Cancer Letters 379 (2016) 206–212

Contents lists available at ScienceDirect

Cancer Letters

journal homepage: www.elsevier.com/locate/canlet

Mini-review

# Immune repertoire: A potential biomarker and therapeutic for hepatocellular carcinoma

Yingxin Han [a,b], Hongmei Li [a,b], Yanfang Guan [a,b], Jian Huang [a,c,d,*]

[a] Binhai Genomics Institute, BGI-Tianjin, and Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, China
[b] BGI-Shenzhen, Shenzhen 518083, China
[c] Key Laboratory of Systems Biomedicine (Ministry of Education) and Collaborative Innovation Center of Systems Biomedicine, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, Shanghai 200240, China
[d] Shanghai-MOST Key Laboratory for Disease and Health Genomics, Chinese National Human Genome Center, Shanghai 201203, China

ARTICLE INFO

*Keywords:*
Hepatocellular carcinoma
Immune repertoire
Complementarity-determining regions
High-throughput sequencing
Immune repertoire sequencing

ABSTRACT

The immune repertoire (IR) refers to the sum of B cells and T cells with functional diversity in the circulatory system of one individual at any given time. Immune cells, which reside within microenvironments and are responsible for protecting the human body, include T cells, B cells, macrophages, and dendritic cells. These dedicated immune cells have a characteristic structure and function. T and B cells are the main lymphocytes and are responsible for cellular immunity and humoral immunity, respectively. The T cell receptor (TCR) and B cell receptor (BCR) are composed of multiple peptide chains with antigen specificity. The amino acid composition and sequence order are more diverse in the complementarity-determining regions (including CDR1, CDR2 and CDR3) of each peptide chain, allowing a vast library of TCRs and BCRs. IR research is becoming increasingly focused on the study of CDR3 diversity. Deep profiling of CDR3s using high-throughput sequencing is a powerful approach for elucidating the composition and distribution of the CDR3s in a given sample, with in-depth information at the sequence level. Hepatocellular carcinoma (HCC) is one of the most common malignancies in the world. To identify novel biomarkers for diagnosis and drug targets for therapeutic interventions, several groups attempted to describe immune repertoire characteristics of the liver in the physiological environment or/and pathological conditions. This paper reviews the recent progress in IR research on human diseases, including hepatocellular carcinoma, attempting to depict the relationships between hepatocellular carcinogenesis and the IR, and discusses the possibility of IR as a potential biomarker and therapeutic for hepatocellular carcinoma.


## Introduction

T and B lymphocytes of the adaptive immune system are selectively activated via specific recognition of an antigen by the variable region of their cell-surface T cell receptors (TCRs) and B cell receptors (BCRs), respectively. The complementarity-determining regions (CDRs), CDR1, CDR2 and CDR3, together with the surrounding framework regions, form the antigen-binding site of the TCR or BCR. CDR1 and CDR2 are encoded by germline sequences, whereas CDR3, the highly polymorphic principal recognition site, is generated when TCR or BCR genomic loci undergo somatic V(D)J recombination, resulting in the vast receptor diversity. The recombination of gene segments of variable (V), diverse (D) and joining (J) results in rearranged genes encoding for variable TCR or BCR regions [1]. The receptor structures determine their antigenic binding, so the nucleotide sequences of the TCRs and BCRs provide primary functional information for each clone [2]. In early efforts to better describe TCR/BCR diversity at the nucleotide sequence level, traditional molecular cloning techniques coupled with Sanger sequencing provided insight into the variation in TCRs [3,4] and BCRs [5,6]. However, the cost of this type of sequencing and its inherently low throughput make the approach impracticable for assessing CDR3 diversity. Some other technologies, such as CDR3 spectratyping and flow cytometry, have also been used to identify particular expansions. Spectratyping strategies, in which the CDR3 length distribution in lymphocyte pools is measured from PCR-amplified VDJ segments, offer a more general view of immune repertoire diversity and are both inexpensive and rapid; however, these methods ignore the actual sequence content of CDR3 regions [7,8]. Flow cytometry often fails to meet the required sensitivity for minimal residual disease (MRD) detection.

High-throughput sequencing (HTS) technologies have enabled millions of TCR or BCR clonotypes to be identified with a high

*Abbreviations:* HCC, hepatocellular carcinoma; IR, immune repertoire; CDR, complementarity-determining regions; TCRs, T cell receptors; BCRs, B cell receptors; MRD, minimal residual disease; HTS, high-throughput sequencing; IR-SEQ, immune repertoire sequencing; TILs, tumor-infiltrating lymphocytes.

* Corresponding author. Tel.: +86 34206059; fax: +86 34206059.
*E-mail address:* huangj@chgc.sh.cn (J. Huang).

http://dx.doi.org/10.1016/j.canlet.2015.06.022
0304-3835/

specificity and sensitivity [9–11]. HTS technologies have been widely used for analysis of immune repertoires. Freeman et al. were among the first scientists to apply HTS in this study field [12]. These authors analyzed the CDR3 diversity in pooled T lymphocytes from 550 normal human peripheral blood samples using an Illumina sequencer. Simultaneously, the TCR distribution in various T-cell compartments was analyzed in a healthy adult by Robins et al. [13]. The immune diversity and expanded clonal lymphocyte populations were analyzed in physiological and pathological contexts by Boyd et al., using massively parallel sequencing of antibody repertoires [14]. Subsequently, immunosequencing was widely used to evaluate the diversity of immune repertoires and to identify disease-related public sequences and track clones [2,15].

## Assessment of the features of immune repertoires using HTS

Currently, the most widely used HTS platforms include Roche 454 sequencing system, Life SOLiD system, Illumina HiSeq and Genome Analyzer platforms, each of which offers advantages and drawbacks. A comprehensive summary and comparison of these platforms have been described in previous reviews [16–20]. In simple terms, the 454 sequencer provides long reads that are able to capture of full-length IgH cDNA sequences, including all somatic hypermutations. The drawback of this system is a high cost per read. Ion Torrent sequencing technology suffers from high indel rates; however, this method is inexpensive and fast. HiSeq 2000/2500 technology is inexpensive, fast and has far fewer issues with indels; however, with this method, it is difficult to span full TCR or BCR genes because of the short read length produced. The Illumina MiSeq platform is also becoming a useful tool for immune repertoire sequencing and is capable of generating tens of millions of paired 250 bp reads in a short time [21].

Importantly, the veracity of immune repertoire analysis is relied on the reliability of HTS data. The reliability of HTS data depends on the accuracy of the sequences and depth of coverage. Warren et al. confirmed that the sensitivity of sequenced-based repertoire profiling is limited by the sequencing accuracy and depth [22]. The distinguishing between real physiological immunogenetic alterations and sequencing errors can be difficult, leading to misinterpretation of immune repertoire HTS data.

Bolotin et al. compared the three platforms mentioned above for individual TCR profiling and proposed advanced platform-specific algorithms to correct sequencing data [16]. Because TCRs do not undergo somatic hypermutation, a typical strategy for identifying and correcting sequencing and late-cycle PCR errors involves generating consensus assemblies from highly similar reads [16,23,24]. Nguyen et al. showed the approaches for filtering sequences based on single nucleotide mismatches compared with high quality or more frequent parental sequences [23]. This correction enhances the accuracy of clonotype identification and quantification as well as overall TCR diversity measurements. Nevertheless, data filtering is not absolute because some “high-quality” molecular errors, such as reverse transcription errors, may be difficult to correct as all derived molecules and sequencing reads contain the same erroneous substitutions. Recently, Shugay et al. reported molecular identifier groups-based error correction (MIGEC), a strategy for high-throughput sequencing data analysis. MIGEC allows for high quality error correction without affecting the natural diversity of complex immune repertoires [24]. With the rapid development of HTS technology, longer, more accurate, and less expensive reads will be provided in the near future.

## Applications of immune repertoire sequencing

Immune repertoire sequencing (IR-SEQ) amplifies the CDRs of TCRs or BCRs using 5′ RACE or multiple-PCR methods, followed by high-throughput sequencing. IR-SEQ is employed to study the diversity of the immune system and the associations between the immune repertoire and diseases.

### *Applications of IR-SEQ in MRD detection and quantification*

The HTS technique is of vital importance in the fields of immunology and immunotherapy because it has helped researchers to better describe and monitor the TCR and BCR repertoires. IR-SEQ had been applied to individualized monitoring of MRD for many years [14,25–33], showing a higher sensitivity and specificity than flow cytometry. In 2009, Boyd et al. analyzed BCR clonality in healthy adults and patients with hematological malignancies and suggested that HTS may be used as a unified MRD detection method, without real-time PCR assays specially designed for clonal malignant sequences of each individual patient [14]. In another study, Wu et al. applied HTS in the diagnosis of T-lineage acute lymphoblastic leukemia/lymphoma and 43 paired patient samples were then used to assess MRD at day 29 after treatment [25]. The results suggested that IR-SEQ was able not only to identify malignant clones in patients with T and B hematologic malignancies at diagnosis but also to detect subsequent MRD by monitoring these clones during and after treatment, showing great prognostic value in clinical application. Compared with other methods, IR-SEQ has more advantages, including excellent sensitivity and the ability to trace all of the clonal composition of a malignant group at the same time [14,29,32,33]. This technology clearly shows powerful potential as a new tool for monitoring many malignancies, including liver cancer.

### *Applications of IR-SEQ in autoimmune disease*

In any individual autoimmune disease, there are multiple autoantibodies with different specificities, a few of which may be able to serve as useful biomarkers. Antibody characterization has enabled the identification of autoantigens, and the detection of autoantibodies has become the foundation of modern clinical diagnosis [34–37].

In recent years, many studies have focused on identifying clonally expanded T cells in autoimmune diseases, as it does not require information of putative autoantigens. To identify potential autoreactive clones, Klarenbeek et al. quantitated the TCR repertoire in the peripheral blood and joints of patients with early or advanced rheumatoid arthritis (RA) using HTS method. Although the dominant clones in the synovium were patient specific, the researchers proposed that inflamed tissue may be an attractive target for the identification and characterization of potentially autoreactive clones, especially in early RA [38]. Recently, Robert Winchester et al. demonstrated that the public Epstein–Barr virus (EBV)-reactive TCR clones enriched in cerebrospinal fluid (CSF) of multiple sclerosis (MS) patients could serve as potential markers [39]. Researchers from the University of California revealed that B cell clones could be exchanged across the blood–brain barrier (BBB) in multiple sclerosis [40]. This study provided a powerful approach for identifying and monitoring B cells in the peripheral blood (PB) that correspond to clonally amplified populations in the central nervous system (CNS) in MS. Additionally, it provided a theoretical basis for the development of a diagnosis and treatment. Several studies on lupus nephritis demonstrated that the intrarenal infiltrated T cells were relatively oligoclonal and that related expanded clones could also be detected in peripheral blood [41,42]. A clonal CD8+ T-cell lineage found in renal tissue and blood samples from a lupus nephritis patient was reported to be detectable over 6 years, indicating the sustained expression of related expanded clones in this disease [42]. A previous study in systemic lupus erythematosus (SLE) patients also indicated that persistent expression of autoantibodies may be an integral feature of SLE [43].