High-Throughput Sequencing Reveals Immunological Characteristics of the *TRB-/IgH-CDR3* Region of Umbilical Cord Blood

Changlong Guo, MD, PhD[1,*], Qidi Wang, MD, PhD[1,*], Xiaofang Cao, MD[1], Ying Yang, MD, PhD[1], Xin Liu, PhD[1,2], Lisha An, MD[1], Ruikun Cai, PhD[1], Meng Du, MD[1,3], Guangyu Wang, MD[1], Yue Qiu, MD[1,3], Zuoqi Peng, MD[1], Jian Han, MD, PhD[4], Shuhua Ni, PhD[5], Xuerui Tan, MD, PhD[5], Li Jin, MD[6], Song Yu, MD[7], Huiying Wang, MD[8], Chunlin Wang, PhD[1,9], Xingyu Wang, MD, PhD[1,2], and Xu Ma, MD, PhD[1,3]

Objective To compare the differences of immunological characteristics between newborn and adults, we performed high-throughput sequencing to reveal the diversity of umbilical cord blood and adult peripheral blood at both T-cell receptor beta chain (TRB) and immunoglobulin heavy chain (IGH) levels.

Study design High-throughput sequencing was performed to analyze the expression of *TRB-CDR3* and *IGH-CDR3* in circulating T and B cells isolated from 20 healthy adults, 56 pregnant women, and 40 newborns.

Results Our results revealed different immunological characteristics between newborn and adults, such as distinctive complementarity determining region 3 (CDR3) lengths, usage bias of variable and joining segments, random nucleotide addition, a large number of unique CDR3 peptides, and a greater repertoire diversity. Moreover, each newborn had a distinctive TRB-/IGH-CDR3 repertoire that was independent of the maternal immune status.

Conclusions This study presents comprehensive, unrestricted profiles of the TRB/IGH-CDR3 repertoire of newborns, pregnant women, and healthy adults at a sequence-level resolution. Our data may contribute to a better understanding of the immune system of newborns and benefit the efficient application of umbilical cord blood transplantation in future. *(J Pediatr 2016;176:69-78)*.

Umbilical cord blood (UCB) is a rich source of hematopoietic stem cells (HSCs), accounting for 0.5%-1.0% of mononuclear cells.[1] Because of prompt availability, ease of collection with no risk to either mother or newborn, less-stringent HLA matching, a decreased risk of transmissible viral infections, and reduced incidence of graft-vs-host disease, UCB has become an established alternative source of HSCs to treat several malignant and nonmalignant diseases.[2-6]

Despite these successes, the use of UCB for HSC transplants still has major challenges. A successful transplant requires a sufficient number of progenitors to participate in rapid short-term neutrophil and platelet recovery and also a sufficient number of long-term repopulating stem cells to sustain hematopoiesis for the lifetime of the individual.[6,7] The vast majority of T and B lymphocytes, however, are antigen-inexperienced in UCB. These naïve CD45RA+ cells have a lower antigen-presenting activity[8,9] with reduced expression of transcription factors associated with T-cell activation (eg, nuclear factor of activated T cell, tumor necrosis factor-α, CD80, CD86), and produce lower levels of effector cytokines (tumor necrosis factor-α, interleukin [IL]-1, IL-2, and IL-12) compared with activated T cells from adult peripheral blood.[10-12] Therefore, it usually takes at least 3-6 months to reconstitute the immune system after UCB transplantation, which greatly increases the risk of infection as well as economic burden compared with bone marrow-mobilized peripheral blood transplantation.[13,14] Moreover, previous studies have revealed obvious differences in the usage of the variable (V), diversity (D), and joining (J) segments of UCB compared with adult peripheral blood mononuclear cells (PBMCs). In addition, T-cell receptor alpha and beta variable regions show polyclonal but restricted expression patterns in UCB.[15,16]

Because of the limitations of technology and/or sample numbers, however, previous studies of T- and B-cell receptor diversity have only provided a descriptive assessment based on a limited number of sequences.[17,18] To fully explore the immunological characteristics of UCB, we applied next-generation sequencing

BM	Bone marrow	IL	Interleukin
CDR3	Complementarity determining region 3	J	Joining
		N addition	Nucleotides addition
D	Diversity	NGS	Next-generation sequencing
D50	Diversity index	nt	Nucleotides
HIGH	Human immunoglobulin heavy chain	PBMC	Peripheral blood mononuclear cell
HSC	Hematopoietic stem cell	PCR	Polymerase chain reaction
HTBI	Human T-cell receptor beta chain	TRB	T-cell receptor beta chain
		UCB	Umbilical cord blood
IGH	Immunoglobulin heavy chain	V	Variable

From the [1]Department of Genetics, National Research Institute for Family Planning; [2]Laboratory of Human Genetics, Beijing Hypertension League Institute; [3]Graduate School of Peking Union Medical College, Beijing, China; [4]HudsonAlpha Institute for Biotechnology, Huntsville, AL; [5]First Affiliated Hospital, Medical College of Shantou University, Shantou, Guangdong, China; [6]Beijing Friendship Hospital, Capital Medical University; [7]Beijing Obstetrics and Gynecology Hospital, Capital Medical University, Beijing Maternal and Child Healthy Care Hospital; [8]Beijing Shijitan Hospital, Capital Medical University, Beijing, China; and [9]Stanford Genome Technology Center, Stanford University, Palo Alto, CA

*Contributed equally.

Supported by International Science & Technology Cooperation Program of China (0102012DFB30130). The authors declare no conflicts of interest.

http://dx.doi.org/10.1016/j.jpeds.2016.05.078

(NGS) and compared the complementarity determining region 3 (CDR3) regions of T-cell receptor beta chain (TRB) and immunoglobulin heavy chain (IGH) of circulating T and B cells isolated from newborns (ie, UCB), pregnant women, and healthy adults. This study presents a broad view of UCB TRB-/IgH-CDR3 characteristics, including greater diversity, less nucleotides (nt) in random nt addition (N addition) and thus shorter CDR3 lengths, and skewed usage of *V* and *J* gene segments at both *TRB* and *IGH* loci compared with pregnant woman and adults. Our results also revealed that the TRB-/IgH-CDR3 repertoire of newborns was independent of their corresponding maternal immune status. Taken together, these results provide a comprehensive understanding of the differences in cellular immune features between newborns and adults and contribute to exploiting the therapeutic potential of UCB.

Methods

This study was approved by the ethics committee of the National Research Institute for Family Planning and adhered to the Declaration of Helsinki. Written informed consent was obtained from all individuals and/or guardians participating in the study. A total of 20 healthy adults, 56 pregnant women, and 40 newborns were involved in this study. Subjects with congenital anomalies, HSC transplantation, or gene therapy were excluded from the study. Pregnant women with antenatal medical or obstetric complications, such as gestational diabetes and hypertension, also were excluded from the study. UCB was collected from healthy full-term newborns with no symptoms of infection. Subjects had not been administered drugs or experienced infections from 1 month before blood collection.

High-Throughput Sequencing

Samples (10 mL) of peripheral blood or UCB from participants were transferred to the laboratory and processed within 4 hours after collection. PBMCs were isolated by density gradient centrifugation via Ficoll Paque Plus (Haoyang Biotech, Tianjin, China) according to manufacturer's instructions. PBMCs were frozen in RNA protec Reagent (QIAGEN, Hilden, Germany) at −80°C until further processing. The PBMCs were placed in lysis buffer, and mRNA was extracted according to the manufacturer's instructions (RNeasy mini kit; QIAGEN).

Two steps were included in the amplification refractory mutation–polymerase chain reaction (PCR).[19,20] In the first step, 1 μg of total RNA was reverse-transcribed to cDNA with a OneStep RT-PCR kit (QIAGEN). cDNA samples were amplified with human TRB (*HTBI*, covers *V* and *C* genes) primers for *TRB-CDR3* amplification, or human IGH (*HIGH*, covers *V* and *C* genes) primers for *IGH-CDR3* amplification according to manufacturer's instructions (iRepertoire Inc, Huntsville, Alabama). Each primer included 3 regions: a common sequence tag located at the 5′ end, which was used for binding and amplification in the second PCR, a barcode tag located in the middle region, allowing donor identification, and a specific primer sequence at the 3′ end, which bound to *V* and/or *J* genes and specifically reverse-transcribed each CDR3 locus.

In the second step, 2 μL of PCR products from the first PCR step was used as the template for direct amplification of the whole CDR read with communal primers (iRepertoire Inc, Huntsville, Alabama) via a multiplex PCR system (Multiplex PCR Kit; QIAGEN). The amplification refractory mutation-PCR products were separated by agarose gel electrophoresis, and then target DNA fragments (250 bp for *HTBI* and 450 bp for *HIGH*) were excised and extracted with a gel-extraction kit. The concentration and quality of each sample were determined with a KAPA SYBR FAST qPCR kit (KAPA Biosystems, Inc, Wilmington, Massachusetts) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California). Ten samples with equal amounts of DNA were pooled together and mixed with Phix Control (Illumina, San Diego, California) at a ratio of 3:1. Finally, 12.5 pM of PCR products was loaded and sequenced with the Illumina platform (PE150 for HTBI and PE250 for HIGH).

Data Analyses

Raw data of CDR3 peptide reads and VDJ segment usage for *TRB* and *IGH* of each sample were separated according to the unique barcode sequence. Trimmed pair-end reads were joined together through overlapping alignment with a modified Needleman-Wunsch algorithm. If paired forward and reverse reads in the overlapping region were not matched perfectly, both forward and reverse reads were discarded. Rearranged mRNA sequences were assigned to their germline VDJ counterparts and translated into corresponding amino acids by the International Immunogenetics database.[20-22] Abundance and cumulative frequency were calculated for CDR3 sequences. Student *t* test and Mann-Whitney *U* test were used to assess differences between groups. Statistical analysis was performed using SPSS version 13.0 (SPSS Inc, Chicago, Illinois).

The diversity index (D50) is a measurement of the diversity of an immune repertoire of J individuals (total number of CDR3s), including S distinct CDR3s in a ranked dominance configuration, where r_i is the abundance of the most abundant CDR3s, r_1 is the abundance of the 20 most abundant CDR3s, r_2 is the abundance of the second most abundant CDR3, and so on. C is the minimum number of distinct CDR3s amounting to ≥50% of total sequencing reads. D50 was calculated by the following formula:

$$\text{Assume that } \frac{r_1 \geq r_2 \cdots \geq r_i \geq r_{i+1} \cdots \geq r_S}{S}, \quad \sum_{i=1}^{S} r_i = J$$

$$\text{if } \sum_{i=1}^{C} r_i \geq J/2 \text{ and } \sum_{i=1}^{C-1} r_i < J/2$$

$$D50 = \frac{C}{X} \times 100$$