



Walter+Eliza Hall  
Institute of Medical Research

Institute Research Publication Repository

***This is the author accepted version of :***

Shi W, Liao Y, Willis SN, Taubenheim N, Inouye M, Tarlinton DM, Smyth GK, Hodgkin PD, Nutt SL, Corcoran LM. Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells. *Nature Immunology*. 2015 16(6):663-673.

***which has been published in final form at***  
doi:[10.1038/ni.3154](https://doi.org/10.1038/ni.3154)

## **Transcriptional profiling of mouse B cell terminal differentiation defines a signature for antibody-secreting plasma cells**

Wei Shi <sup>1,2</sup>, Yang Liao <sup>1,3</sup>, Simon N Willis <sup>1,3</sup>, Nadine Taubenheim <sup>1,3</sup>, Michael Inouye <sup>1,4,5</sup>, David M Tarlinton <sup>1,3</sup>, Gordon K Smyth <sup>1,6</sup>, Philip D Hodgkin <sup>1,3</sup>, Stephen L Nutt <sup>1,3</sup> and Lynn M Corcoran <sup>1,3</sup>

<sup>1</sup>The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria, 3052, Australia. <sup>2</sup>Department of Computing and Information Systems, <sup>3</sup>Department of Medical Biology, <sup>4</sup>Department of Pathology, <sup>5</sup>Department of Microbiology & Immunology, <sup>6</sup>Department of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria, 3010, Australia.

**Correspondence** should be addressed to: W.S. ([shi@wehi.edu.au](mailto:shi@wehi.edu.au)), S.L.N. ([nutt@wehi.edu.au](mailto:nutt@wehi.edu.au)) or L.M.C. ([corcoran@wehi.edu.au](mailto:corcoran@wehi.edu.au)).

## **ABSTRACT**

Upon encounter with antigen, B cells alter their physiological state, anatomical localization, and initiate a differentiation process that ultimately produces antibody-secreting cells (ASCs). We have defined the transcriptomes of many mature B cell populations and stages of plasma cell differentiation from the mouse. We provide a molecular signature of ASCs that highlights the stark transcriptional divide between B cells and plasma cells, and enables the demarcation of ASCs based on location and maturity. The gene expression changes correlated with cell division history and the acquisition of permissive histone modifications and include many regulators not previously implicated in B cell differentiation. These findings both highlight and expand the core program that guides B cell terminal differentiation and the production of antibody.

Antibodies are indispensable effector molecules in the immune system, providing both immediate and longer-term protection against infection, a facet exploited in most currently used vaccines. Antibody-secreting cells (ASCs) comprise a rare population of highly specialized cells that are the end stage of B cell lineage differentiation <sup>1</sup>. The ASC compartment consists of short-lived and cycling plasmablasts (PBs), which are generated early in an immune response in secondary lymphoid organs, and long-lived post-mitotic plasma cells (PCs) that reside in secondary lymphoid organs and also in specialised niches in the bone marrow. In addition to the essential function of B cells and antibody in mediating humoral immunity, pathogenic antibodies are key drivers of some autoimmune diseases such as systemic lupus erythematosus <sup>2</sup>, while multiple myeloma and plasmacytoma result from the malignant transformation of ASCs <sup>3</sup>. Thus understanding the factors that control ASC production, maturation and long-term survival is critical for both improved vaccine design and to provide mechanisms to target pathogenic ASCs.

Mature B cells consist of three distinct subsets, follicular (FoB), marginal zone (MZB) and B1 B cells, all of which contribute to the ASC pool and circulating serum antibody, although with distinct affinities and kinetics <sup>4</sup>. MZ and B1 B cells are specialized types of B cells whose function is largely independent of T cell help. MZ B cells flank the marginal sinuses of the spleen, where they respond to blood-borne antigens, whereas B1 B cells localize to the peritoneal and pleural cavities, where they provide an early line of defence against pathogens that enter through mucosal surfaces. FoB, the numerically dominant mature B cell subset, localize to lymphoid follicles in the spleen and lymph nodes. Upon activation with a protein antigen and in collaboration with cognate T helper cells, FoB cells proliferate extensively, undergo immunoglobulin class switch recombination (CSR) and generate either an extrafollicular PB response, that is both rapid and transient <sup>1</sup>, or enter into a germinal center (GC) reaction where they undergo somatic hypermutation (SHM) and selection for high-

affinity antigen receptors, again under the direction of specialised CD4<sup>+</sup> T cells. The GC provides the majority of the long-term immunity, through the production of long-lived PCs and memory B cells <sup>5,6</sup>.

On a transcriptional level, the differentiation of activated B cells into ASCs requires the coordinated change in the expression of many hundreds of genes. These changes can be characterized into two major categories: the loss of B cell associated transcripts and the acquisition of the ASC gene regulatory network. Research over the past decade has highlighted the roles played by the B cell promoting transcription factors PAX5, BCL6 and BACH2 in maintaining the B cell fate, while a triad of factors, BLIMP1, IRF4 and XBP1, are required to extinguish the B cell genes and activate the ASC program <sup>7,8</sup>. The clear transcriptional distinction between B cells and ASCs is maintained by the mutually antagonistic interactions between these major regulatory factors, although exactly how extrinsic signals such as antigen and T cell derived ‘help’, such as CD40L and cytokines, modulate this process is poorly understood.

Although a number of studies have used microarray technologies to catalogue the transcriptome of various stages of late B cell differentiation from both mouse and human <sup>9-14</sup>, there has been to date no comprehensive, single platform, transcriptional analysis of all aspects of the B cell to PC differentiation process. Here we have utilized RNA sequencing (RNA-seq) technology and the Blimp1-GFP reporter mouse strain, which allows the identification of all ASCs *in vivo* and *in vitro* <sup>15</sup>, to provide a comprehensive description of the dynamic nature of B cell terminal differentiation. These data allowed the derivation of an ASC signature and an improved description of how different sources of mature B cells, activation stimuli, localization, cell division history, histone code or ASC maturation impact on this essential biological process. The analysis also identified many new potential regulators of B cell and ASC differentiation and specialisation.

## RESULTS

### Transcriptional profiling of mouse peripheral B cells

RNA-seq was used to create comprehensive gene expression profiles for mature B cells and terminally differentiated ASCs. We purified the major peripheral B cell populations from C57Bl/6 mice, or from mice on this genetic background that carried a GFP reporter allele within the *Prdm1* (the gene encoding transcriptional regulator Blimp1) locus<sup>15</sup>. Splenic FoB (Small, B220<sup>+</sup>, CD21<sup>+</sup>, CD23<sup>+</sup>), MZB (B220<sup>+</sup>, CD21<sup>hi</sup>, CD23<sup>-</sup>), plasmablasts (SplPB; CD138/Syndecan1<sup>+</sup>, Blimp1-GFP<sup>lo</sup>) and plasma cells (SplPC; Syndecan1<sup>+</sup>, Blimp1-GFP<sup>hi</sup>) were sorted, B1 B cells (B220<sup>+</sup>, Mac1<sup>lo</sup>, CD23<sup>-</sup>) were sorted from peritoneal lavages and plasma cells (BMPC; Syndecan1<sup>+</sup>, Blimp1-GFP<sup>hi</sup>) were sorted from bone marrow (**Fig. 1a**). Germinal center B cells (GCB; CD19<sup>+</sup>, Fas<sup>+</sup>, PNA<sup>+</sup>) were induced by immunization with sheep red blood cells (SRBC). Independent biological replicates were obtained for most populations. Sequencing generated between 30 and 200 million reads from each RNA sample. Reads or read-pairs were mapped to the mouse reference genome using an aligner that is able to map across exon-exon junctions<sup>16</sup>, and the number of reads or read-pairs mapping to the exonic regions of each gene was recorded<sup>17</sup>. The expression level of each gene was summarized by the number of fragments per kilobase of exon length for that gene per million mapped fragments for that sample (FPKM).

First we examined the size of the transcriptome expressed in each cell population. Around two thirds of all sequence reads from ASCs mapped to loci encoding immunoglobulins (Ig; see below), and it might be hypothesized that this might limit the repertoire of other genes expressed by ASCs. By comparing read coverage of annotated exons versus read coverage over the intergenic genomic regions, we were able to estimate the total number of genes expressed in each cell type in a way that accounted for possible mapping errors and background noises and was largely independent of sequence depth. This showed that 64-67%

of all annotated genes were expressed in each cell population including ASCs (**Supplementary Fig. 1**). This indicates that, despite their strong functional specialization, ASCs maintain a highly diverse gene expression repertoire similar to B cells.

To explore the relationships between the cell populations in an unbiased way, we measured the transcriptional distance between any pair of expression profiles in terms of ‘leading fold change’, defined as the average magnitude of fold-change for the top 500 genes most different between the two cell types. A multi-dimensional scaling (MDS) plot was used to display the cell populations in such a way that distances on the plot correspond to log<sub>2</sub>-leading fold change (**Fig. 1b**). The plot shows that FoB and MZ are most closely related, with B1 and GCB cells having more distinct identities. The three ASC populations clustered closely and separately from the three B cell populations. As Ig gene transcripts comprise the majority of the transcriptome of plasma cells (see below), Ig genes were excluded from this MDS analysis and throughout this study unless stated otherwise.

Bioinformatic analysis of all the RNA-seq data allowed us to obtain a gene signature for *ex vivo* ASCs. To be included in the signature, a gene had to be expressed at least 3-fold more highly in SplPB, SplPC and BMPC than in any of the other populations analysed, and achieve a false discovery rate (FDR) of  $\leq 0.05$  in the assessment of differential expression between any ASC population and any of the other populations. Each signature gene was also required to have an expression abundance of  $\geq 32$  FPKMs. The signature contains 301 genes (**Supplementary Table 1**), and includes those encoding the well-known regulators Blimp1, Irf4 and Xbp1, as well as ASC surface markers, CD138 (Syndecan1), Tnfrsf17 (BCMA)<sup>18</sup>, CD28<sup>19</sup>, Ly6C<sup>20</sup> and the amino acid transporter CD98 (**Fig. 1c**). The ASC signature genes were classified by biological function (**Methods** and **Fig. 1d**). They testify to the functionality of ASCs, with ~40% of expressed genes having a role in protein production (translation, modification and trafficking). The signature included novel markers that will be

helpful in the future identification, purification and characterization of ASCs, as well as a number of small non-coding RNAs. Expression profiles of the top ~100 differentially expressed genes (selected from comparing FoB and BMPC) in FoB, MZB, B1, GCB, SplPB, SplPC and BMPC demonstrate the clear demarcation separating the four *ex vivo* B cell and the three ASC populations (**Fig. 1e, f**).

The Blimp-GFP reporter allowed us to distinguish less mature PBs in the spleen from the PCs of spleen and BM<sup>15</sup>. Analysis of the genes most differentially expressed between these three *ex vivo* ASC populations (**Fig. 2a, b**) confirmed the close relationship between the PCs of the spleen and BM, with again, a clear demarcation evident in the heat maps between the less mature SplPBs and their more differentiated counterparts. Validation of these differences by flow cytometry was confirmed for 4 surface markers (**Fig. 2c**). This included the chemokine receptors CCR10 and CXCR3, previously associated with plasma cell migration to mucosal and inflamed sites<sup>21-23</sup>, respectively, the epithelial cell adhesion molecule (EpCam), which mediates homotypic adhesion in epithelia, and Ly6d (CD59), an inhibitor of membrane attack complex formation, protecting cells from complement-mediated lysis. These markers might be useful to identify and distinguish PBs and PCs *in vivo*, and to better understand their biology.

Classification of the differentially expressed genes by biological function shows that the strongest effect was the down-regulation of cell cycle genes in BMPC, followed by repression of genes devoted to receptor expression and signal transduction (**Fig. 2d**, left panel). Genes elevated in terminally differentiated BMPC (**Fig. 2d**, right panel) appear to broadly reinforce the functional properties of ASCs, with enhanced metabolic activity, possibly underlying the greater Ig secreting capacity (unpublished), and changes (chemotaxis, cytoskeleton, survival) that could reflect the physical and biochemical properties that enable longevity in the BM niche.

Closer analysis of RNA-seq data for genes encoding chemokines, cytokines and their receptors highlights their exquisitely tight regulation within the B cell differentiation pathway (**Supplementary Fig. 2**). For example, each B cell subset shows a distinct pattern of Sphingosine-1-phosphate receptor (S1PR) expression, while several chemokine and/or cytokine receptors are strongly regulated at the B cell to ASC transition. In summary, the ASC transcriptional signature we define here can be used to delineate the major populations of ASCs regardless of their maturity or anatomical localization, while also providing insights into the dynamic expression of the effector molecules in distinct biological systems of interest.

### **Immunoglobulin gene transcripts in ASCs**

While it is well known that ASCs actively transcribe Ig genes for large scale antibody secretion, our RNA-seq data allowed us to quantify and contrast the levels of steady-state Ig gene transcripts in B cells and ASCs. Strikingly, more than 70% of the transcriptome of BMPC is derived from the Ig loci (**Fig. 3a**, left panel), with slightly lower levels in SplPC, and lower again in the less mature SplPB. These levels are 30-35 times higher than the Ig gene abundance in peripheral B cell populations, where the frequency is ~2%. GFP<sup>+</sup> (Blimp1<sup>+</sup>) ASCs generated *in vitro* from B cells (see **Methods**) also devoted the majority of their transcriptional efforts to Ig gene expression, regardless of whether they expressed Syndecan1, and independent of the differentiation-inducing signal (**Fig. 3a**, right panel).

The variable region of the *Igh* locus (*Ighv*) encodes approximately 120 annotated transcripts from 16 families in C57Bl/6 mice. The Subjunc aligner<sup>22</sup>, which was used here for mapping *Ighv* reads, was configured to allow the detection of insertions, deletions and

substitutions at any location of the *Ighv* gene segments. Only uniquely mapped reads were included in the analysis.

89 *Ighv* from 14 families were detected at a frequency of at least 0.1% of total *Ighv* reads (**Fig. 3b**). FoB and MZB showed a similar *Ighv* profile, presumably that reflects their common bone marrow origin<sup>24</sup>. B1 cells, in contrast, which derive from a fetal progenitor, showed the expected preferential expression of the *Ighv11-2* and *Ighv12-3* genes<sup>25,26</sup>. A remnant of this B1 *Ighv* transcript pattern is visible in the BMPC compartment, suggesting a small B1 cell contribution to the long-lived ASC population of the BM. *Ighv* usage was more diversified in the SplPC and BMPC populations, probably due to both antigenic selection and affinity maturation, although, on the whole, the repertoire reflected the initial *Ighv* frequency seen in the FoB population. These data indicated that RNA-seq can be used to deduce *Ighv* gene usage and that the ASCs ‘spontaneously’ produced in otherwise naive mice provide a reasonable approximation of the starting *Ighv* gene frequencies.

### **Contrasting ASCs generated *in vivo* and *in vitro***

The capacity for B cells to differentiate into ASCs *in vitro* has been an extremely useful property experimentally, but it is clear from phenotypic and functional studies that such cells generated *in vitro* lack some features of terminally differentiated PC<sup>1</sup>. We performed whole transcriptome analysis of ASCs generated from splenic B cells *in vitro* under T-independent (TI) and T-dependent (TD) via stimulation with lipopolysaccharide (LPS) or CD40L+IL-4 (+/-IL-5) respectively conditions (**Fig. 4a**). During TI stimulation, LPS activates B cells and induces Blimp1-GFP<sup>+</sup> cells that variously express Syndecan1, while TD conditions only produce Syndecan1<sup>+</sup> cells that are also Blimp1-GFP<sup>+</sup><sup>15</sup>. The sorted populations analysed were designated ‘40/4 blasts’ (48h CD40L+IL-4 stimulated B cells; CD19<sup>+</sup>, B220<sup>hi</sup>,

Syndecan1<sup>-</sup>), '40/4/5 PB' (day 5 CD40L+IL-4+IL-5 stimulated B cells; Syndecan1<sup>+</sup>) and cells sorted from day 3 LPS cultures designated 'LPS blasts' (B220<sup>+</sup>, Syndecan1<sup>-</sup>, Blimp1-GFP<sup>-</sup> cells), 'LPS Syn-PB' (B220<sup>+</sup>, Syndecan1<sup>-</sup>, Blimp1-GFP<sup>+</sup>) and 'LPS PB' (B220<sup>+</sup>, Syndecan1<sup>+</sup>, Blimp1-GFP<sup>+</sup>). For cell division experiments, 'Div 0' represented activated but undivided B cells, stimulated for 24h with CD40L+IL-4 (Small, B220<sup>+</sup>, Syndecan1<sup>-</sup>), 'Div1-7' cells (B220<sup>+</sup>, Syndecan1<sup>-</sup>, CFSE<sup>+</sup> cells, sorted by division number) were from day 4 CD40L+IL-4+IL-5 stimulated cells, and '40/4 PB were Syndecan1<sup>+</sup> cells sorted from these same cultures.

Multi-dimensional scaling (MDS) analysis implies a differentiation pathway (dotted line, **Fig. 4b**) that cells take between the FoB B cell state and full differentiation, with all the *in vitro*-derived cell populations lying along it. In this scheme, activated B cell blasts have progressed a short way along this path, and all *in vitro* derived PB lie further along the continuum. It is however striking that the *in vitro* derived ASCs are still as different from fully mature PC as they are from activated B cells. Heat maps highlight the 100 most differentially expressed genes (50-down and 50-up-regulated, comparing FoB and BMPC), and including the *in vitro* populations (**Figure 4c-d**). The genes which are shared between ASC populations, or are unique to them, are indicated using Venn diagrams (**Figure 4e, f** and **Supplementary Table 2**). All ASCs shared a core of differentially expressed genes compared to FoB (903 are repressed and 808 induced). The vast majority of the ASC signature genes (84%, 254 of 301 genes, FDR<0.05, fold change >1.5) were also up-regulated in 40/4/5 PB and LPS PB, confirming the utility of this gene set to identify *in vitro* differentiated ASCs. The comparison also identified the 221 repressed and 350 induced genes that distinguished BMPC from PBs generated *in vitro*, a subset that contained genes required for full PC maturation.

Although the above analyses highlighted the shared gene expression program of ASCs, genes whose expression was selective for TI versus TD differentiation are also of interest. Venn diagrams highlight the gene expression differences between PBs that arose in culture in response to LPS or CD40L+IL-4+IL-5 (**Fig. 5a, b**). A total of 370 repressed genes (214 + 156) and 310 induced genes (152 + 158) distinguished these TI and TD PBs (**Supplementary Table 3**), exemplified by six LPS-PB specific genes (**Fig. 5c**), and six 40/4/5-PB specific genes (**Fig. 5d**). Differential expression of three surface markers was validated by flow cytometry (**Fig. 5e, f**). SOCS2, an inhibitor of growth hormone<sup>27</sup>, was specifically expressed in 40/4/5 PBs (**Fig. 5g**). The physiological relevance of SOCS2 in this setting was suggested by the increased numbers of PBs from *Socs2*<sup>-/-</sup> B cells only in CD40L+IL-4+IL-5 cultures (**Fig 5h**), and the significantly elevated IgG1 (but not IgM) titers in *Socs2*<sup>-/-</sup> mice compared to control mice (**Fig 5i**). These findings suggest that SOCS2 is a selective mediator of TD antibody responses.

### **Gene expression changes with cell division *in vitro***

The probability of becoming an ASC increases with each cell division cycle<sup>28</sup>, implying a sequential staging of the differentiation process following activation. Additional division-linked changes to cell surface molecule expression, and CSR diversify the emerging cell population<sup>14,29,30</sup>. To explore this directly, we labelled resting FoB cells with the cell division tracker carboxyfluorescein diacetate succinimidyl ester (CFSE), activated them with CD40L+IL-4, and sorted Syndecan1<sup>-</sup> activated B cell blasts after 4 days from divisions 1, 3, 5 and 7 (gates illustrated in **Fig. 6a**). Syndecan1<sup>+</sup> ASC cells were sorted at the same time (40/4 PB). Additional populations included unstimulated B cells and cells cultured in CD40L+IL-4 for 24h, at which time cells were activated but had not yet divided (division 0). All samples were profiled by RNA-seq. MDS plotting of the profiles by division number

(**Fig. 6b**) intimated a differentiation pathway (dotted line) that tracked from unstimulated cells through early to late divisions and finally to differentiation. The expression profile of late division non-ASC was very close to the PB, and significantly changed from the early divisions. Furthermore, the biggest transitional step is in the initial 24 h activation from the resting state to the Div 0 sample with successively smaller changes until becoming a PB.

To explore this development program, we calculated the maximum fold change between the highest and lowest expression level in any sample (Unstimulated, Divided samples or PB) for each gene (**Fig. 6c**). 4909 genes altered in expression more than 2.5 fold at some point in the transition from unstimulated B cells to PB. Of these more than half (2717 genes) were accounted for by activation alone (Unstimulated to Division 0), whereas 2095 expressed genes changed within the division sequence itself and 680 were involved in the final step to PB. Within the division sequence the most volatile stage is Div 1 to 3 with later divisions showing progressively fewer expression changes (**Fig. 6d**). Clustering based on the step-wise gene expression changes revealed nine clear expression patterns (**Supplementary Fig. 3**). The individual genes have been normalised to Div 0 and averaged to illustrate the range of gene expression changes involved in the division-linked program (**Fig. 6e**). These patterns include graded and stepwise transitions with successive division, as well as wave like expression, as in patterns 1(a) and 1(b) (**Fig. 6e**). Heat maps illustrate the expression profiles of the top 20 up-regulated and top 20 down-regulated genes, selected by comparing Divisions 0 and 7 (**Fig. 6f**). These expression patterns were likely to be driven by single, or combinations of, transcriptional regulators. Indeed, expression of several well-known transcription factors progressively changed with division (**Fig.6f**). These data confirm that early activation prepares B lymphoblasts to play out a transcriptional program in step with division that both leads to ASCs and also generates considerable cellular heterogeneity along the way.

### **Expression changes correlate with epigenetic modifications**

We also assessed the importance of chromatin modifications on gene expression during B cell terminal differentiation using both published and newly generated data. The histone modifications H3K4me3 and H3K4me1 are enriched in active promoters and distal enhancers in many developmental settings<sup>31</sup>. Analysis for the presence of these modifications using chromatin immunoprecipitation followed by sequencing (ChIP-seq) on the mouse GCB cell line A20 and the plasmacytoma MPC11 showed a strong correlation between the presence of the histone mark and expression of the associated gene in both the A20 and MPC11 cells and the corresponding *in vivo* populations, GCB and BMPCs (**Supplementary Fig. 4a, b**). The transformed cell lines were used as surrogates for the much rarer primary cell populations *in vivo*. With this caveat, mapping of the promoter and enhancer landscape using the histone code was predictive of gene expression changes during the B cell terminal differentiation process.

Although distal enhancers are key regulatory elements in the control of gene expression in general, it has become apparent that many genes involved in cellular identity and disease progression are regulated by large clusters of DNA elements termed ‘super-enhancers’<sup>32,33</sup>. This was particularly the case in multiple myeloma, where genes that were prominent in disease progression, and PC biology in general, were highly enriched for the presence of super-enhancers<sup>34</sup>. Interrogation of the expression of a set of previously annotated super-enhancer containing genes in the B cell terminal differentiation process revealed that super-enhancer containing genes were strongly down-regulated in the transition from either FoB or GCB cells to BMPCs (**Supplementary Fig. 4c, d**), suggesting that inappropriate maintenance of the expression of these genes is likely to contribute to the malignant progression of multiple myeloma. These data further highlight the transcriptional distinction between B cells

and ASCs and highlight the important role of chromatin remodelling in the differentiation process.

### **New transcriptional regulators of B cell differentiation**

Finally, we examined the expression of a particular subset of genes within our RNA-seq data; specifically those involved in the regulation of gene expression, including transcription factors, co-activators or repressors and chromatin modifiers. The expression patterns of 103 such genes that were differentially expressed between FoB and BMPC clustered into 3 major groups (delineated by dotted lines in **Fig. 7a**). A prominent cluster comprised genes that were expressed in all B cell subsets before being down-regulated in ASCs. This cluster includes known B cell regulators such as *Pax5* and *Spib*, as well as a large number of novel potential factors including *Bcl11a*, *Ablim1*, *Tle3* and *Hhex* (**Fig. 7a, b** and **Supplementary Fig. 5a**). The GCB cell program, was highlighted by a second cluster containing the essential GC genes *Bcl6*, *Bach2* and *Pou2af1*, as well as some less characterized potential regulators such as *Phf19* and *Apitd1* (**Fig. 7a, c** and **Supplementary Fig. 5b**). Minor clusters harbored MZB and B1 specific transcriptional regulators (**Fig. 7a, d** and **Supplementary Fig. 5c, d**). The final major cluster contained the genes most highly expressed in ASCs, including all the known regulators *Prdm1*, *Irf4*, *Xbp1*, *Fos*, *Ell2*, and *Ern1* (**Fig. 7a, e** and **Supplementary Fig. 6**). The expression pattern of these established components of the ASC transcriptional network was reflected by those of a further approximately 30 other transcriptional regulators not yet implicated in B cell biology (**Fig. 7a, e** and **Supplementary Fig. 6**). Taken together, this analysis identified a discrete set of genes, many of which have not been implicated in lymphocyte biology, that function together to program B cells terminal differentiation.

## DISCUSSION

Terminal differentiation is an irreversible process resulting in the acquisition of specialized cell functions and often exit from the cell cycle. In B cells, at least three distinct mature cell subsets undergo terminal differentiation to antibody-secreting PBs and PCs. These cells secrete antibodies of a range of isotypes and affinities, reside in multiple anatomical sites in the body and have markedly different lifespans. Despite these phenotypic differences the data presented in this study show clearly that all ASCs share a common transcriptional signature that is distinct from all B cell subsets.

Analysis of the ASC signature revealed that major functional classes of the newly expressed genes encode proteins that are involved in the transcription, translation, intracellular transportation and glycosylation of Ig. Upregulation of the unfolded protein response (UPR) genes was also an expected consequence of the extremely high Ig production. An early event in ASC differentiation is the transition from cycling short-lived PB to post-mitotic and long-lived PCs<sup>1</sup>. In keeping with this transition, genes involved in cell cycle progression were prominent in the PB populations and repressed in PCs. The signals that sustain the longevity of PCs in the bone marrow are only partly understood, but are known to consist of secreted factors such as APRIL and signals elicited by direct cell contact with the other cellular compartments of the survival niche<sup>35</sup>. *Tnfrsf17*, encoding BCMA, the receptor for APRIL is a component of the ASC signature<sup>13,18</sup>, suggesting that this exclusivity provides a component of the specificity of bone marrow PC survival. The genes encoding a variety of other adhesion molecules, including SLAMF7, BST2, EpCAM, ALCAM and FNDC3B are also components of the ASC signature and are therefore potential previously undescribed regulators of ASC homing and survival. As monoclonal antibodies directed against EpCAM are currently in advanced clinical trials in a variety of cancers<sup>36</sup>, our results

suggest that it would be pertinent to examine EpCAM expression on human multiple myeloma samples.

An outcome of this work is that it provides an opportunity to investigate the roles of potential mediators of ASC differentiation and function. In this context the highly regulated expression of many chemokines and their receptors, as well as the S1PRs provides insights into the many, often contradictory signals that facilitate the localization of the B cell and ASC subsets. The function of many of these factors, including the atypical chemokine receptors, CCRL1 (GCB specific) and CCRL2 (ASC specific), remain unknown in B cells. Although most cytokine receptor chains were down-regulated during B cell differentiation, several including *Il2rg*, *Il13ra1* and *Il10rb* remain expressed *in vivo* and might provide insights into the factors that control ASC maturation and survival.

Our data largely confirmed the specific expression patterns of known drivers of B cell, GCB and ASC fate, but more interestingly, identified a number of other transcriptional regulators with parallel patterns of expression. For instance, Phf19, a component of the polycomb repressive complex (PRC2) involved in recruitment to chromatin<sup>37</sup>, is exclusively expressed in GCB cells. Other ASC specific factors include Eaf2, a pro-apoptotic tumor suppressor<sup>38</sup>, and Cited2, a mediator of stem cell maintenance and an oncogene<sup>39</sup>. Further studies of the impact of such genes are likely to provide insights into ASC biology and humoral immunity.

Although their largely overlapping gene expression profiles allowed us to identify a robust ASC signature, it was also apparent that there were considerable transcriptional differences between those ASC generated *in vivo* and *in vitro*, as well as distinct transcriptional consequences of differentiation induced by LPS, a model for TI stimulation, and CD40L+IL-4+IL-5, modelling TD responses. MDS plots suggested that *in vitro* derived PB, although clearly ASCs, were not as fully mature as their *in vivo* counterparts. Despite this

approximately 200 genes were uniquely up-regulated in each of LPS PB and 40/4/5. PB *Socs2* represents one such gene that is specifically induced in 40/4/5 PB and analysis of SOCS2-deficient mice suggests that this protein may negatively regulate ASC differentiation in that setting. SOCS2 is best characterized as a suppressor of growth factor signaling<sup>27</sup>, however there is some evidence for inhibitory control of IL-4 signalling in T cells<sup>40</sup>. Future work will be needed to fully assess whether SOCS2 plays a role in TD immune responses *in vivo*.

On a cellular level, the differentiation of B cells into ASCs unfolds in a cell division-related manner<sup>28</sup>. Our transcriptome data gathered from cells that had undergone a defined number of cell divisions confirmed that with each cell cycle activated B cells acquired a transcription program that more closely resembled that of ASCs. Clustering based on gene expression revealed 9 clusters, suggestive of a common regulatory mechanism for each. Indeed, discrete groups of similarly expressed transcription factors could also be identified that represent candidate regulators of each cluster. These insights suggest that with the application of small scale ChIP-seq protocols to individual cell division cohorts, it will be feasible to define the gene regulatory networks underlying the division linked process of B cell differentiation. Similar approaches have been used to define the transcriptional networks operating in innate immunity<sup>41</sup>, lymphoma cells<sup>42</sup> and in T<sub>H</sub>17 cells<sup>43</sup>.

The stark transcriptional changes we have observed that accompany the differentiation of B cells to ASCs are also manifest in concurrent changes in chromatin state. A small subset of genes that are often major developmental regulators are known to be controlled by the activity of large domains of active chromatin known as super-enhancers<sup>32,33</sup>. In human B cells super-enhancer containing genes include those encoding PAX5, BACH2, IRF8 and MYC, all of which are strongly down-regulated during terminal differentiation<sup>32</sup>. In contrast multiple myeloma is associated with inappropriate maintenance of several B cell super-

enhancers, including the region associated with MYC<sup>34</sup>. In keeping with this, we found super-enhancer containing genes to be enriched in those genes silenced in the ASC differentiation process, suggesting that this is a key process allowing ASC formation. Chromatin modifying enzymes have become attractive targets for pharmaceutical intervention in diseases such as cancer and autoimmunity. Our data support this concept as the large number of transcriptional changes and concurrent chromatin re-modelling we have documented during ASC differentiation suggests that this process will be exquisitely sensitive to therapeutic intervention of the gene regulatory networks driving this process.

## **ACKNOWLEDGEMENTS**

We are grateful to D Emslie, T Kratina, M Everest and S Heinzl for technical assistance, and to J Vasiliadis, J Leahy and T Camilleri for animal care. All experiments were undertaken according to the Walter and Eliza Hall Institute's Animal Ethics Committee guidelines and approval (#AEC2013.014). The institutional flow cytometry facility provided their essential services expertly. This work was supported by the National Health and Medical Research Council (NHMRC) of Australia (NHMRC IRIISS grant #361646, Program Grants #575500 and 1054925, to D.M.T, P.D.H., S.L.N. and L.M.C., Program Grant #1054618 to G.K.S. and Project Grant #1023454 to G.K.S. and W.S.) and a Multiple Myeloma Research Foundation Senior Award to S.L.N. G.K.S., D.M.T, P.D.H., and L.M.C. were supported by fellowships from the NHMRC. S.L.N. was supported by an ARC Future Fellowship. This work was made possible through Victorian State Government Operational Infrastructure Support.

## **ACCESSION CODES**

Raw sequence reads, read counts and normalized expression values have been deposited into the GEO (Gene Expression Omnibus) database under accession number GSE60927.

## **AUTHOR CONTRIBUTIONS**

W.S., Y.L., G.K.S. and M.I. performed the bioinformatic and statistical analyses; S.N.W. and N.T. performed experiments relating to in vitro differentiation and cell division studies; S.L.N. D.M.T, P.D.H., and L.M.C. contributed to experimental design and analysis and provided mouse models; L.M.C. performed many of the experiments; W.S., S.L.N. and L.M.C. wrote the manuscript, to which all authors had editorial input.

## **COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

## FIGURE LEGENDS

### Figure 1 Transcriptome profiling of *ex vivo* B cell and ASC populations

(a) Dot plots indicating the gating strategies used to isolate the *ex vivo* B cell and ASC populations analysed throughout this figure. Tissues used, gates and sorted populations are indicated. For the ASC cell sorts (bottom panels), Syndecan1<sup>+</sup> cells were first enriched using anti-Syndecan1 (CD138) conjugated magnetic beads. (b) MDS clustering of B cell and ASC populations. The top 500 most differentially expressed genes between each pair of samples were selected for sample clustering. Mean expression values (averaged over replicates) for selected genes for each sample were used to calculate the distance (leading *log*<sub>2</sub>-fold-change) between each pair of samples. (c) Functional analysis of genes in the ASC signature. The numbers of genes in each functional category are shown. (d) Surface staining of splenic FoB (Small, B220<sup>+</sup>, CD21<sup>+</sup>, CD23<sup>+</sup>) cells and BMPC (Syndecan1<sup>+</sup>, Blimp1-GFP<sup>hi</sup>), to confirm differential expression of two surface markers (Ly6C and CD98) that are part of the ASC signature. (e) Expression profile of the top 50 down-regulated genes in BMPC compared to FoB. (f) Expression profile of the top 50 up-regulated genes in BMPC compared to FoB, plus 4 extra genes of particular immunological interest (in blue text). *Log*<sub>2</sub>-FPKM expression values of genes are shown in the heatmaps, color-coded according to the legend.

### Figure 2. Comparison of *ex vivo* ASC populations

Expression profiles of the top ~100 genes most differentially expressed in BMPC compared to SplPB (Syndecan1<sup>+</sup>, Blimp1-GFP<sup>lo</sup>) and SplPC (Syndecan1<sup>+</sup>, Blimp1-GFP<sup>hi</sup>). (a) Heat maps of the 50 most strongly down-regulated genes, shown for the three *ex vivo* ASC populations. Genes with the smallest FDRs in all the down-regulated genes were selected. *Log*<sub>2</sub>-FPKM expression values of genes are represented, as shown in the scale. (b) Heat maps of the 50 most strongly up-regulated genes, plus 2 extra genes of particular immunological

interest (Epcam and Ly6d). Genes with the smallest FDRs in all the up-regulated genes were selected. Genes that are members of the plasma cell signature are notated in red. (c) Flow cytometric validation of differential expression of 4 surface molecules (CCR10, CXCR3, EpCam, Ly6d) differentially expressed between PBs and PCs. Splenocytes and BM were from Blimp1-GFP mice, and marker expression is shown for Syndecan1<sup>+</sup>Blimp1-GFP<sup>+</sup> cells, gated as shown in the top panel. (d) Functional annotation of genes differentially expressed between SplPB and BMPC. Left panel annotates genes up-regulated in SplPB while right panel annotates genes up-regulated in BMPC.

### **Figure 3. Expression of immunoglobulin genes in different B cell and ASC populations**

(a) Fraction of reads mapped to annotated immunoglobulin (Ig) genes out of all reads that were successfully mapped to NCBI RefSeq genes, in each cell type. Both heavy chain and light chain Ig genes are included in analysis. Left panel includes *ex vivo* populations and right panel includes *in vitro* populations plus FoB. (b) Fraction of reads mapped to each individual Igh V gene (*Ighv*) out of all the reads mapped to *Ighv* genes in each cell type. *Ighv* genes are ordered according to their fraction values in the FoB (from largest to smallest). The 31 *Ighv* genes with a fraction value less than 0.001 in all samples were not included in the graphs. Names shown for the *Ighv* genes are abbreviated in the figure (eg. '1-26' has the full name '*Ighv1-26*'). Red boxes highlight the *Ighv12-3* and *Ighv11-2* genes known to be enriched in B1 B cells.

### **Figure 4. Transcriptome profiling of *in vitro* B cell and ASC populations**

(a) Staining and sorting strategy used on B cells activated in culture. Splenic B cells cultured for the indicated time and conditions were sorted, either on the basis of the differentiation markers Blimp1-GFP and Syndecan1 (Syn) for LPS cultures, or by Syndecan1 and B220

expression for CD40L+IL-4±IL-5 cultures. The sorted populations are boxed and labelled. (b) MDS clustering of *in vitro* and *ex vivo* B cell and ASC populations. The grey dotted line charts an inferred B cell differentiation pathway. (c) Expression profile of the top 50 genes that are down-regulated as cells differentiate from FoB to BMPC, including *in vitro* populations. Shown are  $\log_2$ -FPKM expression values of genes. (d) Expression profile of the top 50 up-regulated genes in the same comparisons as in (c). Genes that are members of the PC signature are notated in red. (e) Venn diagrams showing overlaps and differences between genes that are significantly down-regulated in BMPC and in *in vitro* PBs, compared to FoB. (f) Venn diagrams showing overlaps and differences between genes up-regulated in BMPC and in *in vitro* PBs, compared to FoB. Complete lists of genes in each Venn category are in **Supplementary Table 2**.

**Figure 5. Gene expression changes that distinguish ASC generated under different conditions *in vitro***

(a) Venn diagrams showing overlaps and differences between genes that are significantly down-regulated in BMPC compared to FoB, and those in *in vitro* PBs from LPS (TI) or CD40L+IL-4±IL-5 (TD) stimulated populations, compared to their undifferentiated, blasting counterparts. (b) Venn diagrams showing overlaps and differences between genes that are significantly induced in BMPC compared to FoB, and those in *in vitro* PBs from LPS or CD40L+IL-4±IL-5, compared to their undifferentiated, blasting counterparts. Complete lists of genes in each Venn category are in **Supplementary Table 3**. (c) RNA-seq data illustrating six examples of genes differentially induced in LPS PBs compared to 40/4/5 PBs. FoB and BMPC are included for comparison. (d) RNA-seq data illustrating six examples of genes differentially induced in 40/4/5 PBs compared to LPS PBs. (e) Flow cytometric data to validate differential marker expression on LPS and 40/4/5 PBs. Day 4 cultures of purified B

cells from Blimp1-GFP mice were used. The analysis gates are as indicated. **(f)** Surface expression of three markers differentially expressed on LPS and 40/4/5 PBs, gated as in **(e)**. **(g)** Immunoblot validation of differential SOCS2 protein expression in day 4 bulk cultures of B cells (in LPS or CD40L+IL-4+IL-5), and in sorted populations of blasts and PBs from such cultures. Syn<sup>-</sup>, Syndecan1<sup>-</sup> PBs. The analysis was performed twice. **(h)** The proportions of Syndecan1<sup>+</sup>CD22<sup>-</sup> PBs from wild type (WT) and *Socs2*<sup>-/-</sup> mice after 4 days in the presence of LPS or CD40L+IL-4+IL-5. Data are the mean ± SD from 3 mice. Dots indicate data from individual mice. **(i)** IgM and IgG1 in serum of naive C57BL/6 and *Socs2*<sup>-/-</sup> mice determined by ELISA. Data are the mean ± SD from 9 mice. Dots indicate individual mice. P values compare WT and *Socs2*<sup>-/-</sup> using an unpaired two-tailed Students *t*-test. ns, not significant, P>0.05; asterisk, P=0.02.

### **Figure 6. Transcriptome profiling during B cell division**

**(a)** Staining and sorting strategy used on B cells activated in culture. Splenic B cells cultured for 4 days in the indicated conditions were sorted, on the basis of CFSE dilution (Divisions 1, 3, 5, 7) and Syndecan1 expression (PB) and subjected to RNA-seq. A sample of cells was processed after cell preparation (unstimulated, ‘Unstim’ in the figure) and after 1 day (undivided cells, ‘Div 0’, for division 0 in the figure). **(b)** MDS clustering of RNA-seq data from cells isolated as described in **(a)**. The grey dotted line charts an inferred B cell differentiation pathway. **(c)** Distribution of gene expression changes at different points in the differentiation process. The number of genes with expression changes of >2.5-fold is shown on each graph. Left panel, the “All stages” include samples Unstim., Div 0, Div 1, Div 3, Div 5, Div 7 and 40/4 PB. The maximum fold change between the highest and lowest expression level in any sample was calculated for each gene and used for density plotting. Right panel (top), the “Activation” stage includes samples Unstim. and Div 0. Absolute log<sub>2</sub>-fold changes

of genes were calculated from comparing the two and used for density plotting. Right panel (middle), the “Division” stage includes samples Div 0, Div 1, Div 3, Div 5 and Div 7. The maximum fold change between the highest and lowest expression level in any of these five samples was calculated for each gene. Right panel (bottom), the “Differentiation” stage includes samples Div 7 and 40/4 PB. Absolute log<sub>2</sub>-fold changes of genes were calculated from comparing the two samples. **(d)** Numbers of genes that were found to have > 2.5-fold change during each division transition. **(e)** Patterns of gene expression changes during cell division. The *k*-mean clustering algorithm was used to classify the genes that have > 2.5-fold changes at the division stage (2095 genes), into nine clusters by using their step-wise expression changes from Div 0 to Div 7. These nine clusters were then further grouped into six distinct patterns. Mean and standard deviation (error bar) are plotted for each division number in each cluster. Numbers of genes included in each cluster are shown at the top left corner of each graph. **(f)** Expression profiles of top 20 up-regulated genes (left panel) and top 20 down-regulated genes (right panel) in cell division data. Genes were selected from comparing Div 0 and Div 7. Log<sub>2</sub>-FPKM expression values of genes are shown in the heatmaps, color-coded according to the legend. Genes that are members of the ASC signature are notated in bold italics. **(g)** Examples of transcription factors that progressively increase (top panel) or decrease (bottom panel) in expression levels during cell division.

### **Figure 7. Cell-type specific changes in transcription factor gene expression**

**(a-e)** RNA-seq data for the indicated *ex vivo* B cell and ASC populations, and selected for genes that show >2.5 fold differential expression between FoB and BMPC, have an expression abundance of at least 32 FPKMs in 1 or more samples and are present on a curated list of factors involved in the regulation of gene expression<sup>45</sup>. **(a)** Expression profile of the 103 differentially expressed transcriptional regulators. Log<sub>2</sub>-FPKM expression values

of genes are shown in the heatmaps, color-coded according to the legend. Genes that are also members of the ASC signature are shown in bold italics. **(b)** Expression data for genes that mimic the pan-B cell restricted pattern displayed by *Pax5*. **(c)** Genes that are expressed in a GC-specific manner. **(d)** Genes that display marginal zone or B1 cell-specific expression. **(e)** Genes whose expression is ASC restricted, reflecting the expression of essential regulators of ASC differentiation, such as *Prdm1*, *Irf4* and *Xbp1*. See **Supplementary Fig. 5** and **6** for an expanded set of genes with such cell-type restricted expression patterns.

## **Online METHODS**

### **Mice**

Blimp1-GFP reporter <sup>21</sup> and *Socs2*<sup>-/-</sup> mice <sup>27</sup> (a kind gift from Dr W Alexander, The Walter and Eliza Hall Institute) have been previously described. All mice were maintained on a C57Bl/6 genetic background and housed in a clean conventional mouse facility. Animal experiments were conducted according to the protocols approved by the Walter and Eliza Hall Institute animal ethics committee.

### **Flow cytometry, antibodies and ELISA**

Cells were analysed on FACSCanto (BD Biosciences) and cell sorting was carried out using FACS Aria (BD Biosciences) or MoFlo (Beckman Coulter Inc.) flow cytometers. Data were analyzed using FlowJo software. Antibodies and reagents used for flow cytometry and immunoblotting were: Anti-CD138/Syndecan1 (clone 281-2); Anti-CD21 (clone 7G6); Anti-CD23 (clone B3B4); Anti-B220 (clone RA3-6B2), Anti-CD95/fas (clone Jo2), Anti-Sca1/Ly6A/E, PE conjugated (clone E13-161.7), Anti-Ly6D, Alexa Fluor 647 conjugated (clone 49-H4), all from BD Biosciences. Anti-CD19 (clone 1D3), eBioscience; PNA, Vector Labs; Anti-CCR10, APC conjugated (#FAB2815A) and Anti-CXCR3, APC conjugated (#FAB1685A), both from R+D Systems; Anti-EpCam/CD326, Alexa Fluor 647 conjugated (clone G8.8) and Anti-CD22, APC conjugated (clone OX-97) both from BioLegend; Anti-TSPAN15, Bioss Antibodies, PE conjugated (#bs-9425R). Serum and supernatant Ig amounts were measured using ELISA as described previously <sup>28</sup>.

### **Preparation of sorted cell populations: *ex vivo* populations.**

Cells were flow cytometrically sorted from lymphoid tissues of adult C57Bl/6 or Blimp1-GFP reporter mice. For most populations, at least two biological replicates were prepared.

For PCs, which are rare *in vivo*, cells from spleens or BM from 3 Blimp1-GFP reporter mice were pooled and Syndecan1<sup>+</sup> cells were magnetically enriched (CD138 MicroBeads, Miltenyi Biotec) before sorting. This was repeated four times. For B1 cells, peritoneal lavages from 3-4 mice were pooled, stained and sorted. Two independent pools were prepared.

**Preparation of sorted cell populations: *in vitro* stimulated cells.**

In each experiment, culture times were chosen to maximise the yields of undifferentiated or differentiated cells of interest. CD40L+IL-4 generates ASC slowly with division, allowing the staged progress to be monitored, as in **Fig. 6**. IL-5 acts selectively to increase the likelihood of differentiation without affecting other cell responses (switching, survival, division), thus enabling greater numbers of ASC to be generated earlier in cultures<sup>28</sup>. For *in vitro* B cell cultures, resting splenic B cells (Negative B cell isolation kit: Miltenyi Biotec) were cultured in medium containing CD40L and IL-4 (1 U/ml) as described<sup>28</sup> for 48 hours. Subsequently, activated CD19<sup>+</sup> B cells that lacked Syndecan1 expression were sorted and total RNA extracted (40/4 blasts). This time point was chosen to examine early transcriptional changes, as minimal differentiation has occurred, with all cells remaining B220<sup>hi</sup>Syndecan1<sup>-</sup> (>99.5%). For *in vitro* PB generation, resting B cells were cultured in medium containing CD40L, IL-4 and IL-5 (5 ng/ml R&D Systems) for 5 days. Total RNA was isolated from sorted Syndecan1<sup>+</sup> PB (40/4/5 PB, Syndecan1<sup>+</sup>). Alternatively, splenic B cells were purified from Blimp1-GFP mice using anti-B220 magnetic beads and cultured for the indicated times in the presence of 10µg/ml lipopolysaccharide (LPS). After 4 days, three populations were isolated: undifferentiated Syndecan1<sup>-</sup>Blimp1-GFP<sup>-</sup> B cell blasts (LPS blasts), Syndecan1<sup>-</sup>Blimp1-GFP<sup>+</sup> PB (LPS Syn<sup>-</sup>PB) and Syndecan1<sup>+</sup>Blimp1-GFP<sup>+</sup> PB (LPS PB).

Finally, small resting B cells from C57Bl/6 mice, purified with magnetic beads followed by density gradients<sup>28</sup>, were labelled with CFSE and placed in cultured with CD40L and IL-4 as described above. After four days, cells were sorted using CFSE dilution into groups that had divided 1, 3, 5 and 7 times, but had not yet differentiated into Syndecan1<sup>+</sup> ASCs. Differentiated Syndecan1<sup>+</sup> ASCs (40/4 PB, Syndecan1<sup>+</sup>) were sorted separately. Again, all populations were independently sorted and two independent pools were prepared for each population.

### **RNA sequencing and data analysis**

RNA was isolated from cells using Qiagen RNeasy Micro or Mini Kits (based on cell number), according to the manufacturer's instructions. RNA-seq was performed using an Illumina HiSeq 2000 or a HiSeq 2500 instrument at the Australian Genome Research Facility (AGRF), Melbourne, Australia or Beijing Genomics Institute (BGI), Shenzhen, China.

Reads were aligned to the GRCm37/mm9 build of the *Mus musculus* genome using the Subread aligner<sup>16</sup>. Ten 'subreads' were extracted from each read and at least three consensus 'subreads' were required for a 'hit' to be reported. Hamming distance or mapping quality score was used to break the tie when more than one best location was found for a read. Only uniquely mapped reads were retained. Genewise counts were obtained using featureCounts<sup>17</sup>. Reads overlapping exons in annotation build 37.2 of NCBI RefSeq database were included. Summary of read mapping and quantification results can be found in **Supplementary Table 4**. Genes were filtered from downstream analysis if they failed to achieve a CPM (counts per million mapped reads) value of 1 or greater in at least one library. Ig genes were excluded from expression analysis when comparing B cell populations with ASC populations, due to

the large difference in fraction of Ig reads between them. Expression analysis for Ig genes was performed separately.

Counts were converted to log<sub>2</sub> counts per million, quantile normalized and precision weighted with the voom function of the limma package<sup>46,47</sup>. A linear model was fitted to each gene, and empirical Bayes moderated t-statistics were used to assess differences in expression<sup>48</sup>. Empirical Bayes moderated-t p-values were computed relative to a fold-change cut-off of 1.5-fold by using treat<sup>49</sup>. P-values were adjusted to control the global false-discovery rate across all comparisons with the ‘global’ option of the limma package. Genes were called differentially expressed if they achieved a false discovery rate of 0.05 or less and also had at least 8 FPKMs in one or both of the two cell types being compared. Heat maps were generated using the gplots package, with negative log<sub>2</sub> FPKM values reset to zero.

### **Mapping of immunoglobulin reads**

The alignment of reads originating from Ig genes was carried out using the Subjunc aligner, which has been found to have an excellent performance in mapping reads spanning multiple short exons<sup>16</sup>. We configured Subjunc to make it be able to detect long insertions and deletions occurring at the start and end of Ig segments, due to imprecise recombination and N nucleotide insertions. Also, we did not require donor/receptor sites to be present when mapping reads that span multiple segments, since VDJ recombination takes place at DNA level.

### **ChIP sequencing and data analysis**

For ChIP-seq samples were prepared according to the Illumina TruSeq ChIP sample preparation protocol. Antibodies used to immunoprecipitate chromatin were anti-H3K4me1

and H3K4me3 both from Abcam. Lymphoma cell line A20 (ATCC<sup>®</sup> TIB-208<sup>™</sup>) and the plasmacytoma MPC11 (ATCC<sup>®</sup> CCL-167<sup>™</sup>) were sourced from ATCC.

Reads were mapped to GRCm37/mm9 build of the *Mus musculus* genome in the same way as that for the mapping of RNA-seq reads. To quantify the degree of histone tail modification in the vicinity of a given gene, we defined a promoter region that included a 3kb upstream sequence and a 2kb downstream sequence from the transcriptional start of the gene. NCBI mouse RefSeq gene annotation was used. Mapped reads were assigned to the promoter region of each gene using the featureCounts program. FPKM values were calculated for each gene based on read counts, size of promoter regions (5kb) and library size.

### **Identification of absolutely expressed genes in RNA-seq data**

To identify absolutely expressed genes, we performed a Poisson test for each gene to test if its expression level was greater than the background noise measured from the intergenic regions of the genome. We divided the intergenic regions of the mouse genome into bins of 2kb and counted the number of mapped reads falling into each bin. The middle position of the read was used to assign the read to a bin. We also calculated the percentage of G and C bases in each bin. We then applied the lowess regression to read counts from each bin to derive a relationship between the background read count and the GC percentage. A GC percentage was also calculated for each gene using all nucleotide bases in the exonic regions of the gene. For each gene, the number of reads mapped to it and the expected background read count for an intergenic region of the same length and GC percentage same as that of the gene were input to the Poisson test to call absolutely expressed genes. A p-value cutoff of 0.05 was applied. NCBI RefSeq annotation for mouse genome mm9 (build 37.2) was used.

### **Gene set test for super-enhancer associated genes**

There were 970 super-enhancer associated RefSeq transcripts described from human blood B cells<sup>32</sup>. We converted the names of these RefSeq transcripts to official gene symbols and then matched them with the gene symbols used in this analysis. We further filtered out those genes that had very low read counts. We were left with 756 genes to be used for the gene set enrichment analysis performed in this study. The gene set enrichment plots were generated using the barcodeplot function in the limma package. Statistical tests of enrichment were carried out using the roast method in limma with 999 rotations<sup>50</sup>. One-sided p-values are reported.

### **Gene functional classifications**

For the biological function classifications shown in **Figures 1c** and **2d**, differentially expressed genes were manually curated using a number of public gene databases including gene ontology as well as published literature.

## REFERENCES

1. Nutt, S.L., Hodgkin, P.D., Tarlinton, D.M. & Corcoran, L.M. The generation of antibody-secreting plasma cells. *Nat Rev Immunol* **in press** (2014).
2. Fujio, K., Okamura, T., Sumitomo, S. & Yamamoto, K. Regulatory cell subsets in the control of autoantibody production related to systemic autoimmunity. *Ann Rheum Dis* **72 Suppl 2**, ii85-9 (2013).
3. Potter, M. Neoplastic development in plasma cells. *Immunol Rev* **194**, 177-95 (2003).
4. Fairfax, K.A. *et al.* Different kinetics of blimp-1 induction in B cell subsets revealed by reporter gene. *J Immunol* **178**, 4104-11 (2007).
5. Radbruch, A. *et al.* Competence and competition: the challenge of becoming a long-lived plasma cell. *Nat Rev Immunol* **6**, 741-50 (2006).
6. Tarlinton, D. & Good-Jacobson, K. Diversity among memory B cells: origin, consequences, and utility. *Science* **341**, 1205-11 (2013).
7. Nutt, S.L., Taubenheim, N., Hasbold, J., Corcoran, L.M. & Hodgkin, P.D. The genetic network controlling plasma cell differentiation. *Semin Immunol* **23**, 341-9 (2011).
8. Shapiro-Shelef, M. & Calame, K. Regulation of plasma-cell development. *Nat. Rev. Immunol* **5**, 230-42 (2005).
9. Shaffer, A.L. *et al.* Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* **17**, 51-62 (2002).
10. Shaffer, A.L. *et al.* XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. *Immunity* **21**, 81-93 (2004).
11. Cocco, M. *et al.* In vitro generation of long-lived human plasma cells. *J Immunol* **189**, 5773-85 (2012).
12. Jourdan, M. *et al.* An in vitro model of differentiation of memory B cells into plasmablasts and plasma cells including detailed phenotypic and molecular characterization. *Blood* **114**, 5173-81 (2009).
13. Peperzak, V. *et al.* Mcl-1 is essential for the survival of plasma cells. *Nat Immunol* **14**, 290-7 (2013).
14. Sciammas, R. *et al.* Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation. *Immunity* **25**, 225-36 (2006).
15. Kallies, A. *et al.* Plasma cell ontogeny defined by quantitative changes in blimp-1 expression. *J Exp Med* **200**, 967-77 (2004).
16. Liao, Y., Smyth, G.K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* **41**, e108 (2013).
17. Liao, Y., Smyth, G.K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-30 (2014).
18. O'Connor, B.P. *et al.* BCMA is essential for the survival of long-lived bone marrow plasma cells. *J Exp Med* **199**, 91-8 (2004).
19. Delogu, A. *et al.* Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells. *Immunity* **24**, 269-81 (2006).

20. Wrammert, J., Kallberg, E., Agace, W.W. & Leanderson, T. Ly6C expression differentiates plasma cells from other B cell subsets in mice. *Eur J Immunol* **32**, 97-103 (2002).
21. Nakayama, T. *et al.* Cutting edge: profile of chemokine receptor expression on human plasma cells accounts for their efficient recruitment to target tissues. *J Immunol* **170**, 1136-40 (2003).
22. Hauser, A.E. *et al.* Chemotactic responsiveness toward ligands for CXCR3 and CXCR4 is regulated on plasma blasts during the time course of a memory immune response. *J Immunol* **169**, 1277-82 (2002).
23. Kunkel, E.J. *et al.* CCR10 expression is a common feature of circulating and mucosal epithelial tissue IgA Ab-secreting cells. *J Clin Invest* **111**, 1001-10 (2003).
24. Cerutti, A., Cols, M. & Puga, I. Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes. *Nat Rev Immunol* **13**, 118-32 (2013).
25. Jeong, H.D. & Teale, J.M. Contribution of the CD5<sup>+</sup> B cell to D-proximal VH family expression early in ontogeny. *J Immunol* **145**, 2725-9 (1990).
26. Herzenberg, L.A., Baumgarth, N. & Wilshire, J.A. B-1 cell origins and VH repertoire determination. *Curr Top Microbiol Immunol* **252**, 3-13 (2000).
27. Metcalf, D. *et al.* Gigantism in mice lacking suppressor of cytokine signalling-2. *Nature* **405**, 1069-73 (2000).
28. Hasbold, J., Corcoran, L.M., Tarlinton, D.M., Tangye, S.G. & Hodgkin, P.D. Evidence from the generation of immunoglobulin G-secreting cells that stochastic mechanisms regulate lymphocyte differentiation. *Nat Immunol* **5**, 55-63 (2004).
29. Hodgkin, P.D., Lee, J.H. & Lyons, A.B. B cell differentiation and isotype switching is related to division cycle number. *J Exp Med* **184**, 277-81 (1996).
30. Rush, J.S., Liu, M., Odegard, V.H., Unniraman, S. & Schatz, D.G. Expression of activation-induced cytidine deaminase is regulated by cell division, providing a mechanistic basis for division-linked class switch recombination. *Proc Natl Acad Sci U S A* **102**, 13242-7 (2005).
31. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-89 (2010).
32. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
33. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-19 (2013).
34. Loven, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-34 (2013).
35. Chu, V.T. & Berek, C. The establishment of the plasma cell survival niche in the bone marrow. *Immunol Rev* **251**, 177-88 (2013).
36. Kurtz, J.E. & Dufour, P. Adecatumumab: an anti-EpCAM monoclonal antibody, from the bench to the bedside. *Expert Opin Biol Ther* **10**, 951-8 (2010).
37. Lu, R. & Wang, G.G. Tudor: a versatile family of histone methylation 'readers'. *Trends Biochem Sci* **38**, 546-55 (2013).
38. Xiao, W. *et al.* U19/Eaf2 knockout causes lung adenocarcinoma, B-cell lymphoma, hepatocellular carcinoma and prostatic intraepithelial neoplasia. *Oncogene* **27**, 1536-44 (2008).
39. Korthuis, P.M. *et al.* CITED2-mediated human hematopoietic stem cell maintenance is critical for acute myeloid leukemia. *Leukemia* (2014).

40. Knosp, C.A. *et al.* SOCS2 regulates T helper type 2 differentiation and the generation of type 2 allergic responses. *J Exp Med* **208**, 1523-31 (2011).
41. Zak, D.E., Tam, V.C. & Aderem, A. Systems-level analysis of innate immunity. *Annu Rev Immunol* **32**, 547-77 (2014).
42. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat Genet* **37**, 382-90 (2005).
43. Yosef, N. *et al.* Dynamic regulatory network controlling TH17 cell differentiation. *Nature* **496**, 461-8 (2013).
44. Perfetti, V. *et al.* Membrane CD22 defines circulating myeloma-related cells as mature or later B cells. *Lab Invest* **77**, 333-44 (1997).
45. Zhang, J.A., Mortazavi, A., Williams, B.A., Wold, B.J. & Rothenberg, E.V. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* **149**, 467-82 (2012).
46. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, doi: 10.1093/nar/gkv007 (2015).
47. Law, C.W., Chen, Y., Shi, W. & Smyth, G.K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29 (2014).
48. Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
49. McCarthy, D.J. & Smyth, G.K. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* **25**, 765-71 (2009).
50. Wu, D. *et al.* ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* **26**, 2176-82 (2010).

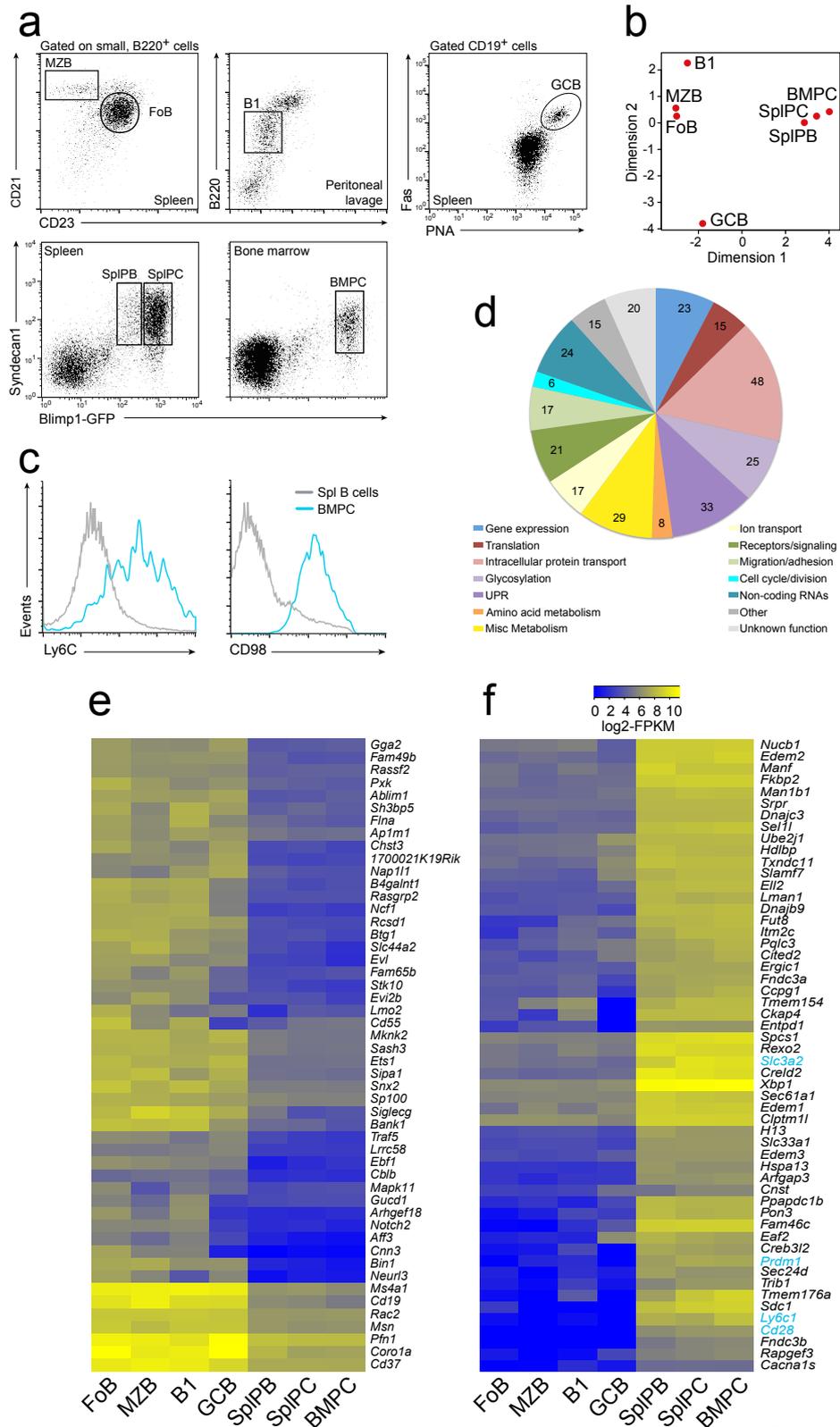


Figure 1. Shi et al.

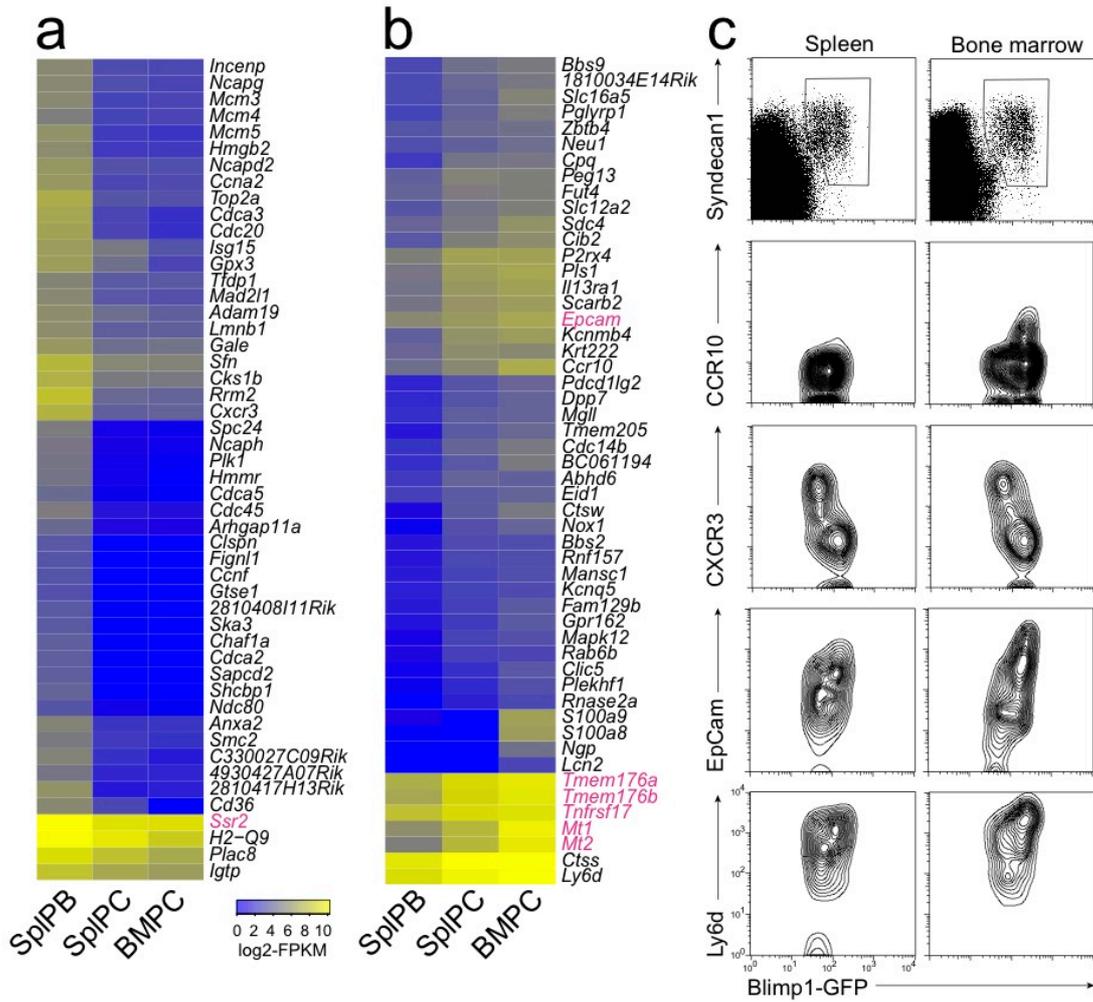


Figure 2. Shi et al.

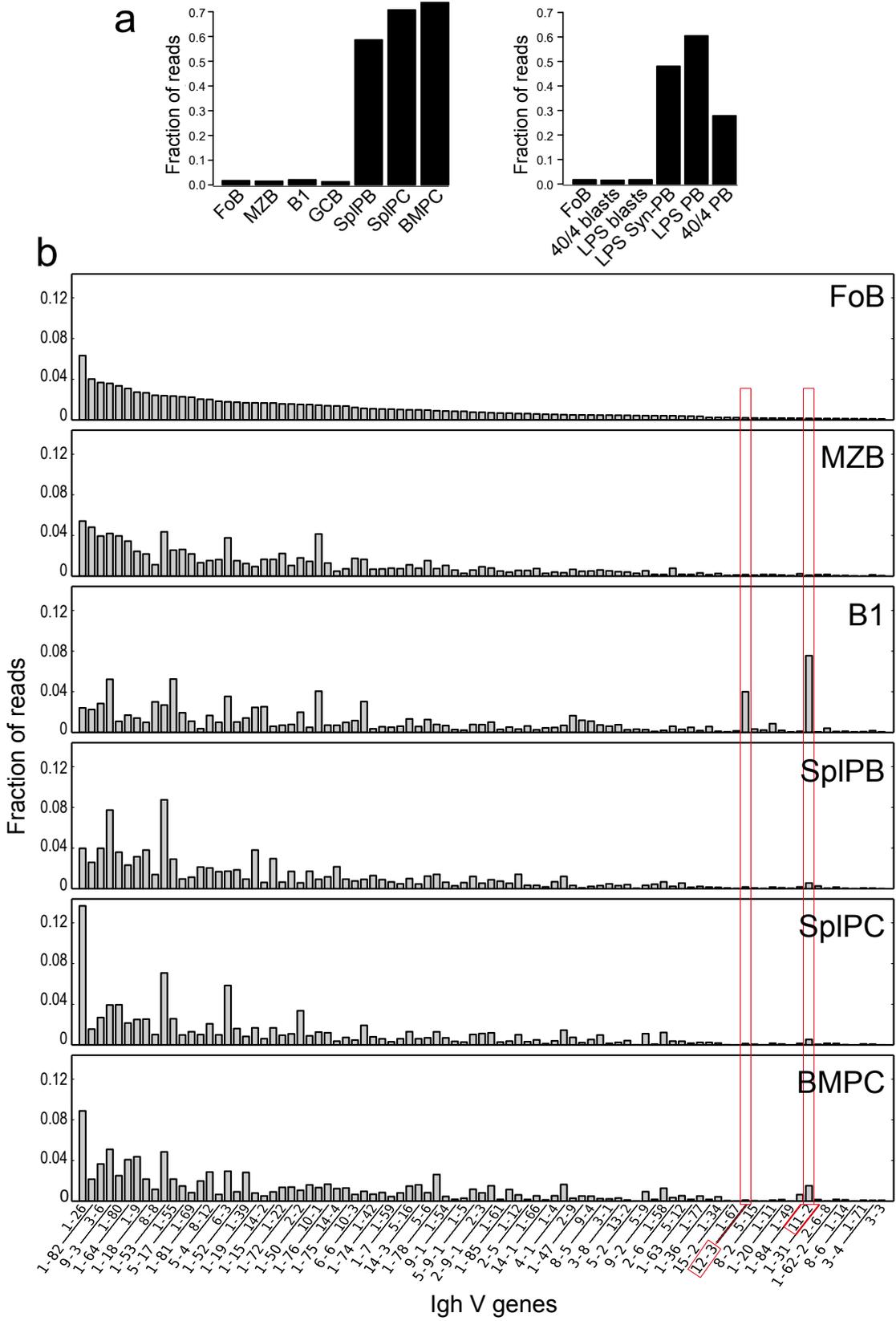


Figure 3. Shi et al.

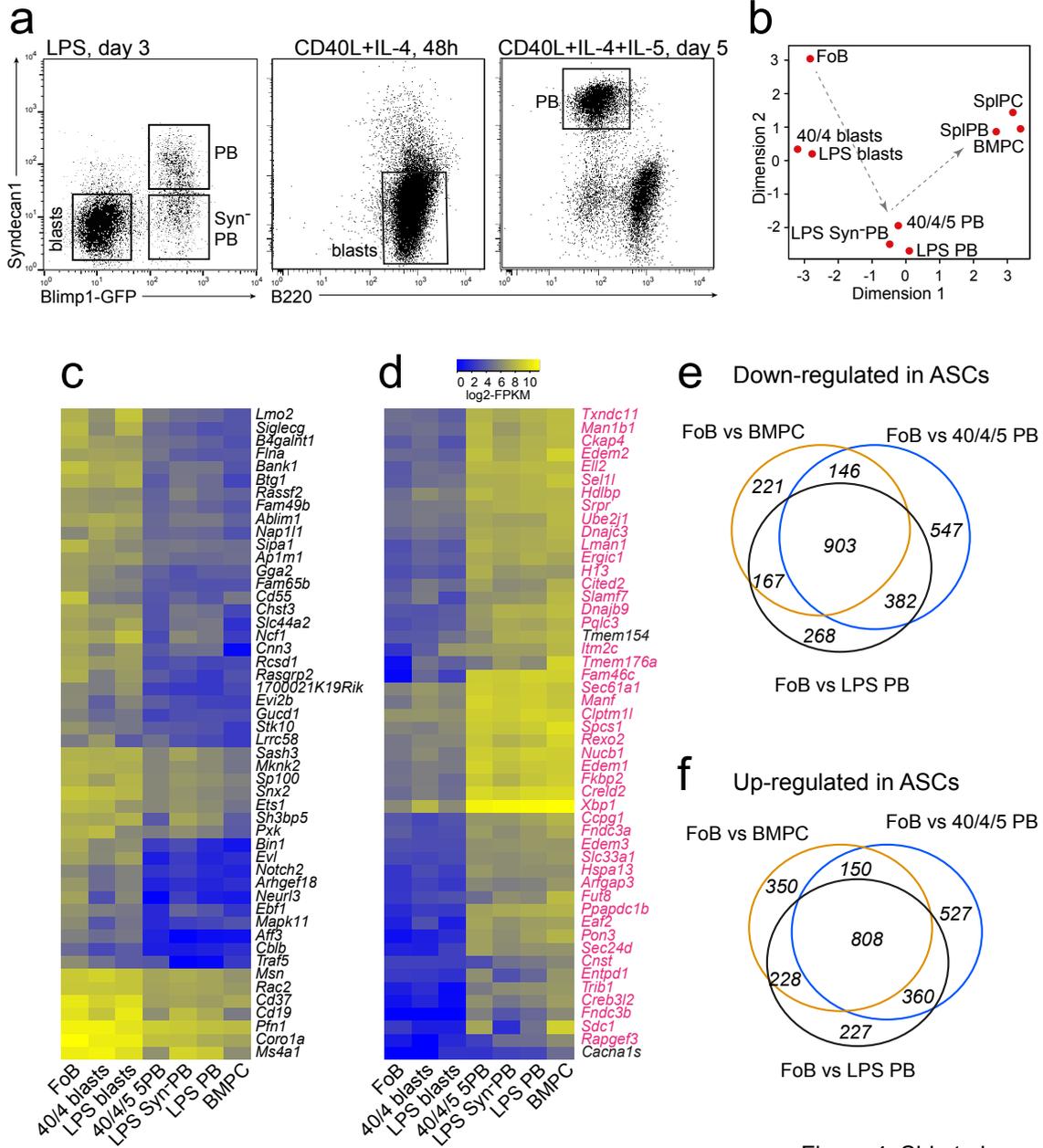


Figure 4. Shi et al.

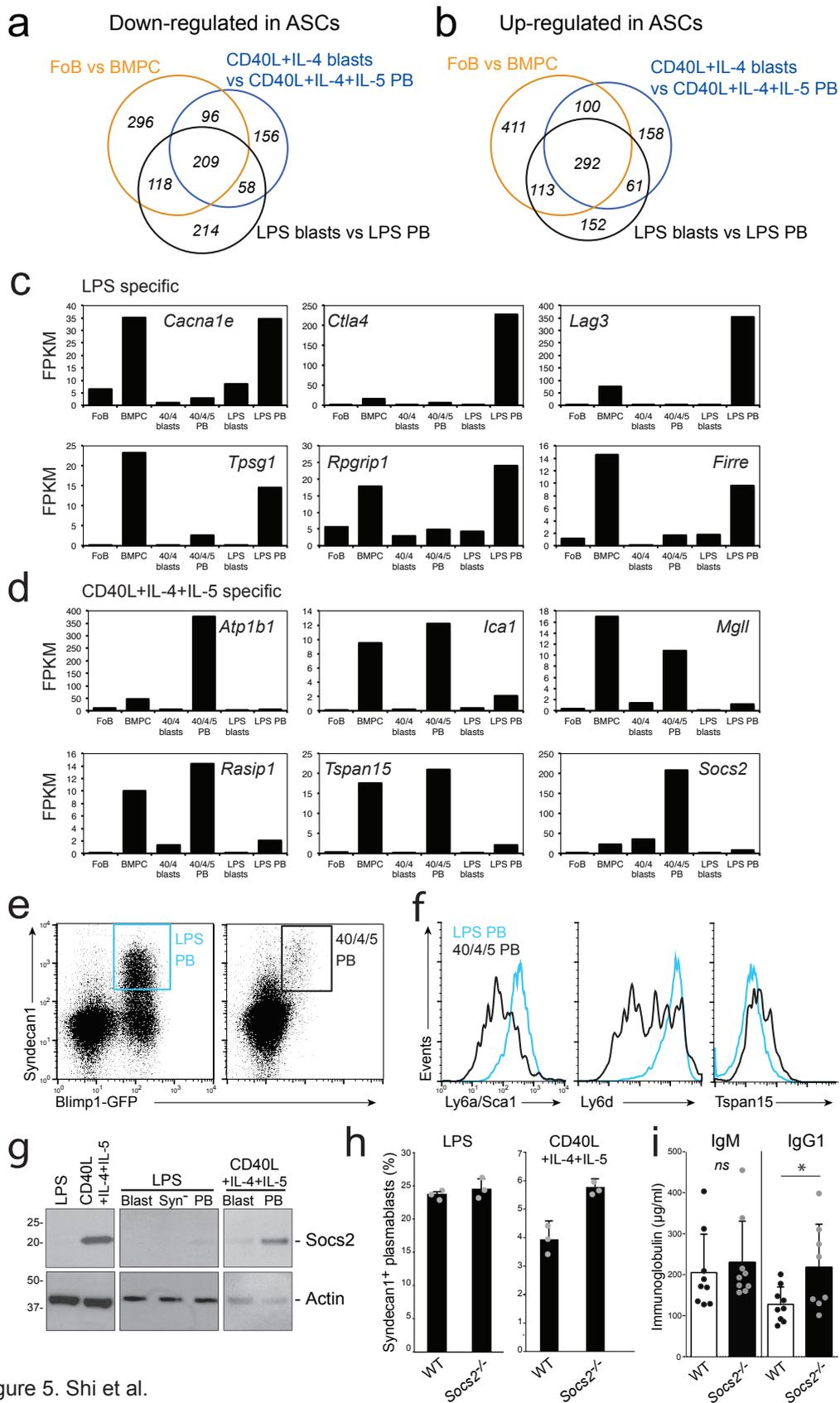


Figure 5. Shi et al.

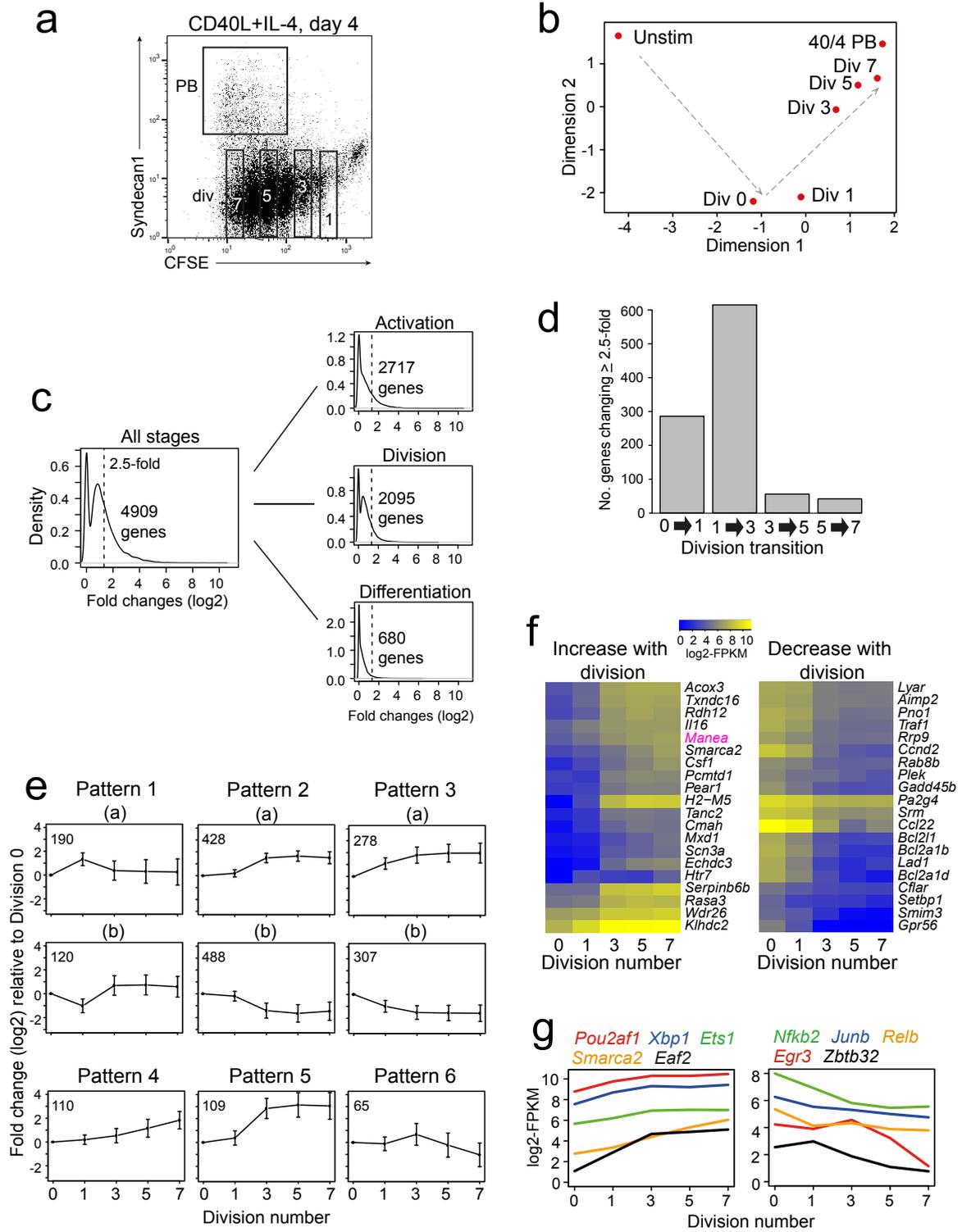


Figure 6. Shi et al.

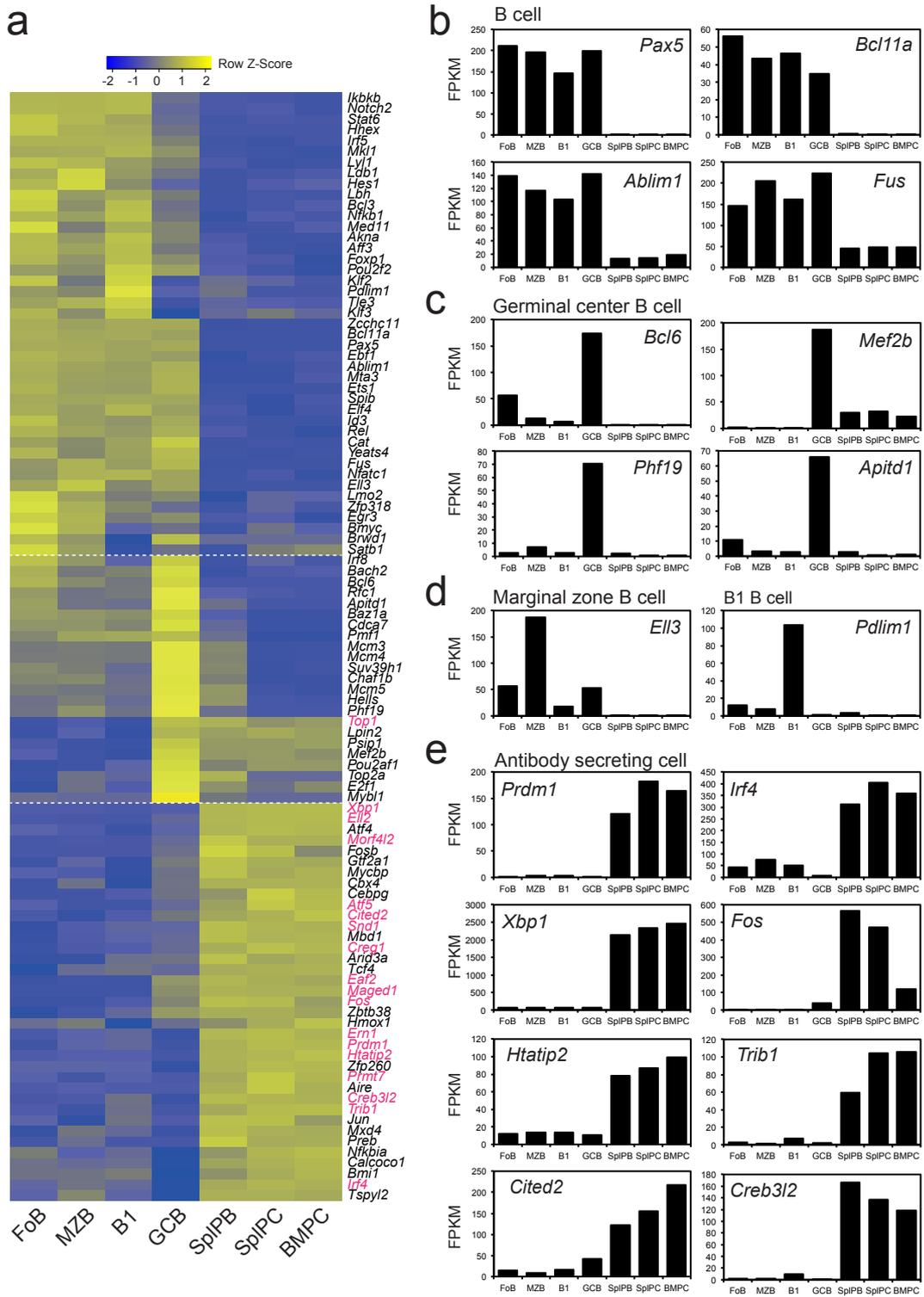


Figure 7. Shi et al.