12152–12169 *Nucleic Acids Research, 2017, Vol. 45, No. 21* *Published online 11 September 2017*
doi: 10.1093/nar/gkx798

Principles for the regulation of multiple developmental pathways by a versatile transcriptional factor, BLIMP1

Tadahiro Mitani[1,2], Yukihiro Yabuta[1,2], Hiroshi Ohta[1,2], Tomonori Nakamura[1,2], Chika Yamashiro[1,2], Takuya Yamamoto[3,4], Mitinori Saitou[1,2,3,5,*] and Kazuki Kurimoto[1,2,*]

[1]Department of Anatomy and Cell Biology, Graduate School of Medicine, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606–8501, Japan, [2]JST, ERATO, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606–8501, Japan, [3]Center for iPS Cell Research and Application, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606–8507, Japan, [4]AMED-CREST, AMED 1-7-1 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan and [5]Institute for Integrated Cell-Material Sciences, Kyoto University, Yoshida-Ushinomiya-cho, Sakyo-ku, Kyoto 606–8501, Japan

Received April 13, 2017; Revised August 25, 2017; Editorial Decision August 26, 2017; Accepted August 30, 2017

ABSTRACT

Single transcription factors (TFs) regulate multiple developmental pathways, but the underlying mechanisms remain unclear. Here, we quantitatively characterized the genome-wide occupancy profiles of BLIMP1, a key transcriptional regulator for diverse developmental processes, during the development of three germ-layer derivatives (photoreceptor precursors, embryonic intestinal epithelium and plasmablasts) and the germ cell lineage (primordial germ cells). We identified BLIMP1-binding sites shared among multiple developmental processes, and such sites were highly occupied by BLIMP1 with a stringent recognition motif and were located predominantly in promoter proximities. A subset of bindings common to all the lineages exhibited a new, strong recognition sequence, a GGGAAA repeat. Paradoxically, however, the shared/common bindings had only a slight impact on the associated gene expression. In contrast, BLIMP1 occupied more distal sites in a cell type-specific manner; despite lower occupancy and flexible sequence recognitions, such bindings contributed effectively to the repression of the associated genes. Recognition motifs of other key TFs in BLIMP1-binding sites had little impact on the expression-level changes. These findings suggest that the shared/common sites might serve as potential reservoirs of BLIMP1 that functions at the specific sites, providing the foundation for a unified understanding of the genome regulation by BLIMP1, and, possibly, TFs in general.

INTRODUCTION

Transcription factors (TFs) recognize short DNA sequences and control the expression of associated genes, contributing to the generation and maintenance of diverse cell types throughout the body based on a single set of genomic information. Remarkably, single TFs can function in the development of many distinct cell types, and clarification of the mechanism underlying this phenomenon remains a fundamental challenge. To understand this mechanism, it will be critical to identify the genome-wide binding profiles of relevant TFs in multiple developmental processes in a systematic and quantitative manner. Studies along this line have been performed on cultured cell lines and a limited number of developmental lineages, and have revealed a number of key regulatory mechanisms for transcriptional activation, including the selection and activation of specific enhancers by collaborative TF interactions at closely spaced DNA recognition motifs [reviewed in (1,2)]. On the other hand, cellular development proceeds under cross-talking signals that may promote irrelevant differentiation or cellular states, and thus repressive transcriptional programs are also vital for appropriate cellular development. Repressive transcriptional programs often play a key role in transient cell populations, but there have been relatively few analyses investigating such programs with regard to TF-binding profiles across multiple cell lineages.

B lymphocyte-induced maturation protein 1 [BLIMP1, also known as PR domain containing 1 (PRDM1)] was originally identified as a key factor for the differentiation of plasma cells from B lymphocytes (3,4). It has been shown to act primarily as a transcriptional repressor and to recognize specific DNA sequences proximal to the transcription start sites (TSSs) in complexes with various co-repressors (3–11). BLIMP1 has subsequently been shown to play critical roles in a wide variety of developmental pathways in embryos and adults, including embryonic derivatives from

*To whom correspondence should be addressed. Tel: +81 75 753 4337; Fax: +81 75 751 7286; Email: kurimoto@anat2.med.kyoto-u.ac.jp
Correspondence may also be addressed to Mitinori Saitou. Tel: +81 75 753 4335; Fax: +81 75 751 7286; Email: saitou@anat2.med.kyoto-u.ac.jp

all three germ layers, the germ line and extraembryonic lineages (12). Thus, BLIMP1 is one of the TFs required for the widest ranges of developmental processes and would be instructive in a comparative analysis of repressive programs. Accordingly, genome-wide BLIMP1-binding profiles have been analyzed in several lineages (13–16), and the function of BLIMP1 as a transcriptional activator has also been documented (15). On the other hand, systematic comparisons of BLIMP1-binding profiles across distinct cell types have been difficult/impractical, due to differences in the technology employed for obtaining such datasets. Thus, key questions related to the mechanism of action of BLIMP1 remain unanswered, including: How do the binding patterns differ among cell types? Which binding sites are cell-type specific or common? How do the binding differences influence gene expression? Is there any function of BLIMP1 common to all cell types?

Using a unified, quantitative ChIP-seq method amenable for a relatively small number of cells (13), we here investigated the BLIMP1-binding profiles and their impacts on gene expression during four distinct developmental processes in mice: (i) differentiation of photoreceptors from their precursors (photoreceptor precursors; PRP cells) (17,18), (ii) maturation of the intestinal epithelium (IE) from its embryonic form (emIE) (19,20), (iii) differentiation of plasmablasts (PBs) from B cells (4,15,21), (iv-a) the specification process of primordial germ cells (PGCs) at embryonic day (E) 6.5 ∼ E9.5 [reconstituted *in vitro* as induction of PGC-like cells (PGCLCs) from embryonic stem cells (ESCs) via epiblast-like cells (EpiLCs)] (13,22), and (iv-b) late PGC development (∼E12.5) (23). Based on the results, we then clarified the mechanisms of action of this highly versatile transcriptional regulator.

MATERIALS AND METHODS

The methods are described in detail in the Supplementary materials and methods section.

Animals

All the animal experiments were performed under the ethical guidelines of Kyoto University. Homozygous *EGFP-Blimp1* knocked-in mice (EGFP-BLIMP1 mice) (Supplementary Figure S1A) were generated as reported previously (13).

Immunofluorescence (IF)

Embryos of EGFP-BLIMP1 mice at various developmental stages were dissected from euthanized pregnant females, fixed in freshly prepared ice-cold 4% PFA (TAAB) for 30 min on ice, and embedded in OCT compound (Sakura Finetek). The frozen samples were sectioned at 10 μm thickness at −20°C, and incubated with blocking solution [0.1% BSA (Sigma) and 0.1% Triton X-100 (Wako)] containing the primary antibodies overnight at 4°C, followed by washing and incubation with secondary antibodies and 4′,6-diamidino-2-phenylindole (DAPI). The samples were then observed under a confocal laser scanning microscope (Olympus FV1000) (Figure 1).

Cell purification

The photoreceptor progenitors (PRPs) [postnatal day (P) 4 for ChIP-seq, P0, P4, P6 and P10 for RNA-seq], embryonic intestinal epithelium (emE) cells (E16.5 for ChIP-seq, and E16.5, P3 and P14 for RNA-seq), and male and female PGCs (PGCMs and PGCFs, respectively) (E12.5 for ChIP-seq) were purified from EGFP-BLIMP1 mice using fluorescence activated cell sorting (FACS) (ARIA III; BD Biosciences) (Figure 2A; Supplementary Figures S3A–D and S6A–D). PBs were induced from splenic B cells of EGFP-BLIMP1 mice, and purified with FACS. The cell suspensions of the respective cell types were prepared as described previously (24–28) with minor modifications. For RNA-seq, we took two biological replicates of the above cell types at each time point. For ChIP-seq, we took two biological replicates for PRP, emIE and PB cells, and single pooled samples for male and female PGCs. The cell types and biological replicates for ChIP-seq and RNA-seq are summarized in Supplementary Table S1.

RNA-seq library construction

Total RNA was purified with an RNeasy Micro Kit (QIAGEN) according to the manufacturer's instructions. One nanogram of purified total RNA was used for synthesis and amplification of cDNAs, and for construction of the cDNA libraries for RNA sequencing; in constructing the cDNA libraries, a previously reported method was used to selectively detect the 3′ ends of mRNA (29,30).

Chromatin immunoprecipitation

ChIP of EGFP-BLIMP1 was performed as described previously (13,31) with reference to the published guidelines (32).

Library preparation for ChIP-seq

Libraries for the next-generation sequencing were prepared as described previously (13) with a modification for application to the Illumina systems. Briefly, ChIP-ed and input DNAs were sheared to an average size of 150 bp by ultra-sonication (Covaris). The sheared DNAs were end-repaired, dA-tailed, ligated to an amplification adaptor (P1-T Adaptor/F and Barcode-Internal+12-mer/R) (Supplementary Table S2), amplified with a 10-cycle polymerase chain reaction (PCR) using Library PCR primer 1 (Life Technologies) and Library PCR primer Barcode001 + Internal adaptor (Supplementary Table S2), and purified with AMPure XP beads supplemented with PEG 8000 and NaCl (AMPure XP purification, hereafter).

For sequencing on the Illumina systems, we then added an adaptor and index to the amplified library with two rounds of additional PCRs. The first round of additional PCR was performed for six cycles with 1 μM each of Read1-P1 and Read2-internal-adaptor primers (Supplementary Table S2), using Platinum Taq DNA Polymerase (Life Technologies) at a 50-μl reaction scale, followed by three rounds of AMPure XP purification. The second round of PCR was performed for four cycles with 1 μM each of Illumina P5-Read1 primer and P7-index N-Read2 primer (Supplementary Table S2) to produce the index-tagged library DNAs,

followed by two rounds of AMPure XP purification (Supplementary Figure S3E for scheme). For the ChIP-seq of d2 and d6 PGCLCs, we performed the two rounds of additional PCRs for the previously amplified libraries (13).

To exclude the initial constant region consisting of the amplification adaptor sequence (P1-T Adaptor) and maximize the efficient sequencing reads, we designed a custom sequencing primer, Custom-primer-29mer (Supplementary Table S2), which met the criteria for sequencing on an Illumina platform (33) in terms of the length, GC content and melting temperature (Supplementary Figure S3E for scheme). The DNAs were then sequenced on the HiSeq 2500 platform (Illumina) in high-throughput mode to generate single-end 100-bp reads, per the manufacturer's instructions (Supplementary Figure S3E).

Mapping and normalization of RNA-seq data

GRCm38/mm10 for the mouse genome and ref_GRCm38 for the mouse transcript annotation were used for mapping of the read data. Read trimming, mapping and estimation of expression levels were performed as described previously (29,30). The reads per million mapped reads (RPM) values were defined as expression levels, and then converted to $\log_2$ (RPM+1) values and averaged for the two biological replicates.

ChIP-seq data analyses

ChIP-seq data analysis and normalization were performed according to the previously described method (13) with modification. The BLIMP1 occupancy levels (see also the Supplementary materials and methods section) were confirmed to be reproducible in the biological replicates of PRP, emIE and PB cells, and d2 and d6 PGCLCs (Figure 2C and D; Supplementary Figure S3F), and were averaged between the replicates for the subsequent analyses. For E12.5 male/female PGCs, we used the occupancy levels of single samples for the analyses. We defined the 'BLIMP1-binding sites' as the peaks with occupancy levels >6.

The classes of the BLIMP1-binding sites detected in PRP cells, emIE cells, PBs and d2 PGCLCs were defined as follows: those exclusively detected in one cell type were defined as the 'specific sites'; those detected in two or three cell types as the 'shared sites'; and those detected in all four cell types as the 'common sites' (Figure 3B and Supplementary Figure S4A). The differential read densities at the specific sites were also confirmed with EdgeR software (Supplementary Figure S4B) (34). The enrichment or depletion of each class of binding sites (Figure 3E) was calculated by comparing the observed and randomly expected distribution of each binding-site class.

We defined BLIMP1-binding sites that were 'proximal to TSSs' (ref_GRCm38) (within 1 Kb), 'around genes' (within 15 Kb from TSSs) and 'at large distances' (>15 Kb) (Figure 4 and Supplementary Figure S4D). We defined those located at a distance of more than 50 Kb as 'far sites'. We defined BLIMP1-associated genes according to binding-site class within 15 Kb of the TSSs. GO terms were analyzed using the DAVID tool (35).

Analyses of sequence properties

We performed a sequence motif search for genomic regions within 1 Kb of the center of BLIMP1-binding sites using the MEME-ChIP program suit (MEME v4.10.1) (36). The best-matched sequences of TF motifs were analyzed for the regions within 200 bp of the center of the binding sites and in random regions, using position frequency matrices (PFMs) (37) from the JASPAR database (Figure 5 and Supplementary Figure S5).

The GGGAAA repeats were considered to be present if this hexamer appeared 10 times or more in a 72-bp window (i.e., 12 × 6 bp) contained in the BLMP1-binding sites or non-overlapping 2-Kbp tiles of the whole genome.

Expression analyses

We identified genes significantly expressed in at least one cell type examined in this study (expressed genes) [maximum $\log_2$ (RPM+1) >4, 12 876 genes], and used these $\log_2$ expression levels in the subsequent analyses. Among the 12 876 significantly expressed genes, we defined genes associated with the three binding-site classes (specific, shared and common) using the same criteria described above (Supplementary Figure S7C).

Differentially expressed genes (DEGs) during the developmental processes examined (photoreceptor differentiation, perinatal IE development, terminal B cell differentiation, germ cell specification *in vitro* and germ cell development *in vivo*) were defined with the following three criteria: (i) $\log_2$ expression levels >4 in at least one cell type (average of biological replicates) during the examination process; (ii) significant variance with one-way ANOVA test ($P < 0.05$) in at least one developmental process; and (iii) more than 2-fold expression-level differences in at least one pairwise comparison among the cell types (average of biological replicates) within each process.

DEGs were classified by k-means clustering using the differences from the mean of expression levels ($k = 2$ for photoreceptor differentiation, perinatal IE development, terminal B cell differentiation and germ cell development; $k = 3$ for germ cell specification). Such clusters were classified into either those positively or those negatively correlated with the *Blimp1* level (Figure 6C).

Integration of ChIP-seq and gene expression

We calculated the enrichment level of DEGs associated with each class of the BLIMP1-binding sites during the developmental processes (Figure 6C and D) as the ratio of BLIMP1-binding-site-associated genes between DEGs and all genes. The enrichment levels of genes up/downregulated by *Blimp1* mutation associated with the BLIMP1-binding sites (Figure 6E) were similarly calculated.

Sequence motifs and gene expression

We calculated enrichment levels of DEGs associated with cell type-specific sites containing recognition motifs of BLIMP1 and other TFs (Figure 6F) as the ratios between the percentages of DEGs associated with the BLIMP1-binding sites bearing TF motifs and those of all DEGs associated with the BLIMP1-binding sites.

Analyses of common BLIMP1 targets

The DEGs were classified on the basis of the number of developmental processes in which the genes were consistently correlated with the *Blimp1* level, and genes associated with the common BLIMP1-binding sites (<15 Kb from TSSs) were identified (Figure 7).

Analyses of BLIMP1-binding sites in the late PGCs

We identified 8776 BLIMP1-binding sites in E12.5 male or female PGCs (late PGCs). For the expression analyses, we selected genes significantly expressed in at least one cell type among EpiLCs, d2-d6 PGCLCs and E9.5-E13.5 male/female PGCs [$\log_2$ (RPM+1) >4, 9339 genes]. Among these genes, 1579 and 1990 were associated with the BLIMP1-binding sites in the d2 PGCLCs and late PGCs, respectively (within 15 Kb from TSSs). This gene set was used in the analyses shown in Figure 8E and F.

RESULTS

Expression of EGFP-BLIMP1 in embryos and adults

We have established homozygous knock-in mice that bear an epitope tag, an enhanced green fluorescence protein (EGFP), at the N terminus of *Blimp1* (EGFP-BLIMP1 mice) (13) and analyzed the EGFP-BLIMP1 expression in embryos and adults (8–10 weeks old) using immunofluorescence (IF) staining with an anti-GFP antibody (Figure 1; Supplementary Figures S1 and 2). We detected EGFP-BLIMP1 in the nuclei of all cells previously shown to express BLIMP1, including cells of the surface ectoderm, ectoderm and mesenchyme of the branchial arch, limb bud mesenchyme and midgut endoderm (emIE), as well as placental spongioblasts, otocysts, myotomes, PGCs, PRP cells in the retina, dental papilla, keratinocytes of the hair shafts and cells of the surface layer of the developing tongue epithelium (Figure 1 and Supplementary Figure S1) (38–40).

Moreover, we detected EGFP-BLIMP1 in the simple squamous epithelium within the embryonic ventricular myocardium, lung mesenchyme, kidney glomerulus and neural epithelium (Figure 1 and Supplementary Figure S1), which were only found in embryos and were most likely premature vascular endothelia, since they were positive for platelet-endothelial cell adhesion molecules (PECAM) and had squamous morphology (Figure 1; Supplementary Figures S1 and 2). We also detected EGFP-BLIMP1 in granular layers of the stratified squamous epithelia of the epidermis, esophagus, and vagina, and migratory cells in the lamina propria of the small intestine, which were most likely plasma cells (Supplementary Figure S2). These results reinforced previous findings and identified novel cell types expressing BLIMP1, demonstrating that the N-terminal EGFP tag was appropriately integrated in all cellular contexts examined.

The landscapes of BLIMP1 bindings

Using a recently developed ChIP-seq method (13), we went on to evaluate the genomic occupancy profiles of BLIMP1 in the following purified cell types from EGFP-BLIMP1 mice: (i) PRP cells [EGFP-positive (+), postnatal day (P) 4], in which BLIMP1 is highly expressed before its downregulation upon differentiation to rod cells (17,18); (ii) emIE cells [EGFP (+)/EpCAM (+), E16.5], in which BLIMP1 is expressed to prevent premature remodeling to adult-type IE before its downregulation upon the suckling-weaning transition (14,19,20); (iii) PBs [EGFP (+)/CD138 (+), day (d) 3 of induction from splenic B cells], in which BLIMP1 represses the B cell program and activates a part of the plasma cell program (15,21); (iv-a) PGCLCs, in which BLIMP1 represses somatic programs (d2 PGCLCs, which recapitulate PGC specification with a transient expression profile similar to that of the early mesoderm at around E6.5; and d6 PGCLCs, which recapitulate established PGCs at E9.5) (13); and (iv-b) E12.5 male/female PGCs [EGFP (+)/SSEA1 (+)], the maintenance of which requires BLIMP1 (23) (Figure 2A and Supplementary Figure S3A–D). A consistent use of the single epitope tag, antibody, library construction method, and sequencing platform allowed a quantitative comparison of the BLIMP1-binding profiles among these cell types (Supplementary Figure S3E and Supplementary Tables S1–4).

We obtained EGFP-BLIMP1 peaks with a sufficient signal-to-noise ratio, and quantified their read-tag densities (normalized IP/input for the regions within 500 bp from the peak centers; BLIMP1 occupancy levels hereafter) ('Materials and Methods' section). Pairwise comparisons revealed reproducibility between the biological replicates (Supplementary Figure S3F), and consistency with the reported data for the whole embryonic intestine and PBs (14,15) (Supplementary Figure S3G and H). Nonetheless, we detected substantial quantitative differences between our datasets and those reported previously, which were likely attributable to the methodological differences, reinforcing the importance of performing analyses under the same experimental conditions. As shown in Figure 2B, ChIP-seq tracks in the genomic loci proximal to *Myc*, a canonical target of BLIMP1 and around the *Hoxa* cluster revealed binding properties characteristic to the four cell types.

We found that the BLIMP1-occupancy levels of PGCLCs and PGCs at E12.5 were highly correlated (Figure 2C and Supplementary Figure S3I), and decided to use the datasets for the d2 PGCLCs as representative of the BLIMP1 binding of PGCs, since they capture the immediate BLIMP1 binding upon PGC specification. On the other hand, PRP cells, emIE cells, PBs and d2 PGCLCs exhibited highly varied occupancies of BLIMP1 as revealed by scatterplots, correlation coefficients and principal component analysis (PCA) (41) (Figure 2C, D and Supplementary Figure S3J), demonstrating that the BLIMP1-binding profiles vary widely among the cell types.

Defining specific, shared and common BLIMP1-binding sites

Specific, shared and common BLIMP1-binding sites. We identified 7203, 3751, 5150 and 6385 BLIMP1-binding sites with significant occupancy levels (>6) in PRP cells, emIE cells, PBs and d2 PGCLCs, respectively (Figure 3A and Supplementary Table S5), and defined three binding-site classes: (i) those exclusively detected in one cell type (spe-

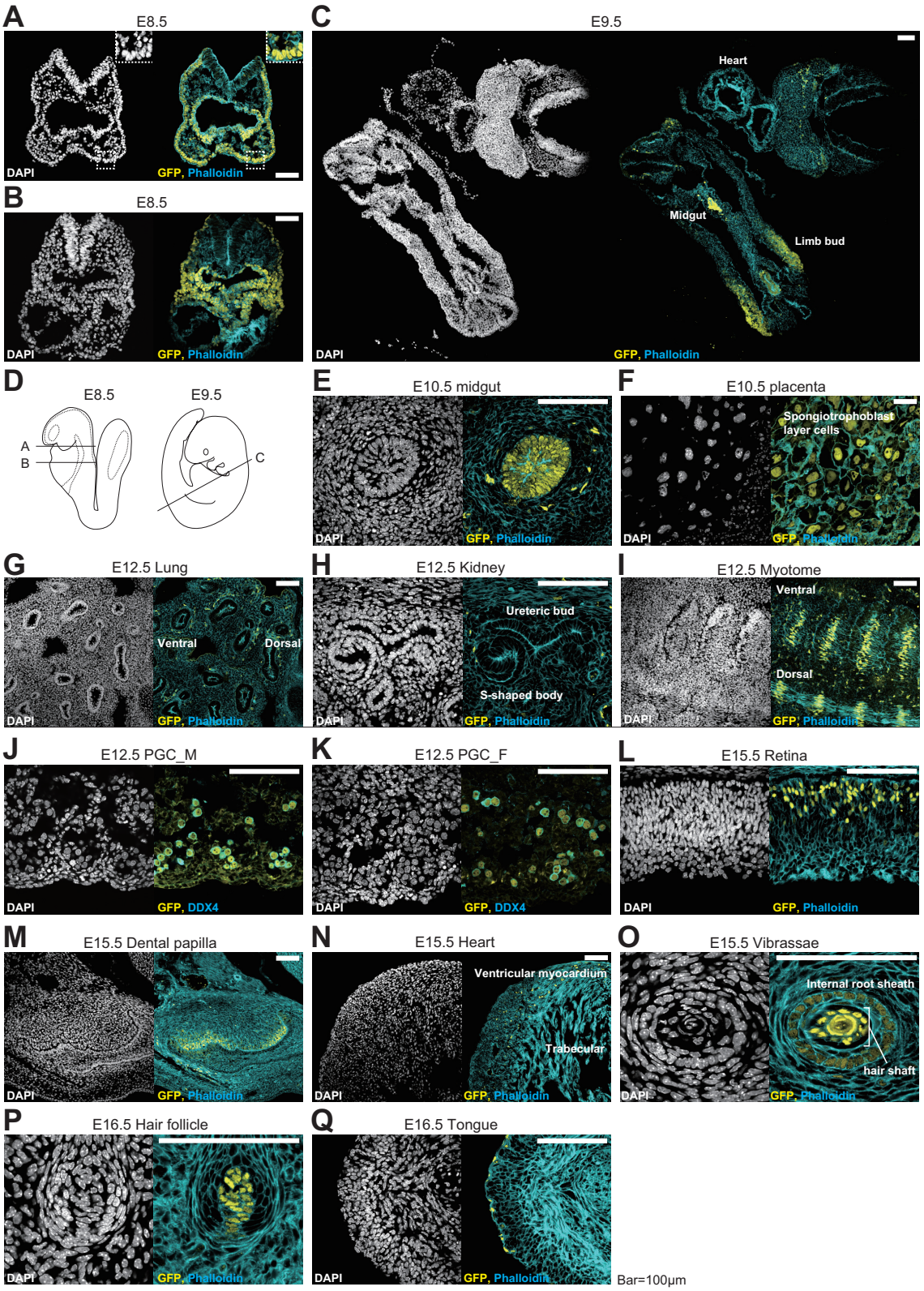

Figure 1. Exploration of EGFP-BLIMP1 protein in the three germ layer derivatives, germ cells and trophoblast cell lineage. (**A**–**D**) IF analyses of GFP in whole embryos at E8.5 (A and B) and E9.5 (C) sectioned as illustrated (D). Panel (C) was prepared by piecing together high resolution images. (**E**–**Q**) IFs of GFP in the E10.5 midgut (E) and placenta (F), the E12.5 lung (G), kidney (H), myotome (I) and male (J) and female (K) PGCs, the E15.5 retina (L), dental papilla (M), heart (N) and vibrassae (O), and the E16.5 hair follicle (P) and tongue (Q). Counterstaining was performed with 4',6-diamidino-2-phenylindole (DAPI) for DNA, and phalloidin for actin (A-I, L-Q) or DDX4 (J and K). Bar: 100 μm.

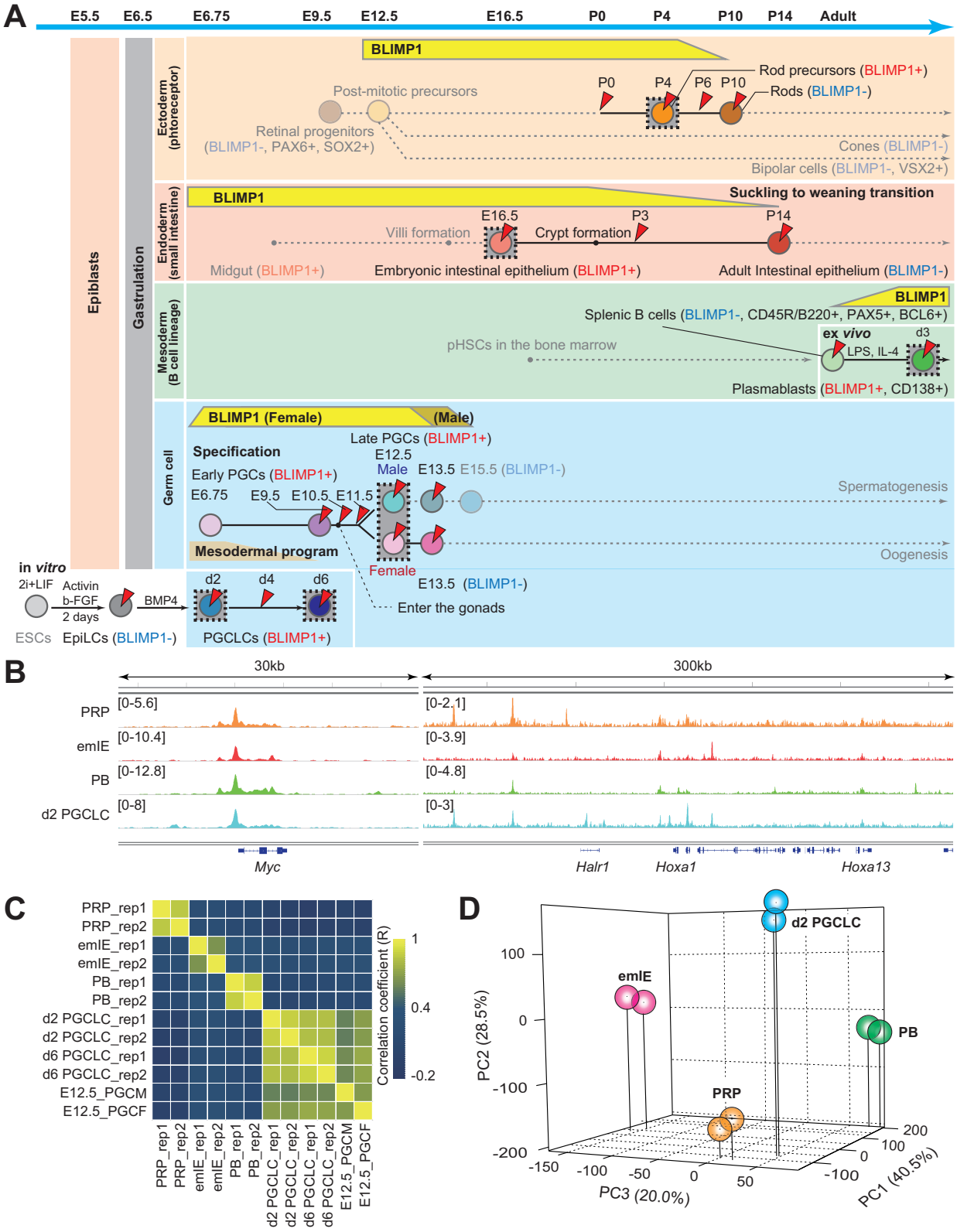

Figure 2. Overview of the BLIMP1-binding profiles. (**A**) Scheme of ChIP-seq and RNA-seq in this study. The cell types used for ChIP-seq analyses are indicated with colored circles enclosed by dotted lines, and those used for the RNA-seq are indicated with red arrowheads. (**B**) ChIP-seq tracks of EGFP-BLIMP1 around the *Myc* (30 Kb) and *Hoxa* cluster (300 Kb). (**C**) Heat map of the correlation coefficients of BLIMP1-occupancy levels among the cell types examined, with color codes as indicated. (**D**) PCA (PC1-PC3) of BLIMP1-occupancy levels in PRP cells, emIE cells, PBs and d2 PGCLCs.

cific sites: 4739, 1436, 2911 and 4051, respectively); (ii) those shared in two or three cell types (shared sites: 1610, 1461, 1385 and 1480, respectively); (iii) and those common to all four cell types (common sites: 854) (Figure 3A, B and Supplementary Figure S4A, 'Materials and Methods' section). The scatterplot comparisons and the PCA of the BLIMP1-occupancy levels clearly delineated the characteristics of the three binding-site classes (Figure 3C and D); in the three-dimensional (3D) space defined by the PC1–3 axes, the specific sites created coherent clusters representing the binding characteristics of individual cell types, the shared sites exhibited a sparse distribution between/among the clusters of the specific sites for the shared cell types, and the common sites were plotted around the origin of the PCA (Figure 3D).

As shown in Figure 3B, while the numbers of the specific sites varied widely, those of the shared sites were relatively uniform (there were between 1385 and 1610 shared sites); consistently, the numbers of the overlapped binding sites in pairwise comparisons were relatively constant (from 1364 to 1638) (Supplementary Figure S4C), indicating that there was essentially no bias in cell types for the shared binding sites. The numbers of specific and common sites were enriched in comparison to those expected from a random distribution among all the binding sites (∼2- and >4-fold, respectively), while the shared sites were largely depleted (Figure 3E, 'Materials and Methods' section), suggesting that the allocation of the binding-site classes reflects a regulatory consequence rather than a random distribution. We found that BLIMP1 exhibited the highest occupancy levels in the common sites, an intermediate level in the shared sites, and the lowest level in the specific sites (Figure 3F). Pairwise comparisons revealed that the BLIMP1 occupancy levels of the shared/common sites were similar between cell types (fold difference <2 for 70–90%) (Figure 3G), indicating that BLIMP1 binds such sites at nearly constant levels in the four cell types.

Genomic distributions of the three binding-site classes. We next investigated the genomic distributions of the three binding-site classes. We explored the distances of the BLIMP1-binding sites from the TSSs, and found that as a whole, they were highly enriched in the genomic regions around TSSs (<15 Kb), particularly in the most proximal regions (<1 Kb) (Figure 4A and Supplementary Figure S4D). The shared/common sites exhibited this trend in a pronounced fashion (>59% of the common sites, and 52–73% of the shared sites were within 15 Kb of the TSSs, respectively), whereas, notably, a substantial proportion (46–66%) of the specific sites were also located at more distal regions (Figure 4B and C). We also examined the local densities of the binding sites, and found that, irrespective of the binding-site classes, the vast majority (77–88%) were distributed in a solitary manner (>10 Kb from the next-nearest neighbor) (Supplementary Figure S4E–G); nonetheless, we found an enrichment of the binding-site clusters in the regions around TSSs (1–15 Kb) (but not in the region most proximal to the TSSs) in all four cell types (Supplementary Figure S4H).

We next examined the association of the BLIMP1-binding sites with the genes (for simplicity of the assignment of the bindings to the genes, we defined the binding sites <15 Kb from the TSSs as those associated with the genes). A majority of the specific, shared, and common binding sites were associated with distinct sets of genes, respectively, in all four cell types (Figure 4D), and notably, the genes associated with the specific sites were highly specific to the respective cell types (Figure 4E) and were enriched with gene ontology (GO) functional terms (42) related to the respective developmental processes: e.g. 'photoreceptor cell development' for PRP cells; 'positive regulation of macromolecule metabolic process' for emIE; 'regulation of lymphocyte activation' for PBs; and 'embryonic morphogenesis', which was repressed in PGCs (43), for d2 PGCLCs (Figure 4F). On the other hand, the genes associated with the shared/common sites exhibited a high enrichment with GO terms for more general processes such as 'transcription, DNA-templated' (Supplementary Figure S4I).

Sequences recognized by BLIMP1. We next went on to investigate the DNA recognition mechanism by BLIMP1, and performed a motif discovery using the MEME-CHIP program (36). We identified highly enriched sequences similar to the canonical recognition motif of BLIMP1 (6,11) [MA0508.1 in the JASPAR database (44)] in all the binding-site classes in the four cell types (Figure 5A). To examine the actual distribution of this motif, we sought for the best-matched sequences in the regions within 200 bp from the center of the binding sites and in random loci, and calculated their similarity to this motif. We consider that, if a sequence contained in a binding site of a TF and sequences abundantly found in the genome showed the same similarity level to the recognition motif of the TF, such a sequence in the TF-binding site would not explain the binding of the TF. Thus, the similarity level to a motif in a binding site was evaluated by the probability of finding sequences that showed such a similarity in the random loci [an occurrence rate in random loci (ORR)]. We then plotted the percentages of the occurrence of the best-matched sequences in the binding sites against their ORR scores, and defined the 'motif-matched sequences' by ORR < 0.1 (Figure 5B) ('Materials and Methods' section).

We found that the distributions of the ORRs were highly varied among the binding-site classes (Figure 5B and C). Notably, the shared/common sites were highly enriched with these sequences (50–74%), particularly those with very low ORRs (ORR < 10^{-3}) (Figure 5B and C). In contrast, the motif-matched sequences were relatively depleted in the PRP- and PB-specific sites (26 and 41%, respectively), while they were more pervasive in the emIE- and d2 PGCLC-specific sites (54 and 55%, respectively) (Figure 5B). Nonetheless, binding sites without the motif-matched sequences (ORR > 0.1) still contained sequences somewhat similar to the BLIMP1 motif (Figure 5D). Interestingly, the presence or absence of the motif-matched sequences had little effect on the BLIMP1 occupancy levels in all binding-site classes (Supplementary Figure S5A), suggesting that sequence recognition by BLIMP1 may be relatively flexible.

We identified a small number of the common binding sites bearing a unique repeat of the GGGAAA hexamer (the hexamer was repeated ≥10 times) with ORRs of 5∼6 × 10^{-3} ($N = 155$) (Figure 5A, C and D; Supplementary Figure

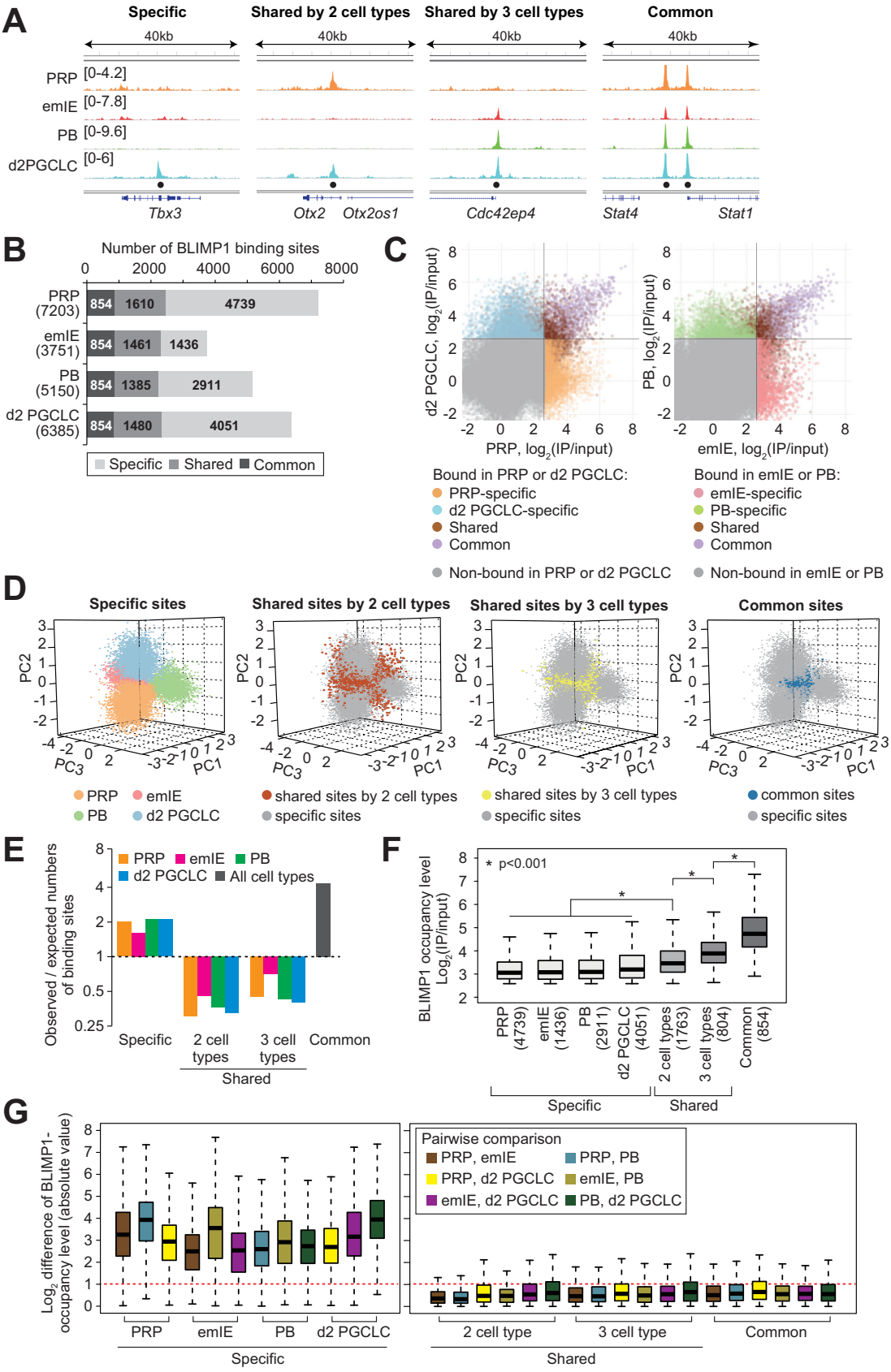

Figure 3. Specific, shared and common BLIMP1-binding sites and their BLIMP1 occupancy levels. (**A**) ChIP-seq track of EGFP-BLIMP1 around *Tbx3*, *Otx2*, *Cdc42ep4* and *Stat1*, which represent BLIMP1-binding sites specific to one cell type (emIE), shared among two (PRP and d2 PGCLCs) or three

S5B). In fact, BLIMP1 very highly occupied the majority of such repeats identified in the genome (58%, 155/268) (Supplementary Figure S5C), and thus these repeats formed a distinct subclass of the common binding sites. This subclass was predisposed to locate at large distances from the TSSs (>15 Kb for 85%) (Supplementary Figure S5D). Nonetheless, a subset of the repeats were distributed around TSSs or within gene bodies of 81 genes, which were enriched with genes related to 'spermatogenesis', including *Prdm9*, a determinant of meiotic recombination hotspots (Supplementary Figure S5E and F).

Next, we found that some sequence motifs discovered in the specific sites were similar to recognition motifs of key TFs for the respective lineages [e.g. Cone-rod homeobox (CRX) in PRP cells, CDX2 and HNF4A in emIE cells, PU.1 (also known as SPI1) in PBs, and TFAP2C and SOX2 in d2 PGCLCs] (Figure 5E and Supplementary Figure S5G). Among these, the CRX motif was abundant in the PRP-specific sites (37%), with an abundance greater than that of the BLIMP1 motif (26%) (Figure 5F). We also found a long terminal repeat (RLTR22), bearing motifs of CDX2 and HNF4A, two key IE regulators, nearly exclusively in the emIE-specific sites (3.7%, 53/1436) (Supplementary Figure S5H and I). The specific distances and orientations between the motifs, which are hallmarks for 'on-DNA TF-TF interactions' (45), were not found in most pairs of BLIMP1 and the other TF motifs, except for a joint motif with PU.1 in the PB-specific sites (1.8%, 51/2911) (Supplementary Figure S5J). Interestingly, the evolutionary conservation of PRP-specific sites was as high as that of all TSSs (Supplementary Figure S5K), implying the importance of their sequence properties. The importance of distal specific binding sites (>50 Kb from TSSs) was highlighted by their evolutionary conservation, which was nearly as high as that proximal to (<1 Kb from) the TSSs (Supplementary Figure S5L).

Collectively, these findings indicate that BLIMP1 binds to the genome in at least three distinct modes. First, BLIMP1 binds to the shared/common sites with constant, relatively high occupancy levels: such sites are similar in number among the cell types, are enriched with sites more proximal to TSSs, and bear relatively stringent recognition sequences. Second, a subset of the common sites covered a majority of GGGAAA repeats in the genome with very high occupancy levels of BLIMP1: such sites are located at large distance from the nearest TSSs. Third, BLIMP1 binds to the specific sites with lower occupancy levels: such sites are diverse in number among the cell types, are located more distal to TSSs, and show less stringent recognition sequences.

The relationship between BLIMP1-binding and gene expression

Gene expression dynamics of the four lineages. Using an RNA-sequencing (RNA-seq) method (29,30), we next determined the gene expression dynamics during the relevant developmental processes of the four lineages: (i) for photoreceptor development, rod cell precursors [P0 EGFP (+), P4 EGFP (+), P4 CD73 (−)] and early rod cells [P4/P6/P10 CD73 (+)]; (ii) for perinatal IE development, IE cells at E16.5, P3 and P14; (iii) for PB differentiation, B cells, Pre-PB and PBs; and (iv) for germ-cell development, epiblast-like cells (EpiLCs), d2 to d6 PGCLCs and E9.5 to E13.5 PGCs (Figure 2A; Supplementary Figure S6A–G and Tables S1–4) (28,46,47) ('Materials and Methods' section). The RNA-seq data were highly reproducible between biological replicates and showed a good dynamic range for the efficient detection of expression dynamics, as previously reported (29,30) (Supplementary Figure S6G). We used genes with expression levels of $\log_2$ [RPM (reads per million mapped reads)+1] > 4, which corresponds to ~10 copies per cell (29), in at least one cell type (12 876 genes) for the subsequent analyses. Figure 6A shows the expression dynamics of key genes for the respective lineages. Note that during photoreceptor and IE development, *Blimp1* was initially expressed at high levels and was rapidly downregulated, whereas during PB and germ-cell differentiation/development, *Blimp1* was acutely upregulated upon their differentiation/development and maintained thereafter (Figure 6A).

The unsupervised hierarchical clustering (UHC) and PCA successfully classified the developing cells of the four lineages (Figure 6B and Supplementary Figure S7A). We identified the DEGs during the development of each lineage (difference of the expression levels >2-fold in at least one pairwise comparison, ANOVA, $P < 0.05$). Interestingly, by *k*-means clustering we also found that the DEGs could be categorized into two major classes, those monotonically and progressively up- or downregulated, for photoreceptor, perinatal IE and PB development, and into three major classes, those monotonically and progressively up- or downregulated or those that show highly acute transient upregulation followed by subsequent downregulation, for germ-cell development (Figure 6C) ('Materials and Methods' section). These findings demonstrate that the four developmental processes reflect a typical cell-state transition in which part of the initial character of the cells is lost and instead a novel character is gained in a relatively straightforward fashion.

BLIMP1 regulates gene expression through specific binding sites. Among the total 12 876 expressed genes, around 15–

(emIE, PBs and d2 PGCLCs) cell types, and common to all cell types, respectively. (**B**) Stacked bar graphs showing the number of specific, shared and common BLIMP1-binding sites in PRP cells, emIE cells, PBs and d2 PGCLCs. (**C**) Scatterplots of BLIMP1-occupancy levels between PRP cells and d2 PGCLCs (left) and between emIE cells and PBs (right). The binding-site classes are color coded as indicated. (**D**) Scatterplots of factor loadings (PC1-PC3) of BLIMP1-occupancy levels for the specific sites (leftmost), the sites shared by two (second leftmost) or three (second rightmost) cell types, and the common binding sites (rightmost). The binding-site classes are color coded as indicated. (**E**) Bar graphs showing the ratios between the observed numbers of BLIMP1-binding sites and those expected from the random distribution for each binding-site class. (**F**) Boxplots for BLIMP1-occupancy levels in the indicated binding-site classes. The numbers of sites are indicated in the brackets. Asterisks indicate statistical significance by *t*-test (two-sided $P < 0.001$). (**G**) Boxplots for the absolute $\log_2$ difference of BLIMP1-occupancy levels in pairwise comparisons of specific (left) and shared/common sites (right).

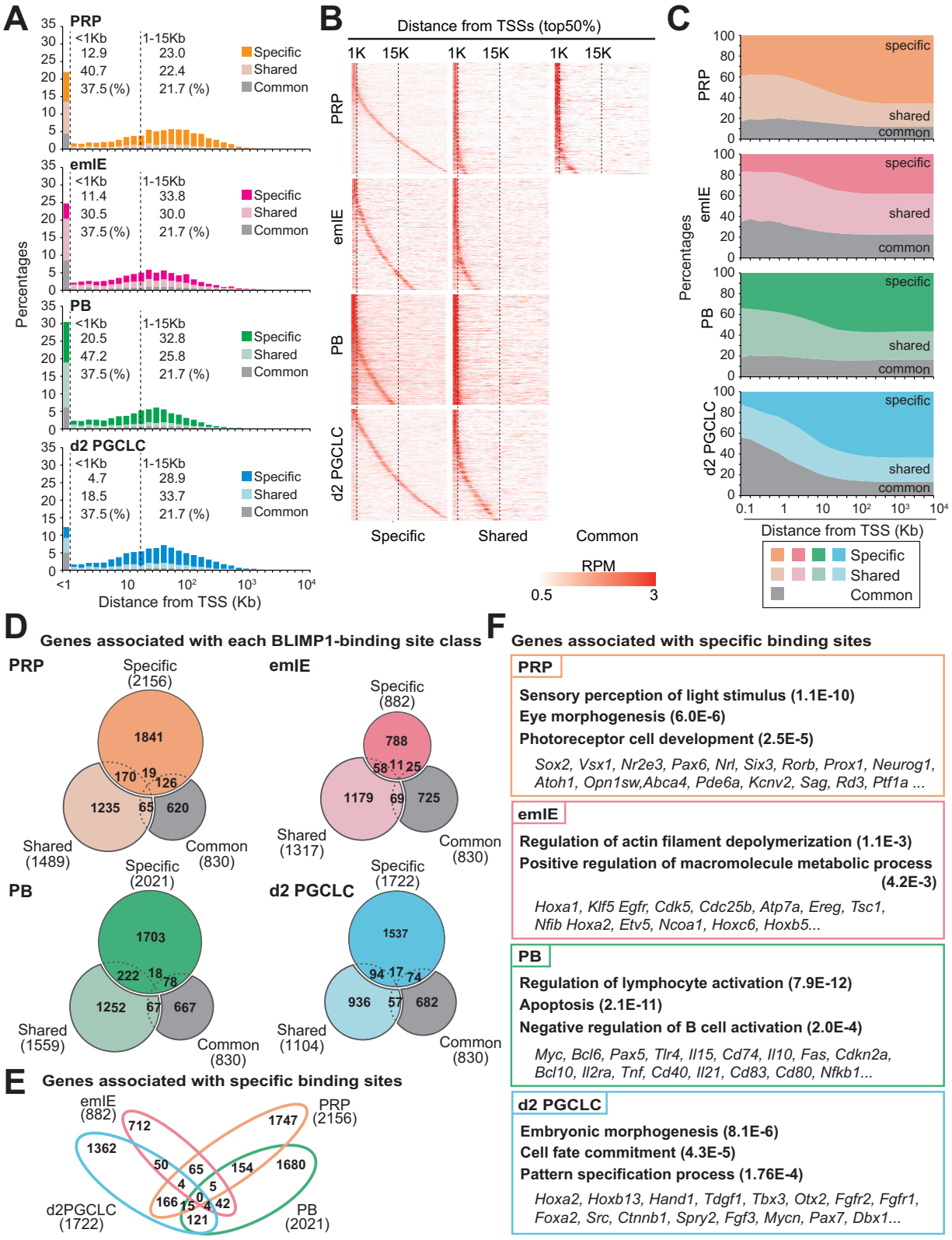

Figure 4. Genomic distribution and gene association of BLIMP1-binding sites. (**A**) Histograms of BLIMP1-binding sites against the $\log_{10}$ distance (Kb) from TSSs. Binding-site classes in the stacked bars are colored as indicated, and the percentages of the sites within 1 Kb and within 1–15 Kb for each class are shown to the left of the color-code symbols. The dotted lines indicate 15 Kb from the TSSs. (**B**) Heat maps of the top 50% of BLIMP1-binding sites sorted by distances from the nearest TSSs. The distances of 1 and 15 Kb are indicated with dotted lines. (**C**) Area charts showing the percentages of binding-site classes located within the indicated distances from TSSs. (**D**) Venn diagrams showing the definition of genes associated with each BLIMP1-binding-site class (solid lines). The numbers of TSSs located within 15 Kb of the nearest binding sites are indicated. (**E**) Venn diagrams showing the overlap of genes associated with the specific binding sites in different cell types. (**F**) Selected GO terms and lists for genes associated with the specific binding sites.

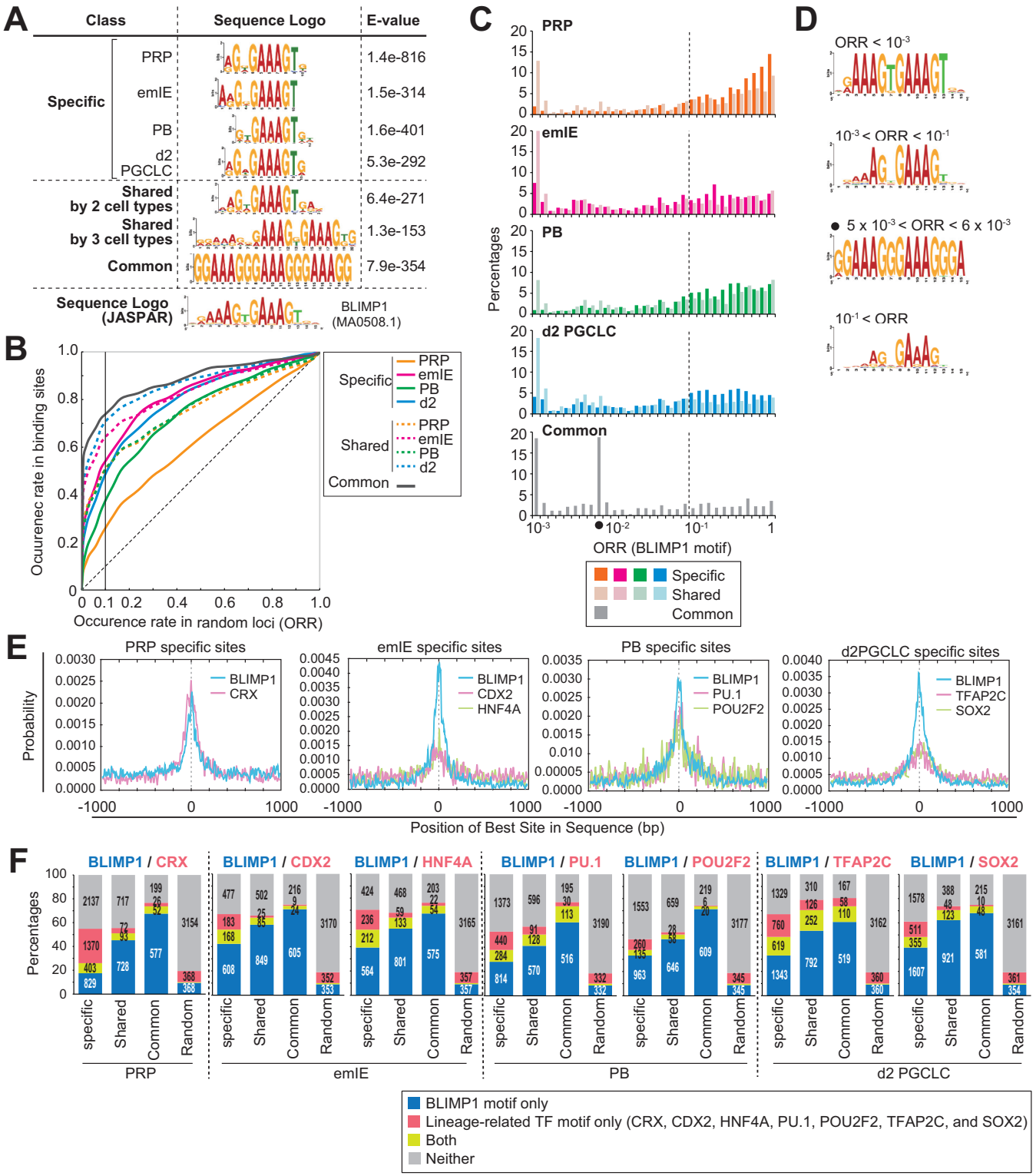

Figure 5. Sequence properties of BLIMP1-binding sites. (**A**) Identified sequence motifs enriched in the respective binding sites and their *E*-values. The BLIMP1 motifs registered in the JASPAR database are shown on the bottom. (**B**) Occurrence-rate plots of the sequences best-matched to the BLIMP1 motif in the binding sites against the ORRs. The vertical solid line indicates ORR 0.1. The dotted diagonal line represents coverages that equal the ORR (i.e. coverages in random genomic loci). Binding-site classes are color coded as indicated. (**C**) Histograms of ORRs for the sequences best-matched to the BLIMP1 motif in the indicated binding-site classes. Dotted lines indicate an ORR of 0.1. The black closed circle indicates ORRs between 5×10^{-3} and 6×10^{-3}, which were highly enriched with the GGGAAA repeat. (**D**) Sequence logos corresponding to the indicated ORR values. (**E**) Enrichment of the indicated TF motifs in specific sites. (**F**) Stacked bar graphs showing the frequencies of the combinations of the BLIMP1 motif and the indicated TF motifs in the binding sites (ORR < 0.1). The numbers of binding sites are indicated.

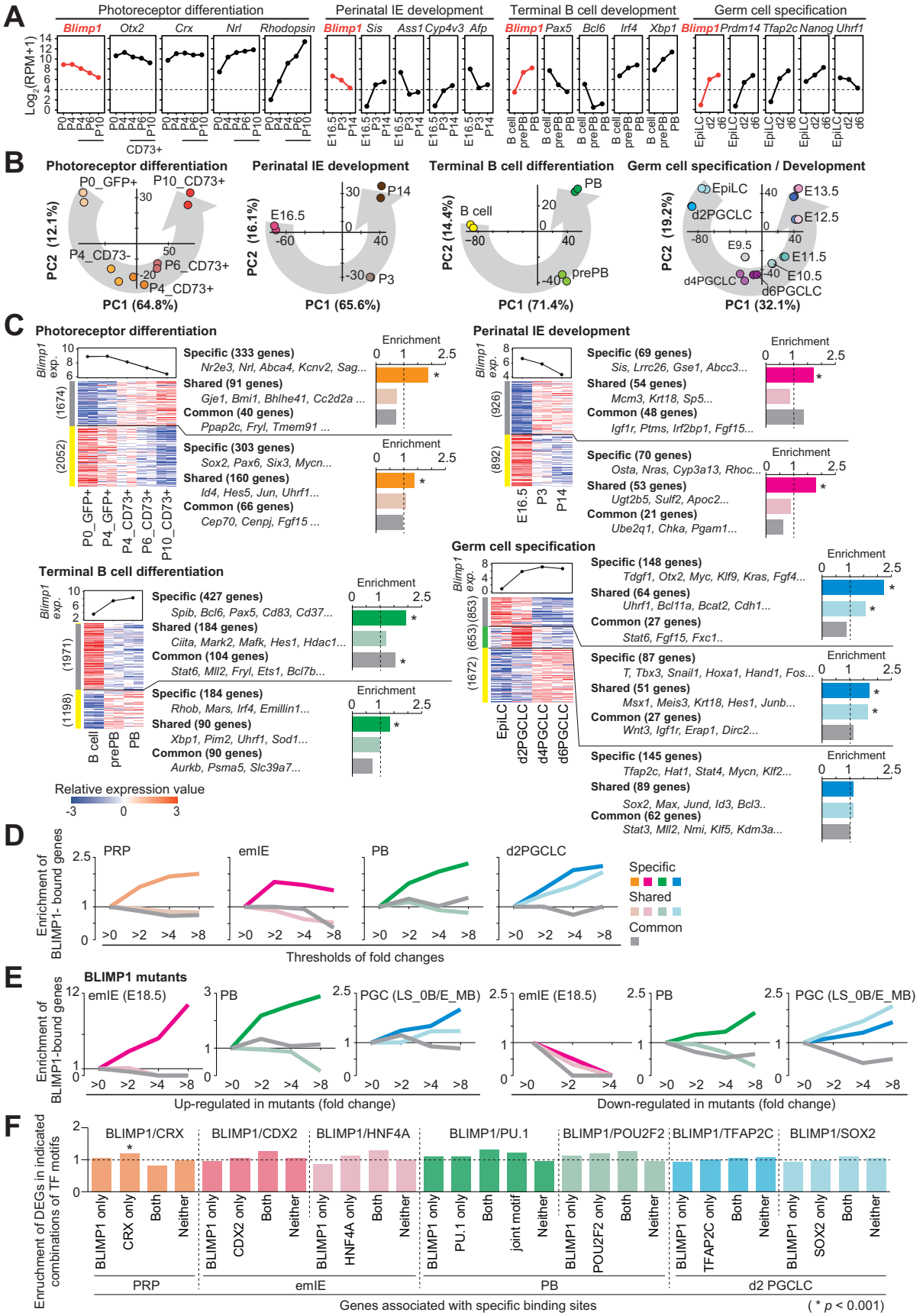

Figure 6. Impact of BLIMP1 binding on gene expression. (**A**) Line graphs showing the expression dynamics of key genes in the indicated developmental processes. (**B**) Scatterplots of PC scores (PC1 and PC2) of the individual developmental processes. The cell types are represented with closed circles and

22% were associated with the BLIMP1-binding sites (<15 Kb from the TSSs) in the four cell types (Supplementary Figure S7B), and those associated with the specific, shared and common binding sites were largely non-overlapped and showed a wide range of expression levels similar to the range for the total group of expressed genes (Supplementary Figure S7C and D). By examining the enrichment levels of genes associated with the BLIMP1-binding sites in DEGs, we made two general findings. First, among the genes associated with the specific, shared and common binding sites, those associated with the specific binding sites were selectively enriched in DEGs (photoreceptor, perinatal IE and PB development) (Figure 6C) ('Materials and Methods' section), and such DEGs were highly specific to the respective developmental processes (Supplementary Figure S7E). Second, while the specific BLIMP1 bindings were associated with DEGs in general, they tended to be more enriched around DEGs negatively correlated with the *Blimp1* level [i.e. genes upregulated when *Blimp1* was repressed (photoreceptor development) or genes downregulated when *Blimp1* was upregulated (PB and germ-cell development)] (Figure 6C). An exception to the first general finding was that genes associated with the shared binding sites in PGCLCs were also enriched in DEGs, and an exception to the second general finding was that during the perinatal IE development, the genes with specific BLIMP1 bindings were both up- and downregulated at similar enrichment levels when *Blimp1* was repressed (Figure 6C). Moreover, we found that the enrichment levels of genes associated with the specific (all four cell types) or shared (PGCLCs) BLIMP1-binding sites in DEGs were progressively increased along with increases in the degree of differential expression (Figure 6D).

We next analyzed the effects of the loss of *Blimp1* on the expression of genes bound by BLIMP1 in emIE cells (19), PBs (15) and PGCs (43) ('Materials and Methods' section). The genes misregulated in the *Blimp1* mutants were highly enriched with those bearing specific, but not shared/common (except for d2 PGCLC shared sites), BLIMP1-binding sites (Figure 6E) and the enrichment levels were increased progressively with the increase of the level of misregulation (Figure 6E): the specific sites were associated with 26% (5/19; IE development), 34% (65/189; PB differentiation) and 12% (74/615; PGC specification) of genes upregulated in the *Blimp1* mutant (>4-fold), and 18% (25/140; PB differentiation) and 10% (30/288; PGC specification) of the downregulated genes (<1/4-fold) (Supplementary Table S6). The shared sites in d2 PGCLCs were linked with 7% (40/615) and 8% (23/288) of the up- and downregulated genes, respectively (Supplementary Table S6). Thus, these findings demonstrate that, primarily, BLIMP1 represses the genes through binding to the specific, but not shared/common, binding sites.

Sequence motifs and gene regulation. We next analyzed the impacts of the motif-matched sequences of BLIMP1 and other TFs in the specific BLIMP1-binding sites on gene expression. Strikingly, the presence/absence and combinations of such sequences did not show a major influence on the association between the specific sites and the DEGs during the normal development (Figure 6F), raising a possibility that BLIMP1 may act on gene expression independent of the stringent sequence recognitions or collaboration with the other TFs. Nonetheless, we found a slight enrichment of the DEGs among genes associated with the emIE- and PB-specific sites that contained motifs of both BLIMP1 and other TFs (Figure 6F). The CRX motifs were frequently detected without a BLIMP1 motif in the PRP-specific sites (29%), and were significantly enriched with DEGs in the photoreceptor differentiation ($P < 0.001$); such sites were highly associated with the genes for photoreception ('Visual perception'; *Rbp3*, *Gucy2f*) (Supplementary Figure S7F), implying that BLIMP1 may block the maturation of CRX-mediated photoreceptor programs. The RLTR22s exhibited no association with the DEGs, but were located in the neighborhood of *B4galnt2* and *Fut8* (8.3 and 37.4 Kb from TSSs, respectively), which encode enzymes for protein glycosylation expressed in the adult IE (Supplementary Figure S7G).

Common BLIMP1 targets. Our analyses so far indicated that genes bearing the common sites were not enriched around the DEGs in normal development, or around the genes affected by *Blimp1* mutation (Figure 6C–E). Nonetheless, among the common sites, 10–19% were associated with DEGs, while 20–30% were associated with stably expressed genes ($\log_2$ RPM+1 > 4) or stably silenced genes ($\log_2$ RPM+1 < 4) (Supplementary Figure S7H). To more directly identify genes that may be commonly regulated by BLIMP1, we looked for genes that were consistently negatively or positively correlated with the *Blimp1* level, and then examined whether such genes harbored the common BLIMP1-binding sites (<15 Kb from TSSs) ('Materials and Methods' section) (Figure 7). This analysis showed that the genes showing a consistent correlation with *Blimp1* in two or more processes were fewer in number than the genes correlated with *Blimp1* in only one process; 3022, 483, 49 and 1 genes were negatively correlated, and 2600, 493, 48 and 5 genes were positively correlated, with *Blimp1* in one, two, three and four developmental processes, respectively (Figure 7). Interestingly, when the genes were consistently neg-

←

colored as indicated. The percentages of variances for PC1 and PC2 are indicated. The time courses are traced as represented by the gray round arrows. (**C**) Line graphs of *Blimp1* expression dynamics (top left), and the heat-maps of relative changes in the expression of DEGs classified by *k*-means clustering ($k = 2$ in photoreceptor differentiation, IE development and terminal B cell differentiation; $k = 3$ in germ cell specification) (bottom left). For each developmental process, 1000 randomly selected DEGs are visualized. Bar graphs for the enrichments of BLIMP1-associated genes and the lists of representative genes in the indicated classes are shown on the right. Asterisks indicate statistical significance by *Chi*-squared test ($P < 0.001$). The color code is the same as in Figure 4A. (**D**) Enrichment plots of the BLIMP1-associated genes against the indicated thresholds of fold change (the greatest changes during each developmental process). (**E**) Enrichment plots of BLIMP1-associated genes against the indicated thresholds of fold changes between the wild-type and *Blimp1* mutants for genes up- (left) and downregulated (right) in the mutants. The color code is the same as in (D). (**F**) Bar graphs of the enrichment levels of DEGs among those associated with the specific sites containing the indicated combinations of motif-matched sequences for BLIMP1 and other TFs. Asterisks indicate statistical significance by *Chi*-squared test ($P < 0.001$).

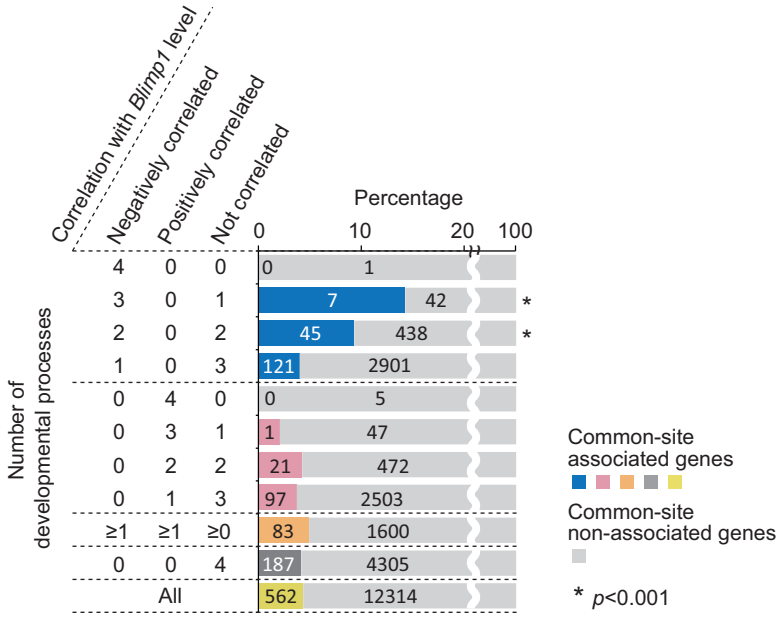

Figure 7. Common targets of BLIMP1. Stacked bar graphs showing the percentages of the common-site associated genes among genes consistently negatively (top 4, blue) and positively (fifth to eighth from the top, pink) correlated with the *Blimp1* level in the indicated number of developmental processes. Genes that showed inconsistent correlation with *Blimp1* (i.e. negatively correlated in one cell type and positively in another; ninth from the top, orange), those not correlated with *Blimp1* in all four processes (10th from the top, dark gray) and all expressed genes (bottom, yellow) are also shown. The numbers of genes are overlapped on the graphs. The asterisks indicate statistical significance by *Chi*-squared test ($P < 0.001$).

atively correlated with *Blimp1* in a greater number of developmental processes, the greater percentages of such genes were associated with the common sites; 4% (121/3022), 9% (45/483), and 14% (7/49) of the genes bore the common sites, when they were negatively correlated with *Blimp1* in one, two and three processes, respectively (Figure 7). In contrast, genes positively correlated with *Blimp1* or those without consistent correlation were associated with common sites at nearly constant percentages that were similar to the percentage for the total group of expressed genes (~4%). Thus, a small number of genes were commonly repressed, not activated, by BLIMP1. *Myc*, a prototypical target of BLIMP1 (5), was included in the seven genes that were negatively correlated with *Blimp1* in three of the four processes (*Myc*, *Klf9*, *Wls*, *Psmb10*, *Plekha1*, *Sept6*, *Mrgbp*) (Supplementary Figure S7I), and was associated with a prominent common binding site (Figure 2B).

BLIMP1-binding profile during germ cell development

Finally, we investigated how the BLIMP1-binding profiles and hence the targets of BLIMP1 might change during the development of a certain cell lineage, using the development of germ cells as a paradigm, since BLIMP1 is expressed consistently during a period of ~7 days from their specification through to sexually dimorphic development (23,48,49). We defined 8776 BLIMP1-binding sites in E12.5 male/female PGCs and compared them with those in d2 PGCLCs. First, consistent with the correlation analyses (Figure 2C), a scatterplot comparison revealed that the BLIMP1-binding profiles were significantly correlated between d2 PGCLCs and E12.5 male/female PGCs, and did not show apparent sexual dimorphism (Figure 8A). However, we identified 348 and 177 BLIMP1-binding sites that exhibited greater occupancy levels (>2-fold) in E12.5 PGCs and in d2 PGCLCs, respectively. Interestingly, the 348 sites with greater occupancy levels in E12.5 PGCs were strongly enriched with the common sites (50%, 174/348), especially those with the GGGAAA repeats (83%, 132/174) (Figure 8B), and such repeats exhibited very high occupancy levels specific to E12.5 PGCs (Figure 8C), whereas the 177 sites with greater occupancy levels in d2 PGCLCs were enriched with the specific sites (Figure 8B).

Second, we examined the genes associated with the BLIMP1-binding sites in d2 PGCLCs and E12.5 PGCs. Among 9399 genes expressed in d2 PGCLCs or in E12.5 PGCs [$\log_2$(RPM+1) > 4], 1579 and 1990 genes were associated with the BLIMP1-binding sites in d2 PGCLCs and in E12.5 PGCs, respectively, among which 1203 genes were common to d2 PGCLCs and E12.5 PGCs (common PGC genes), and 376 and 787 genes were specific to d2 PGCLCs and E12.5 PGCs, respectively (early- and late-PGC specific genes, respectively, see 'Materials and Methods' section) (Figure 8D). The early-PGC, but not late-PGC, specific genes (376 genes) were highly enriched with those bearing the d2 PGCLC-specific binding sites (Figure 8E), reinforcing the case for the involvement of BLIMP1 in a specific gene regulation upon PGC specification (13,43,48). On the other hand, we found that among the 87 genes upregulated when *Blimp1* was conditionally knocked out in late PGCs (knocked out at E10.5 and analyzed at E11.5), 60 genes (~69%) were bound by BLIMP1 at E12.5 and 90% (54/60) of such genes were also bound by BLIMP1 in d2 PGCLCs (Figure 8F). These findings indicate that a majority of key BLIMP1 targets during germ-cell development are occupied by BLIMP1 upon germ cell specification.

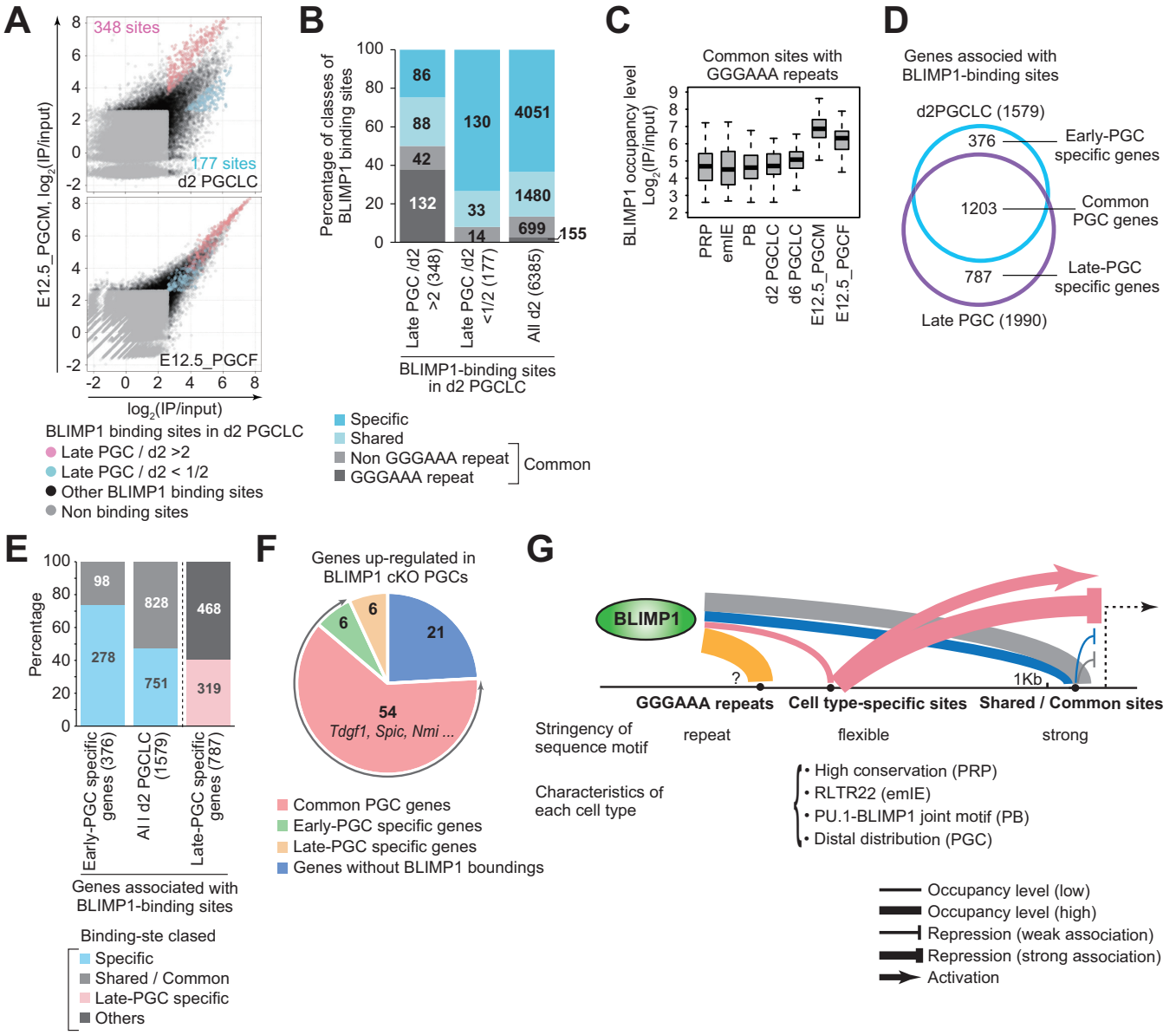

Figure 8. Transition of BLIMP1-binding sites in the late PGCs. (**A**) Scatterplots of BLIMP1-occupancy levels between d2 PGCLCs and E12.5 male and female PGCs. Binding sites in d2 PGCLCs showing higher and lower occupancy levels in late PGCs (>2-fold) are colored red and blue, respectively. (**B**) Stacked bar graphs showing the binding-site classes in d2 PGCLCs for the sites defined in (A). (**C**) Boxplots of the $\log_2$ occupancy levels of GGGAAA repeats. (**D**) Venn diagrams showing overlap of the BLIMP1-associated genes in d2 PGCLCs and late PGCs. Early- and late-PGC specific genes, and common PGC genes are indicated. (**E**) Stacked bar graphs showing the percentages of genes bearing specific sites (blue) among the early-PGC specific genes (leftmost) and all genes associated with BLIMP1-binding sites in d2 PGCLCs (middle). Note that early-PGC specific genes are included in genes associated with BLIMP1-binding sites in d2 PGCLCs. The percentages of genes bearing BLIMP1-binding sites detected in the late PGCs and not detected in the other lineages (PRP, emIE or PBs) (pink, rightmost) are also shown. (**F**) Pie chart showing the percentages of BLIMP1-associated genes among the upregulated genes in *Blimp1* cKO. The classes of the genes are color-coded as indicated. (**G**) Summary of the genomic binding patterns of BLIMP1 and gene regulations.

DISCUSSION

Using the single epitope tag, antibody and ChIP-seq protocol, we have performed a systematic comparison of the genome-wide binding profiles of BLIMP1 across four distinct developing cell types and investigated the regulatory logic of this highly versatile transcriptional regulator. Our key findings included that the specific BLIMP1-binding sites, which are divergent in number among the different cell types, locate more distally to the TSSs, show less stringent recognition sequences and lower BLIMP1 occupancy levels, and contribute more effectively to the regulation of downstream genes, than the shared/common binding sites, which locate more proximally to the TSSs, show stringent recognition sequences and exhibit higher BLIMP1 occupancy levels (Figure 8G). We assume that our estimation of the contribution of the specific binding sites to gene regulation represents a lower limit, since we only included the binding sites

<15 Kb from the TSSs as those associated with the genes for simplicity of the assignment of the bindings to the genes, yet a majority of the specific binding sites locate >15 Kb from the TSSs (Figure 4A–C). For example, there were two PRP-specific sites at 22.8 and 17.7 Kb upstream of *Vsx2* (Supplementary Figure S4E), a critical regulator for the bipolar cell differentiation repressed by BLIMP1 (17,18), and the latter site was reported to be a bipolar-cell enhancer (50), which may thus be a target for repression by BLIMP1.

Previous experiments primarily using cultured cell lines have highlighted that BLIMP1 represses gene expression by recruiting various co-repressors around the BLIMP1-binding sites proximal to the TSSs (7–10). Consistent with such previous findings, we did identify an enrichment of the BLIMP1-binding sites, particularly the shared/common sites, around the TSSs (<1 Kb) (Figure 4A–C). On the other hand, based on our aforementioned findings at the genome-wide level and involving four distinct lineages, the extent to which the shared/common sites contribute to gene regulation during developmental processes *in vivo* may not necessarily be significant, which should be verified experimentally using methodologies such as mutating the binding sequences by a genome-editing technology, etc. In this regard, it is also interesting to note that the shared BLIMP1-binding sites in d2 PGCLCs, which exhibited a significant association with the regulation of downstream genes, were also distributed more distally than those of the other lineages (Figures 4A–C and 6C–E). Collectively, these findings therefore suggest that, in general, BLIMP1 controls gene expression through distally located, cell type-specific binding sites.

In good agreement with the previous notion that BLIMP1 functions primarily as a transcriptional repressor (3–11), specific BLIMP1 bindings exhibited a significant correlation with the repression of associated genes in all four cell types analyzed (both in wild-type and knockout analyses) (Figure 6C–E). On the other hand, in line with the demonstration that BLIMP1 also functions as a transcriptional activator (15), specific BLIMP1 bindings exhibited a significant correlation with the activation of associated genes in perinatal IE development (in wild-type analyses) and in PB and PGC development (in knockout analyses) (Figure 6C and E). We found that the presence of the lineage-related TF motifs in the BLIMP1-binding sites had only a limited impact on the regulation of the associated genes, suggesting that BLIMP1 may exhibit specific functions without intimately cooperating with other TFs. If so, this would stand in contrast to proposed mechanisms of enhancer activation, in which specific TF combinations target cell type-specific enhancers depending on individual recognition sequences located in juxtaposition (~100 bp) (51), or TFs operate collectively through bindings to recognition sequences with flexible orientation or spacing (52) [reviewed in (1,2)], although the motif occurrence patterns in the BLIMP1-binding sites were compatible with these models.

Understanding the mechanisms for gene regulation by BLIMP1 via the specific binding sites with the flexible motif recognition may require extensive investigations of the epigenome, such as analysis of histone modifications. Published data for the distribution of histone H3 lysine 27 trimethylation (H3K27me3) [terminal B cell differentiation (15) and PGC specification *in vitro* (13)] suggested that a substantial elevation of H3K27me3 levels after BLIMP1-bindings to the specific sites may characterize gene repression by BLIMP1, and when genes were activated by BLIMP1 or were expressed stably, the associated binding sites did not show the recruitment of H3K27me3 (Supplementary Figure S8A).

Along with the accumulation of datasets for the binding of TFs across the genome, it has now become well-recognized that the number of TF-binding sites across the genome is much higher than the number of genes that appear to be controlled directly by the relevant TFs [reviewed in (1,53)]. Mutational experiments have shown that the loss of one such element often has only a modest effect on the expression of the associated genes in a cellular context (54–57), suggesting that the TF-binding sites may function in an additive manner or apparently 'spurious' binding sites may play critical roles only upon environmental stresses or mutations in other binding sites [reviewed in (1,53)]. Even in this regard, it has been surprising to find that the common BLIMP1-binding sites—which locate more proximally to the TSSs, show stringent recognition sequences and exhibit higher BLIMP1 occupancy levels—do not essentially influence the expression of associated genes (Figure 6C–E). This might suggest the possibility that the common/shared BLIMP1-binding sites serve as a reservoir of the nuclear BLIMP1 protein, from which BLIMP1 is recruited to specific sites to regulate gene expression. A new sequence motif, a GGGAAA repeat, which was very strongly occupied by BLIMP1, might also play such a role. We note that the specific sites associated with gene regulation by BLIMP1 tend to coexist with the common sites within topologically associated domains (58) (Supplementary Figure S8B), which might support this argument.

In conclusion, our datasets on the correlation of specific BLIMP1 bindings with gene repression/activation in four distinct lineages should serve as a valuable resource for exploring the specific mechanisms underlying BLIMP1-mediated control of gene repression/activation through its distal binding sites. Such studies should be facilitated with analyses on relevant epigenetic features around the BLIMP1-binidng sites, including open chromatin states and histone modifications, as well as 3D genome structures.

DATA AVAILABILITY

The ChIP-seq and RNA-seq data of EGFP-BLIMP1 were deposited to the Gene Expression Omnibus (GEO) (GSE91041).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the members of our laboratory for their helpful input on this study, and R. Kabata, N. Konishi, Y. Sakaguchi, Y. Nagai, M. Kawasaki, T. Sato and M. Kabata for

their technical assistance. We thank K. Hayashi, K. Haniuda and D. Kitamura for their advice on plasmablast induction. We used experimental instruments at the Medical Research Support Center, Graduate School of Medicine, Kyoto University.

FUNDING

Japan Society for the Promotion of Science (JSPS) KAKENHI Grants (JP16H04720, JP16H01216 to K.K., JP25.5081 to T.M., in part); Japan Science and Technology Agency-Exploratory Research for Advanced Technology (JST-ERATO) Grant (JPMJER1104 to M.S.). Funding for open access charge: JSPS KAKENHI Grant (JP16H04720).
Conflict of interest statement. None declared.

REFERENCES

1. Heinz,S., Romanoski,C.E., Benner,C. and Glass,C.K. (2015) The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.*, **16**, 144–154.
2. Long,H.K., Prescott,S.L. and Wysocka,J. (2016) Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell*, **167**, 1170–1187.
3. Keller,A.D. and Maniatis,T. (1991) Identification and characterization of a novel repressor of beta-interferon gene expression. *Genes Dev.*, **5**, 868–879.
4. Turner,C.A. Jr, Mack,D.H. and Davis,M.M. (1994) Blimp-1, a novel zinc finger-containing protein that can drive the maturation of B lymphocytes into immunoglobulin-secreting cells. *Cell*, **77**, 297–306.
5. Lin,Y., Wong,K. and Calame,K. (1997) Repression of c-myc transcription by Blimp-1, an inducer of terminal B cell differentiation. *Science*, **276**, 596–599.
6. Kuo,T.C. and Calame,K.L. (2004) B lymphocyte-induced maturation protein (Blimp)-1, IFN regulatory factor (IRF)-1, and IRF-2 can bind to the same regulatory sites. *J. Immunol.*, **173**, 5556–5563.
7. Ren,B., Chee,K.J., Kim,T.H. and Maniatis,T. (1999) PRDI-BF1/Blimp-1 repression is mediated by corepressors of the Groucho family of proteins. *Genes Dev.*, **13**, 125–137.
8. Yu,J., Angelin-Duclos,C., Greenwood,J., Liao,J. and Calame,K. (2000) Transcriptional repression by blimp-1 (PRDI-BF1) involves recruitment of histone deacetylase. *Mol. Cell. Biol.*, **20**, 2592–2603.
9. Gyory,I., Wu,J., Fejer,G., Seto,E. and Wright,K.L. (2004) PRDI-BF1 recruits the histone H3 methyltransferase G9a in transcriptional silencing. *Nat. Immunol.*, **5**, 299–308.
10. Su,S.T., Ying,H.Y., Chiu,Y.K., Lin,F.R., Chen,M.Y. and Lin,K.I. (2009) Involvement of histone demethylase LSD1 in Blimp-1-mediated gene repression during plasma cell differentiation. *Mol. Cell. Biol.*, **29**, 1421–1431.
11. Doody,G.M., Care,M.A., Burgoyne,N.J., Bradford,J.R., Bota,M., Bonifer,C., Westhead,D.R. and Tooze,R.M. (2010) An extended set of PRDM1/BLIMP1 target genes links binding motif type to dynamic repression. *Nucleic Acids Res.*, **38**, 5336–5350.
12. Hohenauer,T. and Moore,A.W. (2012) The Prdm family: expanding roles in stem cells and development. *Development*, **139**, 2267–2282.
13. Kurimoto,K., Yabuta,Y., Hayashi,K., Ohta,H., Kiyonari,H., Mitani,T., Moritoki,Y., Kohri,K., Kimura,H., Yamamoto,T. *et al.* (2015) Quantitative dynamics of chromatin remodeling during germ cell specification from mouse embryonic stem cells. *Cell Stem Cell*, **16**, 517–532.
14. Mould,A.W., Morgan,M.A., Nelson,A.C., Bikoff,E.K. and Robertson,E.J. (2015) Blimp1/Prdm1 functions in opposition to Irf1 to maintain neonatal tolerance during postnatal intestinal maturation. *PLoS Genet.*, **11**, e1005375.
15. Minnich,M., Tagoh,H., Bonelt,P., Axelsson,E., Fischer,M., Cebolla,B., Tarakhovsky,A., Nutt,S.L., Jaritz,M. and Busslinger,M. (2016) Multifunctional role of the transcription factor Blimp-1 in coordinating plasma cell differentiation. *Nat. Immunol.*, **17**, 331–343.
16. Mackay,L.K., Minnich,M., Kragten,N.A., Liao,Y., Nota,B., Seillet,C., Zaid,A., Man,K., Preston,S., Freestone,D. *et al.* (2016) Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science*, **352**, 459–463.
17. Brzezinski,J.A.t., Lamba,D.A. and Reh,T.A. (2010) Blimp1 controls photoreceptor versus bipolar cell fate choice during retinal development. *Development*, **137**, 619–629.
18. Katoh,K., Omori,Y., Onishi,A., Sato,S., Kondo,M. and Furukawa,T. (2010) Blimp1 suppresses Chx10 expression in differentiating retinal photoreceptor precursors to ensure proper photoreceptor development. *J. Neurosci.*, **30**, 6515–6526.
19. Harper,J., Mould,A., Andrews,R.M., Bikoff,E.K. and Robertson,E.J. (2011) The transcriptional repressor Blimp1/Prdm1 regulates postnatal reprogramming of intestinal enterocytes. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 10585–10590.
20. Muncan,V., Heijmans,J., Krasinski,S.D., Buller,N.V., Wildenberg,M.E., Meisner,S., Radonjic,M., Stapleton,K.A., Lamers,W.H., Biemond,I. *et al.* (2011) Blimp1 regulates the transition of neonatal to adult intestinal epithelium. *Nat. Commun.*, **2**, 452.
21. Shaffer,A.L., Lin,K.I., Kuo,T.C., Yu,X., Hurt,E.M., Rosenwald,A., Giltnane,J.M., Yang,L., Zhao,H., Calame,K. *et al.* (2002) Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity*, **17**, 51–62.
22. Hayashi,K., Ohta,H., Kurimoto,K., Aramaki,S. and Saitou,M. (2011) Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell*, **146**, 519–532.
23. Yamashiro,C., Hirota,T., Kurimoto,K., Nakamura,T., Yabuta,Y., Nagaoka,S.I., Ohta,H., Yamamoto,T. and Saitou,M. (2016) Persistent requirement and alteration of the key targets of PRDM1 during primordial germ cell development in mice. *Biol. Reprod.*, **94**, 7.
24. Furukawa,A., Koike,C., Lippincott,P., Cepko,C.L. and Furukawa,T. (2002) The mouse Crx 5'-upstream transgene sequence directs cell-specific and developmentally regulated expression in retinal photoreceptor cells. *J. Neurosci.*, **22**, 1640–1647.
25. Gracz,A.D., Puthoff,B.J. and Magness,S.T. (2012) Identification, isolation, and culture of intestinal epithelial stem cells from murine intestine. *Methods Mol. Biol.*, **879**, 89–107.
26. Hayashi,K., Nittono,R., Okamoto,N., Tsuji,S., Hara,Y., Goitsuka,R. and Kitamura,D. (2000) The B cell-restricted adaptor BASH is required for normal development and antigen receptor-mediated activation of B cells. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 2755–2760.
27. Nojima,T., Haniuda,K., Moutai,T., Matsudaira,M., Mizokawa,S., Shiratori,I., Azuma,T. and Kitamura,D. (2011) In-vitro derived germinal centre B cells differentially generate memory B or plasma cells in vivo. *Nat. Commun.*, **2**, 465.
28. Kagiwada,S., Kurimoto,K., Hirota,T., Yamaji,M. and Saitou,M. (2013) Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J.*, **32**, 340–353.
29. Nakamura,T., Yabuta,Y., Okamoto,I., Aramaki,S., Yokobayashi,S., Kurimoto,K., Sekiguchi,K., Nakagawa,M., Yamamoto,T. and Saitou,M. (2015) SC3-seq: a method for highly parallel and quantitative measurement of single-cell gene expression. *Nucleic Acids Res.*, **43**, e60.
30. Nakamura,T., Okamoto,I., Sasaki,K., Yabuta,Y., Iwatani,C., Tsuchiya,H., Seita,Y., Nakamura,S., Yamamoto,T. and Saitou,M. (2016) A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature*, **537**, 57–62.
31. Luo,R.X., Postigo,A.A. and Dean,D.C. (1998) Rb interacts with histone deacetylase to repress transcription. *Cell*, **92**, 463–473.
32. Landt,S.G., Marinov,G.K., Kundaje,A., Kheradpour,P., Pauli,F., Batzoglou,S., Bernstein,B.E., Bickel,P., Brown,J.B., Cayting,P. *et al.* (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.*, **22**, 1813–1831.
33. Caporaso,J.G., Lauber,C.L., Walters,W.A., Berg-Lyons,D., Lozupone,C.A., Turnbaugh,P.J., Fierer,N. and Knight,R. (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U.S.A.*, **108**(Suppl. 1), 4516–4522.
34. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
35. Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
36. Machanick,P. and Bailey,T.L. (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, **27**, 1696–1697.

37. Stormo,G.D., Schneider,T.D., Gold,L. and Ehrenfeucht,A. (1982) Use of the 'Perceptron' algorithm to distinguish translational initiation sites in E. coli. *Nucleic Acids Res.*, **10**, 2997–3011.
38. Chang,D.H., Cattoretti,G. and Calame,K.L. (2002) The dynamic expression pattern of B lymphocyte induced maturation protein-1 (Blimp-1) during mouse embryonic development. *Mech. Dev.*, **117**, 305–309.
39. Robertson,E.J., Charatsi,I., Joyner,C.J., Koonce,C.H., Morgan,M., Islam,A., Paterson,C., Lejsek,E., Arnold,S.J., Kallies,A. *et al.* (2007) Blimp1 regulates development of the posterior forelimb, caudal pharyngeal arches, heart and sensory vibrissae in mice. *Development*, **134**, 4335–4345.
40. Ohinata,Y., Sano,M., Shigeta,M., Yamanaka,K. and Saitou,M. (2008) A comprehensive, non-invasive visualization of primordial germ cell development in mice by the Prdm1-mVenus and Dppa3-ECFP double transgenic reporter. *Reproduction*, **136**, 503–514.
41. Yeung,K.Y. and Ruzzo,W.L. (2001) Principal component analysis for clustering gene expression data. *Bioinformatics*, **17**, 763–774.
42. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
43. Kurimoto,K., Yabuta,Y., Ohinata,Y., Shigeta,M., Yamanaka,K. and Saitou,M. (2008) Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes Dev.*, **22**, 1617–1635.
44. Yang,L., Zhou,T., Dror,I., Mathelier,A., Wasserman,W.W., Gordan,R. and Rohs,R. (2014) TFBSshape: a motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Res.*, **42**, D148–D155.
45. Jolma,A., Yin,Y., Nitta,K.R., Dave,K., Popov,A., Taipale,M., Enge,M., Kivioja,T., Morgunova,E. and Taipale,J. (2015) DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature*, **527**, 384–388.
46. Sasaki,K., Yokobayashi,S., Nakamura,T., Okamoto,I., Yabuta,Y., Kurimoto,K., Ohta,H., Moritoki,Y., Iwatani,C., Tsuchiya,H. *et al.* (2015) Robust in vitro induction of human germ cell fate from pluripotent stem cells. *Cell Stem Cell*, **17**, 178–194.
47. Ohta,H., Kurimoto,K., Okamoto,I., Nakamura,T., Yabuta,Y., Miyauchi,H., Yamamoto,T., Okuno,Y., Hagiwara,M., Shirane,K. *et al.* (2017) In vitro expansion of mouse primordial germ cell-like cells recapitulates an epigenetic blank slate. *EMBO J.*, **36**, 1888–1907.
48. Ohinata,Y., Payer,B., O'Carroll,D., Ancelin,K., Ono,Y., Sano,M., Barton,S.C., Obukhanych,T., Nussenzweig,M., Tarakhovsky,A. *et al.* (2005) Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature*, **436**, 207–213.
49. Vincent,S.D., Dunn,N.R., Sciammas,R., Shapiro-Shalef,M., Davis,M.M., Calame,K., Bikoff,E.K. and Robertson,E.J. (2005) The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. *Development*, **132**, 1315–1325.
50. Kim,D.S., Matsuda,T. and Cepko,C.L. (2008) A core paired-type and POU homeodomain-containing transcription factor program drives retinal bipolar cell gene expression. *J. Neurosci.*, **28**, 7748–7764.
51. Kazemian,M., Pham,H., Wolfe,S.A., Brodsky,M.H. and Sinha,S. (2013) Widespread evidence of cooperative DNA binding by transcription factors in Drosophila development. *Nucleic Acids Res.*, **41**, 8237–8252.
52. Junion,G., Spivakov,M., Girardot,C., Braun,M., Gustafson,E.H., Birney,E. and Furlong,E.E. (2012) A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell*, **148**, 473–486.
53. Spivakov,M. (2014) Spurious transcription factor binding: non-functional or genetically redundant? *Bioessays*, **36**, 798–806.
54. Spivakov,M., Akhtar,J., Kheradpour,P., Beal,K., Girardot,C., Koscielny,G., Herrero,J., Kellis,M., Furlong,E.E. and Birney,E. (2012) Analysis of variation at transcription factor binding sites in Drosophila and humans. *Genome Biol.*, **13**, R49.
55. Korkmaz,G., Lopes,R., Ugalde,A.P., Nevedomskaya,E., Han,R., Myacheva,K., Zwart,W., Elkon,R. and Agami,R. (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 192–198.
56. Sanjana,N.E., Wright,J., Zheng,K., Shalem,O., Fontanillas,P., Joung,J., Cheng,C., Regev,A. and Zhang,F. (2016) High-resolution interrogation of functional elements in the noncoding genome. *Science*, **353**, 1545–1549.
57. Ulirsch,J.C., Nandakumar,S.K., Wang,L., Giani,F.C., Zhang,X., Rogov,P., Melnikov,A., McDonel,P., Do,R., Mikkelsen,T.S. *et al.* (2016) Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell*, **165**, 1530–1545.
58. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.