



## Research article

## Single-cell 5' RNA sequencing of camelid peripheral B cells provides insights into cellular basis of heavy-chain antibody production

Li Yi<sup>a,1</sup>, Xin Guo<sup>b,1</sup>, Yuexing Liu<sup>c</sup>, Jirimutu<sup>a,d,\*</sup>, Zhen Wang<sup>b,\*\*</sup><sup>a</sup> Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education, College of Food Science and Engineering, Inner Mongolia Agricultural University, Huhhot 010018, China<sup>b</sup> CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China<sup>c</sup> Guangzhou Laboratory, Guangzhou 510005, China<sup>d</sup> Inner Mongolia China-Kazakhstan Camel Research Institute, Alxa 750306, China

## ARTICLE INFO

## Keywords:

Single-cell sequencing  
Camelid  
B cell subtype  
B-cell receptor  
Heavy-chain antibody

## ABSTRACT

Camelids produce both conventional tetrameric antibodies (Abs) and dimeric heavy-chain antibodies (HCABs). Although B cells that generate these two types of Abs exhibit distinct B cell receptors (BCRs), whether these two B cell populations differ in their phenotypes and developmental processes remains unclear. Here, we performed single-cell 5' RNA profiling of peripheral blood mononuclear cell samples from Bactrian camels before and after immunization. We characterized the functional subtypes and differentiation trajectories of circulating B cells in camels, and reconstructed single-cell BCR sequences. We found that in contrast to humans, the proportion of Tbet<sup>+</sup> B cells was high among camelid peripheral B cells. Several marker genes of human B cell subtypes, including *CD27* and *IGHD*, were expressed at low levels in the corresponding camel B cell subtypes. Camelid B cells expressing variable genes of HACBs (*VHH*) were widely present in various functional subtypes and showed highly overlapping differentiation trajectories with B cells expressing variable genes of conventional Abs (*VH*). After immunization, the transcriptional changes in *VHH*<sup>+</sup> and *VH*<sup>+</sup> B cells were largely consistent. Through structure modeling, we identified a variety of scaffold types among the reconstructed *VHH* sequences. Our study provides insights into the cellular context of HCAB production in camels and lays the foundation for developing single-B cell-based camelid single-domain Ab screening.

## 1. Introduction

Conventional antibodies (Abs) or immunoglobulins (Igs) in vertebrates are heterotetramers composed of two heavy and two light chains [1]. In camelids, homodimeric IgGs contain only two heavy chains and are known as heavy chain antibodies (HCABs) [2]. Despite the absence of light chains, HCABs possess unique mechanisms to generate structural diversity. Compared to conventional Abs, HCABs have a longer third complementarity-determining region (CDR3) and more cysteine residues within the variable regions to form additional intramolecular disulfide bonds [3]. The Ab fragment obtained by retaining only the HCAB variable region is called a single-domain antibody (sdAb, also referred to as a nanobody) because the single variable domain maintains the

complete antigen-binding potential of its HCAB counterpart [3]. sdAbs not only maintain antigen-binding affinity but also have higher stability, hydrophilicity, and tissue permeability; therefore, they have been increasingly used in biomedicine [4]. In 2018, caplacizumab, the first camelid sdAb drug developed by Ablynx (recently acquired by SANOFI), was approved in Europe [5].

HCABs are encoded by dedicated genes that differ from those that encode conventional Abs in two ways. In the constant region genes, *IGHG2* and *IGHG3* encoding HCAB have splice site mutations (GT->AT) at the exon of the first constant domain of the heavy chain (CH1), resulting in the deletion of CH1 connected to the light chain [6–8]. Neither *IGHG1* nor other *IGHC* genes (including *IGHM*, *IGHE*, *IGHA*) encoding conventional Abs possess this mutation [8]. Second, the *IGHV*

\* Corresponding author at: Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education, College of Food Science and Engineering, Inner Mongolia Agricultural University, Huhhot 010018, China.

\*\* Corresponding author.

E-mail addresses: [yeluo1999@vip.163.com](mailto:yeluo1999@vip.163.com) (Jirimutu), [zwang01@sibs.ac.cn](mailto:zwang01@sibs.ac.cn) (Z. Wang).

<sup>1</sup> These authors contributed equally.

<https://doi.org/10.1016/j.csbj.2024.04.041>

Received 19 January 2024; Received in revised form 15 April 2024; Accepted 15 April 2024

Available online 16 April 2024

2001-0370/© 2024 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

gene of HCAB undergoes characteristic amino acid substitutions at four positions in the second framework region (FR2), most of which are V42F, G49E, L50R, and W52G [9–11]. In the variable domain of conventional Abs (VH), these four positions are involved in interaction with the light chain through the hydrophobic effect; therefore, they are highly conserved hydrophobic amino acids. In the variable domain of HCABs (VHH), they are typically replaced with hydrophilic amino acids. Sequencing and assembly of the camelid *IGH* locus showed that both HCABs and conventional Abs are encoded by genes in the same *IGH* locus with a typical *V-D-J-C* organization and that their coding genes are intermixed within the *IGHV* and *IGHC* regions [12,13]. *IGHD* and *IGHJ* genes are common to both conventional Abs and HCABs [12,13]. Immune repertoire sequencing of multiple camelids revealed the genetic diversity of *VHH*, both at the germline and rearranged levels [11,12,14,15].

Studies in animal models have shown that B cells undergo multiple stages of differentiation [16,17], and the functional subsets of these B cells are mainly distinguished by the rearrangement status of B cell receptor (BCR) gene segments and cell surface markers [18]. The coexistence of B cells producing conventional Abs and HCABs in camelids raises the question whether these two populations of B cells have undergone similar developmental stages. Although B cells that generate HCABs were speculated to pass through a naive IgM+ stage, similar to conventional Abs [12], the cellular components remain largely unexplored because of the lack of surface markers and corresponding reagents for camelids.

In recent years, the development of single-cell RNA sequencing (scRNA-seq) technology has advanced the study of cell heterogeneity [19,20]. One of the greatest advantages of scRNA-seq is its ability to define cell types at the genomic level, thus allowing for the unbiased phenotyping of all cells. This is particularly useful for non-model organisms where traditional experimental tools are limited [21–23]. A previous study performed scRNA-seq on peripheral blood mononuclear cells (PBMCs) of an immunized alpaca and described alterations in cell type composition and gene expression [24]. However, BCR types cannot be recovered with 3' RNA sequencing data, making it impossible to distinguish B cells expressing conventional Abs from those expressing HCABs.

In this study, we performed single-cell profiling of PBMCs from four Bactrian camels using 5' RNA sequencing. First, we identified the cell types in camel PBMCs, especially B cell subtypes and differentiation trajectories. We then reconstructed the BCR sequences for each B cell, distinguished between the *IGHV* and *IGHC* types, and revealed their

correspondence with B cell phenotypes. Finally, we compared changes in PBMCs and B cells before and after immunization. Our results showed that, although B cells producing conventional Abs and HCABs displayed a preference for BCR sequences, their differentiation trajectories largely overlapped. Transcriptional changes in the two B cell populations were also highly similar throughout the immunization process.

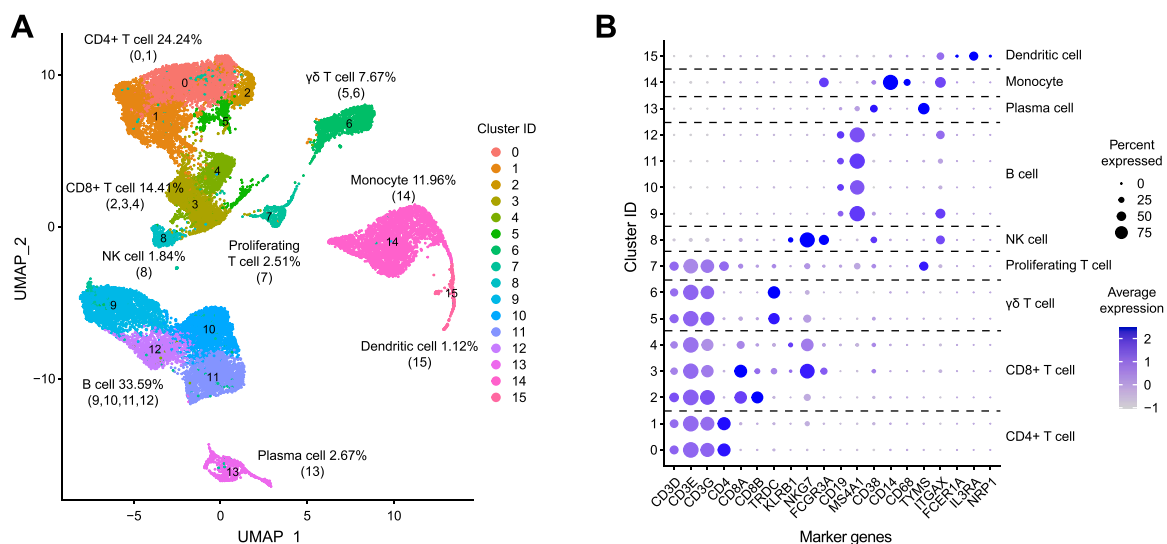
## 2. Results

### 2.1. Cell composition of camel PBMCs

We isolated PBMCs from four healthy adult Bactrian camels (C1–C4, [Supplementary Table S1](#)), and performed scRNA-seq for the PBMCs based on the 10 × Genomics platform. To recover *IGHV* and *IGHC* genes of BCRs, 5' RNA libraries were constructed for sequencing. Approximately 5500–10,000 cells were detected per sample, and the median number of genes detected per cell was 1200–1700 ([Supplementary Table S2](#)). After quality control based on UMI (unique molecular identifier) counts and mitochondrial percentages, 26,759 cells were preserved.

We integrated cells across all samples using Seurat [25] and performed dimensional reduction and unsupervised clustering ([Fig. 1A](#)). Major cell types of camel PBMCs could be recognized with the expression of canonical marker genes as in humans ([Fig. 1B](#)), including CD4 + T cells (*CD3*, *CD4*, 24.24%), CD8 + T cells (*CD3*, *CD8*, 14.41%),  $\gamma\delta$  T cells (*CD3*, *TRDC*, 7.67%), proliferating T cells (*CD3*, *TYMS*, 2.51%), natural killer (NK) cells (*KLRB1*, *NKG7*, *FCGR3A*, 1.84%), B cells (*CD19*, *MS4A1*, 33.59%), plasma cells (*CD38*, *TYMS*, 2.67%), monocytes (*CD14*, *CD68*, 11.96%) and dendritic cells (*ITGAX*, *FCER1A*, *IL3RA*, *NRP1*, 1.12%). All four camels had similar cell compositions ([Supplementary Fig. S1](#)). The lymphocyte composition of CD4 + T,  $\gamma\delta$  T, and B cells was comparable to previous studies based on flow cytometry [26], although other lymphocytes could not be identified using flow cytometry owing to the lack of cross-reactive monoclonal Abs. Notably, camels belong to  $\gamma\delta$ -high mammalian species as other artiodactyls [26], in contrast to  $\gamma\delta$ -low species like humans and mice (<5% in PBMCs [27]).

Our scRNA-seq analysis also enabled us to dissect the heterogeneity of major cell types at high resolution. For example, Cluster 0 (CD4 + T), 2 (CD8 + T), and 5 ( $\gamma\delta$  T) showed high expression of *CCR7* and *SELL* but low expression of *S100A4*, suggesting a phenotype of naive T cells [28] ([Supplementary Fig. S2](#)). Cluster 1 (CD4 + T) and 6 ( $\gamma\delta$  T) showed high expression of *S100A4*, and Cluster 3/4 (CD8 + T) showed high expression of *GZMK*, suggesting a phenotype of memory T cells [28]



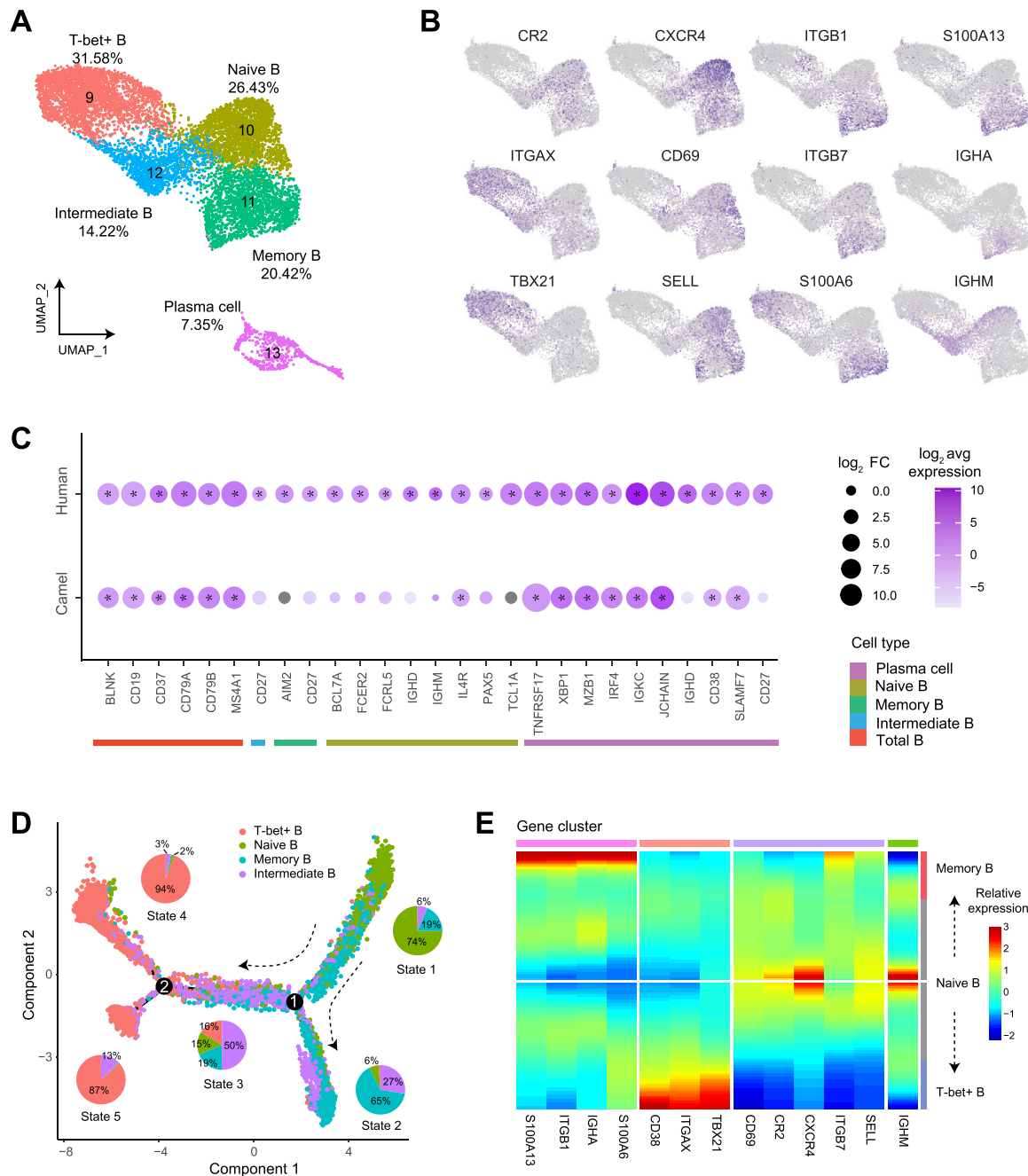
**Fig. 1.** Major cell types of camel PBMCs revealed by scRNA-seq. (A) UMAP visualization of cell clustering. Cell types are annotated based on the expression of marker genes and their proportion in PBMCs is shown. (B) Dot plot depicting average expression and expression percentage of marker genes in each cell cluster.

(Supplementary Fig. S2).

### 2.2. B cell subsets and differentiation trajectories

As Abs or BCRs are expressed by B cells, we performed an in-depth analysis of B cell subtypes. In addition to plasma cells (Cluster 13,

7.35%), unsupervised clustering divided B cells into four subtypes (Cluster 9–12, Fig. 2A) that showed similar proportions in the four camels (Supplementary Fig. S3). We first inferred their functional annotations based on the differentially expressed genes among the subtypes (Fig. 2B). Cluster 9 (31.58%) displayed relatively high expression of *TBX21* and *ITGAX* but relatively low expression of *CR2*, resembling



**Fig. 2.** Characterization of camel B cell subtypes. (A) UMAP plot of B cells colored by their subtypes. The proportion of each subtype in B cells is calculated. (B) Expression of marker genes among the B cell subtypes identified by Seurat FindMarkers. T-bet+ B cells show lower expression of *CR2* but higher expression of *ITGAX* and *TBX21*. Naive B cells show higher expression of *CXCR4*, *CD69* and *SELL*. Memory B cells show higher expression of *ITGB1*, *ITGB7*, *S100A6*, *S100A13* and *IGHA*. Intermediate B cells show higher expression of *IGHM*. (C) Comparison of marker genes for peripheral B cells and their subtypes between humans and camels based on CellMarker 2.0. Only genes significantly over-expressed in humans are shown. Markers for total B cells should be significantly over-expressed compared to other PBMC types, while markers for B cell subtypes should be significantly over-expressed compared to other B cells. Genes with Bonferroni-adjusted  $P < 0.05$  are marked with stars. The log-average expression and log-fold change (logFC) are indicated by colors and sizes of the dots, respectively. Genes without expression detected are colored in grey. (D) B cell state and differentiation trajectories inferred using Monocle2. Cells are colored according to their subtypes, and the proportion of subtypes in each state is displayed by pie plots. The pseudotime is assigned by assuming naive B cells as the root. Plasma cells are excluded from the analysis. (E) Heatmap depicting the relative expression of marker genes along the trajectories. The color bar at the top represents gene clusters with similar expression tendencies, while the color bar at the right represents cell types.

the phenotype of human atypical B cells [29,30]. However, considering that atypical B cells in humans are often associated with chronic infections or autoimmune diseases and may non-specifically integrate multiple B cell populations [31], we designated Cluster 9 T-bet/TBX21 + B cells. In contrast to humans, in which only a small fraction of peripheral T-bet+ B cells were present, camels had a high proportion of T-bet+ B cells, which was observed in other species, such as horses [22]. Cluster 10 (26.43%) expressed higher levels of *CXCR4*, *CD69*, and *SELL*, which was similar to human naive B cells [30,32]. Cluster 11 (20.42%) expressed higher levels of *ITGB1*, *ITGB7*, *S100A6*, *S100A13*, and *IGHA*, which were associated with memory B cell functions [30]. Cluster 12 (14.22%) showed higher expression of *IGHM*. We then mapped the camel B cell transcriptome to the human PBMC reference dataset built by Seurat [25]. The results indicated that, except for the lack of T-bet+ B cells in the human dataset, Cluster 10, 11, and 12 had the highest similarity to human naive, memory, and intermediate B cells, respectively (Supplementary Fig. S4). Therefore, we assigned corresponding cell-type labels to these clusters (Fig. 2A).

We also compared manually curated marker genes for B cells and their subtypes between humans and camels based on CellMarker 2.0 [33] (Fig. 2C). Although most marker genes in humans could also be applied to camels, there were exceptions. Notably, *CD27*, a marker gene for antigen-experienced B cells, including memory B and plasma cells in humans [18], was expressed at low levels in all camel B cell subsets (Supplementary Fig. S5). *IGHD*, a marker gene for human naive B cells [18], also lacked expression and it was previously identified as a pseudogene in camelid genomes [12,13]. Other human marker genes, including *TCL1A*, *BCL7A*, *FCER2*, and *AIM2* were not significantly overexpressed in the corresponding camel B cell subtypes (Fig. 2C).

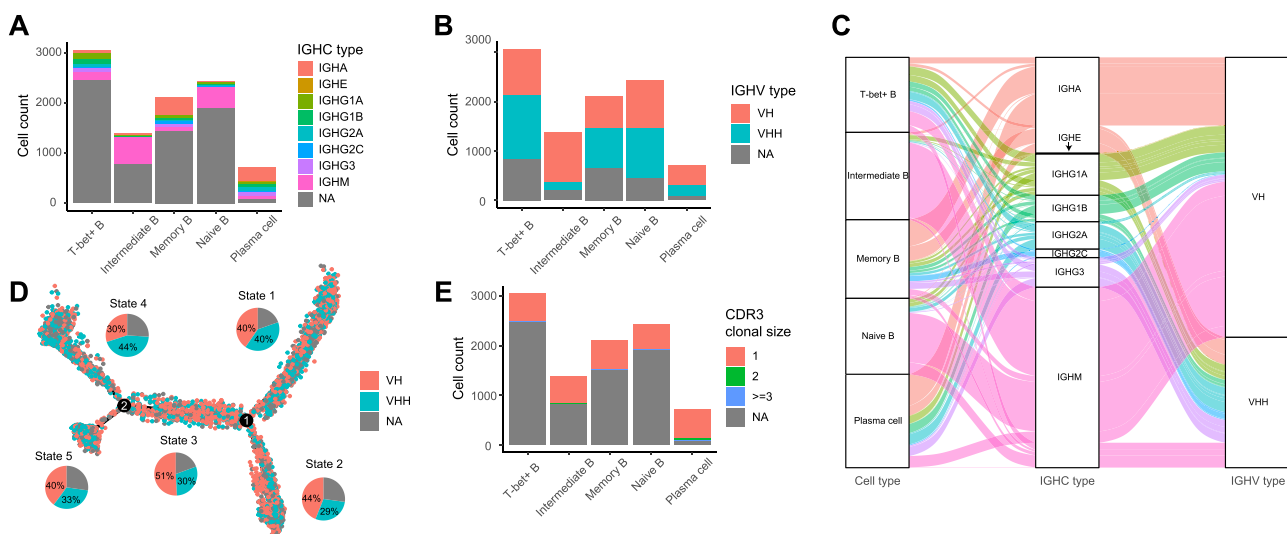
We used Monocle2 [34] to infer single-cell trajectories and explore the relationships between these B cell subtypes. Given that plasma cells formed a separate cluster, our analysis focused solely on non-plasma cells. Consequently, Monocle2 identified five distinct B cell states (Fig. 2D): State 1 consisted mostly of naive (74%) and memory B cells (19%); State 2 consisted mostly of memory (65%) and intermediate B cells (27%); State 3 consisted mostly of intermediate B cells (50%); and both States 4 and 5 were predominantly composed of T-bet+ B cells (94% and 87%, respectively). Moreover, originating from State 1,

Monocle2 delineated two major differentiation trajectories for B cells: (1) naive B cells and memory B cells; (2) naive B cells, intermediate B cells, and T-bet+ B cells. Similar to humans [35], this result suggests that T-bet+ B cell differentiation is an alternative lineage that is distinct from the classical memory lineage in camels. The expression patterns of marker genes for B cell subtypes were consistent with these two differentiation trajectories (Fig. 2E).

### 2.3. BCR types and association with B cell subsets

We first identified the *IGHC* type of BCRs based on the single-cell expression matrix (Fig. 3A and Supplementary Fig. S6). For single B cells that may express more than one *IGHC* gene [36], the UMI count of the major gene should be three times greater than that of the minor gene (Supplementary Fig. S7). Among all B cells, 31.63% had their expressed *IGHC* type recovered, with the highest recovery rate in plasma cells (89.76%) because plasma cells had the highest BCR expression levels. Consistent with the developmental stages of B cells, both naive and intermediate B cells were primarily not class-switched and expressed *IGHM*, while memory B, T-bet+ B, and plasma cells were class-switched and mainly expressed *IGHG* and *IGHA* (Fig. 3A). Plasma cells expressing HCAb genes (*IGHG2/3+*) accounted for 60.79% of all *IGHG+* plasma cells, which is consistent with the fact that HCABs are the major IgG form in the sera of Bactrian camels [37]. We also identified the type of light chain genes expressed in each B cell and found that even in *IGHG2/3+* B cells, there was obvious expression of *IGLC* or *IGKC* (Supplementary Fig. S8). This indicated that the absence of light chains in HCABs was not due to the suppression of light-chain gene recombination and transcription.

To determine the variable region sequences of BCRs, we used TRUST4 [38] to assemble reads mapped to the *IGH* locus at the single-cell level. Reference *IGH* sequences were obtained from our previously published Bactrian camel genome [13]. For the assembled contigs, further annotation was performed based on IgBLAST [39], including *IGHV/D/J/C* genes, FR, and CDR positions and sequences. The *IGHC* types determined using sequence assembly and the UMI counts were highly consistent (Supplementary Fig. S9). We categorized B cells into *VH+* and *VHH+* types based on the most discriminative amino acid



**Fig. 3.** Characterization of camel single-cell BCRs. (A) Number of cells with different *IGHC* types in each B cell subtype. The *IGHC* types are assigned based on gene expression. NA represents cells without *IGHC* expression detected. (B) Number of cells with different *IGHV* types (*VH/VHH*) in each B cell subtype. The *IGHV* types are assigned based on the hallmark substitutions in FR2. NA represents cells without FR2 recovered. (C) Sankey diagram depicting the relationships between B cell subtypes, *IGHC* and *IGHV* types. Although most *IGHG2/3+* cells are associated with *VHH*, and most conventional *IGHC* are associated with *VH*, non-typical associations are widely present. (D) Mapping *IGHV* types to differentiation trajectories of B cells inferred using Monocle2. Cells are colored according to their *IGHV* types, and the proportion in each state is displayed by pie plots. (E) Number of cells with different clonal size in each B cell subtype. The clonotypes are determined by *IGHV* CDR3 sequences. NA represents cells without CDR3 recovered.

substitutions in the FR2 region of *IGHV* genes, namely, G49E/Q and L50R [15,40]. Two other hallmark amino acid substitutions, V42F/Y and W52G/F/W/L, were also detected in the *VH+* and *VHH+* types (Supplementary Fig. S10). Of all B cells, 77.09% had their *IGHV* type recovered (Supplementary Fig. S11), and each B cell subset contained a comparable proportion of *VH+* and *VHH+* cells, with a significantly lower proportion of *VHH+* in intermediate B cells (Benjamini-Hochberg [BH]-adjusted  $P = 4.24 \times 10^{-3}$ , two-sided *t*-test, Fig. 3B). Furthermore, we generated a Sankey diagram showing the relationships between B cell subsets, *IGHC* types, and *IGHV* types (Fig. 3C). As expected, most *IGHG2/3 + B* cells were paired with *VHH*, although a small proportion of *VH+IGHG2/3 + B* cells were also identified. This is consistent with the fact that some HCABs can be isolated with an antigen-binding VH domain [3,41]. Similarly, most *IGHG1 +*, *IGHM+*, and *IGHA+* B cells were paired with *VH*, although *VHH* also contributed to conventional BCRs (Fig. 3C). These non-typical BCRs could be expressed by plasma cells, indicating that they were not only in a transitional state during class switching but might also have specific antigen-binding capabilities. We also mapped the *IGHV* types to the differentiation trajectories of B cells reconstructed using Monocle2 (Fig. 3D). The proportions of *VHH+* and *VH+* cells showed no significant differences in all differentiation states after correction for multiple tests, indicating that they underwent similar developmental trajectories.

We analyzed the diversity of the reconstructed heavy-chain CDR3 sequences. Among all B cells, 29.87% showed complete CDR3 sequence recovery, whereas this proportion reached 90.18% in plasma cells (Fig. 3E and Supplementary Fig. S12). As expected, the CDR3 region of *VHH* was significantly longer than that of *VH* [3] (Supplementary Fig. S13) and the sequence features were similar to those reported in previous studies [42,43] (Supplementary Fig. S14). The vast majority of B cells (95.31%) had unique CDR3 sequences (that is, clone size = 1), and this unique proportion was only slightly decreased in plasma cells (90.05%), indicating that naturally occurring CDR3 sequences are highly diverse in camels (Fig. 3E).

#### 2.4. Structure analysis for *VHH* sequences

As *VHH* sequences are widely adopted in camelid sdAb engineering [3,4], we performed structure analysis of the *VHH* sequences discovered using scRNA-seq. In total, we extracted 640 complete *VHH* sequences covering regions from FR1–FR4. We followed Zimmermann et al. [44] to classify the sequences into three shapes based on the length of the CDR3 region: concave (<10 aa), loop (10–14 aa) and convex (>14 aa). We found the majority of sequences (88.44%) had a convex shape (Fig. 4A), consistent with *VHH* sequences with known structures in the SABDab database [45] (Supplementary Fig. S15). Additionally, there was a significantly higher proportion of sequences with more than three cysteines in the convex and loop shapes than in the concave shape ( $P = 1.29 \times 10^{-3}$ , chi-square test, Fig. 4A). These extra cysteines can

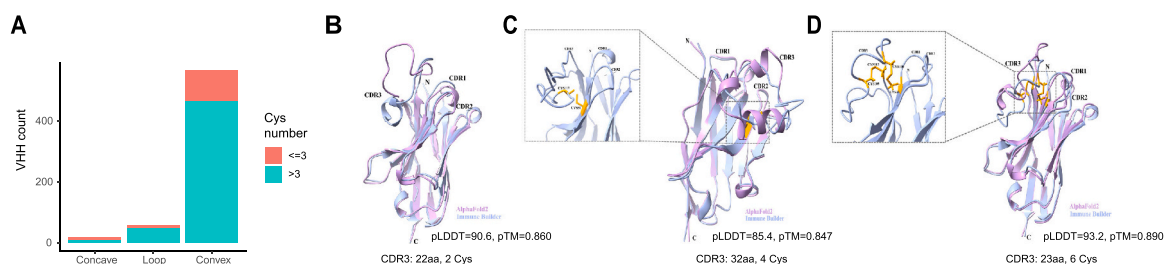
potentially form non-canonical disulfide bonds, which may stabilize the *VHH* domain with a long CDR3 length [3]. However, we did not observe a preference for the shapes among different B cell subtypes ( $P = 0.17$ , chi-square test) or *IGHC* types ( $P = 0.31$ , chi-square test).

We performed structure modeling for representative *VHH* sequences encompassing six scaffold types, which were combinations of three CDR3 shapes and two categories of cysteine numbers. The results demonstrated that both AlphaFold2 [46] and ImmuneBuilder [47] were able to generate high-accuracy structure models for each scaffold type (Supplementary Fig. S16). We specifically focused on the convex shape with a long CDR3 length (>20 aa), which features an extended hydrophobic core to tether CDR3 and exhibits a greater reliance on non-canonical disulfide bonds [48,49]. It has been reported that these non-canonical disulfide bonds can occur between cysteine residues located in CDR1–CDR3, FR2–CDR3, and CDR3–CDR3 [43]. We obtained representative structures for all scenarios: (1) long CDR3 without additional disulfide bonds (Fig. 4B); (2) one additional disulfide bond in FR2–CDR3, folding the loop and restricting flexibility (Fig. 4C); (3) two additional disulfide bonds in CDR1–CDR3 and CDR3–CDR3, which are not folded and maintain the convex shape (Fig. 4D).

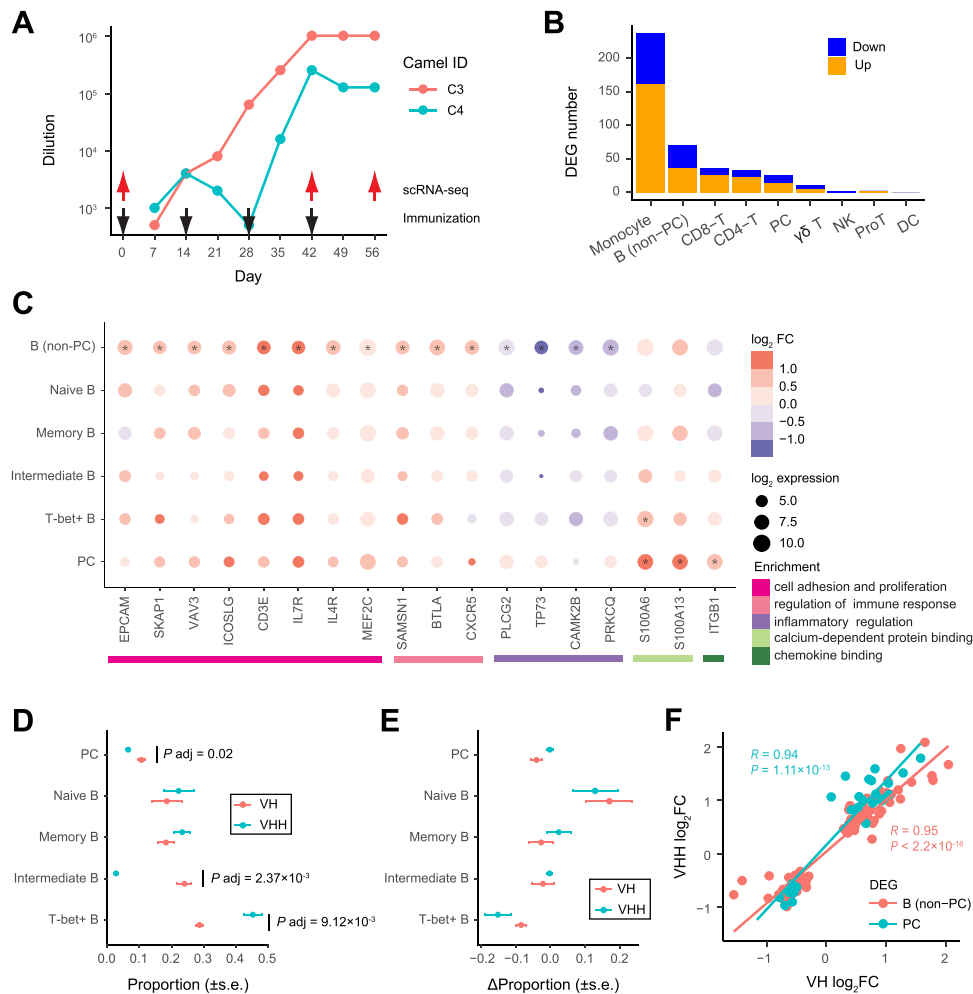
#### 2.5. Comparative analysis during immunization

Camel immunization is an important way to prepare antigen-specific sdAbs. To elucidate the changes in PBMCs and B cells following immunization, we immunized two Bactrian camels (C3 and C4) with bovine serum albumin (BSA). Immunization was performed every 2 weeks, and enzyme-linked immunosorbent assay (ELISA) indicated that the Ab titer reached the highest level after four rounds of immunization (Fig. 5A). We performed scRNA-seq on PBMCs collected at 42 and 56 days post-immunization (Supplementary Table S3) and compared them with pre-immunization samples (0 day). We transferred the cell-type annotations from our previously analyzed samples to the post-immunization samples (Supplementary Fig. S17). The proportions of major cell types showed no significant changes during immunization after multiple testing corrections (Supplementary Fig. S18).

For each PBMC type, we conducted a differential gene expression analysis during immunization using the pseudobulk approach [50]. Because of the higher average gene expression levels in plasma cells, B cells were divided into non-plasma and plasma cells for pseudobulk analysis. We identified the highest number of differentially expressed genes (DEGs, BH-adjusted  $P < 0.05$ ) in monocytes, followed by B, CD8 + T, CD4 + T, and plasma cells, with an overall upregulation of DEGs after immunization (Fig. 5B). Furthermore, we performed functional enrichment analysis of the DEGs. The upregulated DEGs in each cell type were significantly enriched in biological processes and functions related to the immune response, cell activation, and chemotaxis (Supplementary Fig. S19 and Fig. S20). Specifically, the significantly upregulated DEGs in B and plasma cells were enriched in cell adhesion



**Fig. 4.** Structure predictions for *VHH* sequences. (A) Number of complete *VHH* sequences with different shapes based on the length of CDR3 sequences: concave (<10 aa), loop (10–14 aa), and convex (>14 aa). Within each shape, *VHH* sequences with  $\leq 3$  and  $> 3$  cysteines are separated. (B–D) Structure prediction of *VHH* sequences with long convex shape (>20 aa) using AlphaFold2 (purple) and ImmuneBuilder (blue). The AlphaFold2 structure is presented as the top-ranked model. pLDDT scores above 90 indicate very high accuracy, while scores between 70 to 90 indicate reliable prediction. pTM scores greater than 0.5 indicate the same globular fold as the template. Non-canonical disulfide bonds are shown in the inserted panels. (B) No non-canonical disulfide bonds. (C) One non-canonical disulfide bond in FR2–CDR3. (D) Two non-canonical disulfide bonds in CDR1–CDR3 and CDR3–CDR3.



**Fig. 5.** scRNA-seq of post-immunization PBMCs and comparative analysis. (A) Immunization schedule and serum Ab titer. Two camels received four doses of BSA on the days indicated by black arrows, and PBMCs were collected for scRNA-seq before and after immunization on the days indicated by red arrows. The Ab titer was measured as the highest serum dilution using ELISA. (B) Number of DEGs identified in major cell types (PC, plasma cell). Up-regulated genes in post-immunization samples compared to pre-immunization samples are colored in orange, while down-regulated genes are colored in blue. (C) Dot plot of selected DEGs with functional enrichment in B cells and PCs. The color of dots indicates log-fold change ( $\log_2 FC$ ) while the size of dots indicates the normalized mean expression on the pseudobulk level. DEGs with BH-adjusted  $P < 0.05$  in a population are marked with stars. (D, E) Comparison of subtype composition in  $VH+$  and  $VHH+$  B cells. The average proportion before immunization (D) and changes in the proportion after immunization (E) are estimated using a linear model for each subtype. The error bars represent standard errors of estimation.  $P$ -values are calculated using the two-sided  $t$ -test and adjusted using the BH method. (F) Comparison of the  $\log_2 FC$  of DEGs between  $VH+$  and  $VHH+$  B cells during immunization. DEGs are identified separately in non-PCs and PCs, with *IGHC* genes excluded from the analysis. The Pearson's correlation coefficient  $R$  and its  $P$ -value are shown.

and proliferation, immune response regulation, inflammatory regulation, calcium-dependent proteins, and chemokine binding, consistent with the increased serum Ab titer (Fig. 5 C).

A comparison of B cell subtype proportions showed no significant changes during immunization after multiple testing corrections (Supplementary Fig. S21). Our BCR analysis of post-immunization samples did not reveal significant changes in *IGHC* proportions (Supplementary Fig. S22) or clonotype diversity (Supplementary Fig. S23), possibly due to the different kinetics of plasma cell abundance and serum Ab concentrations [24,51]. Although limited DEGs were detected for B cell subtypes, excluding plasma cells, due to insufficient UMI counts, the expression changes of most DEGs in B cells remained consistent across all B cell subtypes (Fig. 5 C).

We also compared the proportion of B cell subtypes in  $VH+$  and  $VHH+$  cells using a linear model, which could simultaneously estimate the average proportion before immunization and changes in the proportion after immunization. Prior to immunization, the proportion of intermediate B cells was significantly lower in  $VHH+$  cells than in  $VH+$  cells (BH-adjusted  $P = 2.37 \times 10^{-3}$ , two-sided  $t$ -test, Fig. 5D).

Accordingly, the proportions of T-bet+ (BH-adjusted  $P = 9.12 \times 10^{-3}$ , two-sided  $t$ -test) and plasma cells (BH-adjusted  $P = 0.02$ , two-sided  $t$ -test) were significantly higher in  $VHH+$  cells than in  $VH+$  cells. However, there were no significant changes in the proportions of any B cell subtype after immunization (Fig. 5E). To determine whether  $VH+$  and  $VHH+$  B cells exhibited consistent changes in gene expression, we performed differential expression analyses using the pseudobulk method for  $VH+$  and  $VHH+$  B cells, respectively. The log-fold changes ( $\log_2 FC$ ) in DEGs between  $VH+$  and  $VHH+$  B cells exhibited a strong positive correlation in both non-plasma and plasma cells (Fig. 5F). These results indicate highly similar transcriptional changes in  $VH+$  and  $VHH+$  B cells during the immune process.

### 3. Discussion

Camelids are the only known mammals capable of producing HCABs. The sequence features and genetic basis of camelid HCABs have been well characterized and are widely applied in sdAb engineering [3,4]. However, the cellular processes underlying HCAB production remain

poorly understood. Certain IgM+ B cells could express *VHH* genes without somatic mutations, suggesting that B cells bearing dimeric IgG might undergo a naive IgM+ stage, similar to conventional IgG-producing cells [12]. However, the phenotypes of camelid B cells have not been fully characterized owing to the absence of surface markers and corresponding reagents [26]. Recently, scRNA-seq was performed on PBMCs of an alpaca during immunization process [24], but 3' RNA sequencing was unable to distinguish the BCR types. In this study, we employed single-cell 5' RNA sequencing to simultaneously reveal the phenotypes and BCR sequences of peripheral B cells in Bactrian camels.

First, we identified cell components in camelid PBMCs, especially B cell subtypes. Compared with humans and mice, camel PBMCs contained a higher proportion of  $\gamma\delta$  T and T-bet+ B cells. Camelid T-bet+ B cells share predominant markers with human atypical B cells, including higher expression of *TBX21* and *ITGAX* and lower expression of *CR2* [31]. Our pseudotime analysis revealed that in addition to the classical differentiation trajectory involving memory B cells, T-bet+ B cells were part of an alternative trajectory of B cells in camels. A similar alternative trajectory for human atypical B cells has been reported [35]. Although human atypical cells are usually associated with chronic infections and inflammation, universal definitions and functions of these cells remain elusive. Therefore, camels provide a unique opportunity to study the functions of T-bet+ cells. Given the limited knowledge of marker genes for camel B cells and their subtypes owing to the lack of camel-specific Abs, our scRNA-seq analysis greatly expanded this knowledge. We found that although many marker genes were shared by humans and camels, there were notable differences. *CD27* and *IGHD*, two widely used marker genes for human memory and naive B cells, respectively [31], were expressed at low levels in all camel B cell subsets. In fact, *CD27* was also found to be absent from most memory B cells in mice [52], and *IGHD* harbors several inactivation mutations in the genomes of both alpaca and Bactrian camels [12,13].

By reconstructing the BCR sequences, we determined the *IGHV* (*VHH*+ and *VH*+) and *IGHC* gene types (*IGHG2/3*+, *IGHG1*, and others) of each B cell. We observed *VHH*+ cells in most B cell subtypes at a proportion comparable to *VH*+ cells, with a significantly lower proportion in intermediate B cells. The differentiation trajectories of *VHH*+ cells largely overlapped with *VH*+ B cells, indicating that both *VHH*+ and *VH*+ cells underwent similar developmental stages. Although *VHH* genes preferentially pair with *IGHG2/3* to form HCABs, and *VH* genes typically pair with other *IGHC* genes to form conventional Abs, we found the widespread presence of BCRs with non-typical connections between *IGHV* and *IGHC* genes. These non-typical connections have two implications. One is the transitional state of *IGHC* class switching. Indeed, in naive B cells, *VHH*-bearing cells can pair with *IGHM*, resulting in an IgM+ state [12]. Additionally, BCRs with non-typical connections can possess specific functions. For camelid sdAb cloning, HCABs with *VH* genes and high antigen affinities have been isolated [41]. We also found that camel plasma cells expressed BCRs in the form of *VHH*+*IGHA*+. The structures and functions of these BCRs require further investigation. It should be noted that although we defined *VH* and *VHH* genes based on the most discriminative amino acid substitutions in FR2 (G49E/Q and L50R) and validated them based on other hallmark substitutions (V42F/Y and W52G/F/W/L) [15,40], the possibility that genetic-origin *VHHs* may have *VH* hallmarks and genetic-origin *VHs* may have *VHH* hallmarks cannot be ruled out. This could be caused by various mechanisms, including somatic hypermutation, gene conversion [53], or secondary rearrangement [54], although the frequency of these mechanisms in camels remains unknown. Another novel finding of our study was the expression of light-chain genes in HCAB-expressing cells. This discovery suggests that the fundamental genetic and transcriptional mechanisms responsible for producing light chains are active in these cells, and that the absence of light chains in HCABs is due to protein assembly issues.

We also investigated dynamic changes in camel PBMCs before and

after BSA immunization. We found widespread signals of transcriptional activation of genes related to immune responses after immunization in the major PBMC types. Although *VHH*+ and *VH*+ cells showed different proportions among the B cell subtypes before immunization, the changes in their proportions were similar after immunization. The DEGs in B cells also exhibited highly consistent patterns between *VHH*+ and *VH*+ cells. These results suggest that the transcriptional regulatory programs of the two B cell populations do not exhibit significant differences. Notably, we sequenced cells at the time of the highest serum Ab titer, which may not necessarily correspond to the highest plasma cell abundance [51]. In fact, Lyu et al. [24] found that the proportion of plasma cells in an immunized alpaca peaked 3 days after the second antigen simulation but rapidly returned to the pre-immunization level after 2 days. Therefore, changes in the proportion and expression of B cell subpopulations depend on the timing of sample collection during immunization.

Antigen-specific B-cell sorting combined with single-cell BCR sequencing is a rapid screening strategy that has been successfully applied for Ab discovery in humans [55] and animals [56]. Although we did not conduct antigen-specific BCR sequencing, our analysis demonstrated a wide scaffold diversity among the *VHH* sequences obtained through scRNA-seq, as evidenced by the predicted structures. The majority of these sequences exhibited a convex shape with a long CDR3 region and were rich in additional cysteines. These observations align with the known characteristics of *VHH* sequences with established structures. Therefore, the camel BCR reconstruction method that we developed holds promise for accelerating the discovery of camelid sdAbs.

## 4. Methods

### 4.1. Camel PBMC collection

We selected two healthy male and two female Bactrian camels aged 4–5 years from Tuzuo Banner, Inner Mongolia, China. For each camel, 50 mL peripheral blood was collected from the jugular vein and placed in a sodium citrate anticoagulation tube after disinfection treatment. The study was reviewed and approved by the Animal Welfare and Ethics Committee of Inner Mongolia Agricultural University (approval no. NND2022078). PBMCs were isolated from the peripheral blood with density gradient centrifugation using the Ficoll-Paque medium. PBMCs (3 mL) were transferred into 5 mL cryotubes with 2 mL cell freezing medium (Yeasten, 40128E550). Cells were stored at  $-80^{\circ}\text{C}$  for 1 day and then transported to sequencing lab on dry ice.

### 4.2. Single-cell library construction and sequencing

The viability of the thawed cells exceeded 80% as determined using trypan blue staining. Single-cell libraries were constructed using the Chromium Next GEM Single Cell V(D)J Reagent Kits v1.1 (10 × Genomics, 5' Library Kit) according to the manufacturer's instructions. Cells, barcoded gel beads, and partitioning oil were combined on a microfluidic chip to form gel beads-in-emulsions (GEMs). Within each GEM, a single cell was lysed, and the transcripts were identically barcoded through reverse transcription. The cDNA libraries were sequenced on an Illumina NovaSeq platform to generate 2 × 150-bp paired-end reads.

### 4.3. scRNA-seq data analysis

Sequencing reads were processed using the Cell Ranger pipeline (v6.1.2, 10 × Genomics). Genome alignment was performed against the assembly BCGSAC\_Cfer\_1.0 (GCF\_009834535.1). Only the annotations of protein-coding genes in RefSeq, including mitochondrial genes, were preserved. Our previous BCR/TCR gene annotations [13] were also incorporated because they were incomplete in RefSeq. The gene-barcode count matrix of UMIs was analyzed using Seurat (v4.0.0) [25].

Specifically, cells with a total UMI count between 1000 and 60,000 and a mitochondrial gene percentage < 5% were preserved for quality control. The count matrix was normalized using *sctransform* [57], and cells across different samples were integrated using canonical correlation analysis. Principal component analysis (PCA) was applied to the integrated matrix, and the top 30 PCs were used for uniform manifold approximation and projection (UMAP) and nearest neighbor graph-based clustering. The clustering resolution was set to 0.5. Marker genes were identified between clusters using *FindMarkers* (Bonferroni-adjusted  $P < 0.05$ , two-sided Wilcoxon test).

#### 4.4. Comparison with human peripheral B cells

For overall transcriptome comparison, we mapped camel B cells to the human reference PBMC dataset using Seurat's reference mapping utility [25]. Anchors between the camel query and human reference were identified using *sctransform* and supervised PCA with the top 50 dimensions. The cell-type labels were then transferred from the human reference to the camel query. For individual cell marker comparisons, we retrieved marker genes of human peripheral B cells and their subtypes from *CellMarker 2.0* [33]. Because these marker genes may not originate from scRNA-seq data, we validated them using the human reference PBMC dataset. Only marker genes showing significant over-expression in B cells or their corresponding subtypes in the human dataset were preserved for comparison (Bonferroni-adjusted  $P < 0.05$ , two-sided Wilcoxon test).

#### 4.5. Differentiation trajectory inference

Differentiation trajectories of B cells (excluding plasma cells) were inferred using *Monocle2* (v2.18.0) [34]. The UMI count matrix was modeled with a negative binomial distribution, and the size factors and dispersions were estimated. Genes expressed in at least 10 cells were preserved. We selected the top 3000 genes with differential expression across B cell subtypes ( $q$ -value < 0.01) to define the trajectory. *DDRTree* was used for dimensional reduction, and the cells were ordered in the reduced dimensional space. Pseudotime was assigned to the trajectories by assuming naive B cells as the root.

#### 4.6. BCR assembly and annotation

The *IGHC* gene type of each B cell was determined if the UMI count of that gene  $\geq 3$ . For cells expressing more than one *IGHC* gene, the UMI count of the major gene should be three times higher than that of the minor gene. We assembled single-cell BCR contigs using *TRUST4* (v1.0.6) [38], with known *IGHV/D/J/C* sequences of Bactrian camels as the reference [13]. The raw FASTQ files were used as the input, and “-barcodeRange 0 15 + -umiRange 16 25 + -read1Range 40 - 1” was set in compliance with the sequence composition of Read 1. Only barcodes in the 10 × Genomics white list were preserved. We re-annotated the assembled contigs using *IgBLAST* (v1.10.0) [39], including the FR2 and CDR3 sequences, in compliance with IMGT standards [58]. Contigs with short alignment lengths (<150 bp) or non-productive sequences were removed. We determined *VH/VHH* contigs based on G49E/Q and L50R in the complete FR2, which are the most discriminative sites [15,40]. For cells with more than one contig containing different FR2 or CDR3 sequences, the contig with highest coverage was selected. If a cell contained both *VH* and *VHH* contigs, it was assigned to the *IGHV* type only when the coverage of the major type was three times larger than that of the minor type. The BCR sequences and B cell transcriptomes were matched based on the same cell barcodes.

#### 4.7. Structure prediction for VHH sequences

Following Zimmermann et al. [44], we defined VHH shapes based on

the length of the CDR3 sequences: concave (<10 aa), loop (10–14 aa), and convex (>14 aa). We also retrieved 630 VHH sequences with known structures from *SABDab* [45] and annotated the CDR3 region using *ANARCI* [59]. The structures of different VHH scaffold types were predicted using *AlphaFold2* [46] and the *ImmuneBuilder* webserver [47]. *AlphaFold2* predictions were obtained using *ColabFold* (v1.5.5) [60], with the top-ranked model retained. We assessed the quality of predicted models using the predicted local distance difference test (pLDDT) score and the predicted template modeling (pTM) score. The pLDDT scores above 90 indicate very high accuracy, while those between 70 and 90 indicate confident prediction [61]. The pTM scores range from 0 to 1, where a score greater than 0.5 is generally interpreted as representing the same globular fold as the template [62].

#### 4.8. Camel immunization

We immunized two camels using BSA as an antigen for four times, with each immunization scheduled 14 days apart. Each immunization used an antigen dose of 0.5 mg (1 mg/mL) mixed with an equal volume of complete Freund's adjuvant (initial immunization) or incomplete Freund's adjuvant, and thoroughly emulsified for subcutaneous multi-point injections in the neck. Before immunization, 10 mL of whole blood was collected for serum titer determination using ELISA.

#### 4.9. ELISA

BSA was diluted in pH 9.6 carbonate-bicarbonate buffer to a concentration of 2 µg/mL, and 100 µL was added per well for coating overnight at 4 °C. After discarding the coating solution, the plate was washed three times with 0.1% PBST (PBS with Tween 20), and then 300 µL of 5% skim milk (dissolved in 0.1% PBST) was added to each well for blocking at 37 °C for 1 h. After discarding the blocking solution, the plate was washed once with 0.1% PBST, and then 100 µL of different dilutions of antiserum (primary Ab, diluted in 5% skim milk) was added to each well at 37 °C for 1 h. Serum from unimmunized animals was used as a negative control, and 5% skim milk was used as a blank control. After washing the plate five times with 0.1% PBST, 100 µL HRP-conjugated goat anti-camel IgG (secondary Ab, *NBbiolab*, diluted 1:10,000) was added to each well and incubated at 37 °C for 1 h. After washing the plate five times with 0.1% PBST, 100 µL TMB substrate solution was added per well for color development and incubated at 37 °C for 7 min. The reaction was stopped by adding 50 µL of stop solution (6 M HCl), and the OD450 value was measured. A positive well was defined as a serum sample with an OD450 value greater than three times that of unimmunized serum and a reading greater than 0.2. The Ab titer was defined as the maximum dilution of positive wells.

#### 4.10. Comparison of cell-type composition during immunization

We processed the post-immunization scRNA-seq data using the same pipeline used for pre-immunization data. Cell-type annotations were transferred from pre-immunization samples using Seurat's reference transfer utility [25]. The proportion of each cell type in each sample was calculated and a linear model “proportion ~ camel + day” was fitted, where the day was scaled from 0 to 1. The intercept of the model represents the average proportion before immunization using the sum contrast, whereas the beta coefficient of the day represents the change in the proportion after immunization. Two-sided *t*-tests were used to test for significant differences from zero of the coefficients, and *P*-values were adjusted using the BH method.

#### 4.11. Differential expression analysis

We performed differential expression analysis between pre- and post-immunization samples using the pseudobulk approach, which accounts for the intrinsic variability of biological replicates and is superior to



single-cell methods [50]. Pseudobulk UMI count matrices for each cell type and *VHH+* /*VH+* population were constructed using Libra (v1.0.0) [50], and DEGs were identified using DESeq2 (v1.30.1) [63]. Genes with low expression, defined as a total UMI count of less than 10 across all samples, were filtered out. The design formula used for DESeq2 analysis was “~ camel + day”, where the day was scaled from 0 to 1 to ensure that the regression coefficient represents the logFC during immunization. The cutoff for determining DEGs was BH-adjusted  $P < 0.05$ . ClusterProfiler (v3.16.0) [64] was used for functional over-representation analyses of DEGs, taking the Gene Ontology (GO) [65] of human orthologs as the database. Up- and down-regulated DEGs were analyzed separately, and the cutoff for significant GO terms was BH-adjusted  $P < 0.05$ .

## Data and code availability

The raw sequencing data generated in this study have been deposited in the Sequence Read Archive (SRA) database under accession code PRJNA997575 (<https://www.ncbi.nlm.nih.gov/sra>). The processed data have been deposited in the Gene Expression Omnibus (GEO) database under accession code GSE238082 (<https://www.ncbi.nlm.nih.gov/geo>). Source code used for analyzing the data are available at a public GitHub repository: <https://github.com/zhenwang100/Abseq>.

## Declaration of Competing Interest

The authors declare no conflict of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (32070570), the Special Fund for Commercialization of Scientific and Research Findings in Inner Mongolia Autonomous Region (2021CG0021) and the National Key Research and Development Project (2020YFE0203300). We thank NEO Bio-technology, Co. Ltd who provided scRNA-seq support for this work. We also thank Elsevier Language Editing Services for its linguistic assistance during the preparation of this manuscript.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.csbj.2024.04.041](https://doi.org/10.1016/j.csbj.2024.04.041).

## References

- Schroeder Jr HW, Cavacini L. Structure and function of immunoglobulins. *J Allergy Clin Immunol* 2010;125(2 Suppl 2): S41–52.
- Hamers-Casterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, et al. Naturally occurring antibodies devoid of light chains. *Nature* 1993;363(6428): 446–8.
- Muyldermans S. Nanobodies: natural single-domain antibodies. *Annu Rev Biochem* 2013;82:775–97.
- Muyldermans S. Applications of nanobodies. *Annu Rev Anim Biosci* 2021;9: 401–21.
- Peyvandi F, Scully M, Kremer Hovinga JA, Cataland S, Knobl P, et al. Caplacizumab for acquired thrombotic thrombocytopenic purpura. *N Engl J Med* 2016;374(6): 511–22.
- Nguyen VK, Hamers R, Wyns L, Muyldermans S. Loss of splice consensus signal is responsible for the removal of the entire C(H)1 domain of the functional camel IGG2A heavy-chain antibodies. *Mol Immunol* 1999;36(8):515–24.
- Woolven BP, Frenken LG, van der Logt P, Nicholls PJ. The structure of the llama heavy chain constant genes reveals a mechanism for heavy-chain antibody formation. *Immunogenetics* 1999;50(1–2):98–101.
- Liang Z, Wang T, Sun Y, Yang W, Liu Z, et al. A comprehensive analysis of immunoglobulin heavy chain genes in the Bactrian camel (*Camelus bactrianus*). *Front Agr Sci Eng* 2015;2(3):249–59.
- Muyldermans S, Atarhouch T, Saldanha J, Barbosa JA, Hamers R. Sequence and structure of VH domain from naturally occurring camel heavy chain immunoglobulins lacking light chains. *Protein Eng* 1994;7(9):1129–35.
- Vu KB, Ghahroudi MA, Wyns L, Muyldermans S. Comparison of llama VH sequences from conventional and heavy chain antibodies. *Mol Immunol* 1997;34(16–17):1121–31.
- Hamers MM, Ruuls RC, Nijman LJ, Niewold TA, Frenken LG, et al. Llama heavy-chain V regions consist of at least four distinct subfamilies revealing novel sequence features. *Mol Immunol* 2000;37(10):579–90.
- Achour I, Cavalier P, Tichit M, Bouchier C, Lafaye P, et al. Tetrameric and homodimeric camelid IgGs originate from the same IgH locus. *J Immunol* 2008; 181(3):2001–9.
- Ming L, Wang Z, Yi L, Batmunkh M, Liu T, et al. Chromosome-level assembly of wild Bactrian camel genome reveals organization of immune gene loci. *Mol Ecol Resour* 2020;20(3):770–80.
- Nguyen VK, Hamers R, Wyns L, Muyldermans S. Camel heavy-chain antibodies: diverse germline V(H)H and specific mechanisms enlarge the antigen-binding repertoire. *EMBO J* 2000;19(5):921–30.
- Liu Y, Yi L, Li Y, Wang Z, Jirimutu. Characterization of heavy-chain antibody gene repertoires in Bactrian camels. *J Genet Genom* 2023;50(1):38–45.
- Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, et al. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat Biotechnol* 2014;32(2):158–68.
- LeBien TW, Tedder TF. B lymphocytes: how they develop and function. *Blood* 2008;112(5):1570–80.
- Kaminski DA, Wei C, Qian Y, Rosenberg AF, Sanz I. Advances in human B cell phenotypic profiling. *Front Immunol* 2012;3:302.
- Morgan D, Tergaonkar V. Unraveling B cell trajectories at single cell resolution. *Trends Immunol* 2022;43(3):210–29.
- Gomes T, Teichmann SA, Talavera-Lopez C. Immunology driven by large-scale single-cell sequencing. *Trends Immunol* 2019;40(11):1011–21.
- Hilton HG, Rubinstein ND, Janki P, Ireland AT, Bernstein N, et al. Single-cell transcriptomics of the naked mole-rat reveals unexpected features of mammalian immunity. *PLoS Biol* 2019;17(11):e3000528.
- Patel RS, Tomlinson JE, Divers TJ, Van de Walle GR, Rosenberg BR. Single-cell resolution landscape of equine peripheral blood mononuclear cells reveals diverse cell types including T-bet(+) B cells. *BMC Biol* 2021;19(1):13.
- Koiwai K, Koyama T, Tsuda S, Toyoda A, Kikuchi K, et al. Single-cell RNA-seq analysis reveals penaeid shrimp hemocyte subpopulations and cell differentiation process. *Elife* 2021;10:e66954.
- Lyu M, Shi X, Liu Y, Zhao H, Yuan Y, et al. Single-cell transcriptome analysis of H5N1-HA-stimulated alpaca PBMCs. *Biomolecules* 2022;13(1):60.
- Hao Y, Hao S, Andersen-Nissen E, Mauck 3rd WM, Zheng S, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184(13):3573–87.
- Hussen J, Schuberth HJ. Recent Advances in Camel Immunology. *Front Immunol* 2020;11:614150.
- Pizzolato G, Kaminski H, Tosolini M, Franchini DM, Pont F, et al. Single-cell RNA sequencing unveils the shared and the distinct cytotoxic hallmarks of human TCRVdelta1 and TCRVdelta2 gamma delta T lymphocytes. *Proc Natl Acad Sci USA* 2019;116(24):11906–15.
- Weng NP, Araki Y, Subedi K. The molecular basis of the memory T cell response: differential gene expression and its epigenetic regulation. *Nat Rev Immunol* 2012; 12(4):306–15.
- Karnell JL, Kumar V, Wang J, Wang S, Vovnova E, et al. Role of CD11c(+) T-bet(+) B cells in human health and disease. *Cell Immunol* 2017;321:40–5.
- Holla P, Dizon B, Ambegaonkar AA, Rogel N, Goldschmidt E, et al. Shared transcriptional profiles of atypical B cells suggest common drivers of expansion and function in malaria, HIV, and autoimmunity. *Sci Adv* 2021;7(22):eabg8384.
- Sanz I, Wei C, Jenks SA, Cashman KS, Tipton C, et al. Challenges and opportunities for consistent classification of human B cell and plasma cell populations. *Front Immunol* 2019;10:2458.
- Zanini F, Robinson ML, Croote D, Sahoo MK, Sanz AM, et al. Virus-inclusive single-cell RNA sequencing reveals the molecular signature of progression to severe dengue. *Proc Natl Acad Sci U S A* 2018;115(52):E12363–9.
- Hu C, Li T, Xu Y, Zhang X, Li F, et al. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res* 2023;51(D1):D870–6.
- Qiu X, Mao Q, Tang Y, Wang L, Chawla R, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 2017;14(10):979–82.
- Sutton HJ, Aye R, Idris AH, Vistein R, Nduati E, et al. Atypical B cells are part of an alternative lineage of B cells that participates in responses to vaccination and infection in humans. *Cell Rep* 2021;34(6):108684.
- Shi Z, Zhang Q, Yan H, Yang Y, Wang P, et al. More than one antibody of individual B cells revealed by single-cell immune profiling. *Cell Discov* 2019;5:64.
- Tillib SV, Vyatchanin AS, Muyldermans S. Molecular analysis of heavy chain-only antibodies of *Camelus bactrianus*. *Biochem (Mosc)* 2014;79(12):1382–90.
- Song L, Cohen D, Ouyang Z, Cao Y, Hu X, et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. *Nat Methods* 2021;18(6): 627–30.
- Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* 2013;41(Web Server issue):W34–40.
- Deszynski P, Mlokosiewicz J, Volanakis A, Jaszczyszyn I, Castellana N, et al. INDI-integrated nanobody database for immunoinformatics. *Nucleic Acids Res* 2022;50(D1):D1273–81.
- Deschacht N, De Groeve K, Vincke C, Raes G, De Baetselier P, et al. A novel promiscuous class of camelid single-domain antibody contributes to the antigen-binding repertoire. *J Immunol* 2010;184(10):5696–704.

- [42] McMahon C, Baier AS, Pascolutti R, Wegrecki M, Zheng S, et al. Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nat Struct Mol Biol* 2018;25(3):289–96.
- [43] Melarkode Vattekatte A, Shinada NK, Narwani TJ, Noel F, Bertrand O, et al. Discrete analysis of camelid variable domains: sequences, structures, and in-silico structure prediction. *PeerJ* 2020;8:e8408.
- [44] Zimmermann I, Egloff P, Hutter CA, Arnold FM, Stohler P, et al. Synthetic single domain antibodies for the conformational trapping of membrane proteins. *Elife* 2018;7:e34317.
- [45] Schneider C, Raybould MJ, Deane CM. SABDab in the age of biotherapeutics: updates including SABDab-nano, the nanobody structure tracker. *Nucleic Acids Res* 2022;50(D1):D1368–72.
- [46] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596(7873):583–9.
- [47] Abanades B, Wong WK, Boyles F, Georges G, Bujotzek A, et al. ImmuneBuilder: deep-learning models for predicting the structures of immune proteins. *Commun Biol* 2023;6(1):575.
- [48] Sircar A, Sanni KA, Shi J, Gray JJ. Analysis and modeling of the variable region of camelid single-domain antibodies. *J Immunol* 2011;186(11):6357–67.
- [49] Govaert J, Pellis M, Deschacht N, Vincke C, Conrath K, et al. Dual beneficial effect of interloop disulfide bond for single domain antibody fragments. *J Biol Chem* 2012;287(3):1970–9.
- [50] Squair JW, Gautier M, Kathe C, Anderson MA, James ND, et al. Confronting false discoveries in single-cell differential expression. *Nat Commun* 2021;12(1):5692.
- [51] Blanchard-Rohner G, Pulickal AS, Jol-van der Zijde CM, Snape MD, Pollard AJ. Appearance of peripheral blood plasma cells and memory B cells in a primary and secondary immune response in humans. *Blood* 2009;114(24):4998–5002.
- [52] Xiao Y, Hendriks J, Langerak P, Jacobs H, Borst J. CD27 is acquired by primed B cells at the centroblast stage and promotes germinal center formation. *J Immunol* 2004;172(12):7432–41.
- [53] Arakawa H, Hauschild J, Buerstedde JM. Requirement of the activation-induced deaminase (AID) gene for immunoglobulin gene conversion. *Science* 2002;295(5558):1301–6.
- [54] Sun A, Novobrantseva TI, Coffre M, Hewitt SL, Jensen K, et al. VH replacement in primary immunoglobulin repertoire diversification. *Proc Natl Acad Sci U S A* 2015;112(5):E458–66.
- [55] Cao Y, Su B, Guo X, Sun W, Deng Y, et al. Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* 2020;182(1):73–84.
- [56] Goldstein LD, Chen YJ, Wu J, Chaudhuri S, Hsiao YC, et al. Massively parallel single-cell B-cell receptor sequencing enables rapid discovery of diverse antigen-reactive antibodies. *Commun Biol* 2019;2:304.
- [57] Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 2019;20(1):296.
- [58] Lefranc MP, Pommie C, Ruiz M, Giudicelli V, Foulquier E, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev Comp Immunol* 2003;27(1):55–77.
- [59] Dunbar J, Deane CM. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* 2016;32(2):298–300.
- [60] Mirdita M, Schutze K, Moriwaki Y, Heo L, Ovchinnikov S, et al. ColabFold: making protein folding accessible to all. *Nat Methods* 2022;19(6):679–82.
- [61] Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, et al. AlphaFold protein structure database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2022;50(D1):D439–44.
- [62] Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins* 2004;57(4):702–10.
- [63] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15(12):550.
- [64] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 2012;16(5):284–7.
- [65] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25(1):25–9.