RESOURCE

nature immunology

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

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

When B cells encounter an antigen, they alter their physiological state and 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 in mice. 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 on the basis of location and maturity. Changes in gene expression correlated with cell-division history and the acquisition of permissive histone modifications, and they included many regulators that had not been 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 antibodies.

Antibodies are indispensable effector molecules of the immune system that provide both immediate and longer-term protection against infection, a facet exploited by most currently used vaccines. Antibody-secreting cells (ASCs) are a rare population of highly specialized cells representing the end stage of B cell–lineage differentiation[1]. 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 in specialized niches in the bone marrow. Although B cells and antibodies have essential functions in mediating humoral immunity, pathogenic antibodies are also key drivers of some autoimmune diseases such as systemic lupus erythromatosus[2], and multiple myeloma and plasmacytoma result from the malignant transformation of ASCs[3]. Thus understanding the factors that control the production, maturation and long-term survival of ASCs is critical both for improved vaccine design and to provide mechanisms to target pathogenic ASCs.

Mature B cells are categorized into three distinct subsets, follicular B cells (FoBs), marginal-zone B cells (MZBs) and B1 cells, all of which contribute to the ASC pool and circulating serum antibody, although with distinct affinities and kinetics[4]. MZBs and B1 cells are specialized types of B cells that function largely independently of T cells. MZBs flank the marginal sinuses of the spleen, where they respond to blood-borne antigens, whereas B1 cells localize to the peritoneal and pleural cavities, where they provide an early line of defense against pathogens that enter through mucosal surfaces. FoBs, the numerically dominant subset of mature B cells, localize to lymphoid follicles in the spleen and lymph nodes. After activation by a protein antigen and in collaboration with cognate helper T cells, FoBs proliferate extensively, undergo immunoglobulin class-switch recombination and either generate an extrafollicular PB response that is both rapid and transient[1] or enter into a germinal center (GC) reaction, where they undergo somatic hypermutation and selection for high-affinity antigen receptors, again under the direction of specialized CD4$^+$ T cells. The GC provides the majority of the long-term immunity, through the production of long-lived PCs and memory B cells[5,6].

On a transcriptional level, the differentiation of activated B cells into ASCs requires coordinated changes in the expression of many hundreds of genes. These changes fall 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 of the B cell–promoting transcription factors PAX5, BCL6 and BACH2 in maintaining the fate of B cells, whereas a different triad of factors—BLIMP1, IRF4 and XBP1—are required to extinguish the B cell genes and activate the ASC program[7,8]. 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 from, for example, antigens and T cell–derived 'helpers' such as CD40L and cytokines modulate this process is poorly understood.

Although a number of studies have used microarray technologies to catalog the transcriptomes of various stages of late B cell differentiation

[1]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. [2]Department of Computing and Information Systems, The University of Melbourne, Parkville, Victoria, Australia. [3]Department of Medical Biology, The University of Melbourne, Parkville, Victoria, Australia. [4]Department of Pathology, The University of Melbourne, Parkville, Victoria, Australia. [5]Department of Microbiology & Immunology, The University of Melbourne, Parkville, Victoria, Australia. [6]Department of Mathematics and Statistics, The University of Melbourne, Parkville, Victoria, Australia. Correspondence should be addressed to W.S. (shi@wehi.edu.au), S.L.N. (nutt@wehi.edu.au) or L.M.C. (corcoran@wehi.edu.au).

Received 21 September 2014; accepted 24 March 2015; published online 20 April 2015; doi:10.1038/ni.3154

in both mice and humans[9–14], so far there has been no comprehensive, single-platform, transcriptional analysis of all aspects of the process of differentiation from B cell to PC. Here we used RNA sequencing (RNA-seq) technology and the Blimp1-GFP reporter mouse strain, which allows for the identification of all ASCs *in vivo* and *in vitro*[15], to provide a comprehensive description of the dynamic nature of B cell terminal differentiation. These data enabled 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 and ASC maturation affect this essential biological process. The analysis also identified many new potential regulators of B cell and ASC differentiation and specialization.

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 that genetic background that carried a GFP reporter allele within the *Prdm1* (the gene encoding the transcriptional regulator Blimp1) locus[15]. Splenic FoBs (small, B220$^+$CD21$^+$CD23$^+$), MZBs (B220$^+$CD21^{hi}CD23$^-$), PBs (SplPBs; Syndecan-1$^+$Blimp1-GFP^{lo}) and PCs (SplPCs; Syndecan-1$^+$Blimp1-GFP^{hi}) were sorted; B1 cells (B220$^+$Mac1^{lo}CD23$^-$) were sorted from peritoneal lavages; and plasma cells (Syndecan-1$^+$Blimp1-GFP^{hi}) were sorted from bone marrow (BMPCs) (**Fig. 1a**). GC B cells (CD19$^+$Fas$^+$PNA$^+$) were induced by immunization with sheep red blood cells. 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[16], and the number of reads or read pairs mapping to the exonic regions of each gene was recorded[17]. 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, and it might be hypothesized that this could limit the repertoire of other genes expressed by ASCs. By comparing the read coverage of annotated exons with the 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 noise 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.

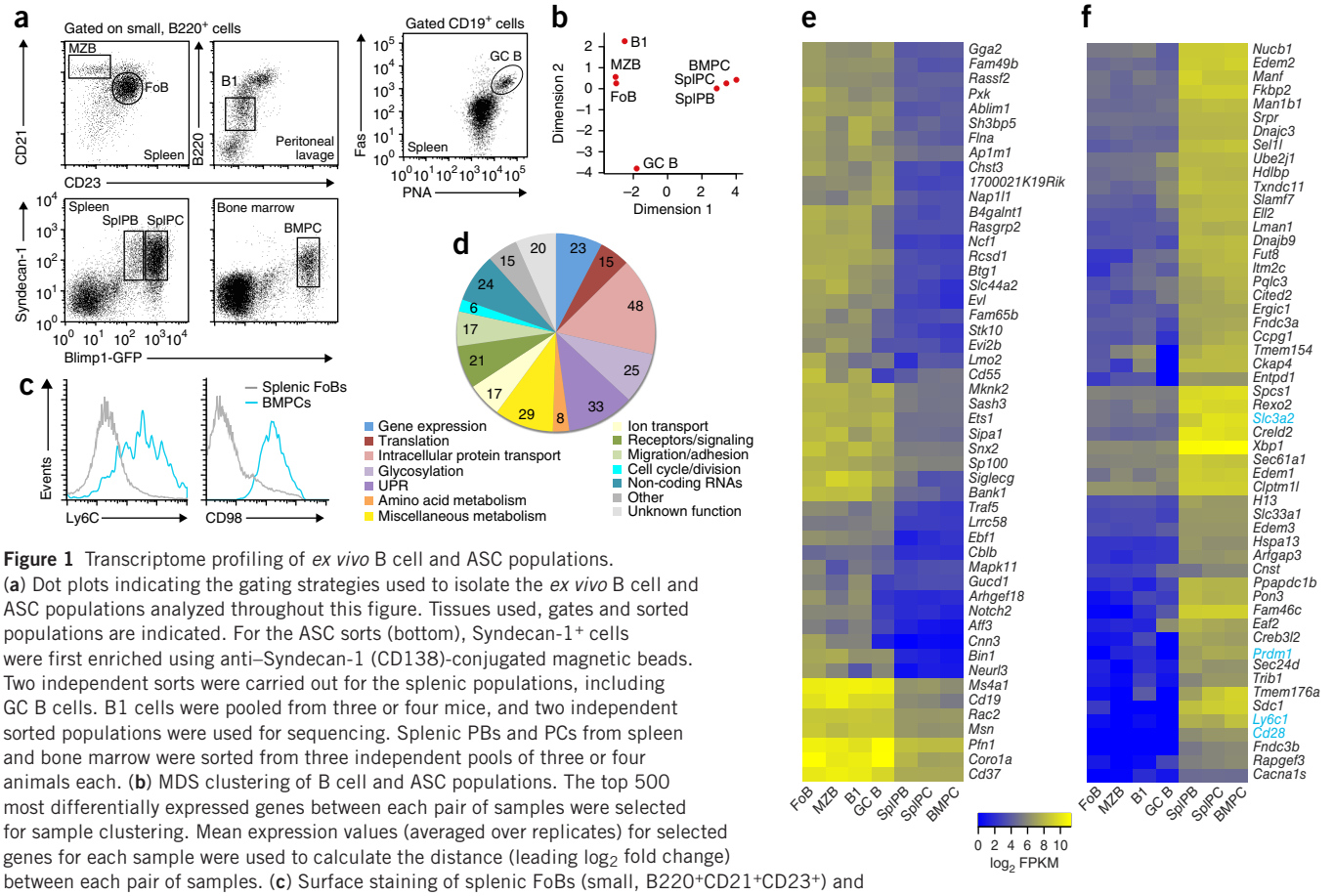

Figure 1 Transcriptome profiling of *ex vivo* B cell and ASC populations. (**a**) Dot plots indicating the gating strategies used to isolate the *ex vivo* B cell and ASC populations analyzed throughout this figure. Tissues used, gates and sorted populations are indicated. For the ASC sorts (bottom), Syndecan-1$^+$ cells were first enriched using anti–Syndecan-1 (CD138)-conjugated magnetic beads. Two independent sorts were carried out for the splenic populations, including GC B cells. B1 cells were pooled from three or four mice, and two independent sorted populations were used for sequencing. Splenic PBs and PCs from spleen and bone marrow were sorted from three independent pools of three or four animals each. (**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 $\log_2$ fold change) between each pair of samples. (**c**) Surface staining of splenic FoBs (small, B220$^+$CD21$^+$CD23$^+$) and bone marrow plasma cells (BMPCs; Syndecan-1$^+$Blimp1-GFP^{hi}) to confirm differential expression of two surface markers (Ly6C and CD98) that are part of the ASC signature. Data are representative of five independent experiments. (**d**) Functional analysis of genes in the ASC signature. The numbers of genes in each functional category are shown. UPR, unfolded protein response. (**e**) Expression profile of the top 50 downregulated genes in BMPCs compared with FoBs. (**f**) Expression profile of the top 50 upregulated genes in BMPCs compared with FoBs, plus 4 additional genes of particular immunological interest (blue lettering). Log_2 FPKM expression values of genes are shown in the heat maps, color-coded according to the legend.

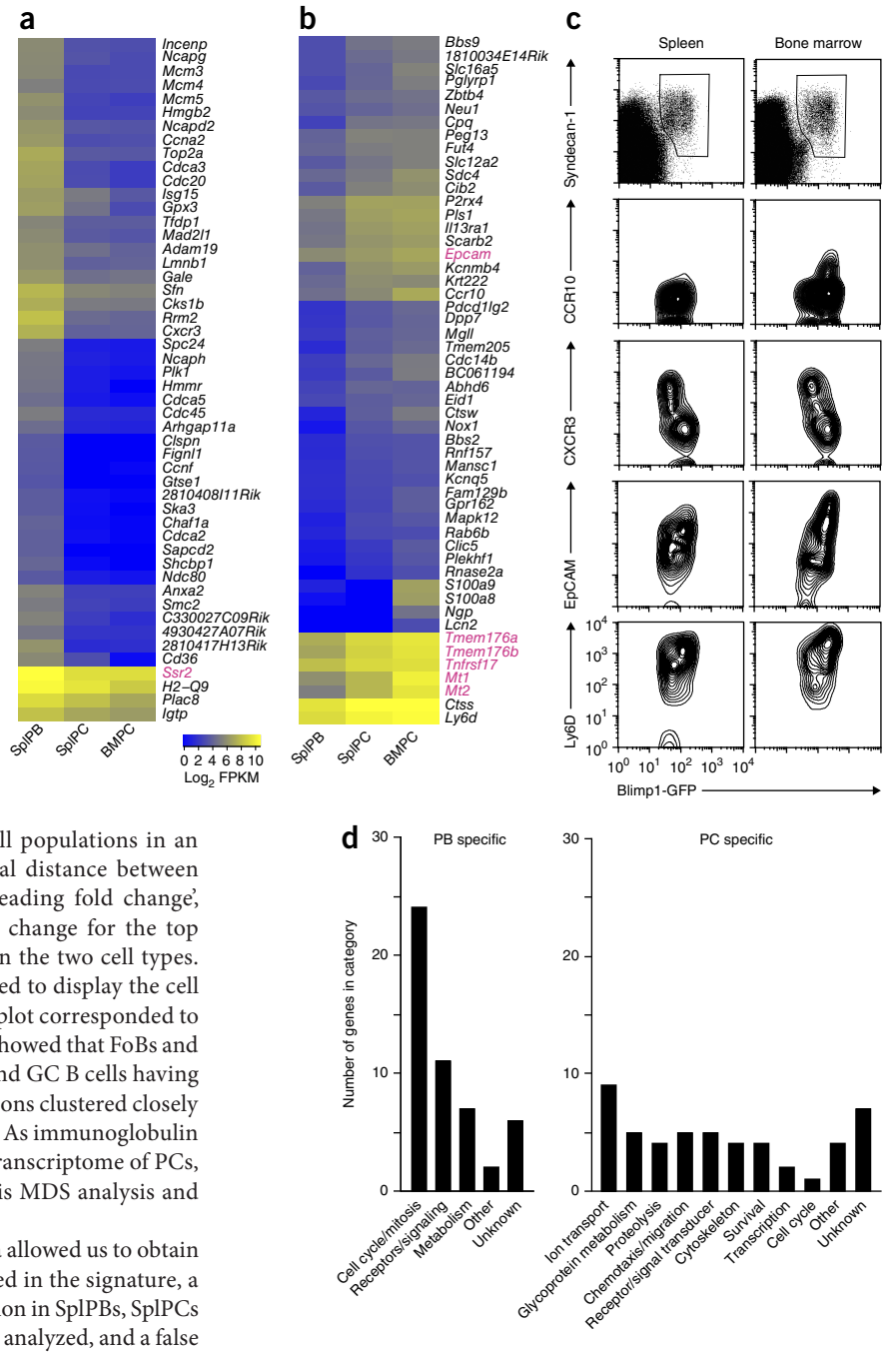

Figure 2 Comparison of *ex vivo* ASC populations. Expression profiles of the top ~100 genes most differentially expressed in BMPCs compared with SplPBs (Syndecan-1$^+$Blimp1-GFP^{lo}) and SplPCs (Syndecan-1$^+$Blimp1-GFP^{hi}). (**a**) Heat map of the 50 most strongly downregulated genes, shown for the three indicated *ex vivo* ASC populations. Genes with the smallest FDRs in all the downregulated genes were selected. Log_2 FPKM expression values of genes are represented, as defined in the legend. (**b**) Heat map of the 50 most strongly upregulated genes, plus 2 additional genes of particular immunological interest (*Epcam* and *Ly6d*). Genes with the smallest FDRs in all the upregulated genes were selected. Genes that were members of the PC signature are denoted by pink lettering. (**c**) Flow cytometric validation of four surface molecules (CCR10, CXCR3, EpCAM and Ly6D) differentially expressed between PBs and PCs. Splenocytes and bone marrow cells were from Blimp1-GFP mice, and marker expression is shown for Syndecan-1$^+$ Blimp1-GFP$^+$ cells, gated as shown in the top panels. Data are representative of three or four independent experiments for each marker. (**d**) Functional annotation of genes differentially expressed between SplPBs and BMPCs. The graph on the left annotates genes upregulated in SplPBs, and that on the right annotates genes upregulated in BMPCs.

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 the fold change for the top 500 genes showing the most difference between the two cell types. A multidimensional scaling (MDS) plot was used to display the cell populations in such a way that distances on the plot corresponded to the log_2 leading fold change (**Fig. 1b**). The plot showed that FoBs and MZBs were most closely related, with B1 cells and GC B cells having more distinct identities. The three ASC populations clustered closely and separately from the three B cell populations. As immunoglobulin gene transcripts constitute the majority of the transcriptome of PCs, immunoglobulin genes were excluded from this MDS analysis and the rest of 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 have at least threefold higher expression in SplPBs, SplPCs and BMPCs than in any of the other populations analyzed, and a false discovery rate (FDR) of ≤0.05 had to be achieved 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 ≥32 FPKMs in each ASC population. The signature contained 301 genes (**Supplementary Table 1**) and included those encoding the well-known regulators Blimp1, Irf4 and Xbp1, as well as ASC surface markers, Syndecan-1 (CD138), Tnfrsf17 (BCMA)[18], CD28 (ref. 19), Ly6C[20] and the amino acid transporter CD98 (**Fig. 1c**). The ASC signature genes were classified by biological function (Online Methods and **Fig. 1d**). These functions testified to the functionality of ASCs, with ~40% of expressed genes having a role in protein production (translation, modification and trafficking). The signature included new markers that will be helpful in the future identification, purification and characterization of ASCs, as well as a number of small noncoding RNAs. The expression profiles of the top ~100 differentially expressed genes (selected from comparisons of FoBs and BMPCs) in FoBs, MZBs, B1 cells, GC B cells, SplPBs, SplPCs and BMPCs demonstrated the clear demarcation separating the four *ex vivo* B cell and three ASC populations (**Fig. 1e**,**f**).

The Blimp-GFP reporter allowed us to distinguish less mature PBs in the spleen from the PCs in spleen and bone marrow[15]. Analysis of the genes most differentially expressed among these three *ex vivo* ASC populations (**Fig. 2a**,**b**) confirmed the close relationship between PCs from the spleen and those from bone marrow, with heat maps showing a clear demarcation between the less mature SplPBs and their more differentiated counterparts. Validation of these differences was obtained by flow cytometry for four surface markers (**Fig. 2c**). These included the chemokine receptors CCR10 and CXCR3, previously associated with PC migration to mucosal and inflamed

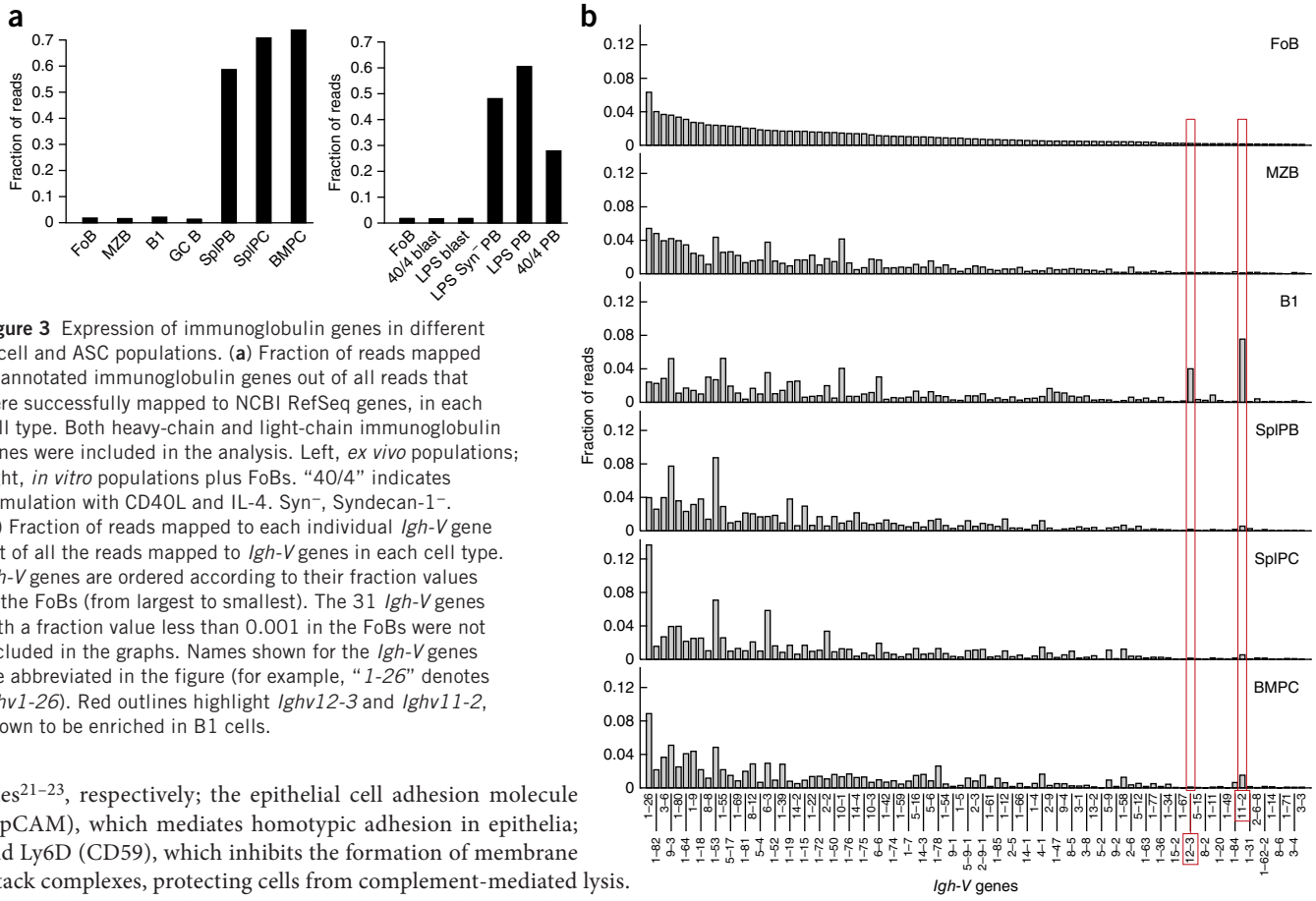

Figure 3 Expression of immunoglobulin genes in different B cell and ASC populations. (**a**) Fraction of reads mapped to annotated immunoglobulin genes out of all reads that were successfully mapped to NCBI RefSeq genes, in each cell type. Both heavy-chain and light-chain immunoglobulin genes were included in the analysis. Left, *ex vivo* populations; right, *in vitro* populations plus FoBs. "40/4" indicates stimulation with CD40L and IL-4. Syn⁻, Syndecan-1⁻. (**b**) Fraction of reads mapped to each individual *Igh-V* gene out of all the reads mapped to *Igh-V* genes in each cell type. *Igh-V* genes are ordered according to their fraction values in the FoBs (from largest to smallest). The 31 *Igh-V* genes with a fraction value less than 0.001 in the FoBs were not included in the graphs. Names shown for the *Igh-V* genes are abbreviated in the figure (for example, "*1-26*" denotes *Ighv1-26*). Red outlines highlight *Ighv12-3* and *Ighv11-2*, known to be enriched in B1 cells.

sites[21–23], respectively; the epithelial cell adhesion molecule (EpCAM), which mediates homotypic adhesion in epithelia; and Ly6D (CD59), which inhibits the formation of membrane attack complexes, protecting cells from complement-mediated lysis. These markers might be useful for identifying and distinguishing PBs and PCs *in vivo*, and they might enable better understanding of the biology of those cells.

Classification of the differentially expressed genes by biological function showed that the strongest effect was the downregulation of cell-cycle genes in BMPCs, followed by repression of genes devoted to receptor expression and signal transduction (**Fig. 2d**). Genes elevated in terminally differentiated BMPCs (**Fig. 2d**) seemed to broadly reinforce the functional properties of ASCs, with enhanced metabolic activity, possibly underlying the greater immunoglobulin-secreting capacity (unpublished data), and changes (e.g., in chemotaxis, cytoskeleton and survival) that could reflect the physical and biochemical properties that enable longevity in the bone marrow niche.

Closer analysis of RNA-seq data for genes encoding chemokines, cytokines and their receptors highlighted their exquisitely tight regulation within the B cell differentiation pathway (**Supplementary Fig. 2**). For example, each B cell subset showed a distinct expression pattern for sphingosine-1-phosphate receptors, and several chemokine and/or cytokine receptors were strongly regulated at the transition from B cell to ASC. 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, and it also provides insights into the dynamic expression of the effector molecules in distinct biological systems of interest.

Immunoglobulin gene transcripts in ASCs

It is well known that ASCs actively transcribe immunoglobulin genes for large-scale antibody secretion; with our RNA-seq data we were able to quantify and contrast the amounts of steady-state immunoglobulin gene transcripts in B cells and ASCs. More than 70% of the transcriptome of BMPCs was derived from immunoglobulin loci (**Fig. 3a**); this percentage was slightly lower in SplPCs, and even lower in the less mature SplPBs. These levels are 30–35 times higher than the immunoglobulin gene abundance in peripheral B cell populations, where the frequency is ~2%. GFP⁺ (Blimp1⁺) ASCs generated *in vitro* from B cells (Online Methods) also devoted the majority of their transcriptional efforts to immunoglobulin gene expression, regardless of whether they expressed Syndecan-1, and independent of the differentiation-inducing signal (**Fig. 3a**).

The variable region of the *Igh* locus (*Igh-V*) encodes approximately 120 annotated transcripts from 16 families in C57BL/6 mice. The Subjunc aligner[16], which was used here for mapping *Igh-V* reads, was configured to allow for the detection of insertions, deletions and substitutions at any location of the *Igh-V* gene segments. Only uniquely mapped reads were included in the analysis.

We detected 89 *Igh-V* genes from 14 families at a frequency of at least 0.1% of total *Igh-V* reads in the FoBs (**Fig. 3b**). FoBs and MZBs showed a similar *Igh-V* profile, which presumably reflected their common bone marrow origin[24]. B1 cells, in contrast, which derive from a fetal progenitor, showed the expected preferential expression of *Ighv11-2* and *Ighv12-3* (refs. 25,26). A remnant of this B1 *Igh-V* transcript pattern is visible in the BMPC compartment, which suggests a small B1 cell contribution to the long-lived ASC population of the bone marrow. *Igh-V* usage was more diversified in the SplPC and BMPC populations, probably as a result of both antigenic selection and affinity maturation, although on the whole the repertoire reflected the initial *Igh-V* frequency seen in the FoB population. These data indicated that

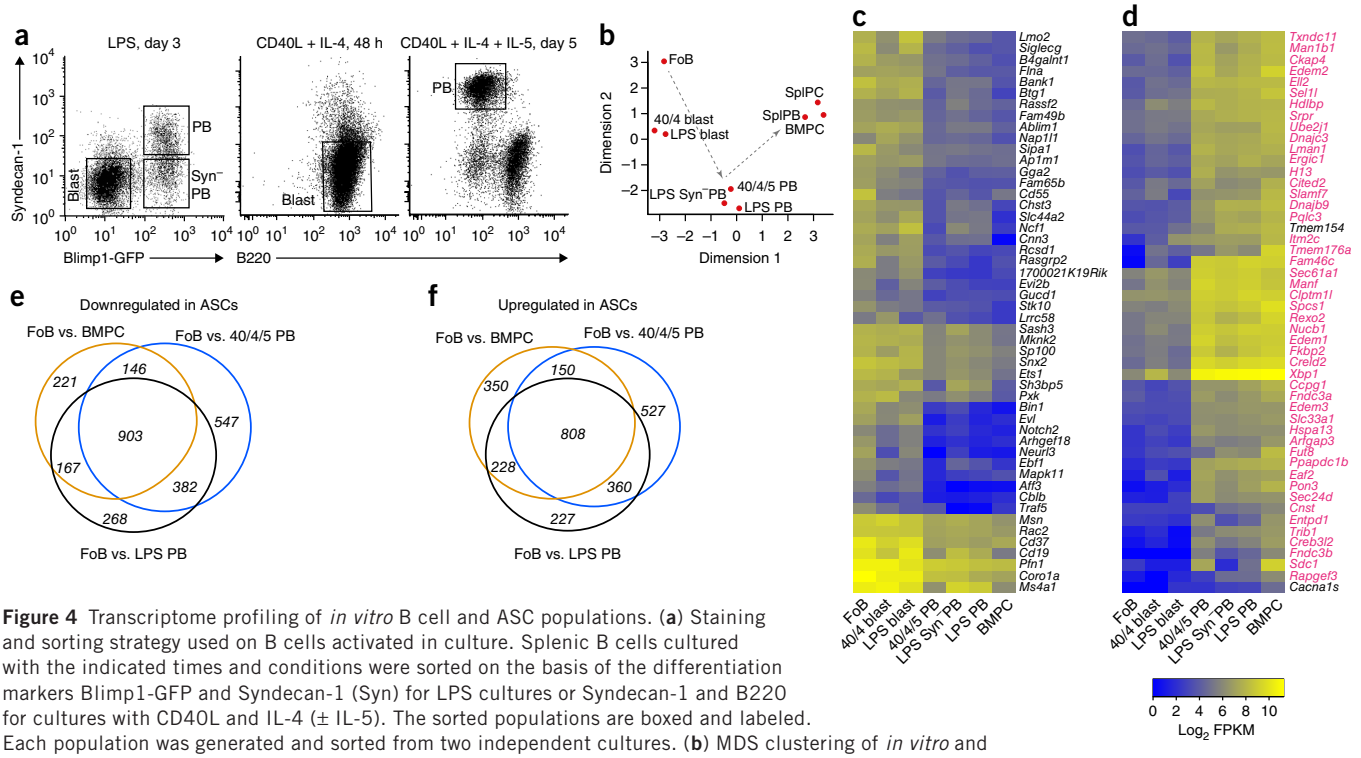

Figure 4 Transcriptome profiling of *in vitro* B cell and ASC populations. (**a**) Staining and sorting strategy used on B cells activated in culture. Splenic B cells cultured with the indicated times and conditions were sorted on the basis of the differentiation markers Blimp1-GFP and Syndecan-1 (Syn) for LPS cultures or Syndecan-1 and B220 for cultures with CD40L and IL-4 (± IL-5). The sorted populations are boxed and labeled. Each population was generated and sorted from two independent cultures. (**b**) MDS clustering of *in vitro* and *ex vivo* B cell and ASC populations. The gray dashed arrows chart an inferred B cell differentiation pathway. (**c**) Expression profile of the top 50 genes that were downregulated as cells differentiated from FoBs to BMPCs, including *in vitro* populations. Shown are $\log_2$ FPKM expression values of genes. (**d**) Expression profile of the top 50 upregulated genes in the same comparisons as in **c**. Genes that are members of the PC signature are denoted by pink lettering. (**e**) Venn diagram showing overlaps and differences between genes that were significantly downregulated in BMPCs and in *in vitro* PBs compared with FoBs. (**f**) Venn diagram showing overlaps and differences between genes upregulated in BMPCs and in *in vitro* PBs compared with FoBs. Complete lists of genes in each Venn category are in **Supplementary Table 2**.

RNA-seq can be used to deduce *Igh-V* gene usage and that the ASCs 'spontaneously' produced in otherwise naive mice provide a reasonable approximation of the starting *Igh-V* 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 PCs[1]. We carried out whole-transcriptome analyses of ASCs generated from splenic B cells *in vitro* under T cell–independent (TI) or T cell–dependent (TD) conditions via stimulation with lipopolysaccharide (LPS) or CD40L and interleukin 4 (IL-4) (± IL-5), respectively (**Fig. 4a**). During TI stimulation, LPS activates B cells and induces Blimp1-GFP$^+$ cells that variously express Syndecan-1, whereas TD conditions produce only Syndecan-1$^+$ cells that are also Blimp1-GFP$^+$ (ref. 15). The sorted populations analyzed were designated 40/4 blasts (48-h CD40L + IL-4–stimulated B cells; CD19$^+$B220^{hi}Syndecan-1$^-$) or 40/4/5 PBs (day-5 CD40L + IL-4 + IL-5–stimulated B cells; Syndecan-1$^+$), and cells sorted from day-3 LPS cultures were designated LPS blasts (B220$^+$Syndecan-1$^-$Blimp1-GFP$^-$), LPS Syn-PBs (B220$^+$Syndecan-1$^-$Blimp1-GFP$^+$) or LPS PBs (B220$^+$Syndecan-1$^+$Blimp1-GFP$^+$). For cell-division experiments, 'Div 0' represented activated but undivided B cells stimulated for 24 h with CD40L and IL-4 (small, B220$^+$Syndecan-1$^-$), division (Div) 1–7 cells (B220$^+$Syndecan-1$^-$ cells positive for the division-tracking dye CFSE, sorted by division number) were from day-4 CD40L, IL-4 and IL-5–stimulated cells, and 40/4 PBs were Syndecan-1$^+$ cells sorted from these same cultures.

MDS analysis implied a cell-differentiation pathway (**Fig. 4b**) between the FoB 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 PBs lie further along the continuum. We did, however, find it striking that the *in vitro*–derived ASCs were still as different from fully mature PCs as they were from activated B cells. Heat maps highlighted the 100 most differentially expressed genes (50 downregulated and 50 upregulated, comparing FoBs and BMPCs), including the *in vitro* populations (**Fig. 4c**,**d**). We created Venn diagrams to identify genes that were shared between ASC populations (or unique to them) (**Fig. 4e**,**f** and **Supplementary Table 2**). All ASCs shared a core of differentially expressed genes compared with expression in FoBs (903 repressed and 808 induced). The vast majority of the genes in the ASC signature (84% (254 of 301 genes)) were also upregulated in 40/4/5 PBs and LPS PBs (FDR < 0.05, fold change > 1.5), which confirmed the utility of this gene set for identifying *in vitro*–differentiated ASCs. The comparison also identified the 221 repressed and 350 induced genes that distinguished BMPCs from PBs generated *in vitro*, a subset that contained genes required for full PC maturation.

Although these results highlighted the shared gene expression program of ASCs, genes whose expression was selective for TI versus TD differentiation were also of interest. We used Venn diagrams to identify the differences in gene expression in PBs that arose in culture in response to LPS and to CD40L with IL-4 and 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

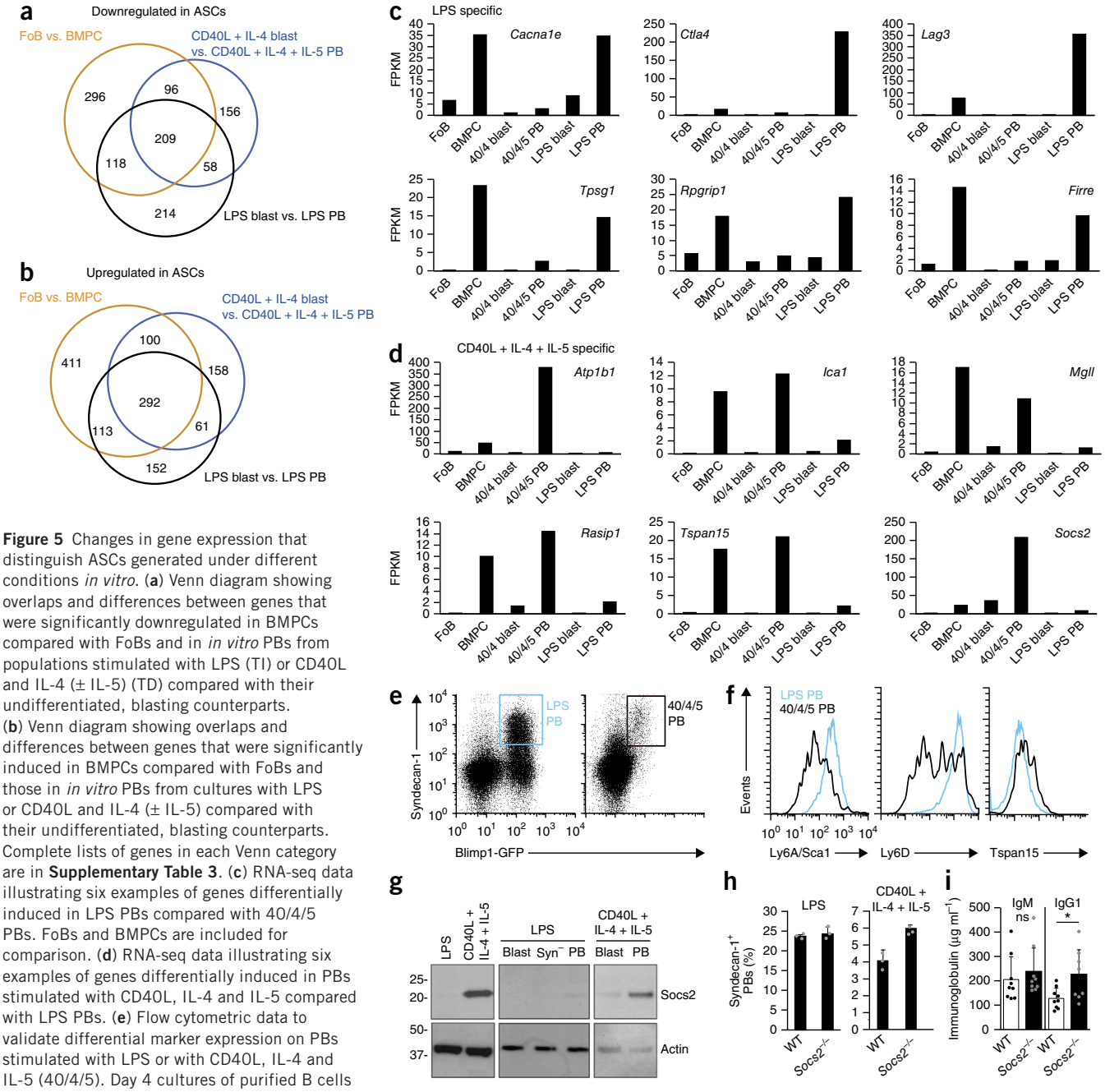

Figure 5 Changes in gene expression that distinguish ASCs generated under different conditions *in vitro*. (**a**) Venn diagram showing overlaps and differences between genes that were significantly downregulated in BMPCs compared with FoBs and in *in vitro* PBs from populations stimulated with LPS (TI) or CD40L and IL-4 (± IL-5) (TD) compared with their undifferentiated, blasting counterparts. (**b**) Venn diagram showing overlaps and differences between genes that were significantly induced in BMPCs compared with FoBs and those in *in vitro* PBs from cultures with LPS or CD40L and IL-4 (± IL-5) compared with 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 with 40/4/5 PBs. FoBs and BMPCs are included for comparison. (**d**) RNA-seq data illustrating six examples of genes differentially induced in PBs stimulated with CD40L, IL-4 and IL-5 compared with LPS PBs. (**e**) Flow cytometric data to validate differential marker expression on PBs stimulated with LPS or with CD40L, IL-4 and IL-5 (40/4/5). Day 4 cultures of purified B cells from Blimp1-GFP mice were used. The analysis was gated as indicated. (**f**) Surface expression of three markers differentially expressed on PBs stimulated with LPS and with CD40L, IL-4 and IL-5 (40/4/5), gated as in **e**. Data in **e** and **f** are representative of three independent experiments. (**g**) Immunoblot validation of differential Socs2 protein expression in day 4 bulk cultures of B cells (in LPS or in CD40L, IL-4 and IL-5) and in sorted populations of blasts and PBs from such cultures. Syn⁻, Syndecan-1⁻ PBs. The analysis was done twice. (**h**) The proportions of Syndecan-1⁺CD22⁻ PBs[44] from wild-type (WT) and *Socs2*$^{-/-}$ mice after 4 d in the presence of LPS or a combination of CD40L, IL-4 and IL-5. Data are the mean ± s.d. from three mice. Dots indicate data from individual mice. (**i**) IgM and IgG1 in serum of naive C57BL/6 and *Socs2*$^{-/-}$ mice determined by ELISA. Data are the mean ± s.d. from nine mice. Dots indicate data from individual mice. *P* values derived via comparison of wild-type and *Socs2*$^{-/-}$ mice using an unpaired two-tailed Student's *t*-test. ns, not significant; $P > 0.05$; $^*P = 0.02$. The analysis was done twice.

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[27], 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*$^{-/-}$ B cells only in cultures with CD40L, IL-4 and IL-5 (**Fig. 5h**) and the significantly elevated immunoglobulin G1 (IgG1) (but not IgM) titers in *Socs2*$^{-/-}$ mice compared with 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[28], which implies sequential staging of the differentiation process after activation. Additional division-linked changes in cell

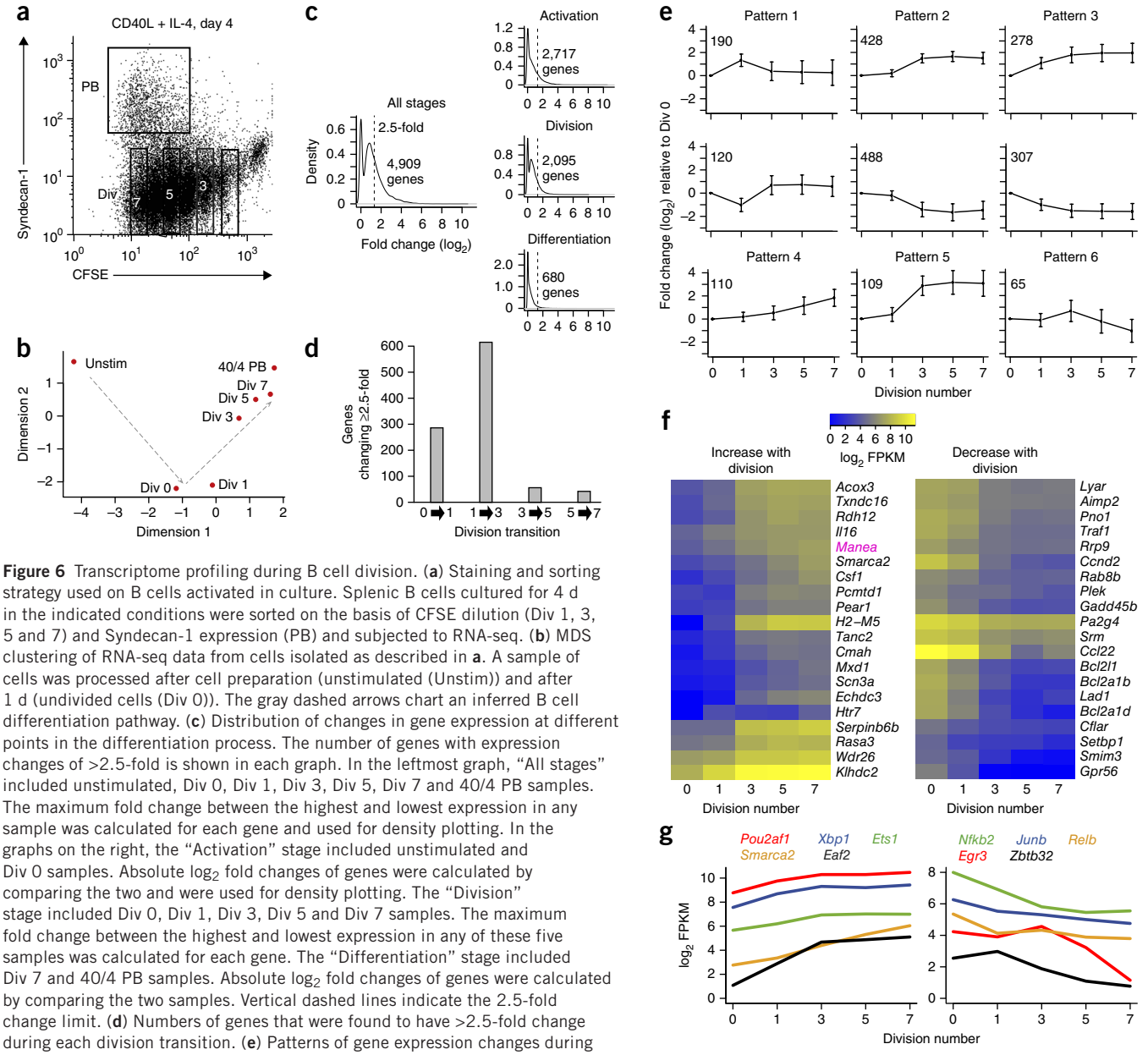

Figure 6 Transcriptome profiling during B cell division. (**a**) Staining and sorting strategy used on B cells activated in culture. Splenic B cells cultured for 4 d in the indicated conditions were sorted on the basis of CFSE dilution (Div 1, 3, 5 and 7) and Syndecan-1 expression (PB) and subjected to RNA-seq. (**b**) MDS clustering of RNA-seq data from cells isolated as described in **a**. A sample of cells was processed after cell preparation (unstimulated (Unstim)) and after 1 d (undivided cells (Div 0)). The gray dashed arrows chart an inferred B cell differentiation pathway. (**c**) Distribution of changes in gene expression at different points in the differentiation process. The number of genes with expression changes of >2.5-fold is shown in each graph. In the leftmost graph, "All stages" included unstimulated, Div 0, Div 1, Div 3, Div 5, Div 7 and 40/4 PB samples. The maximum fold change between the highest and lowest expression in any sample was calculated for each gene and used for density plotting. In the graphs on the right, the "Activation" stage included unstimulated and Div 0 samples. Absolute $\log_2$ fold changes of genes were calculated by comparing the two and were used for density plotting. The "Division" stage included Div 0, Div 1, Div 3, Div 5 and Div 7 samples. The maximum fold change between the highest and lowest expression in any of these five samples was calculated for each gene. The "Differentiation" stage included Div 7 and 40/4 PB samples. Absolute $\log_2$ fold changes of genes were calculated by comparing the two samples. Vertical dashed lines indicate the 2.5-fold change limit. (**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. We used the *k*-means clustering algorithm to classify the genes that had changes of >2.5-fold at the division stage (2,095 genes) into nine clusters by using their stepwise expression changes from Div 0 to Div 7. These nine clusters were then further grouped into six distinct patterns. The mean ± s.d. (error bars) is plotted for each division number in each cluster. Numbers of genes included in each cluster are shown at the top left of each graph. (**f**) Expression profiles of top 20 upregulated genes (left) and top 20 downregulated genes (right) in cell-division data. Genes were selected by comparing Div 0 with Div 7. Log_2 FPKM expression values of genes are shown in the heat maps, color-coded according to the legend. Genes that are members of the ASC signature are denoted by pink lettering. (**g**) Examples of transcription factors that progressively increased (left) or decreased (right) in expression during cell division. (Data extracted from the total RNA-seq data in **b**–**g**.)

surface-molecule expression and class-switch recombination diversify the emerging cell population[14,29,30]. To explore this directly, we labeled resting FoBs with CFSE, activated them with CD40L and IL-4 and then sorted Syndecan-1$^-$ activated B cell blasts after 4 d from divisions 1, 3, 5 and 7 (**Fig. 6a**). Syndecan-1$^+$ ASCs were sorted at the same time (40/4 PBs). Additional populations included unstimulated B cells and cells cultured with CD40L and IL-4 for 24 h, at which time cells were activated but had not yet divided (Div 0). All samples were profiled by RNA-seq. MDS plotting of the profiles by division number (**Fig. 6b**) intimated a differentiation pathway that tracked from unstimulated cells through early to late divisions and finally to differentiation. The expression profile of late-division non-ASCs was very close to that of PBs and was significantly different from that seen in the early divisions. Furthermore, the biggest transitional step was in the initial 24-h activation from the resting state to Div 0, with successively smaller changes until cells became PBs.

To explore this development program, we calculated the maximum fold change between the highest and lowest expression for each gene in any sample (unstimulated, divided samples or PBs) (**Fig. 6c**). We found 4,909 genes with expression altered by more than 2.5-fold at some point in the transition from unstimulated B cells to PBs. Of these, more than half (2,717 genes) were accounted for by activation alone

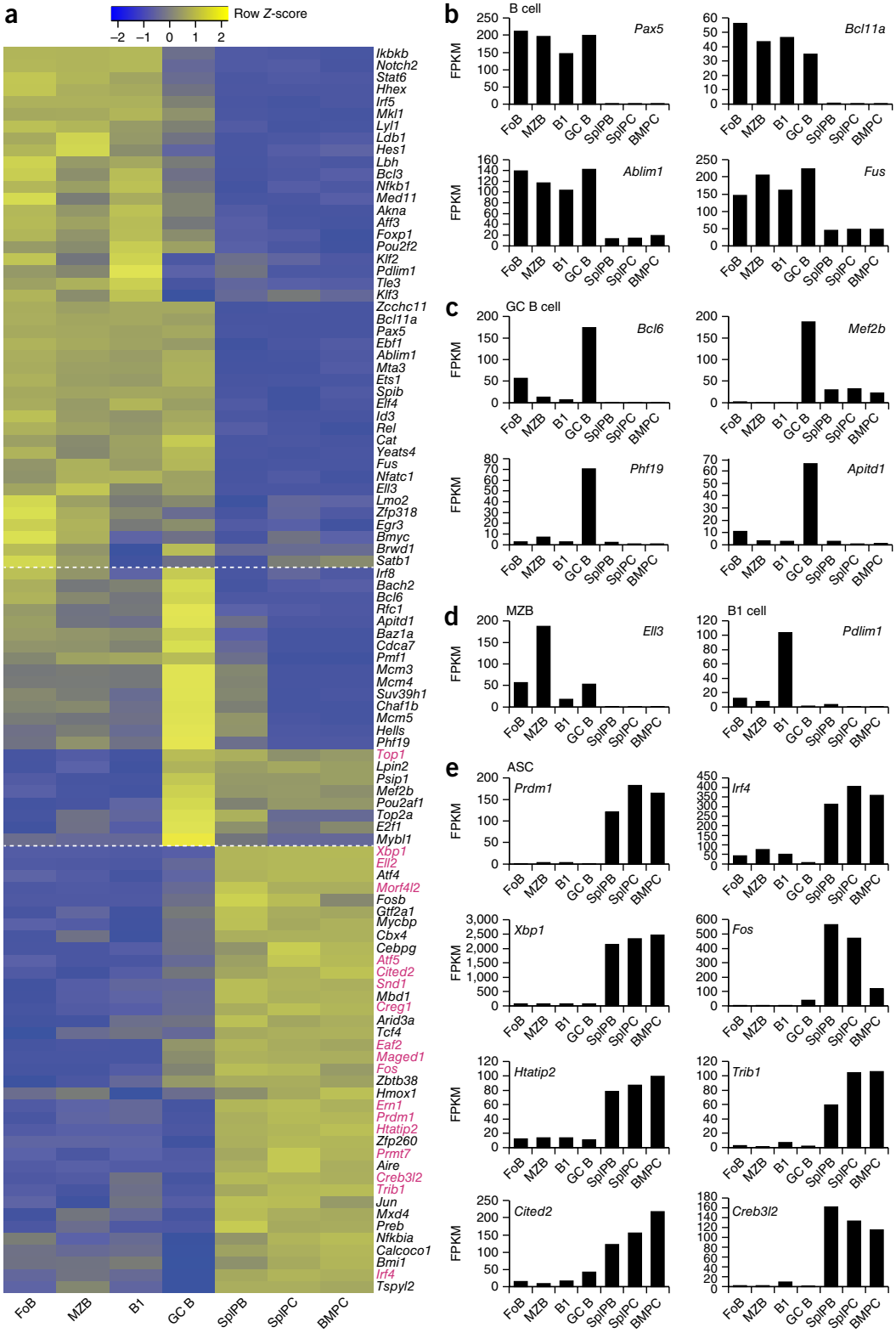

Figure 7 Cell-type-specific changes in transcription factor gene expression. (**a**–**e**) RNA-seq data for the indicated *ex vivo* B cell and ASC populations, selected for genes that showed >2.5-fold differential expression between FoBs and BMPCs, had an expression abundance of at least 32 FPKMs in one or more samples and were present on a curated list of factors involved in the regulation of gene expression[45]. (**a**) Expression profile of the 103 differentially expressed transcriptional regulators. Relative expression levels (*Z*-scores) of genes are shown in the heat maps, color-coded according to the legend. Rows are scaled to have a mean of 0 and an s.d. of 1. Genes that are also members of the ASC signature are denoted by pink lettering. (**b**) Expression data for genes that mimic the pan–B cell restricted pattern shown by *Pax5*. (**c**) Genes that were expressed in a GC-specific manner. (**d**) Genes that showed MZB- or B1 cell–specific expression. (**e**) Genes whose expression was ASC restricted, reflecting the expression of essential regulators of ASC differentiation, such as *Prdm1*, *Irf4* and *Xbp1*. See **Supplementary Figures 5** and **6** for an expanded set of genes with such cell-type-restricted expression patterns. (Data extracted from the total RNA-seq data.)

(unstimulated to Div 0), whereas 2,095 expressed genes changed within the division sequence itself, and 680 were involved in the final change to PBs. Within the division sequence, the most volatile stage was that from Div 1 to Div 3, with later divisions showing progressively fewer expression changes (**Fig. 6d**). Clustering based on stepwise changes in gene expression showed nine expression patterns (**Supplementary Fig. 3**). The individual genes were normalized to Div 0 and averaged to illustrate the range of gene expression changes involved in the division-linked program (**Fig. 6e**). These patterns included graded and stepwise transitions with successive division, as well as wave-like expression (**Fig. 6e**). We used heat maps to illustrate the expression profiles of the top 20 upregulated and top 20 downregulated genes, selected by comparing Div 0 with Div 7 (**Fig. 6f**). These expression patterns were likely to be driven by transcriptional regulators, either singly or in combination. Indeed, the expression of several well-known transcription factors changed progressively with division (**Fig. 6g**). 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 generates considerable cellular heterogeneity along the way.

Expression changes correlate with epigenetic modifications

We also assessed the importance of chromatin modifications to gene expression during B cell terminal differentiation using both published and newly generated data. The histone modifications H3K4me3 and H3K4me1 (histone H3 trimethylated and singly methylated at Lys4, respectively) are enriched in active promoters and distal enhancers in many developmental settings[31]. Analysis for the presence of these modifications using chromatin immunoprecipitation followed by sequencing (ChIP-seq) on the mouse GC B cell line A20 and the plasmacytoma MPC-11 showed a strong correlation between the presence of the histone mark and expression of the associated gene in both A20 and MPC-11 cells and the corresponding *in vivo* populations, GC B cells 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 changes in gene expression 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[32,33]. This is particularly the case in multiple myeloma, where genes that are prominent in disease progression, and in PC biology in general, are highly enriched for the presence of super-enhancers[34]. Investigation of the expression of a set of previously annotated super-enhancer–containing genes in the B cell terminal differentiation process showed that super-enhancer–containing genes were strongly downregulated in the transition from either FoBs or GC B cells to BMPCs (**Supplementary Fig. 4c**,**d**), which suggested that inappropriate maintenance of the expression of these genes would be 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 remodeling in the differentiation process.

New transcriptional regulators of B cell differentiation

Finally, we examined the expression of a particular subset of genes in 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 FoBs and BMPCs clustered into three major groups (**Fig. 7a**). A prominent cluster comprised genes that were expressed in all B cell subsets before being downregulated in ASCs. This cluster included known B cell regulators such as *Pax5* and *Spib*, as well as a large number of previously unrecognized potential factors, including *Bcl11a*, *Ablim1*, *Tle3* and *Hhex* (**Fig. 7a**,**b** and **Supplementary Fig. 5a**). The GC B 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 cell–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 approximately 30 other transcriptional regulators not yet implicated in B cell biology (**Fig. 7a**,**e** and **Supplementary Fig. 6**). This analysis thus identified a discrete set of genes, many of which have not been implicated in lymphocyte biology, that function together to program B cell terminal differentiation.

DISCUSSION

Terminal differentiation is an irreversible process that results 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 with a range of isotypes and affinities, reside in several anatomical sites in the body and have markedly different life spans. Despite these phenotypic differences, our data show that all ASCs investigated in this study shared a common transcriptional signature that was distinct from the signatures of all B cell subsets.

Analysis of the ASC signature showed that major functional classes of the newly expressed genes encoded proteins involved in the transcription, translation, intracellular transport and glycosylation of immunoglobulins. Upregulated expression of unfolded protein response genes was also an expected consequence of the extremely high immunoglobulin production. An early event in ASC differentiation is the transition from cycling, short-lived PBs to post-mitotic, long-lived PCs[1]. 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 they 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[35]. *Tnfrsf17*, encoding BCMA, the receptor for APRIL, is a component of the ASC signature[13,18], which suggests that this exclusivity is a component of the specificity of BMPC survival. The genes encoding a variety of other adhesion molecules, including SLAMF7, BST2, EpCAM, ALCAM and FNDC3B, were also found to be components of the ASC signature and are therefore potential previously undescribed regulators of ASC homing and survival. As monoclonal antibodies to EpCAM are currently in advanced clinical trials for the treatment of a variety of cancers[36], 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 of the sphingosine-1-phosphate

receptors, provides insights into the many, often contradictory signals that facilitate the localization of the B cell and ASC subsets. The functions of many of these factors, including the atypical chemokine receptors CCRL1 (GC B cell specific) and CCRL2 (ASC specific), remain unknown in B cells. Although most cytokine-receptor chains were downregulated during B cell differentiation, several, including *Il2rg*, *Il13ra1* and *Il10rb*, remained 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, GC B cell and ASC fate, but they also identified a number of other transcriptional regulators with parallel patterns of expression. For instance, Phf19, a component of the polycomb repressive complex involved in recruitment to chromatin[37], was exclusively expressed in GC B cells. Other ASC-specific factors included Eaf2, a pro-apoptotic tumor suppressor[38], and Cited2, a mediator of stem cell maintenance and an oncogene[39]. 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 ASCs generated *in vivo* and *in vitro*, as well as distinct transcriptional consequences of differentiation induced by LPS, a model for TI stimulation, and by the combination of CD40L, IL-4 and IL-5, modeling TD responses. MDS plots suggested that *in vitro*–derived PBs, although clearly ASCs, were not as fully mature as their *in vivo* counterparts. Despite this, approximately 200 genes were uniquely upregulated in each of LPS PBs and 40/4/5 PBs. *Socs2* represents one such gene that was specifically induced in 40/4/5 PBs, and analysis of cells from SOCS2-deficient mice suggested that this protein may negatively regulate ASC differentiation in that setting. SOCS2 is best characterized as a suppressor of growth factor signaling[27]; however, there is some evidence for inhibitory control of IL-4 signaling in T cells[40]. Future work will be needed to fully assess whether SOCS2 has 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[28]. 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 showed nine clusters, suggestive of a common regulatory mechanism for each. Indeed, discrete groups of similarly expressed transcription factors were also identified that represented candidate regulators of each cluster. These insights suggest that with the application of small-scale ChIP-seq protocols to individual cell division cohorts, it could be possible 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[41], lymphoma cells[42] and IL-17–producing helper T cells[43].

The stark transcriptional changes that we observed in association with the differentiation of B cells to ASCs were 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; as mentioned, these are known as super-enhancers[32,33]. In human B cells, super-enhancer–containing genes include those encoding PAX5, BACH2, IRF8 and c-MYC, all of which are strongly downregulated during terminal differentiation[32]. In contrast, multiple myeloma is associated with inappropriate maintenance of several B cell super-enhancers, including the region associated with *MYC*[34]. In keeping with this, we found super-enhancer–containing genes to be enriched for those genes silenced in the ASC-differentiation process, which suggests 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 idea, as the large number of transcriptional changes and the concurrent chromatin remodeling we documented during ASC differentiation suggest that this process will be exquisitely sensitive to therapeutic intervention in the gene regulatory networks driving this process.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. GEO: raw sequence reads, read counts and normalized expression values, GSE60927.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank D. Emslie, T. Kratina, M. Everest and S. Heinzel for technical assistance and J. Vasiliadis, J. Leahy and T. Camilleri for animal care. *Socs2*$^{-/-}$ mice were a gift from W. Alexander (The Walter and Eliza Hall Institute, Parkville, Victoria, Australia). The institutional flow cytometry facility provided essential services. 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.; Project Grant 1023454 to G.K.S. and W.S.; and fellowships to G.K.S., D.M.T., P.D.H. and L.M.C.), the Multiple Myeloma Research Foundation (Senior Award to S.L.N.), the Australian Research Council (ARC Future Fellowship to S.L.N.) and the Victorian State Government through an Operational Infrastructure Support grant.

AUTHOR CONTRIBUTIONS

W.S., Y.L., G.K.S. and M.I. did the bioinformatic and statistical analyses; S.N.W. and N.T. did 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 the experimental design and analysis, and provided mouse models; L.M.C. carried out many of the experiments; W.S., S.L.N. and L.M.C. wrote the manuscript; all authors had editorial input.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Nutt, S.L., Hodgkin, P.D., Tarlinton, D.M. & Corcoran, L.M. The generation of antibody-secreting plasma cells. *Nat. Rev. Immunol.* **15**, 160–171 (2015).
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–ii89 (2013).
3. Potter, M. Neoplastic development in plasma cells. *Immunol. Rev.* **194**, 177–195 (2003).
4. Fairfax, K.A. *et al.* Different kinetics of blimp-1 induction in B cell subsets revealed by reporter gene. *J. Immunol.* **178**, 4104–4111 (2007).
5. Radbruch, A. *et al.* Competence and competition: the challenge of becoming a long-lived plasma cell. *Nat. Rev. Immunol.* **6**, 741–750 (2006).
6. Tarlinton, D. & Good-Jacobson, K. Diversity among memory B cells: origin, consequences, and utility. *Science* **341**, 1205–1211 (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**, 341–349 (2011).
8. Shapiro-Shelef, M. & Calame, K. Regulation of plasma-cell development. *Nat. Rev. Immunol.* **5**, 230–242 (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**, 5773–5785 (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**, 5173–5181 (2009).
13. Peperzak, V. *et al.* Mcl-1 is essential for the survival of plasma cells. *Nat. Immunol.* **14**, 290–297 (2013).
14. Sciammas, R. *et al.* Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation. *Immunity* **25**, 225–236 (2006).
15. Kallies, A. *et al.* Plasma cell ontogeny defined by quantitative changes in blimp-1 expression. *J. Exp. Med.* **200**, 967–977 (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**, 923–930 (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**, 91–98 (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**, 269–281 (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**, 1136–1140 (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**, 1277–1282 (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**, 1001–1010 (2003).
24. Cerutti, A., Cols, M. & Puga, I. Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes. *Nat. Rev. Immunol.* **13**, 118–132 (2013).
25. Jeong, H.D. & Teale, J.M. Contribution of the CD5$^+$ B cell to D-proximal VH family expression early in ontogeny. *J. Immunol.* **145**, 2725–2729 (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**, 1069–1073 (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**, 277–281 (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. USA* **102**, 13242–13247 (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**, 576–589 (2010).
32. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
33. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
34. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
35. Chu, V.T. & Berek, C. The establishment of the plasma cell survival niche in the bone marrow. *Immunol. Rev.* **251**, 177–188 (2013).
36. Kurtz, J.E. & Dufour, P. Adecatumumab: an anti-EpCAM monoclonal antibody, from the bench to the bedside. *Expert Opin. Biol. Ther.* **10**, 951–958 (2010).
37. Lu, R. & Wang, G.G. Tudor: a versatile family of histone methylation 'readers'. *Trends Biochem. Sci.* **38**, 546–555 (2013).
38. Xiao, W. *et al.* U19/Eaf2 knockout causes lung adenocarcinoma, B-cell lymphoma, hepatocellular carcinoma and prostatic intraepithelial neoplasia. *Oncogene* **27**, 1536–1544 (2008).
39. Korthuis, P.M. *et al.* CITED2-mediated human hematopoietic stem cell maintenance is critical for acute myeloid leukemia. *Leukemia* **29**, 625–635 (2015).
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**, 1523–1531 (2011).
41. Zak, D.E., Tam, V.C. & Aderem, A. Systems-level analysis of innate immunity. *Annu. Rev. Immunol.* **32**, 547–577 (2014).
42. Basso, K. *et al.* Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37**, 382–390 (2005).
43. Yosef, N. *et al.* Dynamic regulatory network controlling TH17 cell differentiation. *Nature* **496**, 461–468 (2013).
44. Perfetti, V. *et al.* Membrane CD22 defines circulating myeloma-related cells as mature or later B cells. *Lab. Invest.* **77**, 333–344 (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**, 467–482 (2012).

ONLINE METHODS

Mice. Blimp1-GFP reporter[21] and *Socs2*$^{-/-}$ mice[27] (a kind gift from W. Alexander, The Walter and Eliza Hall Institute) have been described before. All mice were maintained on a C57BL/6 genetic background and housed in a clean conventional mouse facility. All animal experiments were conducted according to the protocols and with the approval of the Walter and Eliza Hall Institute Animal Ethics Committee (AEC2013.014).

Flow cytometry, antibodies and ELISA. Cells were analyzed on FACSCanto (BD Biosciences), and cell sorting was carried out using FACSAria (BD Biosciences) or MoFlo (Beckman Coulter) flow cytometers. Data were analyzed using FlowJo software. The following antibodies and reagents were used for flow cytometry and immunoblotting: anti–Syndecan-1 (CD138; 281-2), anti-CD21 (7G6), anti-CD23 (B3B4), anti-B220 (RA3-6B2), anti-CD95 (Jo2), phycoerythrin-conjugated Ly6A/E (Sca-1) (E13-161.7) and Alexa Fluor 647–conjugated anti-Ly6D (49-H4), all from BD Biosciences; anti-CD19 (1D3; eBioscience); PNA (Vector Labs); allophycocyanin-conjugated anti-CCR10 (FAB2815A) and allophycocyanin-conjugated anti-CXCR3 (FAB1685A), both from R&D Systems; Alexa Fluor 647–conjugated anti-EpCAM (CD326) (G8.8) and allophycocyanin-conjugated anti-CD22 (OX-97), both from BioLegend; and phycoerythrin-conjugated anti-TSPAN15 (bs-9425R; Bioss Antibodies). Serum and supernatant immunoglobulin amounts were measured using ELISA as described[28].

Preparation of sorted cell populations: *ex vivo* populations. Flow cytometry was used to sort cells 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 bone marrow from three Blimp1-GFP reporter mice were pooled, and Syndecan-1$^+$ cells were magnetically enriched (CD138 MicroBeads, Miltenyi Biotec) before sorting. This was repeated four times. For B1 cells, peritoneal lavages from three or four 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 maximize the yields of undifferentiated or differentiated cells of interest. Culturing with CD40L and IL-4 generates ASCs slowly with division, allowing the staged progress to be monitored, as in **Figure 6**. IL-5 acts selectively to increase the likelihood of differentiation without affecting other cell responses (e.g., switching, survival or division), thus enabling greater numbers of ASCs to be generated earlier in cultures[28]. 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[28] for 48 h. Subsequently, activated CD19$^+$ B cells that lacked Syndecan-1 expression were sorted, and total RNA was extracted (40/4 blasts). This time point was chosen to examine early transcriptional changes, as minimal differentiation had occurred, with all cells remaining B220^{hi}Syndecan-1$^-$ (>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 d. Total RNA was isolated from sorted Syndecan-1$^+$ PBs (40/4/5 PBs, Syndecan-1$^+$). Alternatively, splenic B cells were purified from Blimp1-GFP mice using anti-B220 magnetic beads and cultured for the indicated times (**Fig. 4a**) in the presence of 10 μg/ml LPS. After 3 d, three populations were isolated: undifferentiated Syndecan-1$^-$Blimp1-GFP$^-$ B cell blasts (LPS blasts), Syndecan-1$^-$Blimp1-GFP$^+$ PBs (LPS Syn$^-$ PBs) and Syndecan-1$^+$Blimp1-GFP$^+$ PBs (LPS PBs).

Finally, small resting B cells from C57BL/6 mice, purified with magnetic beads and density gradients[28], were labeled with CFSE and cultured with CD40L and IL-4 as described above. After 4 d, cells were sorted using CFSE dilution into groups that had divided one, three, five and seven times but had not yet differentiated into Syndecan-1$^+$ ASCs. Differentiated Syndecan-1$^+$ ASCs (40/4 PBs, Syndecan-1$^+$) 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 with Qiagen RNeasy Micro or Mini Kits (on the basis of cell number), according to the manufacturer's instructions. RNA-seq was done with an Illumina Hiseq 2000 or a HiSeq 2500 instrument at the Australian Genome Research Facility, Melbourne, Australia, or at Beijing Genomics Institute, Shenzhen, China.

Reads were aligned to the GRCm37/mm9 build of the *Mus musculus* genome with the Subread aligner[16]. 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. Gene-wise counts were obtained with featureCounts[17]. Reads overlapping exons in annotation build 37.2 of the NCBI RefSeq database were included. A summary of read mapping and quantification results can be found in **Supplementary Table 4**. Genes were filtered from downstream analysis if they did not have a CPM (counts per million mapped fragments) value of at least 1 in at least one library. Immunoglobulin genes were excluded from expression analysis when we compared B cell populations with ASC populations, because of the large difference in the fraction of immunoglobulin reads between them. Expression analysis for immunoglobulin genes was done separately.

Counts were converted to $\log_2$ counts per million, quantile normalized and precision weighted with the 'voom' function of the limma package[46,47]. A linear model was fitted to each gene, and empirical Bayes moderated *t*-statistics were used to assess differences in expression[48]. Empirical Bayes moderated-*t* *P* values were computed relative to a fold-change cutoff of 1.5-fold using treat[49]. *P* values were adjusted to control the global FDR across all comparisons with the 'global' option of the limma package. Genes were considered differentially expressed if they had an FDR 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 $\log_2$ FPKM values reset to zero.

Mapping of immunoglobulin reads. The alignment of reads originating from immunoglobulin genes was carried out using the Subjunc aligner, which has been found to have excellent performance in mapping reads spanning several short exons[16]. We configured Subjunc to make it be able to detect long insertions and deletions occurring at the start and end of immunoglobulin segments, because of imprecise recombination and N nucleotide insertions. Also, we did not require that donor or receptor sites be present when we mapped reads that spanned more than one segment, as V(D)J recombination takes place at the 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 (ab8895) and anti-H3K4me3 (ab8580), both from Abcam. Lymphoma cell line A20 (ATCC TIB-208) and the plasmacytoma MPC-11 (ATCC CCL-167) were sourced from ATCC.

Reads were mapped to the GRCm37/mm9 build of the *Mus musculus* genome in the same way as 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 3-kb upstream sequence and a 2-kb downstream sequence relative to the transcriptional start of the gene. NCBI mouse RefSeq gene annotation was used. Mapped reads were assigned to the promoter region of each gene with the featureCounts program. FPKM values were calculated for each gene on the basis of read counts, size of promoter regions (5 kb) and library size.

Identification of absolutely expressed genes in RNA-seq data. To identify absolutely expressed genes, we used Poisson testing for each gene to determine whether 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 2 kb 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 with the same length and GC percentage as the gene were input into the Poisson

doi:10.1038/ni.3154

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[32]. 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 done in this study. The gene set enrichment plots were generated with the 'barcodeplot' function in the limma package. Statistical tests of enrichment were carried out using the 'roast' method in limma with 999 rotations[50]. One-sided *P* values are reported.

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

46. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **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**, 765–771 (2009).
50. Wu, D. *et al.* ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics* **26**, 2176–2182 (2010).