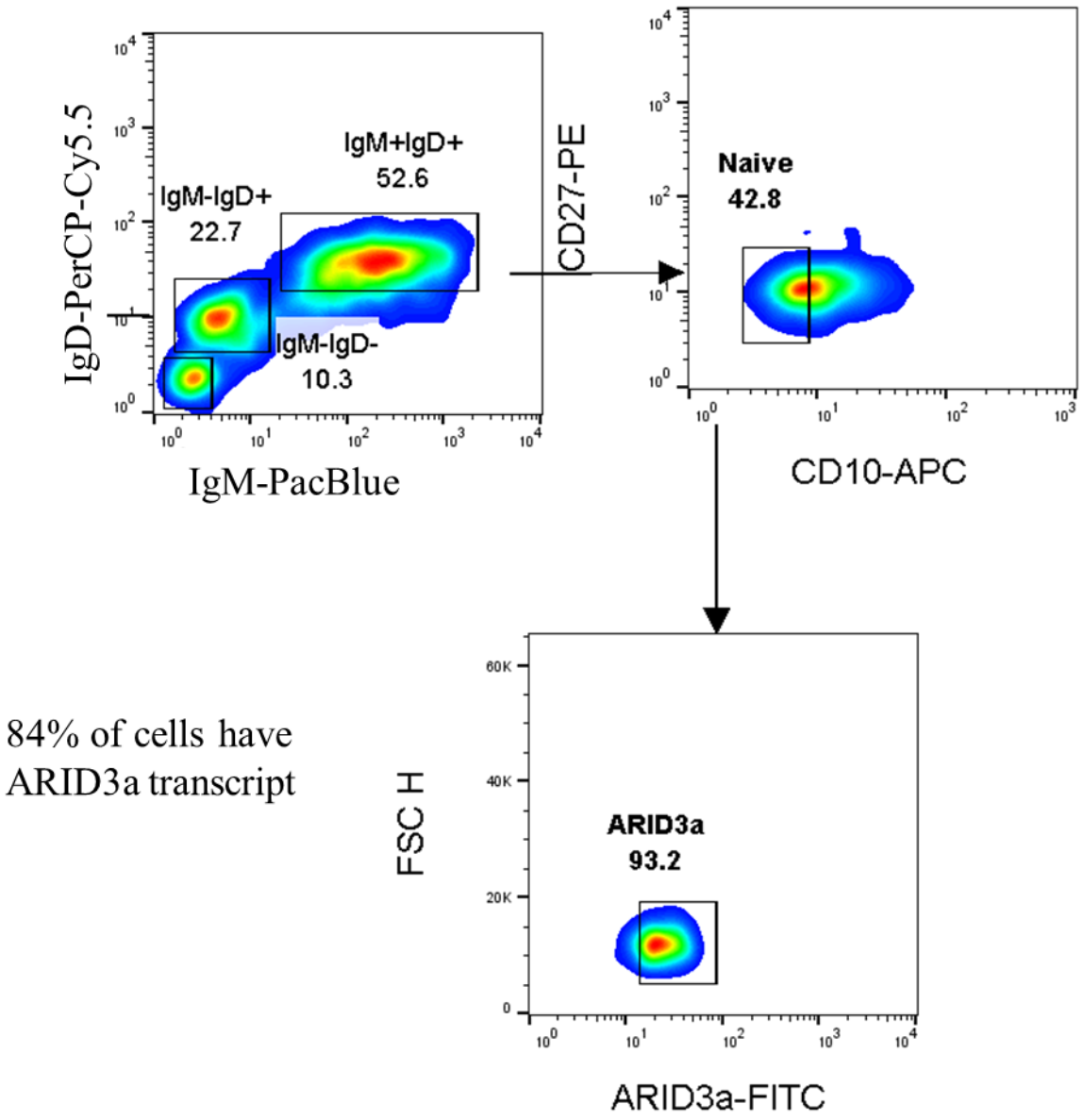


Figure 19. Percentage of ARID3a$^+$ B cells levels were measured by flow cytometry. Naïve (IgM^+ IgD^+) B cells show a tight correlation between transcript (84% of cells ARID3a$^+$, not shown) and protein levels (bottom right panel).

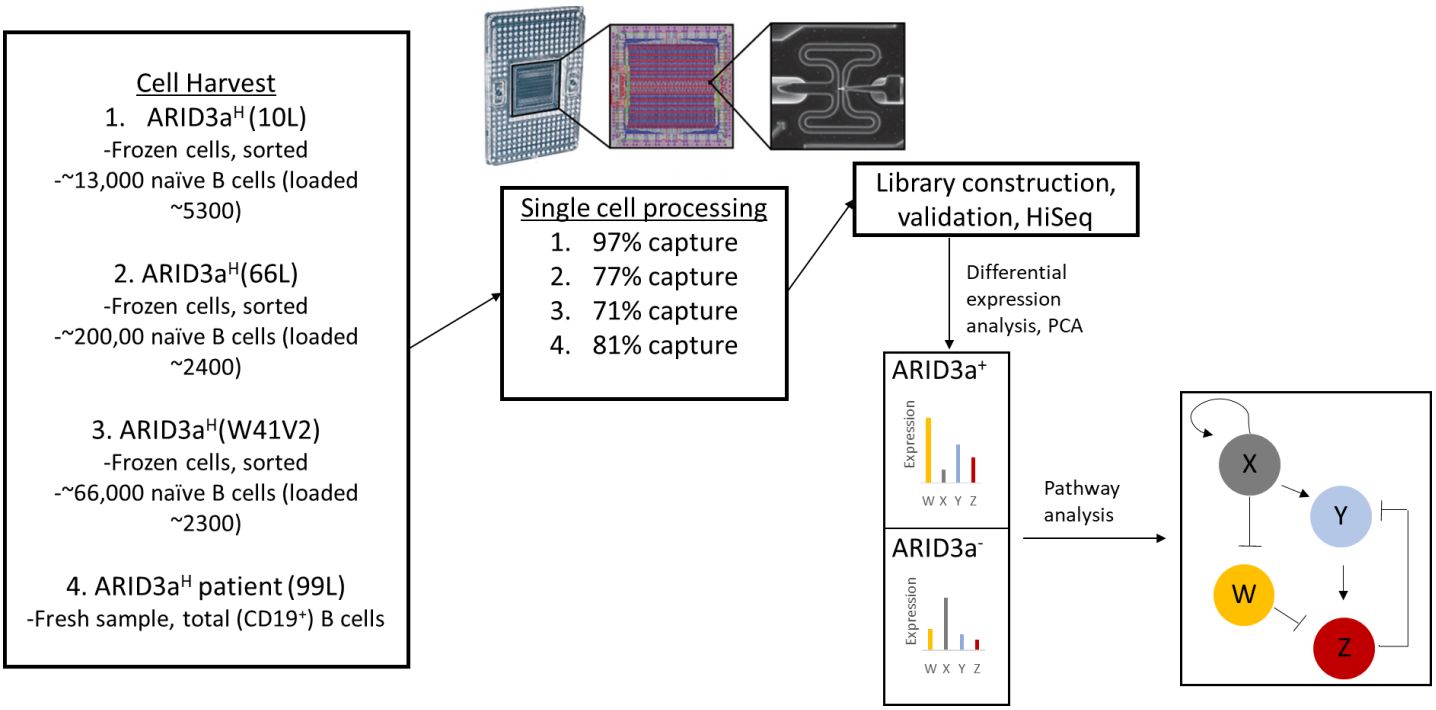


Figure 20. Schematic diagram of isolation, sequencing, and analysis of scRNA-seq. Capture rate of single cells was confirmed by visual inspection using a standard light microscope prior to construction of sequencing libraries. Preprocessing of raw sequencing reads for each cell was performed prior to binning single cells based on ARID3a expression. Principal component analysis was performed on all detected genes of high-quality cells. Differential expression analysis was then performed on ARID3a^{+} vs ARID3a^{-} cells. Differentially expressed genes were then subjected to pathway and network analysis using Ingenuity Pathway Analysis.

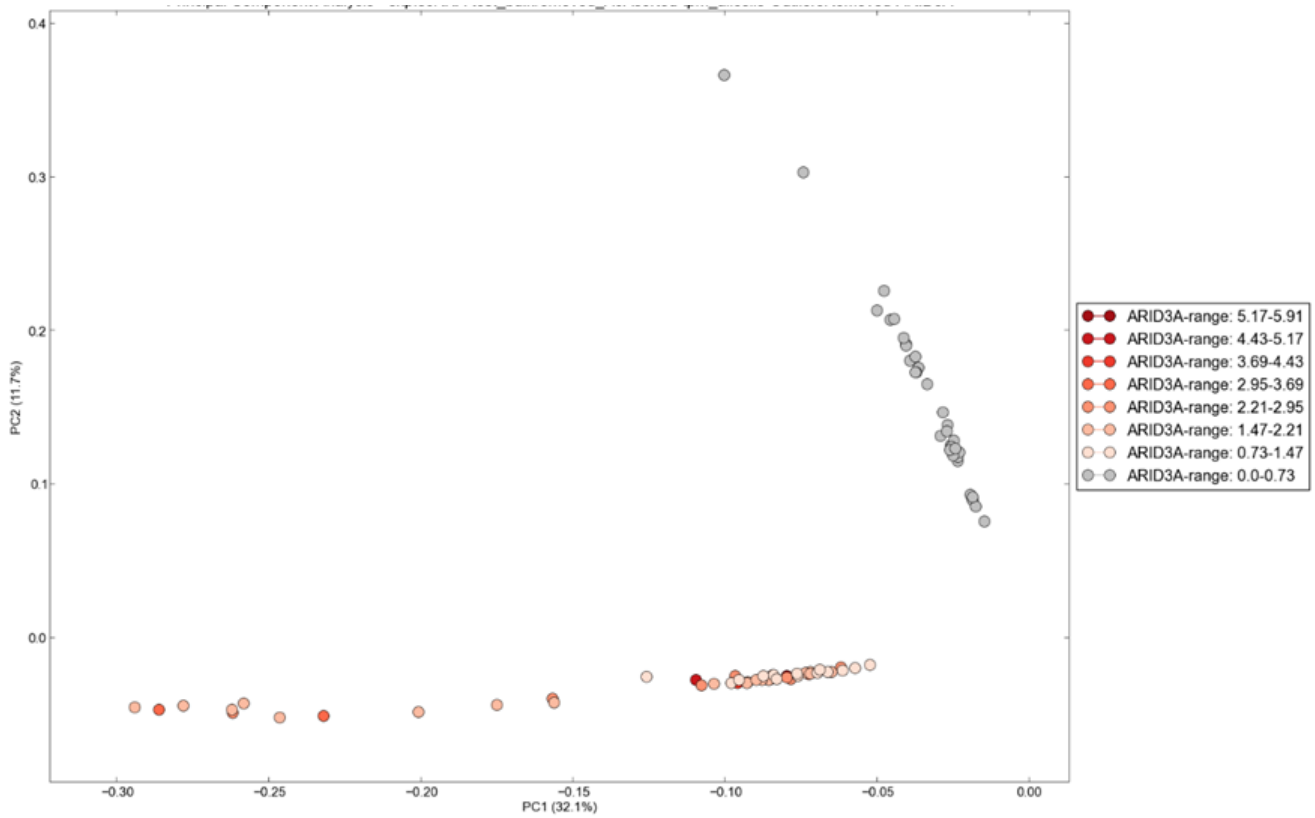


Figure 21. PCA plot using all detected genes in high-quality single cells from one CD19$^+$ (n=58) and three naïve (CD19$^+$ IgM$^+$ IgD$^+$)(n=249) SLE patient reveals ARID3a$^+$ cells cluster together and away from ARID3a$^-$ cells. ARID3a expression for each individual cell is indicated by the red-orange colors.

Principal component analysis was performed on both the total (CD19$^+$) and the naïve (CD19$^+$ IgM$^+$ IgD$^+$) B cells to determine how ARID3a$^+$ cells compare to ARID3a$^-$ cells (Figure 21). Since PCA groups cells with similar expression patterns and ARID3a$^+$ cells cluster together, we hypothesize that we will identify DEGs that differ between ARID3a$^+$ and ARID3a$^-$ B cells. This analysis reveals that ARID3a-expressing B cells cluster together and away from all other cells, suggesting that ARID3a expression is associated with a distinct transcriptome or cell type. We then chose to focus our analysis on naïve B cells because this is the cell population in SLE patients that have increased expression of ARID3a when compared to healthy controls (44,46). A total of 248 naïve SLE B cells from three SLE patients were captured and analyzed together.

Quality control was performed on these cells prior to PCA and differential expression analyses. Cells with 25% or greater reads mapping to mitochondrial genes were removed, leaving 179 cells for downstream analysis (Figure 22).

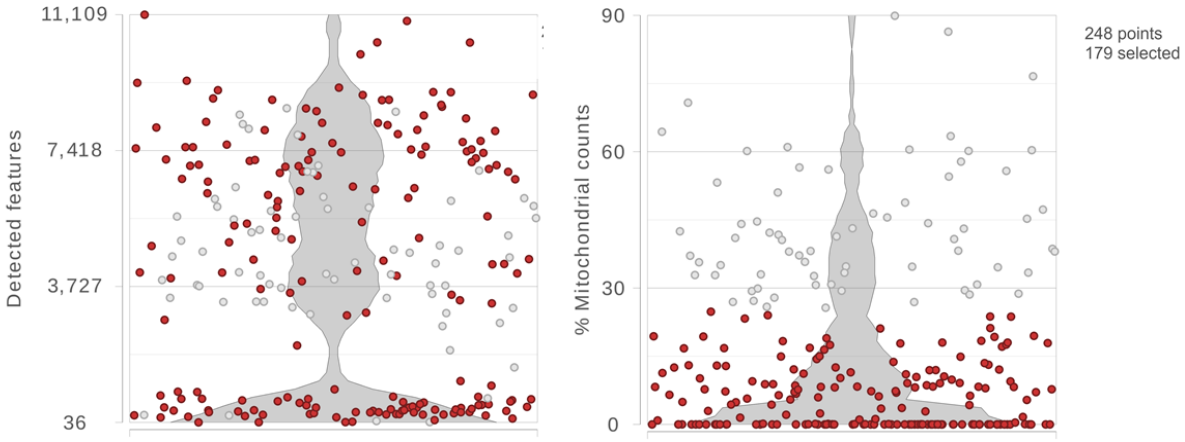


Figure 22. Quality control of scRNA-seq data from 3 SLE naïve B cell samples. Violin plots showing the density of the numbers of detected genes (left panel) and the percentage of sequencing reads mapping to mitochondrial genes (right panel) for each of the 248 cells isolated from 3 SLE naïve B cell samples. Cells with less than 25% mitochondrial reads were retained, leaving 179 cells for downstream analysis (red circles).

PCA analysis on the remaining 179 naïve B cells from three SLE patients also indicate ARID3a expression is associated with a unique transcriptome (Figure 23). Therefore, these data suggest that it is possible to identify cells with ARID3a protein by measure transcript levels with RNA-seq. Because the PCA analyses showed clustering of cells that expressed ARID3a transcript, we performed differential expression analyses on ARID3a^{+} (transcripts per million (TPM) > 0.5) vs ARID3a^{-} (TPM < 0.5) to determine how ARID3a-expressing cells differ at the transcription level from naïve B cells that do not express ARID3a in the same subsets of SLE patients. This classification resulted in 90 ARID3a^{+} cells and 89 ARID3a^{-} cells.

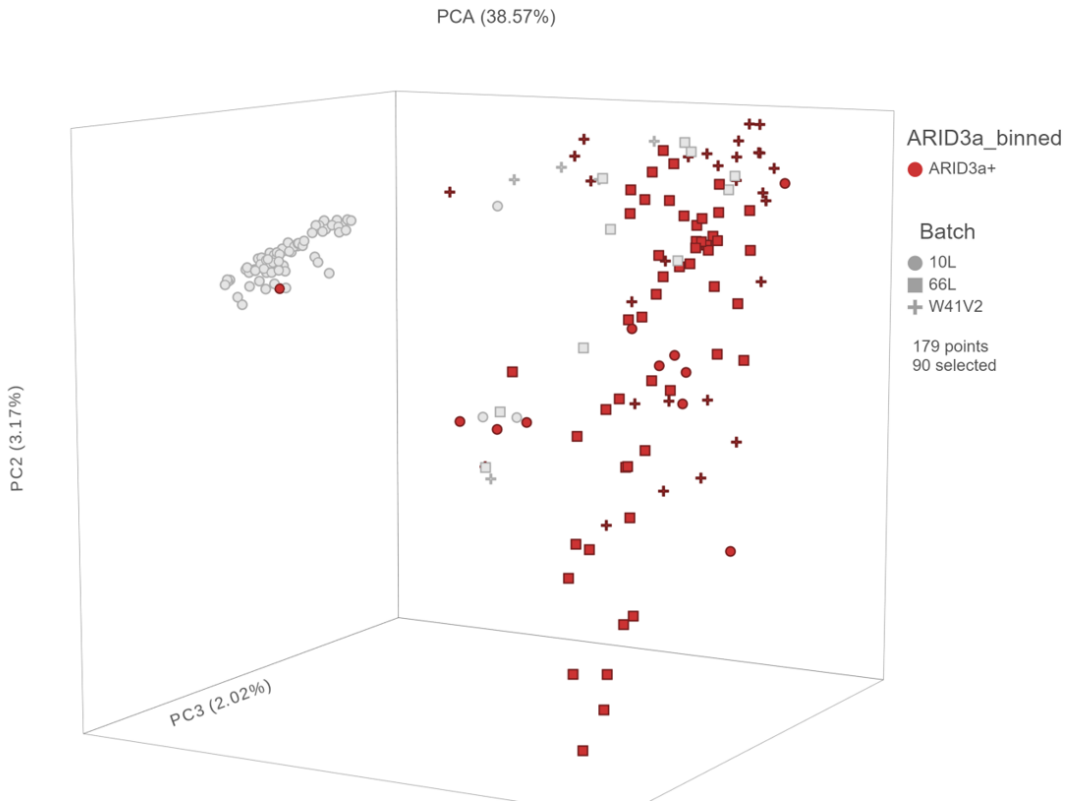


Figure 23. PCA plot using all detected genes in single naïve B cells from 3 SLE patients reveals ARID3a^{+} cells cluster together and away from ARID3a^{-} cells. Individual naïve B cells that passed quality control are plotted. Red color indicates ARID3a^{+} cells, gray color indicates ARID3a- cells, and the shape of each cell indicates the SLE patient they were isolated from.

To investigate transcripts associated with ARID3a expression, differential expression analysis was performed on ARID3a^{+} (n=90) vs ARID3a^{-} (n=89) naïve B cells from 3 SLE patients. This analysis identified 3,006 differentially expressed genes (DEGs) (FDR < 0.05, FC ≥ ±2) (Figure 24). The top 10 up- and downregulated genes are displayed in Table 3.

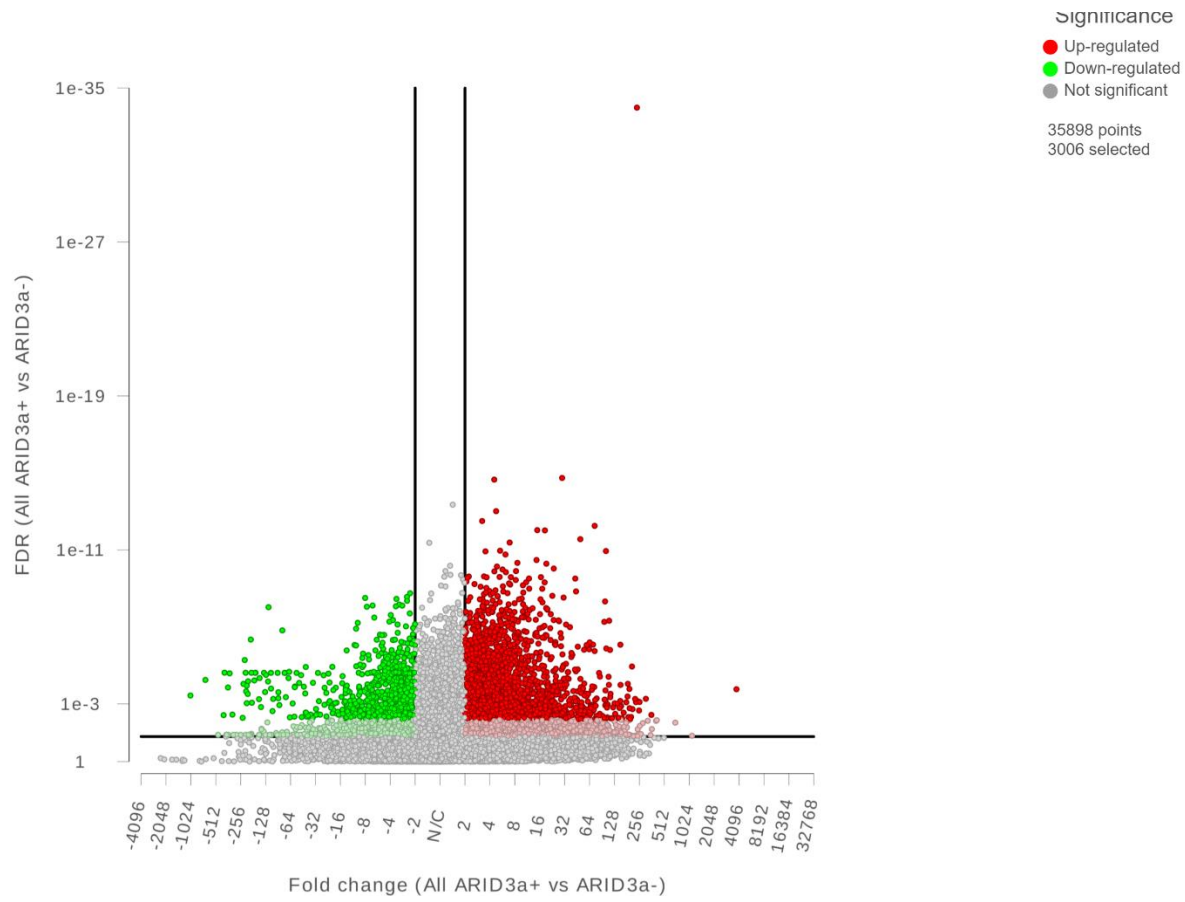


Figure 24. Volcano plot of differentially expressed genes (FDR < 0.05, FC > 2) identified by analyzing ARID3a+ vs ARID3a- naïve B cells from 3 SLE patients. 3,006 genes were identified by this analysis.

Table 3. Top DEGs identified by analyzing ARID3a^{+} vs ARID3a^{-} naïve B SLE cells

Gene	FDR	Fold Change
MTUS1	3.23E-03	3,807
ATP5L	0.03	357
ECSIT	7.60E-03	306
NPM1	8.00E-04	257
CASZ1	2.21E-03	253
AC119751.2	7.39E-04	245
ARID3A	1.05E-34	239
AC140113.2	8.14E-04	233
CD83	5.54E-04	227
HNRNPA1P30	1.88E-03	211
EDF1	5.69E-03	-233
FAM98A	1.42E-03	-228
CHMP2A	1.99E-03	-237
AC0055692.1	4.13E-03	-242
RPPH1	3.10E-03	-322
AL449404.2	2.87E-03	-348
AC010533.1	2.74E-03	-364
AL139824.1	2.49E-03	-401
NFASC	2.43E-03	-411
AC107905.1	1.46E-03	-682
NCAN	5.69E-03	-1,036

Among the most significant pathways were oxidative phosphorylation, iron homeostasis pathways, B cell receptor signaling, SLE and TLR signaling pathways (not shown). TLR7 was identified to be 3-fold upregulated in ARID3a^{+} cells and signaling through this receptor results in transcription of pro-inflammatory genes, such as IFNa.

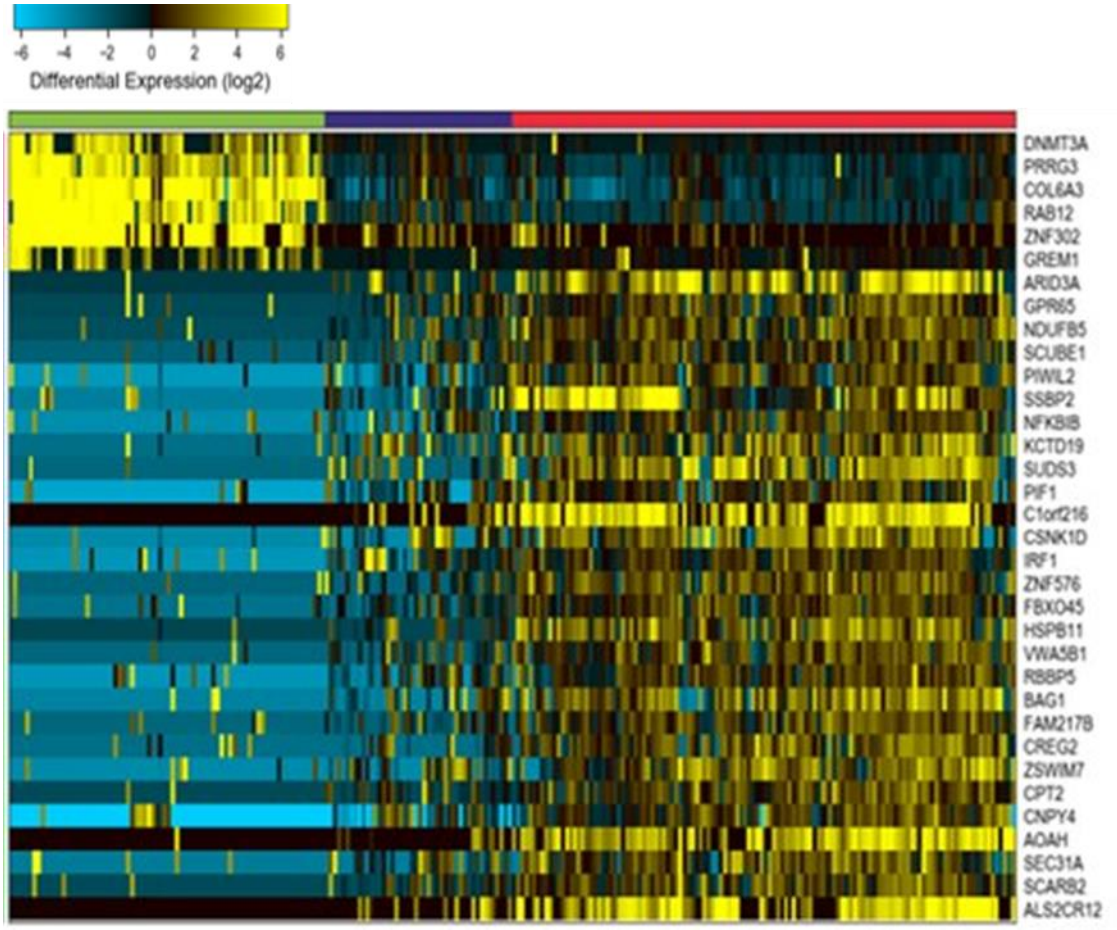


Figure 25. ICGS analysis of 179 CD19+IgM+IgD+ B cells from three SLE patients. Heatmap of genes delineated by ICGS (excluding cell-cycle genes) in scRNA-seq data (n=179 cells). Gene expression clusters were generated using hierarchical-ordered partitioning and collapsing hybrid (HOPAC) algorithm. Three clusters were identified (top green, blue, and red bars). ARID3a was highly expressed in naïve B cells within cluster 3 (red bar).

Hierarchical-ordered partitioning and collapsing hybrid analysis identified 3 clusters within the naïve single cell data (Figure 25). Cluster 1 (Figure 25, green bar) was enriched for DNA methyltransferase, DNMT3A, PRRG3, RAB12, and ZNF302. ARID3a was enriched in cluster 3 (Figure 25, red bar), along with the NFKB inhibitor, NFKBIB, IRF1, and BAG1. It is important to mention that these data are preliminary and additional samples are required to increase statistical power.

Interestingly, Ingenuity Pathway Analysis (IPA) identified the EHMT1 gene network to be inhibited in ARID3a^{+} naïve B cells (Figure 26). Our ATAC-seq data on ARID3a$^{-/-}$ KO K562 cells identified significant upregulation of EHMT1 (Chapter 3, manuscript submitted). This suggests ARID3a may directly control the expression of EHMT1 in multiple cell types.

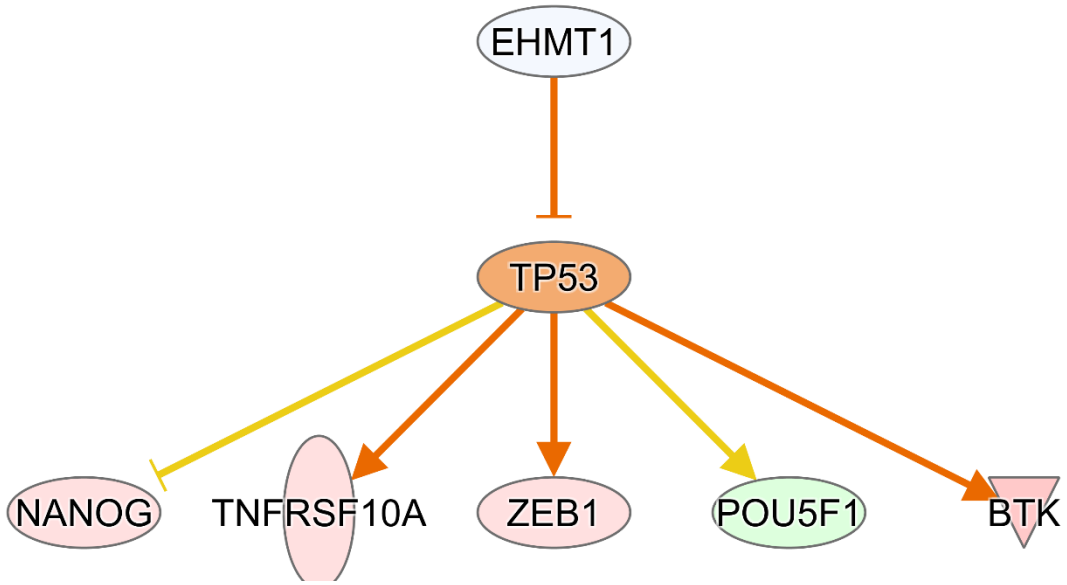


Figure 26. Casual network analysis using IPA identified EHMT1 network to be significantly inhibited in ARID3a^{+} naïve B cells. Orange color indicates upregulation, blue (EHMT1) and green (POU5F1) indicate downregulation. Orange lines indicate the gene upregulates the downstream targets. Yellow lines indicate unknow interaction between the downstream targets.

DISCUSSION

It has previously been shown that ARID3a is differentially expressed in naïve B cells of SLE patients (44,46). Patients with SLE have major blood B lymphocyte alterations, including expansion of activated naïve B cells (231-233), PCs, and extrafollicular DN2 cells (229). This study demonstrates that ARID3a^{+} naïve B cells have a distinct transcriptome when compared to naïve B cells that do not express ARID3a. At the time of writing this chapter, we expect to

process 3 additional SLE patient samples to obtain sufficient numbers of cells for statistical validity. Although this work is ongoing, I hypothesize that ARID3a-expressing cells represent precursors of immune cells fated to contribute to autoimmunity, or new subsets of cells that are not properly deleted in lupus compared to healthy individuals. To date, these data support this hypothesis and we show that ARID3a^{+} naïve B cells have increased expression of pro-inflammatory mediators and differential expression of TFs involved in cell fate commitment.

Differential expression analysis between ARID3a^{+} vs ARID3a^{-} identified the upregulation of TLR7 and IRF4. IRF4 is involved in differentiation of naïve B cells into antibody secreting cells and receptor editing (234). Normally ignorant naïve B cells, which have not been exposed to antigen, may be activated when their autoantigens are also ligands for TLRs. Some mice models have been shown to the ability to be activated through a TLR7-dependent process (235). It could be that ARID3a^{+} naïve B cells are more inclined to escape tolerance and induce clonal expansion of autoreactive naïve B cells into plasma cells through increased TLR7 signaling (236). B cells can internalize CpG sequences, which results in production of anti-chromatin autoantibodies. Autoantibodies against DNA, chromatin, and ribonucleoproteins are produced in SLE. It could be that ARID3a expression activates TLR7, lowering the threshold for activation by self-antigen. The increased expression of IRF4 in ARID3a^{+} could also increase the differentiation of naïve B cells activated by self-antigens. This suggests that increased ARID3a levels in naïve B cells of SLE patients affects editing of autoreactive BCRs or predisposes naïve B cells to become activated by self-antigen. but it is unclear why these autoreactive B cells escape anergy and clonal deletion. One possibility is that ARID3a^{+} cells either expand rapidly to bypass tolerance or by inhibiting apoptosis pathways In support of this notion that ARID3a-expressing naïve B cells escape clonal deletion, suppressors of apoptotic pathways, such as

IKBKB and XIAP, were upregulated in ARID3a^{+} naïve B cells (Figure 27). Double negative switched memory (DN2) (CD19+ IgD- CD27- CXCR5-) B cells were recently found to be expanded in SLE patients and are hypersensitive to TLR7 stimulation (229). Since this work focused solely on precursor naïve B cells, we were not able to identify if ARID3a is enriched in DN2 cell populations. We hypothesize that ARID3a^{+} naïve B cells in SLE patients may be poised to differentiate into DN2 cells. Our data support this as many of the genes found to be significantly upregulated in DN2 cells, such as IRF4, were also upregulated in ARID3a^{+} naïve B cells from SLE patients. In addition, ARID3a^{+} naïve B cells had high expression of CD38, TLR7, IL10RA, FCRL2, FCRL3, FCRL5, FGR, TFEC, and CD9 (Figure 27), supporting the notion of its extrafollicular, potentially autoreactive precursor DN2 phenotype (230).

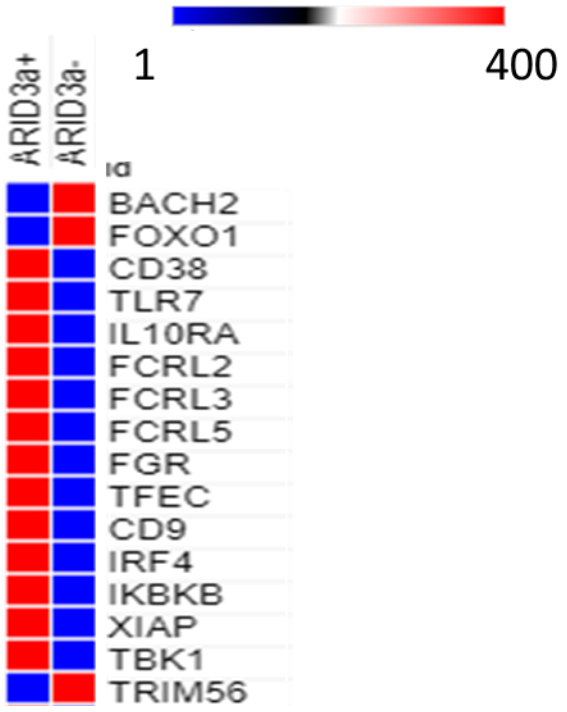


Figure 27. Gene expression in ARID3a^{+} and ARID3a^{-} naïve B cells. Expression, in TPM, of select transcripts in scRNA-seq data.

Although, we did not see high expression of T-bet (TBX21), which is uniquely expressed at high levels in DN2 cells (229), but it may be that ARID3a^{+} naïve B cells are a precursor to DN2 cells. To further support this hypothesis, our data show decreased expression of FOXO1 and BACH2, a TF that inhibits terminal differentiation of plasma cells (PCs), and increased expression of ZEB2, a TF involved in effector cell differentiation (not shown). This finding is consistent with our ATAC data presented in Chapter 3 identified increased accessibility of the BACH2 locus upon deletion of ARID3a in K562 cells. Additional studies are needed to confirm if ARID3a^{+} naïve B cells preferentially differentiate into DN2 and subsequently into autoreactive PC and/or plasmablasts that generate autoantibodies.

It is unclear if ARID3a directly causes increased expression of TLR7 or IRF4. Decreases in DNA methylation has previously been observed in PBMCs with high levels of ARID3a levels in patients with SLE (50). Indeed, ARID3a^{+} naïve B cells are enriched in clusters with decreased expression of DNMT3a (Figure 26). It is possible that ARID3a causes changes in expression of these genes by altering the chromatin landscape by regulating known histone methyltransferases, such as DNMT3a and EHMT1, which adds repressive methyl marks to H3K9 histone subunits (188). Our data also identified EHMT1 networks to be inhibited, which is consistent with our ATAC-seq data from ARID3a$^{-/-}$ K562 cells where we identified significant increase in accessibility of the EHMT1 locus. We have previously shown that deletion of ARID3a in K562 cells results in altered chromatin accessibility of developmentally controlled genes essential for erythrocyte differentiation, some of which is inhibited by EHMT1. We are currently awaiting additional data from 4 SLE naïve B cell samples which increase statistical validity and will support the genes already discussed. Further studies are needed to determine if ARID3a acts as an epigenetic regulator in naïve B cells.

CHAPTER 6

SUMMARY

ARID3a was originally identified to be required for proper expression of immunoglobulin heavy chain in B cells (reviewed in Chapter 1). Further investigation discovered that ARID3a is required for B cell development and is highly expressed in naïve B cells of SLE patients compared to healthy controls. The cause of the observed IFN signature in PBMCs from SLE patients remains unknown (119). This led us to investigate the role of ARID3a in two other cell types important in IFNa production and inflammatory responses in SLE, pDCs and LDNs. In Chapter 2, I performed differential expression analysis, which led to the discovery that ARID3a protein, but not transcript, is strongly associated with increased disease activity in both pDCs and LDNs. Hierarchical clustering on DEGs show that patient samples cluster based on ARID3a protein levels, but not transcript levels in pDCs and LDNs. These data show for the first time that ARID3a is expressed in pDCs and LDNs and that ARID3a protein can be used as a biomarker for increased disease activity in SLE. In addition, my work reveals that ARID3a functions as an epigenetic regulator in these cell types, by repressing and activating multiple genes. We then hypothesized that ARID3a functions in a cell-type specific fashion and is important for other cell types.

The rationale for this came from early studies on ARID3a knockout mice, which proved to be embryonic lethal for >90% of littermates (14). The rare survivors exhibited parlor and lacked erythrocytes. Using the erythroid model cell line, K562, I show that ARID3a is essential for hemin-induced differentiation and fetal globin expression. Knockout of ARID3a blocks erythrocyte development and inhibits fetal globin expression. Fetal globin expression is induced by a GATA2-to-GATA1 switch and a study by Buenestro et al identified ARID3a TF binding

motifs to be enriched in genes regulated by GATA1/GATA2 (141). My RNA-seq analysis on K562 cells identified 227 genes that require ARID3a. This study also identified significant inhibition of erythroid master regulators, GATA1, GATA2, and TAL1. Moreover, significant enrichment of these TFs was found in the 227 genes associated with ARID3a expression. The ATAC-seq data that I generated from CRISPR/Cas9 ARID3a$^{-/-}$ K562 clones also show a significant overlap of transcription factor binding sites for ARID3a and GATA1, GATA2, and TAL1. These erythroid-specific master regulators are known to be responsible for changing chromatin structure such that erythroid genes are transcribed in a developmentally controlled manner (148,184,186,187). Furthermore, ATAC-seq on ARID3a KO K562 cells revealed drastic changes to chromatin accessibility when compared to wild type. Among the differentially accessible regions identified in the ATAC-seq data, many contained overlapping TFBSs for GATA1 and GATA2. These data show for the first time that ARID3a functions as an epigenetic regulator and is necessary for differentiation of erythroid precursors. In support of this notion, we identified significant increase in accessibility of the histone lysine methyltransferase, EHMT1, which is responsible for adding repressive methylation marks to histones at the erythroid-specific LCR (188). Inhibition of EHMT1 is required for induction of fetal globin genes (188). EHMT1 pathways were significantly inhibited in ARID3a^{+} naïve B cells (discussed below) (Figure 26). However, the role of EHMT1 in B cells is unknown. Together, these data reveal that ARID3a functions as an epigenetic regulator, either directly through binding cell type specific enhancer regions or indirectly through controlling expression of epigenetic regulators.

In Chapter 4, we provide proof of concept that ARID3a-expressing cells within the same individual and sorted with the same surface markers showed differential gene expression. Additionally, these data raise the possibility that current surface marker identification of cell

subsets may be insufficient to identify cells with similar transcriptomes. It is not possible to isolate ARID3a$^+$ cells as intracellular staining requires fixation and affects RNA integrity. Bulk RNA-seq masks true cellular heterogeneity by averaging transcripts in a cell population. For example, not all cells express ARID3a and bulk RNA-seq studies could make it difficult to unravel the functions of ARID3a. scRNA-seq gets around this by allowing for the differential expression analysis of ARID3a$^+$ (CPM $\geq$ 0.5) vs ARID3a$^-$ (CPM < 0.5) cells, within and between patient samples. More work is needed to determine how cells with similar surface markers have different transcriptomes and what factors cause some cells to differentiate into different cell types. This will be important for B cells where some cells go on to display autoimmune characteristics and indicate that ARID3a-expressing cells could be direct precursors of those B cells. The data presented in Chapter 4 also reveal that ARID3a$^+$ cells in aged individuals differ from those in young individuals and these differences could be the result of expansion of precursors of different cell types. scRNA-seq analysis revealed that aged HSCs differ from young, even within defined subsets important for B cell differentiation. ARID3a levels cause changes in lineage decisions of HSCs into different cell types. Additionally, the differences found with B1 lineage cells explain why immune responses to some vaccines are reduced.

SLE is a complicated autoimmune disease with many clinical manifestations. It has been previously demonstrated that nearly half of SLE patients have expanded numbers of ARID3a-expressing B cells (46). ARID3a$^+$ B cells can also produce IFNa upon TLR stimulation with CpG, and may represent a new type of B effector cell that is expanded in SLE (46). Although ARID3a can be expressed in large numbers of B lymphocytes in SLE patients, and total numbers of ARID3a-expressing B cells are increased in patients with increased disease activity, no individual subset of B cells expresses ARID3a in one-hundred percent of the

B cells in that subset. Because ARID3a is uniquely expressed in naïve SLE B compared to healthy B cells, further studies to better define the characteristics of these cells are needed. The goal of Chapter 5 was to determine how ARID3a affects gene expression in ARID3a^{+} and ARID3a^{-} SLE B cells. ARID3a is an intracellular protein, and isolation of ARID3a-expressing B cells requires permeabilization of cells and interferes with RNA integrity. In Chapter 5, I took advantage of single cell technology to capture single ARID3a^{+} and ARID3a^{-} naïve B cells. Differential expression analysis identified 3,006 DEGs and identified upregulation of genes involved in TLR and SLE signaling as well as IFNa production, which strongly correlates to increased disease activity in patients with SLE (13,119). Out of the 3,006 DEGs, 680 genes are uncharacterized, highlighting how little is still known about the role ARID3a in SLE. The work presented in Chapter 5 will lead to additional information regarding genes with undiscovered function. To gain a better understanding of the complex interplay within SLE naive B cells, it will be important to use transcriptomic, ATAC-seq (DNA accessibility), and ChIP-seq (i.e. histone marks) data to build a model that faithfully illuminates functions of ARID3a at distinct stages of B cell development in SLE patients at the systems level, and to find new targets to treat SLE for which only one treatment has been developed in the past 60 years. More work is needed to identify if ARID3a^{+} naïve B cells are precursors to effector cell types that exacerbate SLE by producing autoantibody secreting cells, or if they activate other effector cells that go on to cause a feedforward loop of interferon production.

More detailed studies on enhancer activity in ARID3a^{+} vs ARID3a^{-} cells will be important for determining if ARID3a functions through regulating enhancer accessibility. New high-throughput assays that measure enhancer activity, such as self-transcribing active regulatory region sequencing (STARR-seq), would be able to detect enhancers genome-wide in a

quantitative (237). In combination with scATAC-seq and scRNA-seq, a model that faithfully represents the genes/enhancers that are regulated by ARID3a and will identify new drug targets for diseases, such as SLE.

LIST OF REFERENCES

1. Schatz, M.C. and Langmead, B. (2013) The DNA data deluge. *IEEE Spectrum*, **50**, 28-33.
2. Wu, A.R., Neff, N.F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M.E., Mburu, F.M., Mantalas, G.L., Sim, S., Clarke, M.F. *et al.* (2014) Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods*, **11**, 41-46.
3. Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D. *et al.* (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, **498**, 236-240.
4. Yu, P. and Lin, W. (2016) Single-cell Transcriptome Study as Big Data. *Genomics Proteomics Bioinformatics*, **14**, 21-30.
5. Garton, J., Barron, M.D., Ratliff, M.L. and Webb, C.F. (2019) New Frontiers: ARID3a in SLE. *Cells*, **8**.
6. Wilsker, D., Patsialou, A., Dallas, P.B. and Moran, E. (2002) ARID proteins: a diverse family of DNA binding proteins implicated in the control of cell growth, differentiation, and development. *Cell Growth Differ*, **13**, 95-106.
7. Wilsker, D., Probst, L., Wain, H.M., Maltais, L., Tucker, P.W. and Moran, E. (2005) Nomenclature of the ARID family of DNA-binding proteins. *Genomics*, **86**, 242-251.
8. Webb, C., Zong, R.T., Lin, D., Wang, Z., Kaplan, M., Paulin, Y., Smith, E., Probst, L., Bryant, J., Goldstein, A. *et al.* (1999) Differential regulation of immunoglobulin gene transcription via nuclear matrix-associated regions. *Cold Spring Harb Symp Quant Biol*, **64**, 109-118.
9. Webb, C.F., Das, C., Eaton, S., Calame, K. and Tucker, P.W. (1991) Novel protein-DNA interactions associated with increased immunoglobulin transcription in response to antigen plus interleukin-5. *Molecular and cellular biology*, **11**, 5197-5205.
10. Webb, C.F., Das, C., Eneff, K.L. and Tucker, P.W. (1991) Identification of a matrix-associated region 5' of an immunoglobulin heavy chain variable region gene. *Molecular and cellular biology*, **11**, 5206-5211.
11. An, G., Miner, C.A., Nixon, J.C., Kincade, P.W., Bryant, J., Tucker, P.W. and Webb, C.F. (2010) Loss of Bright/ARID3a Function Promotes Developmental Plasticity. *Stem Cells*, **28**, 1560-1567.
12. Popowski, M., Templeton, T.D., Lee, B.K., Rhee, C., Li, H., Miner, C., Dekker, J.D., Orlanski, S., Bergman, Y., Iyer, V.R. *et al.* (2014), *Stem Cell Reports*, Vol. 2, pp. 26-35.
13. Ratliff, M.L., Garton, J., Garman, L., Barron, M.D., Georgescu, C., White, K.A., Chakravarty, E., Wren, J.D., Montgomery, C.G., James, J.A. *et al.* (2019) ARID3a gene profiles are strongly associated with human interferon alpha production. *J Autoimmun*, **96**, 158-167.
14. Webb, C.F., Bryant, J., Popowski, M., Allred, L., Kim, D., Harriss, J., Schmidt, C., Miner, C.A., Rose, K., Cheng, H.L. *et al.* (2011) The ARID family transcription factor bright is required for both hematopoietic stem cell and B lineage development. *Molecular and cellular biology*, **31**, 1041-1053.
15. Carter, E.E., Barr, S.G. and Clarke, A.E. (2016) The global burden of SLE: prevalence, health disparities and socioeconomic impact. *Nat Rev Rheumatol*, **12**, 605-620.
16. Bombardier, C., Gladman, D.D., Urowitz, M.B., Caron, D. and Chang, C.H. (1992) Derivation of the SLEDAI. A disease activity index for lupus patients. The Committee on Prognosis Studies in SLE. *Arthritis Rheum*, **35**, 630-640.
17. Gladman, D.D., Ibañez, D. and Urowitz, M.B. (2002) Systemic lupus erythematosus disease activity index 2000. *J Rheumatol*, **29**, 288-291.
18. Golbus, J. and McCune, W.J. (1994) Lupus nephritis. Classification, prognosis, immunopathogenesis, and treatment. *Rheum Dis Clin North Am*, **20**, 213-242.

19. Contreras, G., Lenz, O., Pardo, V., Borja, E., Cely, C., Iqbal, K., Nahar, N., de La Cuesta, C., Hurtado, A., Fornoni, A. *et al.* (2006) Outcomes in African Americans and Hispanics with lupus nephritis. *Kidney Int*, **69**, 1846-1851.
20. Lech, M. and Anders, H.-J. (2013) The Pathogenesis of Lupus Nephritis. *Journal of the American Society of Nephrology*, **24**, 1357-1366.
21. Lin, D., Ippolito, G.C., Zong, R.T., Bryant, J., Koslovsky, J. and Tucker, P. (2007) Bright/ARID3A contributes to chromatin accessibility of the immunoglobulin heavy chain enhancer. *Mol Cancer*, **6**, 23.
22. Herrscher, R.F., Kaplan, M.H., Lelsz, D.L., Das, C., Scheuermann, R. and Tucker, P.W. (1995) The immunoglobulin heavy-chain matrix-associating regions are bound by Bright: a B cell-specific trans-activator that describes a new DNA-binding protein family. *Genes & Development*, **9**, 3067-3082.
23. Rajaiya, J., Nixon, J.C., Ayers, N., Desgranges, Z.P., Roy, A.L. and Webb, C.F. (2006) Induction of Immunoglobulin Heavy-Chain Transcription through the Transcription Factor Bright Requires TFII-I. *Molecular and cellular biology*, **26**, 4758-4768.
24. Rajaiya, J., Hatfield, M., Nixon, J.C., Rawlings, D.J. and Webb, C.F. (2005) Bruton's Tyrosine Kinase Regulates Immunoglobulin Promoter Activation in Association with the Transcription Factor Bright. *Molecular and cellular biology*, **25**, 2073-2084.
25. Webb, C.F., Das, C., Coffman, R.L. and Tucker, P.W. (1989) Induction of immunoglobulin mu mRNA in a B cell transfectant stimulated with interleukin-5 and a T-dependent antigen. *The Journal of Immunology*, **143**, 3934-3939.
26. Kortschak, R.D., Tucker, P.W. and Saint, R. (2000) ARID proteins come in from the desert. *Trends in Biochemical Sciences*, **25**, 294-299.
27. Ratliff, M.L., Templeton, T.D., Ward, J.M. and Webb, C.F. (2014) The Bright Side of Hematopoiesis: Regulatory Roles of ARID3a/Bright in Human and Mouse Hematopoiesis. *Frontiers in immunology*, **5**, 113.
28. Nixon, J.C., Rajaiya, J.B., Ayers, N., Evetts, S. and Webb, C.F. (2004) The transcription factor, Bright, is not expressed in all human B lymphocyte subpopulations. *Cell Immunol*, **228**, 42-53.
29. Hayakawa, K., Li, Y.-S., Shinton, S.A., Bandi, S.R., Formica, A.M., Brill-Dashoff, J. and Hardy, R.R. (2019) Crucial Role of Increased Arid3a at the Pre-B and Immature B Cell Stages for B1a Cell Generation. *Frontiers in immunology*, **10**.
30. Meffre, E. (2011) The establishment of early B cell tolerance in humans: lessons from primary immunodeficiency diseases. *Annals of the New York Academy of Sciences*, **1246**, 1-10.
31. De Groof, A., Hémon, P., Mignen, O., Pers, J.-O., Wakeland, E.K., Renaudineau, Y. and Lauwerys, B.R. (2017) Dysregulated Lymphoid Cell Populations in Mouse Models of Systemic Lupus Erythematosus. *Clinical Reviews in Allergy & Immunology*, **53**, 181-197.
32. Nixon, J.C., Ferrell, S., Miner, C., Oldham, A.L., Hochgeschwender, U. and Webb, C.F. (2008) Transgenic Mice Expressing Dominant-Negative Bright Exhibit Defects in B1 B Cells. *The Journal of Immunology*, **181**, 6913-6922.
33. Webb, C.F., Smith, E.A., Medina, K.L., Buchanan, K.L., Smithson, G. and Dou, S. (1998) Expression of bright at two distinct stages of B lymphocyte development. *J Immunol*, **160**, 4747-4754.
34. Zhou, Y., Li, Y.-S., Bandi, S.R., Tang, L., Shinton, S.A., Hayakawa, K. and Hardy, R.R. (2015) Lin28b promotes fetal B lymphopoiesis through the transcription factor Arid3a. *Journal of Experimental Medicine*, **212**, 569-580.
35. Oldham, A.L., Miner, C.A., Wang, H.-C. and Webb, C.F. (2011) The transcription factor Bright plays a role in marginal zone B lymphocyte development and autoantibody production. *Molecular Immunology*, **49**, 367-379.

36. Shankar, M., Nixon, J.C., Maier, S., Workman, J., Farris, A.D. and Webb, C.F. (2007) Anti-Nuclear Antibody Production and Autoimmunity in Transgenic Mice That Overexpress the Transcription Factor Bright. *The Journal of Immunology*, **178**, 2996-3006.
37. Wang, John H., Li, J., Wu, Q., Yang, P., Pawar, R.D., Xie, S., Timares, L., Raman, C., Chaplin, D.D., Lu, L. *et al.* (2010) Marginal Zone Precursor B Cells as Cellular Agents for Type I IFN–Promoted Antigen Transport in Autoimmunity. *The Journal of Immunology*, **184**, 442-451.
38. Wu, Y.-Y., Georg, I., Díaz-Barreiro, A., Varela, N., Lauwerys, B., Kumar, R., Bagavant, H., Castillo-Martín, M., El Salem, F., Marañón, C. *et al.* (2015) Concordance of Increased B1 Cell Subset and Lupus Phenotypes in Mice and Humans Is Dependent on BLK Expression Levels. *The Journal of Immunology*, **194**, 5692-5702.
39. Bendelac, A., Bonneville, M. and Kearney, J.F. (2001) Autoreactivity by design: innate B and T lymphocytes. *Nature Reviews Immunology*, **1**, 177-186.
40. Molano-González, N., Rojas, M., Monsalve, D.M., Pacheco, Y., Acosta-Ampudia, Y., Rodríguez, Y., Rodríguez-Jimenez, M., Ramírez-Santana, C. and Anaya, J.-M. (2019) Cluster analysis of autoimmune rheumatic diseases based on autoantibodies. New insights for polyautoimmunity. *Journal of Autoimmunity*, **98**, 24-32.
41. Dieudonné, Y., Gies, V., Guffroy, A., Keime, C., Bird, A.K., Liesveld, J., Barnas, J.L., Poindron, V., Douiri, N., Soulas-Sprauel, P. *et al.* (2019) Transitional B cells in quiescent SLE: An early checkpoint imprinted by IFN. *Journal of Autoimmunity*, **102**, 150-158.
42. Jacobs, H.M., Thouvenel, C.D., Leach, S., Arkatkar, T., Metzler, G., Scharping, N.E., Kolhatkar, N.S., Rawlings, D.J. and Jackson, S.W. (2016) Cutting Edge: BAFF Promotes Autoantibody Production via TACI-Dependent Activation of Transitional B Cells. *The Journal of Immunology*, **196**, 3525-3531.
43. Martin, F. and Kearney, J.F. (2002) Marginal-zone B cells. *Nature Reviews Immunology*, **2**, 323-335.
44. Ward, J.M., Rose, K., Montgomery, C., Adrianto, I., James, J.A., Merrill, J.T. and Webb, C.F. (2014) Disease activity in systemic lupus erythematosus correlates with expression of the transcription factor AT-rich-interactive domain 3A. *Arthritis Rheumatol.*, **66**, 3404-3412.
45. Ward, J.M., James, J.A., Zhao, Y.D. and Webb, C.F. (2015) Antibody Reactivity of B Cells in Lupus Patients with Increased Disease Activity and ARID3a Expression. *Antibodies*, **4**, 354-368.
46. Ward, J.M., Ratliff, M.L., Dozmorov, M.G., Wiley, G., Guthridge, J.M., Gaffney, P.M., James, J.A. and Webb, C.F. (2016) Human effector B lymphocytes express ARID3a and secrete interferon alpha. *Journal of Autoimmunity*, **75**, 130-140.
47. James, J.A. and Robertson, J.M. (2012) Lupus and Epstein-Barr. *Current Opinion in Rheumatology*, **24**, 383-388.
48. Sundar, K., Jacques, S., Gottlieb, P., Villars, R., Benito, M.-E., Taylor, D.K. and Spatz, L.A. (2004) Expression of the Epstein-Barr virus nuclear antigen-1 (EBNA-1) in the mouse can elicit the production of anti-dsDNA and anti-Sm antibodies. *Journal of Autoimmunity*, **23**, 127-140.
49. Boreström, C., Forsman, A., Rüetschi, U. and Rymo, L. (2012) E2F1, ARID3A/Bright and Oct-2 factors bind to the Epstein–Barr virus C promoter, EBNA1 and oriP, participating in long-distance promoter–enhancer interactions. *Journal of General Virology*, **93**, 1065-1075.
50. Ward, J.M., Ratliff, M.L., Dozmorov, M.G., Wiley, G., Guthridge, J.M., Gaffney, P.M., James, J.A. and Webb, C.F. (2016) Expression and methylation data from SLE patient and healthy control blood samples subdivided with respect to ARID3a levels. *Data in Brief*, **9**, 213-219.
51. Calame, K. and Atchison, M. (2007) YY1 helps to bring loose ends together. *Genes & Development*, **21**, 1145-1152.

52. Scharer, C.D., Blalock, E.L., Barwick, B.G., Haines, R.R., Wei, C., Sanz, I. and Boss, J.M. (2016) ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naïve SLE B cells. *Scientific Reports*, **6**, 27030.
53. Scharer, C.D., Blalock, E.L., Mi, T., Barwick, B.G., Jenks, S.A., Deguchi, T., Cashman, K.S., Neary, B.E., Patterson, D.G., Hicks, S.L. *et al.* (2019) Epigenetic programming underpins B cell dysfunction in human SLE. *Nature Immunology*, **20**, 1071-1082.
54. Nascimbeni, M., Perié, L., Chorro, L., Diocou, S., Kreitmann, L., Louis, S., Garderet, L., Fabiani, B., Berger, A., Schmitz, J. *et al.* (2009) Plasmacytoid dendritic cells accumulate in spleens from chronically HIV-infected patients but barely participate in interferon-α expression. *Blood*, **113**, 6112-6119.
55. Garcia-Romo, G.S., Caielli, S., Vega, B., Connolly, J., Allantaz, F., Xu, Z., Punaro, M., Baisch, J., Guiducci, C., Coffman, R.L. *et al.* (2011) Netting Neutrophils Are Major Inducers of Type I IFN Production in Pediatric Systemic Lupus Erythematosus. *Science Translational Medicine*, **3**, 73ra20-73ra20.
56. Liu, Y.-J. (2005) IPC: Professional Type 1 Interferon-Producing Cells and Plasmacytoid Dendritic Cell Precursors. *Annual Review of Immunology*, **23**, 275-306.
57. Denny, M.F., Yalavarthi, S., Zhao, W., Thacker, S.G., Anderson, M., Sandy, A.R., McCune, W.J. and Kaplan, M.J. (2010) A Distinct Subset of Proinflammatory Neutrophils Isolated from Patients with Systemic Lupus Erythematosus Induces Vascular Damage and Synthesizes Type I IFNs. *The Journal of Immunology*, **184**, 3284-3297.
58. Lindau, D., Mussard, J., Rabsteyn, A., Ribon, M., Kötter, I., Igney, A., J Adema, G., Boissier, M.-C., Rammensee, H.-G. and Decker, P. (2014) TLR9 independent interferon α production by neutrophils on NETosis in response to circulating chromatin, a key lupus autoantigen. *Annals of the Rheumatic Diseases*, **73**, 2199-2207.
59. Landolt-Marticorena, C., Bonventi, G., Lubovich, A., Ferguson, C., Unnithan, T., Su, J., Gladman, D.D., Urowitz, M., Fortin, P.R. and Wither, J. (2009) Lack of association between the interferon-α signature and longitudinal changes in disease activity in systemic lupus erythematosus. *Annals of the Rheumatic Diseases*, **68**, 1440-1446.
60. Villanueva, E., Yalavarthi, S., Berthier, C.C., Hodgin, J.B., Khandpur, R., Lin, A.M., Rubin, C.J., Zhao, W., Olsen, S.H., Klinker, M. *et al.* (2011) Netting Neutrophils Induce Endothelial Damage, Infiltrate Tissues, and Expose Immunostimulatory Molecules in Systemic Lupus Erythematosus. *The Journal of Immunology*, **187**, 538-552.
61. Carmona-Rivera, C. and Kaplan, M.J. (2013) Low-density granulocytes: a distinct class of neutrophils in systemic autoimmunity. *Seminars in Immunopathology*, **35**, 455-463.
62. Lande, R., Ganguly, D., Facchinetti, V., Frasca, L., Conrad, C., Gregorio, J., Meller, S., Chamilos, G., Sebasigari, R., Riccieri, V. *et al.* (2011) Neutrophils Activate Plasmacytoid Dendritic Cells by Releasing Self-DNA–Peptide Complexes in Systemic Lupus Erythematosus. *Science Translational Medicine*, **3**, 73ra19-73ra19.
63. Lande, R., Gregorio, J., Facchinetti, V., Chatterjee, B., Wang, Y.-H., Homey, B., Cao, W., Wang, Y.-H., Su, B., Nestle, F.O. *et al.* (2007) Plasmacytoid dendritic cells sense self-DNA coupled with antimicrobial peptide. *Nature*, **449**, 564-569.
64. Gestermann, N., Di Domizio, J., Lande, R., Demaria, O., Frasca, L., Feldmeyer, L., Di Lucca, J. and Gilliet, M. (2018) Netting Neutrophils Activate Autoreactive B Cells in Lupus. *The Journal of Immunology*, **200**, 3364-3371.
65. Puga, I., Cols, M., Barra, C.M., He, B., Cassis, L., Gentile, M., Comerma, L., Chorny, A., Shan, M., Xu, W. *et al.* (2012) B cell–helper neutrophils stimulate the diversification and production of immunoglobulin in the marginal zone of the spleen. *Nature Immunology*, **13**, 170-180.

66. Coquery, C.M., Wade, N.S., Loo, W.M., Kinchen, J.M., Cox, K.M., Jiang, C., Tung, K.S. and Erickson, L.D. (2014) Neutrophils Contribute to Excess Serum BAFF Levels and Promote CD4+ T Cell and B Cell Responses in Lupus-Prone Mice. *PLOS ONE*, **9**, e102284.
67. Scapini, P., Bazzoni, F. and Cassatella, M.A. (2008) Regulation of B-cell-activating factor (BAFF)/B lymphocyte stimulator (BLyS) expression in human neutrophils. *Immunology Letters*, **116**, 1-6.
68. Scapini , P., Nardelli , B., Nadali , G., Calzetti , F., Pizzolo , G., Montecucco , C. and Cassatella , M.A. (2003) G-CSF–stimulated Neutrophils Are a Prominent Source of Functional BLyS. *Journal of Experimental Medicine*, **197**, 297-302.
69. Parsa, R., Lund, H., Georgoudaki, A.-M., Zhang, X.-M., Ortlieb Guerreiro-Cacais, A., Grommisch, D., Warnecke, A., Croxford, A.L., Jagodic, M., Becher, B. *et al.* (2016) BAFF-secreting neutrophils drive plasma cell responses during emergency granulopoiesis. *Journal of Experimental Medicine*, **213**, 1537-1553.
70. Marini, O., Costa, S., Bevilacqua, D., Calzetti, F., Tamassia, N., Spina, C., De Sabata, D., Tinazzi, E., Lunardi, C., Scupoli, M.T. *et al.* (2017) Mature CD10+ and immature CD10− neutrophils present in G-CSF–treated donors display opposite effects on T cells. *Blood*, **129**, 1343-1356.
71. Palanichamy, A., Bauer, J.W., Yalavarthi, S., Meednu, N., Barnard, J., Owen, T., Cistrone, C., Bird, A., Rabinovich, A., Nevarez, S. *et al.* (2014) Neutrophil-Mediated IFN Activation in the Bone Marrow Alters B Cell Development in Human and Murine Systemic Lupus Erythematosus. *The Journal of Immunology*, **192**, 906-918.
72. Suzuki, N., Hirano, K., Ogino, H. and Ochi, H. (2019) Arid3a regulates nephric tubule regeneration via evolutionarily conserved regeneration signal-response enhancers. *eLife*, **8**, e43186.
73. Rhee, C., Lee, B.-K., Beck, S., Anjum, A., Cook, K.R., Popowski, M., Tucker, H.O. and Kim, J. (2014) Arid3a is essential to execution of the first cell fate decision via direct embryonic and extraembryonic transcriptional regulation. *Genes & Development*, **28**, 2219-2232.
74. Rhee, C., Edwards, M., Dang, C., Harris, J., Brown, M., Kim, J. and Tucker, H.O. (2017) ARID3A is required for mammalian placenta development. *Developmental Biology*, **422**, 83-91.
75. Fairhurst, A.-M., Xie, C., Fu, Y., Wang, A., Boudreaux, C., Zhou, X.J., Cibotti, R., Coyle, A., Connolly, J.E., Wakeland, E.K. *et al.* (2009) Type I Interferons Produced by Resident Renal Cells May Promote End-Organ Disease in Autoantibody-Mediated Glomerulonephritis. *The Journal of Immunology*, **183**, 6831-6838.
76. Ratliff, M.L., Mishra, M., Frank, M.B., Guthridge, J.M. and Webb, C.F. (2016) The Transcription Factor ARID3a Is Important for In Vitro Differentiation of Human Hematopoietic Progenitors. *J Immunol*, **196**, 614-623.
77. Ratliff, M.L., Ward, J.M., Merrill, J.T., James, J.A. and Webb, C.F. (2015) Differential Expression of the Transcription Factor ARID3a in Lupus Patient Hematopoietic Progenitor Cells. *The Journal of Immunology*, **194**, 940-949.
78. Moonen, J.R.A.J., de Leeuw, K., van Seijen, X.J.G.Y., Kallenberg, C.G.M., van Luyn, M.J.A., Bijl, M. and Harmsen, M.C. (2007) Reduced number and impaired function of circulating progenitor cells in patients with systemic lupus erythematosus. *Arthritis Research & Therapy*, **9**, R84.
79. Westerweel, P.E., Luijten, R.K.M.A.C., Hoefer, I.E., Koomans, H.A., Derksen, R.H.W.M. and Verhaar, M.C. (2007) Haematopoietic and endothelial progenitor cells are deficient in quiescent systemic lupus erythematosus. *Annals of the Rheumatic Diseases*, **66**, 865-870.
80. Huang, X., Chen, W., Ren, G., Zhao, L., Guo, J., Gong, D., Zeng, C., Hu, W. and Liu, Z. (2019) Autologous Hematopoietic Stem Cell Transplantation for Refractory Lupus Nephritis. *Clinical Journal of the American Society of Nephrology*, **14**, 719-727.
81. Marmont du Haut Champ, A.M. (2012) Hematopoietic Stem Cell Transplantation for Systemic Lupus Erythematosus. *Clinical and Developmental Immunology*, **2012**, 380391.

82. Essers, M.A.G., Offner, S., Blanco-Bose, W.E., Waibler, Z., Kalinke, U., Duchosal, M.A. and Trumpp, A. (2009) IFNα activates dormant haematopoietic stem cells in vivo. *Nature*, **458**, 904-908.
83. Sato, T., Onai, N., Yoshihara, H., Arai, F., Suda, T. and Ohteki, T. (2009) Interferon regulatory factor-2 protects quiescent hematopoietic stem cells from type I interferon–dependent exhaustion. *Nature Medicine*, **15**, 696-700.
84. Pietras, E.M., Lakshminarasimhan, R., Techner, J.-M., Fong, S., Flach, J., Binnewies, M. and Passegué, E. (2014) Re-entry into quiescence protects hematopoietic stem cells from the killing effect of chronic exposure to type I interferons. *Journal of Experimental Medicine*, **211**, 245-262.
85. Ivashkiv, L.B. and Donlin, L.T. (2014) Regulation of type I interferon responses. *Nature Reviews Immunology*, **14**, 36-49.
86. Herrada, A.A., Escobedo, N., Iruretagoyena, M., Valenzuela, R.A., Burgos, P.I., Cuitino, L. and Llanos, C. (2019) Innate Immune Cells' Contribution to Systemic Lupus Erythematosus. *Frontiers in immunology*, **10**.
87. He, J., Tsai, Louis M., Leong, Yew A., Hu, X., Ma, Cindy S., Chevalier, N., Sun, X., Vandenberg, K., Rockman, S., Ding, Y. *et al.* (2013) Circulating Precursor CCR7^{lo}PD-1^{hi} CXCR5⁺ CD4⁺ T Cells Indicate Tfh Cell Activity and Promote Antibody Responses upon Antigen Reexposure. *Immunity*, **39**, 770-781.
88. Klarquist, J., Zhou, Z., Shen, N. and Janssen, E.M. (2016) Dendritic Cells in Systemic Lupus Erythematosus: From Pathogenic Players to Therapeutic Tools. *Mediators of Inflammation*, **2016**, 5045248.
89. Orme, J. and Mohan, C. (2012) Macrophage subpopulations in systemic lupus erythematosus. *Discov Med*, **13**, 151-158.
90. Spada, R., Rojas, J.M. and Barber, D.F. (2015) Recent findings on the role of natural killer cells in the pathogenesis of systemic lupus erythematosus. *Journal of Leukocyte Biology*, **98**, 479-487.
91. Suárez-Fueyo, A., Bradley, S.J. and Tsokos, G.C. (2016) T cells in Systemic Lupus Erythematosus. *Current Opinion in Immunology*, **43**, 32-38.
92. Fike, A.J., Elcheva, I. and Rahman, Z.S.M. (2019) The Post-GWAS Era: How to Validate the Contribution of Gene Variants in Lupus. *Current Rheumatology Reports*, **21**, 3.
93. Callery, E.M., Smith, J.C. and Thomsen, G.H. (2005) The ARID domain protein dril1 is necessary for TGFβ signaling in Xenopus embryos. *Developmental Biology*, **278**, 542-559.
94. Shandala, T., Kortschak, R.D., Gregory, S. and Saint, R. (1999) The Drosophila dead ringer gene is required for early embryonic patterning through regulation of argos and buttonhead expression. *Development*, **126**, 4341-4349.
95. Ren, J., Panther, E., Liao, X., Grammer, A.C., Lipsky, P.E. and Reilly, C.M. (2018) The Impact of Protein Acetylation/Deacetylation on Systemic Lupus Erythematosus. *Int J Mol Sci*, **19**.
96. White, C.A., Pone, E.J., Lam, T., Tat, C., Hayama, K.L., Li, G., Zan, H. and Casali, P. (2014) Histone Deacetylase Inhibitors Upregulate B Cell microRNAs That Silence AID and Blimp-1 Expression for Epigenetic Modulation of Antibody and Autoantibody Responses. *The Journal of Immunology*, **193**, 5933-5950.
97. Chamilos, G., Gregorio, J., Meller, S., Lande, R., Kontoyiannis, D.P., Modlin, R.L. and Gilliet, M. (2012) Cytosolic sensing of extracellular self-DNA transported into monocytes by the antimicrobial peptide LL37. *Blood*, **120**, 3699-3707.
98. Puissegur, M.P., Eichner, R., Quelen, C., Coyaud, E., Mari, B., Lebrigand, K., Broccardo, C., Nguyen-Khac, F., Bousquet, M. and Brousset, P. (2012) B-cell regulator of immunoglobulin heavy-chain transcription (Bright)/ARID3a is a direct target of the oncomir microRNA-125b in progenitor B-cells. *Leukemia*, **26**, 2224-2232.

99. Honarpisheh, M., Köhler, P., von Rauchhaupt, E. and Lech, M. (2018) The Involvement of MicroRNAs in Modulation of Innate and Adaptive Immunity in Systemic Lupus Erythematosus and Lupus Nephritis. *Journal of Immunology Research*, **2018**, 4126106.
100. YOON, G., PARK, J.Y., KIM, H.J., CHOI, G.S., KIM, J.G., KANG, B.W., KANG, M.K. and SEO, A.N. (2019) ARID3A Positivity Correlated With Favorable Prognosis in Patients With Residual Rectal Cancer After Neoadjuvant Chemoradiotherapy. *Anticancer Research*, **39**, 2845-2853.
101. Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*, **403**, 503-511.
102. Chirshev, E., Oberg, K.C., Ioffe, Y.J. and Unternaehrer, J.J. (2019) Let-7 as biomarker, prognostic indicator, and therapy for precision medicine in cancer. *Clinical and Translational Medicine*, **8**, 24.
103. Samotij, D. and Reich, A. (2019) Biologics in the Treatment of Lupus Erythematosus: A Critical Literature Review. *BioMed Research International*, **2019**, 8142368.
104. Khamashta, M., Merrill, J.T., Werth, V.P., Furie, R., Kalunian, K., Illei, G.G., Drappa, J., Wang, L. and Greth, W. (2016) Sifalimumab, an anti-interferon-α monoclonal antibody, in moderate to severe systemic lupus erythematosus: a randomised, double-blind, placebo-controlled study. *Annals of the Rheumatic Diseases*, **75**, 1909-1916.
105. Lauwerys, B.R., Ducreux, J. and Houssiau, F.A. (2013) Type I interferon blockade in systemic lupus erythematosus: where do we stand? *Rheumatology*, **53**, 1369-1376.
106. Sanz, I. (2017) New Perspectives in Rheumatology: May You Live in Interesting Times: Challenges and Opportunities in Lupus Research. *Arthritis & Rheumatology*, **69**, 1552-1559.
107. Bénard, A., Sakwa, I., Schierloh, P., Colom, A., Mercier, I., Tailleux, L., Jouneau, L., Boudinot, P., Al-Saati, T. and Lang, R. (2018) B cells producing type I IFN modulate macrophage polarization in tuberculosis. *American journal of respiratory and critical care medicine*, **197**, 801-813.
108. Nascimbeni, M., Perié, L., Chorro, L., Diocou, S., Kreitmann, L., Louis, S., Garderet, L., Fabiani, B., Berger, A. and Schmitz, J. (2009) Plasmacytoid dendritic cells accumulate in spleens from chronically HIV-infected patients but barely participate in interferon-α expression. *Blood, The Journal of the American Society of Hematology*, **113**, 6112-6119.
109. Hamilton, J.A., Wu, Q., Yang, P., Luo, B., Liu, S., Hong, H., Li, J., Walter, M.R., Fish, E.N. and Hsu, H.-C. (2017) Cutting edge: endogenous IFN-β regulates survival and development of transitional B cells. *The Journal of Immunology*, **199**, 2618-2623.
110. de Padilla, C.M.L. and Niewold, T.B. (2016) The type I interferons: Basic concepts and clinical relevance in immune-mediated inflammatory diseases. *Gene*, **576**, 14-21.
111. Stifter, S.A. and Feng, C.G. (2015) Interfering with immunity: detrimental role of type I IFNs during infection. *The Journal of Immunology*, **194**, 2455-2465.
112. Baechler, E.C., Batliwalla, F.M., Karypis, G., Gaffney, P.M., Ortmann, W.A., Espe, K.J., Shark, K.B., Grande, W.J., Hughes, K.M. and Kapur, V. (2003) Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proceedings of the National Academy of Sciences*, **100**, 2610-2615.
113. Bengtsson, A., Sturfelt, G., Truedsson, L., Blomberg, J., Alm, G., Vallin, H. and Rönnblom, L. (2000) Activation of type I interferon system in systemic lupus erythematosus correlates with disease activity but not with antiretroviral antibodies. *Lupus*, **9**, 664-671.
114. Kirou, K.A., Lee, C., George, S., Louca, K., Peterson, M.G. and Crow, M.K. (2005) Activation of the interferon-α pathway identifies a subgroup of systemic lupus erythematosus patients with distinct serologic features and active disease. *Arthritis & Rheumatism*, **52**, 1491-1503.
115. Niewold, T., Hua, J., Lehman, T., Harley, J. and Crow, M. (2007) High serum IFN-α activity is a heritable risk factor for systemic lupus erythematosus. *Genes & Immunity*, **8**, 492-502.

116. Shrivastav, M. and Niewold, T.B. (2013) Nucleic acid sensors and type I interferon production in systemic lupus erythematosus. *Frontiers in immunology*, **4**, 319.
117. Kyogoku, C., Smiljanovic, B., Grün, J.R., Biesen, R., Schulte-Wrede, U., Häupl, T., Hiepe, F., Alexander, T., Radbruch, A. and Grützkau, A. (2013) Cell-specific type I IFN signatures in autoimmunity and viral infection: what makes the difference? *PloS one*, **8**.
118. Bezalel, S., Guri, K.M., Elbirt, D., Asher, I. and Sthoeger, Z.M. (2014) Type I interferon signature in systemic lupus erythematosus. *The Israel Medical Association journal: IMAJ*, **16**, 246-249.
119. Banchereau, R., Hong, S., Cantarel, B., Baldwin, N., Baisch, J., Edens, M., Cepika, A.-M., Acs, P., Turner, J. and Anguiano, E. (2016) Personalized immunomonitoring uncovers molecular networks that stratify lupus patients. *Cell*, **165**, 551-565.
120. Fairhurst, A.M., Mathian, A., Connolly, J.E., Wang, A., Gray, H.F., George, T.A., Boudreaux, C.D., Zhou, X.J., Li, Q.Z. and Koutouzov, S. (2008) Systemic IFN-α drives kidney nephritis in B6. Sle123 mice. *European journal of immunology*, **38**, 1948-1960.
121. Liu, Z., Bethunaickan, R., Huang, W., Lodhi, U., Solano, I., Madaio, M.P. and Davidson, A. (2011) Interferon-α accelerates murine systemic lupus erythematosus in a T cell–dependent manner. *Arthritis & Rheumatism*, **63**, 219-229.
122. Ramanujam, M., Kahn, P., Huang, W., Tao, H., Madaio, M.P., Factor, S.M. and Davidson, A. (2009) Interferon-α treatment of female (NZW× BXSB) F1 mice mimics some but not all features associated with the Yaa mutation. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, **60**, 1096-1101.
123. Rowland, S.L., Riggs, J.M., Gilfillan, S., Bugatti, M., Vermi, W., Kolbeck, R., Unanue, E.R., Sanjuan, M.A. and Colonna, M. (2014) Early, transient depletion of plasmacytoid dendritic cells ameliorates autoimmunity in a lupus model. *Journal of Experimental Medicine*, **211**, 1977-1991.
124. Sisirak, V., Ganguly, D., Lewis, K.L., Couillault, C., Tanaka, L., Bolland, S., D'Agati, V., Elkon, K.B. and Reizis, B. (2014) Genetic evidence for the role of plasmacytoid dendritic cells in systemic lupus erythematosus. *Journal of Experimental Medicine*, **211**, 1969-1976.
125. Felten, R., Dervovic, E., Chasset, F., Gottenberg, J.-E., Sibilia, J., Scher, F. and Arnaud, L. (2018) The 2018 pipeline of targeted therapies under clinical development for Systemic Lupus Erythematosus: a systematic review of trials. *Autoimmunity reviews*, **17**, 781-790.
126. Crow, M.K. (2014) Type I interferon in the pathogenesis of lupus. *The Journal of Immunology*, **192**, 5459-5468.
127. Elkon, K.B. and Wiedeman, A. (2012) Type I IFN system in the development and manifestations of SLE. *Current opinion in rheumatology*, **24**, 499-505.
128. Radic, M. and Marion, T.N. (2013), *Seminars in immunopathology*. Springer, Vol. 35, pp. 465-480.
129. Villanueva, E., Yalavarthi, S., Berthier, C.C., Hodgin, J.B., Khandpur, R., Lin, A.M., Rubin, C.J., Zhao, W., Olsen, S.H. and Klinker, M. (2011) Netting neutrophils induce endothelial damage, infiltrate tissues, and expose immunostimulatory molecules in systemic lupus erythematosus. *The Journal of Immunology*, **187**, 538-552.
130. Hacbarth, E. and Kajdacsy-Balla, A. (1986) Low density neutrophils in patients with systemic lupus erythematosus, rheumatoid arthritis, and acute rheumatic fever. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, **29**, 1334-1342.
131. Hochberg, M.C. (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, **40**, 1725-1725.
132. Gillis, C., Gouel-Chéron, A., Jönsson, F. and Bruhns, P. (2014) Contribution of human FcγRs to disease with evidence from human polymorphisms and transgenic animal studies. *Frontiers in immunology*, **5**, 254.

133. Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-2120.
134. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods*, **9**, 357-359.
135. Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
136. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, **43**, e47-e47.
137. Hua, J., Kirou, K., Lee, C. and Crow, M.K. (2006) Functional assay of type I interferon in systemic lupus erythematosus plasma and association with anti–RNA binding protein autoantibodies. *Arthritis & Rheumatism*, **54**, 1906-1916.
138. Carvalheiro, T., Rodrigues, A., Lopes, A., Inês, L., Velada, I., Ribeiro, A., Martinho, A., Silva, J.A., Pais, M.L. and Paiva, A. (2012) Tolerogenic versus inflammatory activity of peripheral blood monocytes and dendritic cells subpopulations in systemic lupus erythematosus. *Clinical and Developmental Immunology*, **2012**.
139. Hakkim, A., Fürnrohr, B.G., Amann, K., Laube, B., Abed, U.A., Brinkmann, V., Herrmann, M., Voll, R.E. and Zychlinsky, A. (2010) Impairment of neutrophil extracellular trap degradation is associated with lupus nephritis. *Proceedings of the National Academy of Sciences*, **107**, 9813-9818.
140. Lin, C., Song, W., Bi, X., Zhao, J., Huang, Z., Li, Z., Zhou, J., Cai, J. and Zhao, H. (2014) Recent advances in the ARID family: focusing on roles in human cancer. *OncoTargets and therapy*, **7**, 315.
141. Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y. and Greenleaf, W.J. (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, **523**, 486-490.
142. Grossman, S.R., Engreitz, J., Ray, J.P., Nguyen, T.H., Hacohen, N. and Lander, E.S. (2018) Positional specificity of different transcription factor classes within enhancers. *Proceedings of the National Academy of Sciences*, **115**, E7222-E7230.
143. Xavier-Ferrucio, J. and Krause, D.S. (2018) Concise Review: Bipotent Megakaryocytic-Erythroid Progenitors: Concepts and Controversies. *Stem Cells*, **36**, 1138-1145.
144. Kingsley, P.D., Greenfest-Allen, E., Frame, J.M., Bushnell, T.P., Malik, J., McGrath, K.E., Stoeckert, C.J. and Palis, J. (2013) Ontogeny of erythroid gene expression. *Blood*, **121**, e5-e13.
145. Rivella, S. (2009) Ineffective erythropoiesis and thalassemias. *Current opinion in hematology*, **16**, 187-194.
146. Perino, M. and Veenstra, Gert Jan C. (2016) Chromatin Control of Developmental Dynamics and Plasticity. *Developmental Cell*, **38**, 610-620.
147. Strouboulis, J., Dillon, N. and Grosveld, F. (1992) Developmental regulation of a complete 70-kb human beta-globin locus in transgenic mice. *Genes & Development*, **6**, 1857-1864.
148. Rodriguez, P., Bonte, E., Krijgsveld, J., Kolodziej, K.E., Guyot, B., Heck, A.J., Vyas, P., de Boer, E., Grosveld, F. and Strouboulis, J. (2005) GATA-1 forms distinct activating and repressive complexes in erythroid cells. *Embo j*, **24**, 2354-2366.
149. Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. and Grosveld, F. (1991) Importance of globin gene order for correct developmental expression. *Genes & Development*, **5**, 1387-1394.
150. Gomes, A.M., Kurochkin, I., Chang, B., Daniel, M., Law, K., Satija, N., Lachmann, A., Wang, Z., Ferreira, L., Ma'ayan, A. *et al.* (2018) Cooperative Transcription Factor Induction Mediates Hemogenic Reprogramming. *Cell Rep*, **25**, 2821-2835 e2827.

151. Bonifer, C. and Cockerill, P.N. (2017) Chromatin priming of genes in development: Concepts, mechanisms and consequences. *Exp Hematol*, **49**, 1-8.
152. Forrester, W.C., Thompson, C., Elder, J.T. and Groudine, M. (1986) A developmentally stable chromatin structure in the human beta-globin gene cluster. *Proc Natl Acad Sci U S A*, **83**, 1359-1363.
153. Thein, S.L. (2018) Molecular basis of beta thalassemia and potential therapeutic targets. *Blood Cells Mol Dis*, **70**, 54-65.
154. Patrinos, G.P., de Krom, M., de Boer, E., Langeveld, A., Imam, A.M., Strouboulis, J., de Laat, W. and Grosveld, F.G. (2004) Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes Dev*, **18**, 1495-1509.
155. Pimkin, M., Kossenkov, A.V., Mishra, T., Morrissey, C.S., Wu, W., Keller, C.A., Blobel, G.A., Lee, D., Beer, M.A., Hardison, R.C. *et al.* (2014) Divergent functions of hematopoietic transcription factors in lineage priming and differentiation during erythro-megakaryopoiesis. *Genome Res*, **24**, 1932-1944.
156. Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., Trompouki, E., Bowman, T.V., Zon, L.I., Yuan, G.-C. *et al.* (2016) Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. *Developmental Cell*, **36**, 9-23.
157. Kortschak, R.D., Tucker, P.W. and Saint, R. (2000) ARID proteins come in from the desert. *Trends Biochem.Sci.*, **25**, 294-299.
158. Ratliff, M.L., Templeton, T.D., Ward, J.M. and Webb, C.F. (2014) The Bright Side of Hematopoiesis: Regulatory Roles of ARID3a/Bright in Human and Mouse Hematopoiesis. *Front Immunol.*, **5**, 113.
159. Gillespie, M.A., Palii, C.G., Sanchez-Taltavull, D., Shannon, P., Longabaugh, W.J.R., Downes, D.J., Sivaraman, K., Espinoza, H.M., Hughes, J.R., Price, N.D. *et al.* (2020) Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. *Molecular Cell*.
160. Lozzio, B.B., Lozzio, C.B., Bamberger, E.G. and Feliu, A.S. (1981) A Multipotential Leukemia Cell Line (K-562) of Human Origin. *Proceedings of the Society for Experimental Biology and Medicine*, **166**, 546-550.
161. Andersson, L.C., Jokinen, M. and Gahmberg, C.G. (1979) Induction of erythroid differentiation in the human leukaemia cell line K562. *Nature*, **278**, 364-365.
162. Lozzio, B.B. and Lozzio, C.B. (1979) Properties and usefulness of the original K-562 human myelogenous leukemia cell line. *Leukemia Research*, **3**, 363-370.
163. Rutherford, T., Clegg, J.B., Higgs, D.R., Jones, R.W., Thompson, J. and Weatherall, D.J. (1981) Embryonic erythroid differentiation in the human leukemic cell line K562. *Proceedings of the National Academy of Sciences*, **78**, 348-352.
164. Addya, S., Keller, M.A., Delgrosso, K., Ponte, C.M., Vadigepalli, R., Gonye, G.E. and Surrey, S. (2004) Erythroid-induced commitment of K562 cells results in clusters of differentially expressed genes enriched for specific transcription regulatory elements. *Physiological Genomics*, **19**, 117-130.
165. Mizuta, S., Minami, T., Fujita, H., Kaminaga, C., Matsui, K., Ishino, R., Fujita, A., Oda, K., Kawai, A., Hasegawa, N. *et al.* (2014) CCAR1/CoCoA pair-mediated recruitment of the Mediator defines a novel pathway for GATA1 function. *Genes to Cells*, **19**, 28-51.
166. Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011*, **17**, 3.
167. Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A. *et al.* (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res*, **39**, D876-882.

168. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.
169. Rajaiya, J., Nixon, J.C., Ayers, N., Desgranges, Z.P., Roy, A.L. and Webb, C.F. (2006) Induction of immunoglobulin heavy-chain transcription through the transcription factor Bright requires TFII-I. *Mol Cell Biol*, **26**, 4758-4768.
170. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*, **9**, R137.
171. Huo, X.-F., Yu, J., Peng, H., Du, Z.-W., Liu, X.-L., Ma, Y.-N., Zhang, X., Zhang, Y., Zhao, H.-L. and Zhang, J.-W. (2006) Differential expression changes in K562 cells during the hemin-induced erythroid differentiation and the phorbol myristate acetate (PMA)-induced megakaryocytic differentiation. *Molecular and Cellular Biochemistry*, **292**, 155.
172. Rutherford, T.R., Clegg, J.B. and Weatherall, D.J. (1979) K562 human leukaemic cells synthesise embryonic haemoglobin in response to haemin. *Nature*, **280**, 164-165.
173. Smith, R.D., Malley, J.D. and Schechter, A.N. (2000) Quantitative analysis of globin gene induction in single human erythroleukemic cells. *Nucleic acids research*, **28**, 4998-5004.
174. Hu, P., Nebreda, A.R., Hanenberg, H., Kinnebrew, G.H., Ivan, M., Yoder, M.C., Filippi, M.-D., Broxmeyer, H.E. and Kapur, R. (2018) P38α/JNK signaling restrains erythropoiesis by suppressing Ezh2-mediated epigenetic silencing of Bim. *Nature Communications*, **9**, 3518.
175. Yu, X., Martella, A., Kolovos, P., Stevens, M., Stadhouders, R., Grosveld, F.G. and Andrieu-Soler, C. (2019) The dynamic emergence of GATA1 complexes identified in in vitro ES differentiation and in vivo mouse fetal liver. *Haematologica*.
176. Li, Y., Guo, S., Liu, W., Jin, T., Li, X., He, X., Zhang, X., Su, H., Zhang, N. and Duan, C. (2019) Silencing of SNHG12 Enhanced the Effectiveness of MSCs in Alleviating Ischemia/Reperfusion Injuries via the PI3K/AKT/mTOR Signaling Pathway. *Frontiers in Neuroscience*, **13**.
177. Lessard, S., Beaudoin, M., Benkirane, K. and Lettre, G. (2015) Comparison of DNA methylation profiles in human fetal and adult red blood cell progenitors. *Genome Med*, **7**, 1.
178. Bartholdy, B., Lajugie, J., Yan, Z., Zhang, S., Mukhopadhyay, R., Greally, J.M., Suzuki, M. and Bouhassira, E.E. (2018) Mechanisms of establishment and functional significance of DNA demethylation during erythroid differentiation. *Blood Adv*, **2**, 1833-1852.
179. Weiss, M.J., Keller, G. and Orkin, S.H. (1994) Novel insights into erythroid development revealed through in vitro differentiation of GATA-1 embryonic stem cells. *Genes Dev*, **8**, 1184-1197.
180. Fujiwara, Y., Browne, C.P., Cunniff, K., Goff, S.C. and Orkin, S.H. (1996) Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc Natl Acad Sci U S A*, **93**, 12355-12358.
181. Pevny, L., Lin, C.S., D'Agati, V., Simon, M.C., Orkin, S.H. and Costantini, F. (1995) Development of hematopoietic cells lacking transcription factor GATA-1. *Development*, **121**, 163-172.
182. Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H. and Costantini, F. (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature*, **349**, 257-260.
183. Tsai, F.Y. and Orkin, S.H. (1997) Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation. *Blood*, **89**, 3636-3643.
184. Leonard, M., Brice, M., Engel, J.D. and Papayannopoulou, T. (1993) Dynamics of GATA transcription factor expression during erythroid differentiation. *Blood*, **82**, 1071-1079.
185. Bresnick, E.H., Lee, H.Y., Fujiwara, T., Johnson, K.D. and Keles, S. (2010) GATA switches as developmental drivers. *J Biol Chem*, **285**, 31087-31093.

186. Bresnick, E.H., Martowicz, M.L., Pal, S. and Johnson, K.D. (2005) Developmental control via GATA factor interplay at chromatin domains. *J Cell Physiol*, **205**, 1-9.
187. Xu, J., Shao, Z., Li, D., Xie, H., Kim, W., Huang, J., Taylor, J.E., Pinello, L., Glass, K., Jaffe, J.D. *et al.* (2015) Developmental control of polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. *Mol Cell*, **57**, 304-316.
188. Renneville, A., Van Galen, P., Canver, M.C., McConkey, M., Krill-Burger, J.M., Dorfman, D.M., Holson, E.B., Bernstein, B.E., Orkin, S.H., Bauer, D.E. *et al.* (2015) EHMT1 and EHMT2 inhibition induces fetal hemoglobin expression. *Blood*, **126**, 1930-1939.
189. Webb, C.F., Das, C., Coffman, R.L. and Tucker, P.W. (1989) Induction of immunoglobulin mu mRNA in a B cell transfectant stimulated with interleukin-5 and a T-dependent antigen. *J Immunol*, **143**, 3934-3939.
190. Webb, C.F., Yamashita, Y., Ayers, N., Evetts, S., Paulin, Y., Conley, M.E. and Smith, E.A. (2000) The transcription factor *Bright* associates with Bruton's tyrosine kinase, the defective protein in immunodeficiency disease. *J.Immunol.*, **165**, 6956-6965.
191. Alvarez-Dominguez, J.R., Knoll, M., Gromatzky, A.A. and Lodish, H.F. (2017) The Super-Enhancer-Derived alncRNA-EC7/Bloodlinc Potentiates Red Blood Cell Development in trans. *Cell Rep*, **19**, 2503-2514.
192. Ward, J.M., Ratliff, M.L., Dozmorov, M.G., Wiley, G., Guthridge, J.M., Gaffney, P.M., James, J.A. and Webb, C.F. (2016) Human effector B lymphocytes express ARID3a and secrete interferon alpha. *J Autoimmun*, **75**, 130-140.
193. Colby, S.L. and Ortman, J.M. (2015). US Census Bureau. 4600 Silver Hill Road, Washington, DC 20233. Tel: 800-923-8282; Tel: 301-763-4636; e-mail: Census.in.Schools@census.gov; Web site: http://www.census.gov/, pp. 13.
194. Pang, W.W., Schrier, S.L. and Weissman, I.L. (2017) Age-associated changes in human hematopoietic stem cells. *Semin Hematol*, **54**, 39-42.
195. Laurenti, E. and Gottgens, B. (2018) From haematopoietic stem cells to complex differentiation landscapes. *Nature*, **553**, 418-426.
196. Dykstra, B., Olthof, S., Schreuder, J., Ritsema, M. and de Haan, G. (2011) Clonal analysis reveals multiple functional defects of aged murine hematopoietic stem cells. *J Exp Med*, **208**, 2691-2703.
197. Kuranda, K., Vargaftig, J., de la Rochere, P., Dosquet, C., Charron, D., Bardin, F., Tonnelle, C., Bonnet, D. and Goodhardt, M. (2011) Age-related changes in human hematopoietic stem/progenitor cells. *Aging Cell*, **10**, 542-546.
198. Pang, W.W., Price, E.A., Sahoo, D., Beerman, I., Maloney, W.J., Rossi, D.J., Schrier, S.L. and Weissman, I.L. (2011) Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc Natl Acad Sci U S A*, **108**, 20012-20017.
199. Sun, D., Luo, M., Jeong, M., Rodriguez, B., Xia, Z., Hannah, R., Wang, H., Le, T., Faull, K.F., Chen, R. *et al.* (2014) Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. *Cell Stem Cell*, **14**, 673-688.
200. Fali, T., Fabre-Mersseman, V., Yamamoto, T., Bayard, C., Papagno, L., Fastenackels, S., Zoorab, R., Koup, R.A., Boddaert, J., Sauce, D. *et al.* (2018) Elderly human hematopoietic progenitor cells express cellular senescence markers and are more susceptible to pyroptosis. *JCI Insight*, **3**.
201. Flach, J., Bakker, S.T., Mohrin, M., Conroy, P.C., Pietras, E.M., Reynaud, D., Alvarez, S., Diolaiti, M.E., Ugarte, F., Forsberg, E.C. *et al.* (2014) Replication stress is a potent driver of functional decline in ageing haematopoietic stem cells. *Nature*, **512**, 198-202.
202. Kortschak, R.D., Tucker, P.W. and Saint, R. (2000) ARID proteins come in from the desert. *Trends Biochem Sci*, **25**, 294-299.

203. Patsialou, A., Wilsker, D. and Moran, E. (2005) DNA-binding properties of ARID family proteins. *Nucleic Acids Res*, **33**, 66-80.
204. Popowski, M., Templeton, T.D., Lee, B.K., Rhee, C., Li, H., Miner, C., Dekker, J.D., Orlanski, S., Bergman, Y., Iyer, V.R. *et al.* (2014) Bright/Arid3A acts as a barrier to somatic cell reprogramming through direct regulation of Oct4, Sox2, and Nanog. *Stem Cell Reports*, **2**, 26-35.
205. Rajaiya, J., Hatfield, M., Nixon, J.C., Rawlings, D.J. and Webb, C.F. (2005) Bruton's tyrosine kinase regulates immunoglobulin promoter activation in association with the transcription factor Bright. *Mol Cell Biol*, **25**, 2073-2084.
206. Ward, J.M., Ratliff, M.L., Dozmorov, M.G., Wiley, G., Guthridge, J.M., Gaffney, P.M., James, J.A. and Webb, C.F. (2016) Expression and methylation data from SLE patient and healthy control blood samples subdivided with respect to ARID3a levels. *Data Brief*, **9**, 213-219.
207. Nixon, J.C., Ferrell, S., Miner, C., Oldham, A.L., Hochgeschwender, U. and Webb, C.F. (2008) Transgenic mice expressing dominant-negative bright exhibit defects in B1 B cells. *J Immunol*, **181**, 6913-6922.
208. Oldham, A.L., Miner, C.A., Wang, H.C. and Webb, C.F. (2011) The transcription factor Bright plays a role in marginal zone B lymphocyte development and autoantibody production. *Mol Immunol*, **49**, 367-379.
209. Baumgarth, N. (2017) A Hard(y) Look at B-1 Cell Development and Function. *J Immunol*, **199**, 3387-3394.
210. Sanz, I., Wei, C., Jenks, S.A., Cashman, K.S., Tipton, C., Woodruff, M.C., Hom, J. and Lee, F.E. (2019) Challenges and Opportunities for Consistent Classification of Human B Cell and Plasma Cell Populations. *Frontiers in immunology*, **10**, 2458.
211. Ratliff, M.L., Ward, J.M., Merrill, J.T., James, J.A. and Webb, C.F. (2015) Differential expression of the transcription factor ARID3a in lupus patient hematopoietic progenitor cells. *J Immunol*, **194**, 940-949.
212. Hayakawa, K., Li, Y.S., Shinton, S.A., Bandi, S.R., Formica, A.M., Brill-Dashoff, J. and Hardy, R.R. (2019) Crucial Role of Increased Arid3a at the Pre-B and Immature B Cell Stages for B1a Cell Generation. *Frontiers in immunology*, **10**, 457.
213. Li, Y.S., Zhou, Y., Tang, L., Shinton, S.A., Hayakawa, K. and Hardy, R.R. (2015) A developmental switch between fetal and adult B lymphopoiesis. *Ann N Y Acad Sci*, **1362**, 8-15.
214. Zhou, Y., Li, Y.S., Bandi, S.R., Tang, L., Shinton, S.A., Hayakawa, K. and Hardy, R.R. (2015) Lin28b promotes fetal B lymphopoiesis through the transcription factor Arid3a. *J Exp Med*, **212**, 569-580.
215. Lepus, C.M., Gibson, T.F., Gerber, S.A., Kawikova, I., Szczepanik, M., Hossain, J., Ablamunits, V., Kirkiles-Smith, N., Herold, K.C., Donis, R.O. *et al.* (2009) Comparison of human fetal liver, umbilical cord blood, and adult blood hematopoietic stem cell engraftment in NOD-scid/gammac-/-, Balb/c-Rag1-/-gammac-/-, and C.B-17-scid/bg immunodeficient mice. *Hum Immunol*, **70**, 790-802.
216. Harrison, D.E. and Astle, C.M. (1997) Short- and Long-Term Multilineage Repopulating Hematopoietic Stem Cells in Late Fetal and Newborn Mice: Models for Human Umbilical Cord Blood. *Blood*, **90**, 174-181.
217. Holodick, N.E. and Rothstein, T.L. (2015) B cells in the aging immune system: time to consider B-1 cells. *Ann N Y Acad Sci*, **1362**, 176-187.
218. Montecino-Rodriguez, E. and Dorshkind, K. (2012) B-1 B cell development in the fetus and adult. *Immunity*, **36**, 13-21.
219. Cancro, M.P. (2020) Age-Associated B Cells. *Annu Rev Immunol*.
220. Hao, Y., O'Neill, P., Naradikian, M.S., Scholz, J.L. and Cancro, M.P. (2011) A B-cell subset uniquely responsive to innate stimuli accumulates in aged mice. *Blood*, **118**, 1294-1304.

221. Rubtsov, A.V., Rubtsova, K., Fischer, A., Meehan, R.T., Gillis, J.Z., Kappler, J.W. and Marrack, P. (2011) Toll-like receptor 7 (TLR7)-driven accumulation of a novel CD11c(+) B-cell population is important for the development of autoimmunity. *Blood*, **118**, 1305-1315.
222. de Haan, G. and Lazare, S.S. (2018) Aging of hematopoietic stem cells. *Blood*, **131**, 479-487.
223. Rundberg Nilsson, A., Soneji, S., Adolfsson, S., Bryder, D. and Pronk, C.J. (2016) Human and Murine Hematopoietic Stem Cell Aging Is Associated with Functional Impairments and Intrinsic Megakaryocytic/Erythroid Bias. *PLoS One*, **11**, e0158369.
224. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)*, **29**, 15-21.
225. Hardy, R.R. (2006) B-1 B Cell Development. *The Journal of Immunology*, **177**, 2749.
226. Grammer, A.C. and Lipsky, P.E. (2003) B cell abnormalities in systemic lupus erythematosus. *Arthritis Res Ther*, **5**, S22.
227. Nashi, E., Wang, Y. and Diamond, B. (2010) The role of B cells in lupus pathogenesis. *Int J Biochem Cell Biol*, **42**, 543-550.
228. Navarra, S.V., Guzmán, R.M., Gallacher, A.E., Hall, S., Levy, R.A., Jimenez, R.E., Li, E.K., Thomas, M., Kim, H.-Y. and León, M.G. (2011) Efficacy and safety of belimumab in patients with active systemic lupus erythematosus: a randomised, placebo-controlled, phase 3 trial. *The Lancet*, **377**, 721-731.
229. Jenks, S.A., Cashman, K.S., Zumaquero, E., Marigorta, U.M., Patel, A.V., Wang, X., Tomar, D., Woodruff, M.C., Simon, Z. and Bugrovsky, R. (2018) Distinct effector B cells induced by unregulated toll-like receptor 7 contribute to pathogenic responses in systemic lupus erythematosus. *Immunity*, **49**, 725-739. e726.
230. Nehar-Belaid, D., Hong, S., Marches, R., Chen, G., Bolisetty, M., Baisch, J., Walters, L., Punaro, M., Rossi, R.J., Chung, C.-H. *et al.* (2020) Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nature Immunology*.
231. Arce, E., Jackson, D.G., Gill, M.A., Bennett, L.B., Banchereau, J. and Pascual, V. (2001) Increased frequency of pre-germinal center B cells and plasma cell precursors in the blood of children with systemic lupus erythematosus. *The Journal of Immunology*, **167**, 2361-2369.
232. Jacobi, A.M., Reiter, K., Mackay, M., Aranow, C., Hiepe, F., Radbruch, A., Hansen, A., Burmester, G.R., Diamond, B. and Lipsky, P.E. (2008) Activated memory B cell subsets correlate with disease activity in systemic lupus erythematosus: delineation by expression of CD27, IgD, and CD95. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, **58**, 1762-1773.
233. Tipton, C.M., Fucile, C.F., Darce, J., Chida, A., Ichikawa, T., Gregoretti, I., Schieferl, S., Hom, J., Jenks, S. and Feldman, R.J. (2015) Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus. *Nature immunology*, **16**, 755.
234. Shukla, V. and Lu, R. (2014) IRF4 and IRF8: governing the virtues of B lymphocytes. *Frontiers in Biology*, **9**, 269-282.
235. Berland, R., Fernandez, L., Kari, E., Han, J.H., Lomakin, I., Akira, S., Wortis, H.H., Kearney, J.F., Ucci, A.A. and Imanishi-Kari, T. (2006) Toll-like receptor 7-dependent loss of B cell tolerance in pathogenic autoantibody knockin mice. *Immunity*, **25**, 429-440.
236. Suurmond, J. and Diamond, B. (2015) Autoantibodies in systemic autoimmune diseases: specificity and pathogenicity. *The Journal of Clinical Investigation*, **125**, 2194-2202.
237. Muerdter, F., Boryń, Ł.M. and Arnold, C.D. (2015) STARR-seq — Principles and applications. *Genomics*, **106**, 145-150.