Journal of Autoimmunity 97 (2019) 100–107

Contents lists available at ScienceDirect

Journal of Autoimmunity

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

The landscape and diagnostic potential of T and B cell repertoire in Immunoglobulin A Nephropathy

Chen Huang[a,1], Xuemei Li[b,c,1], Jinghua Wu[b,c,1], Wei Zhang[b,c,1], Shiren Sun[a,1], Liya Lin[b,c], Xie Wang[b,c], Hongmei Li[d], Xiaolei Wu[d], Peng Zhang[a], Guoshuang Xu[a], Hanmin Wang[a], Hongbao Liu[a], Yuzhen Liu[a], Dapeng Chen[e], Li Zhuo[e], Wenge Li[e], Huanming Yang[b,f], Jian Wang[b,f], Ling Wang[g,**], Xiao Liu[b,c,*]

[a] Department of Nephrology, Xijing Hospital, The Fourth Military Medical University, Xi'an 710032, China
[b] BGI-Shenzhen, Shenzhen 518083, China
[c] China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China
[d] Tianjin Medical Laboratory, BGI-Tianjin, BGI-Shenzhen, Tianjin 300308, China
[e] Department of Nephrology, China-Japan Friendship Hospital, Beijing 100029, China
[f] James D. Watson Institute of Genome Sciences, Hangzhou 310058, China
[g] Department of Vascular and Endocrine Surgery, Xijing Hospital, Forth Military Medical University, Xi'an 710032, China

ARTICLE INFO

Keywords:
Immunoglobulin A Nephropathy (IgAN)
TCRB repertoire
IGH repertoire
Non-invasive
Biomarkers

ABSTRACT

Immunoglobulin A Nephropathy (IgAN) is the most common glomerulonephritis worldwide. The pathologic hallmark of IgAN is immune complex deposited in glomerular mesangium, which induces inflammation and affects the kidney's normal functions. The exact pathogenesis of IgAN, however, remains obscure. Further, in current clinical practice, the diagnosis relies on needle biopsy of renal tissue. Therefore, a non-invasive method for diagnosis and prognosis surveillance of the disease is highly desirable. To this end, we investigated the T cell receptor beta chain (TCRB) and immunoglobulin heavy chain (IGH) repertoire in circulating lymphocytes and compared them with kidney infiltrating lymphocytes using immune repertoire high throughput sequencing. We found that some features of TCRB and IGH in renal tissues were remarkably different from that in the blood, including decreased repertoire diversity, increased IgA and IgG frequency, and more antigen-experienced B cells. The complementarity-determining region 3 (CDR3) length of circulating TCRB and IGH in IgAN patients was significantly shorter than that in healthy controls, which is the result of both VDJ rearrangement and clonal selection. The IgA1 frequency in the blood of IgAN patients is significantly higher than that in other Nephropathy (NIgAN) patients and healthy control. Importantly we identified a set of TCRB and IGH clones, which can be used to distinguish IgAN from NIgAN and healthy controls with high accuracy. These results indicated that the TCRB and IGH repertoire can potentially serve as non-invasive biomarkers for the diagnosis of IgAN. The characteristics of the kidney infiltrating and circulating lymphocytes repertoires shed light on IgAN detection, treatment and surveillance.

1. Introduction

Immunoglobulin A Nephropathy (IgAN) is the most common glomerulonephritis, especially in South East Asia [1,2]. In Singapore, it accounts for about 40% of the primary glomerulonephritis [2]; in China, IgAN accounts for about 40–50% of the primary glomerular disease [3,4]. With current therapies, only 5–30% of the IgAN patients could get a complete remission, and about 30–50% of the patients will develop an end-stage renal disease within 20 year of the disease initiation [1], which damages the glomeruli and brings in considerable physical and economic burden to the patients and the families.

The disease derives its name from the deposition of immunoglobulin A (IgA), especially IgA1 subclass, in the kidney, which distinguishes IgAN from other glomerulonephritis without IgA deposits (NIgAN). However, the exact pathogenesis of IgAN is still incompletely understood. Lai [5] and Suzuki et al. [6] reviewed the previous studies and

* Corresponding author. BGI-Shenzhen, Shenzhen 518083, China.
** Corresponding author.
E-mail addresses: vascular@fmmu.edu.cn (L. Wang), liuxiao@genomics.cn (X. Liu).
[1] These authors contribute equally.

https://doi.org/10.1016/j.jaut.2018.10.018
Received 3 August 2018; Received in revised form 20 October 2018; Accepted 22 October 2018
Available online 29 October 2018

discussed the pathogenesis of IgAN in details, and proposed the four-hits pathway to explain the pathogenesis of IgAN. The four hits include the production of galactose-deficient IgA1 (Gd-IgA1), the production of anti-Gd-IgA1 antibodies, the formation of IgA immune complex, and the activation of mesangial cells by the immune complex. It is widely accepted now that the IgAN is an immune related disease with the involvement of the whole immune system.

The diagnosis of IgAN relies on the histopathology of a renal biopsy [7], which shows IgA deposition on the glomerular mesangium using immunofluorescence and electron microscopy. However, this is an invasive procedure and could bring risks, such as infection and potential damage to the kidney. In addition, it is impractical to perform renal biopsy repeatedly during and after treatment to monitor the disease progress. Therefore, there is an urgent need for reliable noninvasive biomarkers to diagnose subclinical IgAN, to evaluate the disease stages and to assess the treatment efficacy [8,9]. In fact, some novel biomarkers have been developed in plasma or urine, such as, using serum Gd-IgA1 [10], serum Gd-IgA1-specific IgG [11], or using urine IgA–uromodulin complex [12]. Other biomarkers include various interleukins and urinary podocalyxin [8]. Although these biomarkers show some potential, there are still considerable false positives and false negatives with them.

Immune repertoire high throughput sequencing (HTS) has become a powerful tool to quantitatively analyze the distribution of large number of T/B cells in a biological sample [13,14]. Immune repertoire HTS has been used in the basic scientific research and translational research of all kinds of immune related disease, such as infection diseases [15], autoimmune diseases [16,17], tumor [18] and allergy [19]. Disease specific T/B cells have also been used as non-invasive biomarkers in disease diagnosis and prognosis, such as tumor immunotherapy [20], leukemia minimal residual disease monitoring [21,22], and status of cytomegalovirus infection [23]. Investigating the diversity of T/B lymphocytes by sequencing the repertoire of T cell receptors (TCR) and B cell receptors (BCR), and identifying the phenotype/antigen associated T/B cells in IgAN is critical to understand the disease pathogenesis. In addition, the knowledge can facilitate the development of novel diagnostic methods and therapies for the IgAN.

In this study, we analyzed the T cell receptor beta chain (TCRB) and immunoglobulin heavy chain (IGH) repertoire of kidney infiltrating and circulating lymphocytes of IgAN patients and NIgAN using immune repertoire HTS. We compared multiple aspects of antigen receptor characteristics, including the diversity and CDR3 length of the repertoire, the IGH isotype distribution and their somatic hypermutation rate, among IgAN patients, NIgAN patients and healthy controls. We also explored the possibilities to distinguish patients with IgAN from NIgAN and healthy controls based on the disease associated clones in peripheral blood mononuclear cell (PBMC).

2. Methods

2.1. Samples

This study was approved by the medical ethics committee of Xijing hospital and all specimen donors signed the written informed consent. We recruited fifty-two patients with nephropathy, including thirty-nine IgAN patients, and thirteen NIgAN patients as a comparative group. Clinical characteristics of these patients are showed in Table S1. The NIgAN disease category included membranous nephropathy, capillary disease, IgM nephropathy, glomerulosclerosis and hyperplasia glomerulonephritis. The 24-hour urine was collected in order to estimate the amount of protein excretion, and the renal function was evaluated by Urea Nitrogen, Creatinine and eGFR level. Lee's pathological classification was used to assess the state of the IgAN patients. The kidney tissue biopsies and peripheral blood were collected during the disease diagnosis. We also recruited 60 healthy volunteers without history of cancer, autoimmune disorder and surgery, and 2 ml blood were drawn from each healthy volunteer. All renal tissues were frozen in liquid nitrogen immediately and were stored in −80 °C until usage. For peripheral blood samples, the PBMC were isolated immediately after the blood drawing using Ficoll-Paque and were frozen in liquid nitrogen and then stored in −80 °C.

2.2. High throughput sequencing of IGH and TCRB repertoire

Total RNA was extracted from the kidney tissue and PBMC using TRIzol™ Reagent (Invitrogen, 15596026), following the manufacturer's guidelines. We used Agilent RNA 6000 Pico Kit (Agilent, 5067-1513) to assess the quality and quantity of the total RNA. RNA samples with RIN > 6.5 were used for sequencing analysis. RNA of all nephropathy patients and randomly selected forty-two healthy controls were processed for *IGH* repertoire HTS. Seventeen IgAN patients with enough RNA left were processed for *TCRB* repertoire HTS, together with eighteen healthy individuals used for comparison. For all samples, 200 ng of total RNA was used for cDNA synthesis with oligo(dT) primer using SuperScript™ II Reverse Transcriptase (Invitrogen, 18064014). In order to enrich the completely arranged IGH sequences, the first-strand cDNA was used as templates for a multiplex PCR with forward primer mix annealed to all functional variable genes and reverse primer mix annealed to junction region [22] or reverse primer mix annealed to constant region Cα (IgA), Cμ (IgM), Cγ (IgG), Cδ (IgD) and Cε (IgE) [24]. In order to amplify the completely arranged *TCRB* fragments, the first-strand cDNA was used as templates for a multiplex PCR with forward primer mix specific for all functional V genes and reverse primer mix specific to all functional junction genes. The PCR reaction conditions were the same in our previous paper [22,25]. The target amplified production (150-300bp for *IGH* and 100-200bp for *TCRB*) was purified by 2% agarose gel electrophoresis. The Illumina Hiseq sequence adaptors were ligated to construct sequencing libraries, which then were sequenced on Illumina Hiseq2500 platform.

2.3. Bioinformatics analysis of the immune repertoire HTS data

Sequencing data were analyzed using the pipeline *IMonitor* [26]. For comparing the diversity of different samples, we randomly chose a subset of 2.5 million mapped reads for all the continuous analysis. We used the CDR3s with correct open reading frame throughout this paper unless stated otherwise. If two samples were sequenced in the same sequencing lane, cross contamination for low-frequency sequences was possible due to errors in index synthesis and sequencing. These contaminated sequences are problematic, especially for the disease-associated-clones analysis. To ensure the data quality, we added two steps to filter the potentially contaminated sequences before further analysis. First, if two sequences have the same CDR3 nucleotides sequences and above 98% similarity for the non-CDR3 part, and meanwhile the abundance ratio of this sequence between the two samples was larger than 2000:1, the sequence was removed from the sample with lower abundance. Secondly, we counted the number of samples contain a certain CDR3 in a sequencing lane, and then calculated the percentage of this number in the total samples number sequenced in that lane. If this percentage in one sequencing lane is twice or more than the average percentage of other sequencing lanes, the CDR3s was removed from the sample with the lowest frequency until it is nearly equal to the average percentage of other lanes.

2.4. Definition of antigen-experienced IGH clones

In this study, a clone or clonotype was defined as a unique *TCRB* or *IGH* CDR3 nucleotide sequence. We defined antigen-experienced IGH clones as those having IgA or IgG with high mutation rate (above 3%) as reported previously [27]. The SHM rate of each IGH clone was calculated by dividing the number of total mapped nucleotides by the number of mismatching in the FR3 and FR4 region.

2.5. Classification of the IgAN patients according to the disease associated clones

2.5.1. Disease-associated-clones definition

Disease associated clones were defined as those *TCRB* or *IGH* presenting in at least four renal tissues of IgAN patients and less than three healthy blood samples, or the 10 most abundant clones in renal tissue of each patient and presenting in less than three healthy blood samples. For TCRB, the frequency of disease associated clones is restricted to be higher than 0.001% in renal tissue.

2.5.2. Training classification model and leave-one-out cross validation

We used the logistic regression model in R package based on two features, the proportion of unique disease associated clones presented in the sample (the number of disease associated clones divided by the total number of clones in the sample) and the proportion of total disease associated clones presented in the sample (the sum frequency of disease associated clones in the sample). Exhausted leave-one-out cross validation was used to assess the identifier's performance during model training. Concretely, given there are N IgAN patients, N-1 patients were used as training data and processed the above classification model. The left one patient was used as testing data to perform the classification. The cross validations were repeated for 5N times until every sample was used as testing data for 5 times.

2.6. Statistical analysis

For comparison of the two groups, the Mann-Whitney *U* test was used unless otherwise noted. The correlation coefficient was calculated using the cor.test in R tools. In box and whisker plots, the box extends from the 25th to 75th percentiles, the line in the middle represents the median and outliers represent the minimum and maximum value.

3. Result

3.1. Clinical and pathological characteristics of the patients

The clinicopathological data of the patients is summarized in Table 1. For IgAN patients, 30 of them are male, and 9 are female. As to NIgAN patients, 8 are male and 5 are female (Table 1). Most of the IgAN patients were diagnosed as grade 2, 3 or 4 according to the Lee's classification. The average 24-hours proteinuria was 2169 mg for patients with IgAN, and varied largely among patients, with 5 patients had a urine protein above 5 g/day. The urea nitrogen, creatinine and estimated glomerular filtration rate (eGFR) also varied extensively among IgAN patients, with an average level of 6.17 mmol/L, 111.51 μmol/l and 81.56 ml/min/1.73 m^2 respectively (Table S1). We did not observe any correlations between those renal-function-related clinical indices with the Lee' classification for IgAN patients (Table 1). When we compared IgAN with NIgAN, no significant difference was observed for those clinical indices (Fig. S1), which reflected the similar extent of renal damage between patients with IgAN and NIgAN.

3.2. Increased IgA1 frequency and decreased IgG frequency in PBMC for IgAN patients

We compared the distribution of IgA, IgG and IgM isotypes between patients and healthy controls in PBMC. For IgA, we subdivided it into two different IgA transcripts, IgA1 and IgA2. Because we just sequenced the first 40bp of the constant region for IgG, we cannot subdivide it into IgG subtypes (IgG1, IgG2, IgG3 and IgG4). The frequency of IgA1 in PBMC was significantly higher in IgAN patients than in NIgAN patients and healthy control (Fig. 1A). In contrast to IgA1, the total IgG frequency in PBMC was significantly lower in IgAN patients than that in NIgAN patients and healthy control (Fig. 1C). For IgA2 and IgM, no significant difference was observed in their frequencies between patients and healthy controls (Fig. 1B, D). Then we compared the distribution of the above four isotypes in renal tissue between IgAN patients and NIgAN patients. The frequency of IgA1 in renal tissue in most IgAN patients was higher than in NIgAN patients, but was not statistically significant (Fig. 1A). There is no significant difference for the IgA2, IgG, and IgM frequencies in renal tissues between IgAN patients and NIgAN patients (Fig. 1B, C, D). Importantly, the IgA1 and IgG frequencies in the renal tissue were significantly higher than that in PBMC for both IgAN and NIgAN patients (Fig. 1A, C).

3.3. More antigen-experienced B cells in the renal tissue of patients

Class switching and somatic hypermutation (SHM) are two biological mechanisms through which B cells adapt to pathogens. During B cell activation and responding to antigens, the naïve IgM B cells switch to IgA, IgG or IgE B cells, and accumulate SHMs. SHM rate in FR3 and FR4 of IgA did not show significant difference between any two groups (Fig. 1E and F). IgG in kidney tissue had higher SHM rate in FR3 and FR4 than that in PBMC in IgAN patients, but not in NIgAN patients (Fig. 1G). IgM from renal tissue of patients displayed a higher average SHM rate in FR3 and FR4 than that from PBMC (Fig. 1H). When comparing the SHM rate between the PBMC of patients and healthy controls, no significant difference was observed for any isotype. We then investigated the antigen-experienced clone frequency (defined in methods) in renal tissue and PBMC of patients, as well as in healthy controls. The result demonstrated that the frequency of antigen-experienced IGH clones was significantly higher for B cells infiltrating in the renal tissue than in PBMC (Fig. 1I). We also found that the antigen-experienced IGH clones proportion in the PBMC of both IgAN and NIgAN was significantly lower than that in the PBMC of healthy control. We did not observe obvious difference for the frequency of antigen-experienced IGH clones between IgAN patients and NIgAN patients in both the renal tissue and PBMC (Fig. 1I).

3.4. IGH and TCRB in the renal tissue of the IgAN patients showed decreased repertoire diversity compared with that in PBMC

VDJ recombination and other factors introduce an astronomical diversity in the CDR3 sequences. Therefore, the CDR3 sequence could

Table 1
Summary of clinical and laboratory information of the recruited patients.

Label	Lee's C	Sex (M vs F)	Age	Proteinuria (mg/24 h)	S. Cre. (μmol/l)	S. Urea N. (mmol/l)	eGFR (ml/min/1.73 m^2)
			Average (range)				
IgAN	1	2 vs 1	23.7 (15–29)	5824.7 (3364–7810)	110.7 (58–152)	8.9 (4.0–12.9)	82.3 (58–119)
IgAN	2	10 vs 2	29.8 (15–68)	1849.4 (185–8250)	101.7 (76–168)	5.3 (3.3–10.4)	89.9 (60.4–130.2)
IgAN	3	8 vs 4	29.8 (9–52)	2016.5 (700–4030)	108.2 (54–212)	6.6 (3.0–10.7)	81.1 (37.3–145.6)
IgAN	4	8 vs 2	35.2 (21–51)	2124.9 (157–6570)	126.3 (64–295)	6.0 (3.39–8.7)	73.0 (22.7–132)
IgAN	5	2 vs 0	35.5 (33–38)	1568.5 (97–3040)	112 (97–127)	5.2 (4.2–6.3)	80.1 (63.5–96.7)
NIgAN	NA	8 vs 5	36.8 (14–57)	1698 (95–4510)	90.0 (62–132)	5.1 (3.3–7.3)	86.6 (59.6–127.3)

Lee's C, Lee's classification; 24 h Prot., S. Cre., Serum Creatinine; S. Urea N., Serum Urea Nitrogen.

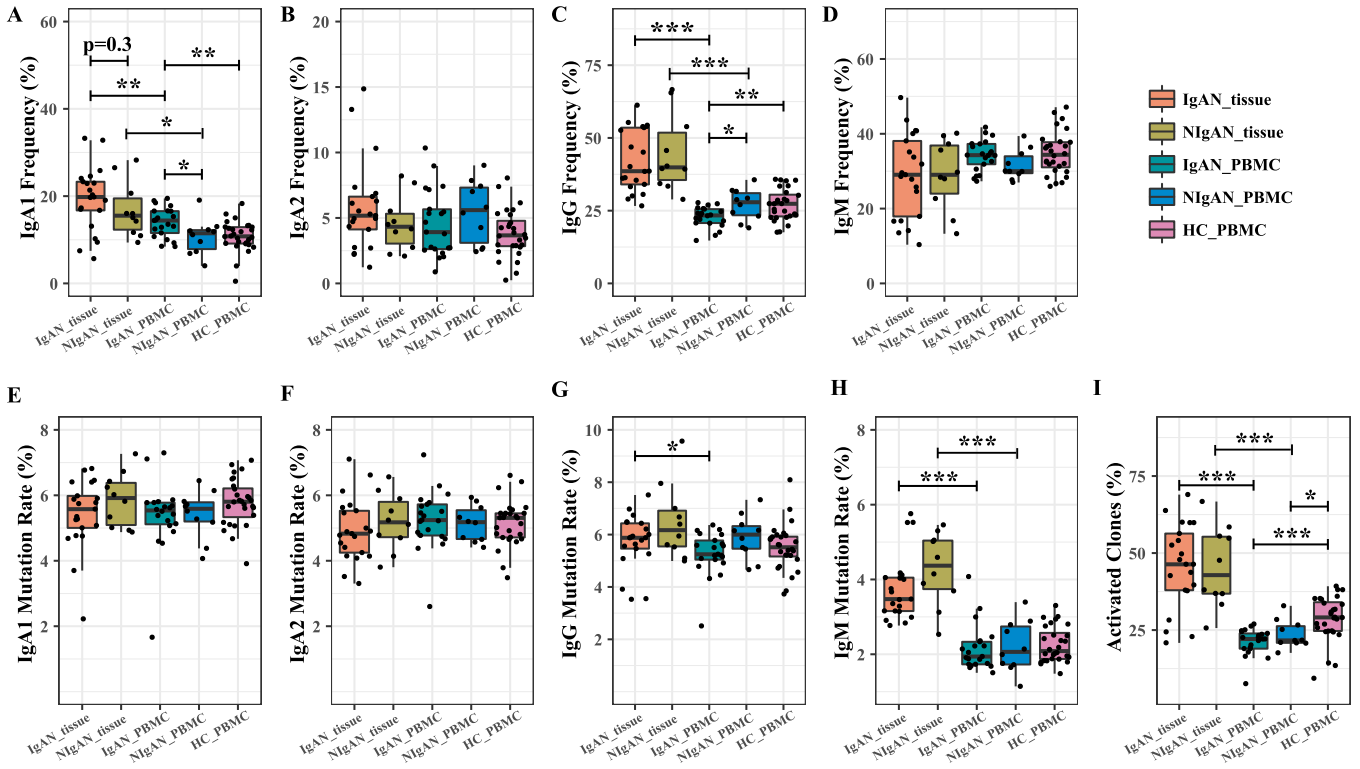

Fig. 1. The isotype distribution and somatic hypermutation of each IGH isotype in renal tissue and PBMC of patients and healthy controls. IgA1 (A), IgA2 (B), IgG (C) and IgM (D) frequency distribution in renal tissue and PBMC of patients and healthy controls. IgA1 (E), IgA2 (F), IgG (G) and IgM (H) somatic hypermutation rate in renal tissue and PBMC of patients and healthy controls. The antigen-experienced IGH clone frequency (I) in renal tissue and PBMC of patients and healthy controls. HC, healthy controls. The P value of IgAN_tissue vs IgAN_PBMC, NIgAN_tissue vs NIgAN_PBMC, IgAN_tissue vs NIgAN_tissue, IgAN_PBMC vs NIgAN_PBMC, IgAN_PBMC vs HC_PBMC, and NIgAN_PBMC vs HC_PBMC was calculated. *, ** and *** indicate P values less than 0.05, 0.01 and 0.001 respectively.

serve as a natural barcode for a particular *IGH* or *TCRB* clone. We measured the *IGH* and *TCRB* diversity by two indices, the Shannon index and the Gini coefficient of CDR3. We firstly compared the *TCRB* and *IGH* diversity between renal tissue and PBMC of the same IgAN patients. Both the *IGH* and *TCRB* in renal tissue of IgAN patients were less diverse than that in PBMC of IgAN patients. Concretely, the Shannon index of CDR3s (Fig. 2A, C) in renal tissue was significantly lower than that in PBMC; the Gini coefficient of CDR3s, which is mostly reflected by the global imbalances of CDR3 frequencies and barely affected by the number of cells in samples, was significantly higher in renal tissue than that in PBMC (Fig. 2B, D), which indicated antigenic stimulation and clonal expansion of both B and T cells in the kidney. When comparing the *TCRB* and *IGH* diversity in PBMC between IgAN patients and healthy controls, no significant difference was observed (Fig. 2). We also performed *IGH* repertoire analysis for the NIgAN patients; however, no significant difference for *IGH* diversity was observed between IgAN patients and NIgAN patients in either renal tissue or PBMC (Fig. 2B, D).

3.5. TCRB and IGH reveal shorter CDR3 length in IgAN patients

We observed a significant reduction in both *TCRB* and *IGH* CDR3s length in PBMC of IgAN patients comparing with healthy controls (Fig. 3A and B, right panel). The shorter *TCRB* and *IGH* CDR3s could either be a feature of the initial VDJ rearrangement, or the result of negative and positive selection during T/B cell development/maturation and antigenic selection in periphery, or the combination of both. To investigate the origination of the shorter CDR3s in patients, we analyzed the CDR3 length of the out of frame *TCRB* and *IGH* sequences, which did not go through the negative and positive selection, as described by previous studies [17,28,29]. The out-of-frame *TCRB* CDR3s in IgAN patients and heathy control had similar length, (Fig. 3A, left panel), indicating the shorter *TCRB* CDR3s in IgAN patients is mainly the result of the positive and negative selection during T cell development in the thymus. For *IGH,* the pre-selected out-of-frame CDR3s in IgAN patients were also shorter than that in healthy controls (Fig. 3B, left panel), however, the in-frame functional clones showed more marked distinction between IgAN patients and healthy controls (Fig. 3B, right panel).

3.6. Disease associated clones

We speculated that a set of pathogenic or disease associated T and B cells existing in the renal tissue, and we are particularly interested in their appearance in the blood and the potentiality to serve as non-invasive biomarkers for IgAN diagnosis. First, we evaluated the repertoire overlap between the renal tissue and the blood. The overlap rate was calculated by adding together the total number of CDR3s identified in the both repertoires, and then dividing it by the sum of CDR3 number of the two repertoires. The overlap rate between the renal tissue and PBMC repertoire of the same patient was significantly higher than the inter-individual rate for both *TCRB* and *IGH* (Fig. S2A and C). Additionally, an average of 34.13% of *TCRB* CDR3s and 30.21% of *IGH* CDR3 from the renal tissue were found in the PBMC of the same patient (Fig. S2B and D). In several samples, more than 60% of *TCRB* and *IGH* CDR3 in renal tissue could be detected in PBMC. The frequent appearance of renal tissue lymphocytes in PBMC suggested the diagnostic potential of the blood. Considering the antigenic complexity and the background noise in the blood, we intended to search the disease associated CDR3 clonotypes in the renal tissue, and used those in the

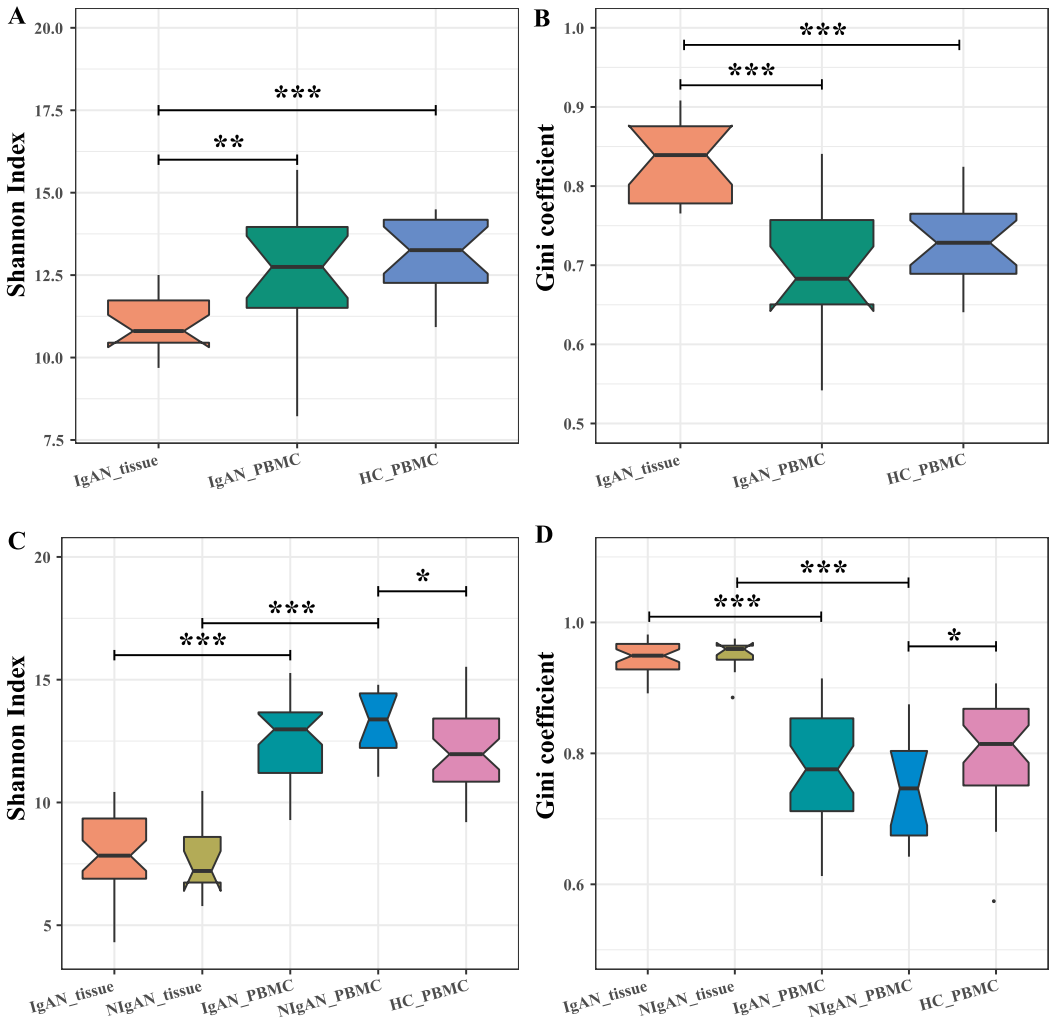

Fig. 2. The *TCRB* (A–B) and *IGH* (C–D) CDR3 diversity in renal tissue and PBMC of patients and healthy controls. The diversity was measured by the Shannon index (A, C) and Gini coefficient (B, D). HC, healthy controls. The P value of IgAN_tissue vs IgAN_PBMC, NIgAN_tissue vs NIgAN_PBMC, IgAN_tissue vs NIgAN_tissue, IgAN_PBMC vs NIgAN_PBMC, IgAN_PBMC vs HC_PBMC, and NIgAN_PBMC vs HC_PBMC was calculated. *, ** and *** indicate P values less than 0.05, 0.01 and 0.001 respectively.

blood to do the diagnosis.

Disease associated clones are usually widely and abundantly represented in the pathogenic tissues but rarely appear in the healthy tissues [23]. Accordingly, IgAN associated T and B cell clonotypes were selected as described in the methods. We identified 279 *TCRB* and 471 *IGH* clones suggested to be associated with IgAN (Tables S2 and S3). Similar to previous reports about the CDR3 length of autoimmune T and B cells [17,30], we also identified that the CDR3s of IgAN associated *TCRB* and *IGH* clones were significant shorter than the other CDR3s (Fig. 3C and D). Interestingly, we found that the unique frequencies (the number of unique disease associated clones divided by the number of total unique clones) of disease associated *TCRB* and *IGH* clones were positively correlated, in both the renal tissue ($r = 0.91$; $p = 2.95 \times 10^{-7}$) and the blood ($r = 0.67$; $p = 0.003$) (Fig. S3A and B), which may due to the specificity of *TCRB* and *IGH* clones when recognizing the IgAN associated antigens.

3.7. Noninvasive classification of IgAN patients with the disease associated CDR3s

Next, we intended to use the identified IgAN associated clones to do the noninvasive classification of IgAN using the PBMC data. We constructed a binary classifier to distinguish the IgAN patients from healthy controls according to the unique and total frequency of disease associated *TCRB* or *IGH* clones as features (Fig. 4A). As a supporting evidence, we found that both the unique and total disease associated *TCRB* clones frequencies in PBMC were significantly higher in IgAN patients than in healthy controls (Fig. 4B and C). In the end, the disease associated *TCRB* clones separated the PBMC of IgAN patients from healthy control with a sensitivity of 94.1%, a specificity of 94.1% and an accuracy of 94.1%, (AUC (the area under the ROC curve) = 96.1%) (Fig. 4F), supporting that those disease associated *TCRB* clones could be used as biomarkers for IgAN. There were also obvious differences for the unique frequency and total frequency of disease associated *IGH* clones in PBMC between IgAN patients and healthy controls, and between IgAN patients and NIgAN patients (Fig. 4D and E). With the disease associated *IGH* clones, the classifier displayed a less satisfying performance with a sensitivity of 61.5%, a specificity of 76.9% and an accuracy of 69.2% (AUC = 73.8$) (Fig. 4G). We also tried to see if those disease associated *IGH* clones can be used to distinguish the IgAN patients from NIgAN patients, and the result showed the accuracy of discrimination was 78.5% (Fig. 4H), which indicated that those IgAN associated *IGH* clones were more prone to exist in patients with IgAN than NIgAN. We tried to investigate if the unique frequency and total

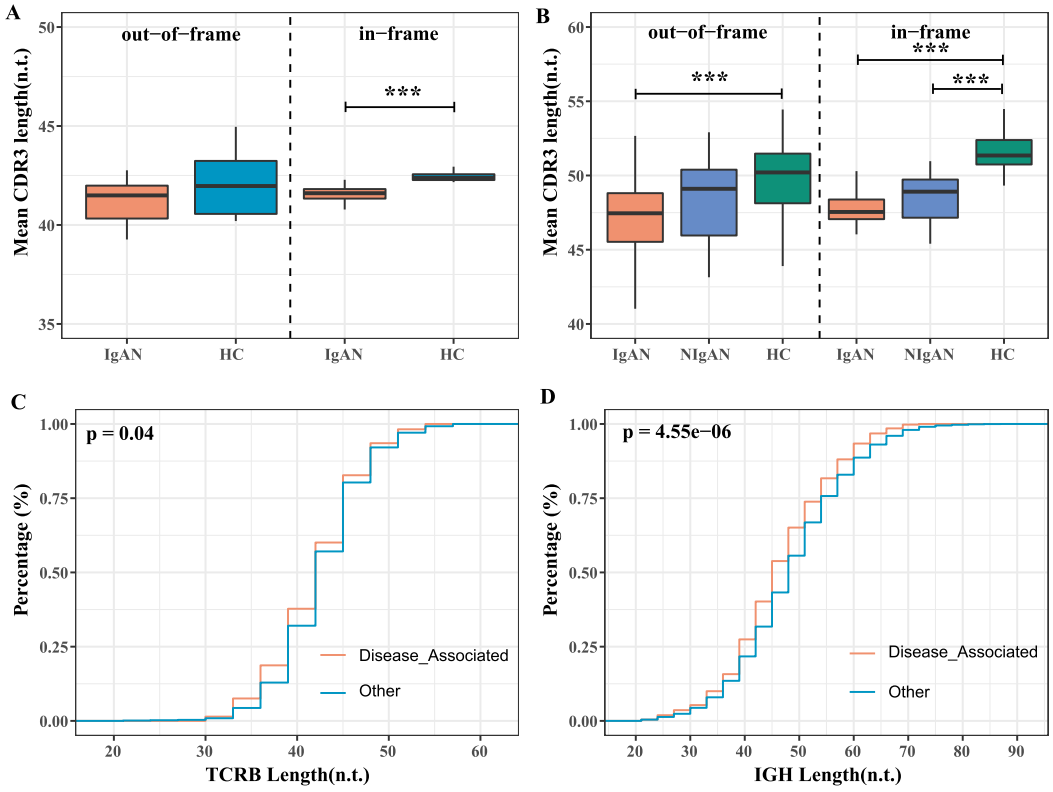

Fig. 3. *TCRB* and *IGH* CDR3s length distribution. (A) The mean CDR3 length of out-of-frame (left panel) and in-frame (right panel) *TCRB* clones in the PBMC of patients and healthy controls. (B) The mean CDR3 length of out-of-frame (left panel) and in-frame (right panel) *IGH* clones in the PBMC of patients and healthy controls. (C) Cumulative CDR3 length plots of the disease associated *TCRB* and other *TCRB* in the renal tissue of IgAN patients. (D) Cumulative CDR3 length plots of the disease associated *IGH* and other *IGH* in the renal tissue of IgAN patients. HC, healthy control. n.t., nucleotide. *, ** and *** indicate P values less than 0.05, 0.01 and 0.001 respectively.

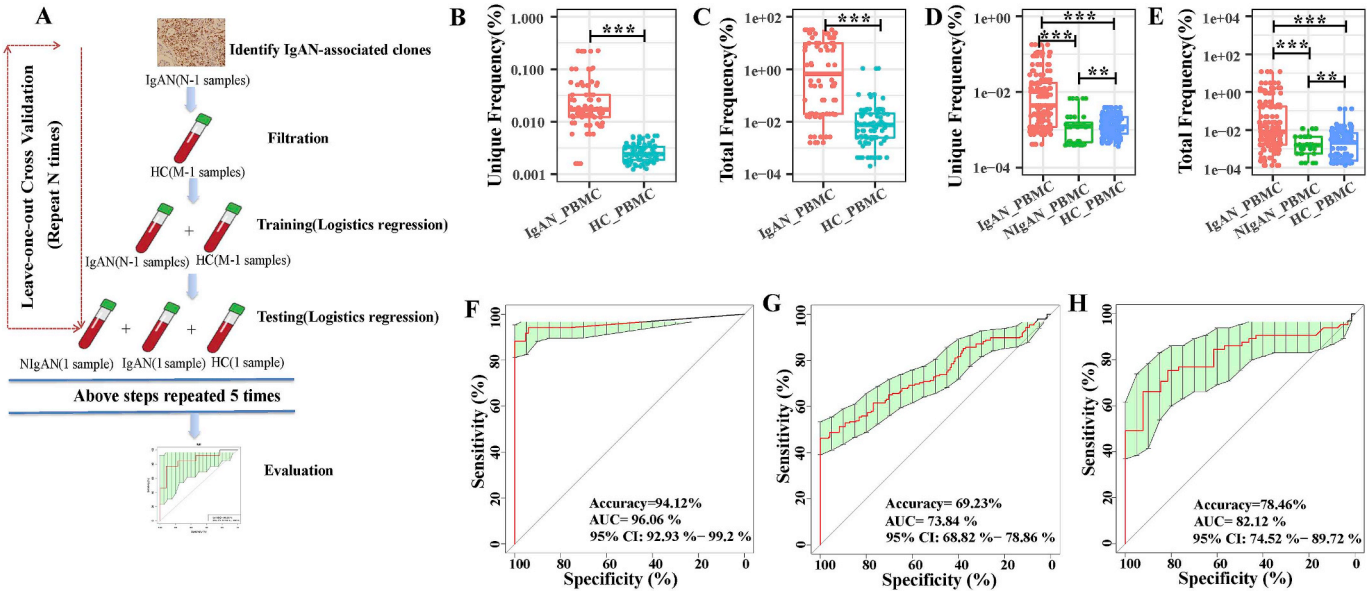

Fig. 4. Classifying different groups according to *TCRB* and *IGH* repertoire data. (A) Analytical overview, which was described in detail in the methods. (B–E) The burden of disease associated clones in PBMC of different groups. The unique frequency (B) and total frequency (C) of disease associated *TCRB* clones in PBMC of IgAN patients and healthy controls. The unique frequency (D) and total frequency (E) of disease associated *IGH* clones in PBMC of IgAN patients, NIgAN patients and healthy controls. (F–H) The leave-out-one cross validation ROC curves of classification according to the disease associated clones. The ROC curve of classification of IgAN patients and healthy controls according to disease associated *TCRB* clones (F) and disease associated *IGH* clones (G). The ROC curve of classification of patients with IgAN and NIgAN according to disease associated *IGH* clones (H). HC, healthy controls. *, ** and *** indicate P values less than 0.05, 0.01 and 0.001 respectively.

frequency of those disease associated clones is related to the clinical indices, including 24-hours proteinuria, serum creatinine, urea nitrogen, eGFR and Lee's grade, however no obvious correlation was observed (data not shown). Overall, we demonstrated that IgAN could be well classified non-invasively by *TCRB* and *IGH* repertoire data.

4. Discussion

In this study, we performed *TCRB* and *IGH* repertoire HTS of paired renal tissue and PBMC of nephropathy patients, as well as of PBMC of healthy controls, to investigate the pathogenesis of IgAN and develop new biomarkers for the disease. Comparing with the lymphocytes in PBMC, renal tissue infiltrating lymphocytes was less diverse (Fig. 2), which resulted from both the clonotypes reduction and their imbalanced frequencies due to antigen stimulation in renal tissue compared with PBMC. The higher frequency of IgA and IgG, and the increased IgM SHM rate in FR3 and FR4 comparing with PBMC (Fig. 1) also supports the high accumulation of antigen-experienced lymphocytes in the tissue infiltrating lymphocytes.

A recent study indicated aberration in *TCR* VDJ recombination conferring risks to autoimmune diseases such as type 1 diabetes, in which the diabetes autoantigen specific T cells were found to have shorter *TCRB* CDR3 than viral specific T cells and global T cells [17]. We also observed that the average lengths of both *TCRB* and *IGH* CDR3 in IgAN patients were shorter than those in healthy controls, from the PBMC data (Fig. 3). By separately investigating the out-of-frame and the in-frame CDR3s, we conclude that VDJ recombination and selection during T/B cell development and in periphery play different roles in *TCRB* and *IGH*. The shorter *IGH* CDR3 in patients is the combined result of increased production of clones with shorter CDR3 during VDJ rearrangement and the clone selection during lymphocytes maturation in periphery. Additionally, the IgAN associated *TCRB* and *IGH* clonotypes also showed significantly shorter CDR3 length. Our findings provide further evidence that the generation of a self-reactive prone T and B cell repertoire accounts for the susceptibility of common autoimmune diseases.

Numerous studies have demonstrated that the IgA deposited in the glomeruli of IgAN patients is the subclass IgA1 [5,31,32], which is a functionally and structurally distinct immunoglobulin from IgA2. A number of literatures have reported a higher serum IgA level in IgAN [33,34], and the association of the serum Gd-IgA1 level with disease progression [35]. The possible reasons of the increased serum IgA1 including the increased percentage of IgA1 producing B cells, the increased transcription of IgA1 mRNA and the declined degradation of IgA1 antibody. Several studies have reported that the percentage of IgA1 producing B cell was higher in blood of patients with IgAN than health controls [31,36]. In this study, we found that the IgA1 mRNA frequency in PBMC of IgAN patients was significantly higher than healthy controls and also NIgAN patients (Fig. 1), which may due to more IgA1 producing B cells or higher transcription activity of those B cells.

Because IgAN is an immune related disease, there should be abundant disease associated T/B cells targeting the disease related autoantigens infiltrating in the injured tissues, and our data (Fig. S2B and D) and published paper [31] suggested those disease associated T/B cells derived from the bone marrow and circulate in the blood. In this paper, we identified a set of disease associated *TCRB* and *IGH* clones from the renal tissues which then was used to distinguish the PBMC of IgAN patients from healthy individuals. With the disease associated *IGH* clones panel, the leave-out-one cross validation showed an accuracy of 69.2% when classifying IgAN patients and healthy individuals (Fig. 4G); more importantly, this IgAN associated *IGH* clones panel can also distinguish patients with IgAN from NIgAN with an accuracy of 78.5% (Fig. 4H). What is striking is that the disease associated *TCRB* clones panel can predict the IgAN status from healthy controls accurately with an accuracy of 94.1%, a sensitivity 94.1% and a specificity 94.1% (Fig. 4F), which is more accurate than the biomarkers based on serum Gd-IgA1 (76.5% sensitivity with 94% specificity) [10], serum Gd-IgA1-specific IgG (89% sensitivity and 92% specificity) [11] and urine IgA–uromodulin complex (Sensitivity 81.7% and specificity 73.4%) [12]. However, due to the limited RNA samples, we did not sequence the TCRB repertoire of the NIgAN patients; therefore, we could not assess the performance of the disease associated *TCRB* CDR3 in distinguishing IgAN from NIgAN. In fact, some studies have reported the involvement of T cell in the pathogenesis of IgAN, such as increased circulating αβ and γδ T cells in PBMC of IgAN patients [37,38], involvement of T cell in the progression of IgAN [39], and the existence of conserved TCRB CDR3 in the IgAN biopsies [40]. Overall, our study further demonstrates that disease associated *TCRB* and *IGH* clones could work as the potential biomarker to help the diagnosis of IgAN, at least in screening the IgAN. However, due to the limited number of samples in our study, validation of the performance of those biomarkers and demonstrating the accuracy and reproducibility of this method by more studies from other laboratories is needed.

5. Conclusion

In summary, we present here a comprehensive landscape of T and B cell repertoire in IgAN, including the restricted repertoire size and increased frequency of antigen-experienced T/B cells in kidney infiltrating lymphocytes, the increment of IgA1 frequency, and shorter CDR3 length for both *TCRB* and *IGH* of IgAN patients, which may be related to the autoimmune nature of those lymphocytes. The most important discovery is that using our defined disease associated *TCRB* and *IGH* clones, we can distinguish IgAN patients from healthy controls using PBMC. These findings demonstrate the potential of disease related *TCRB* and *IGH* clonotypes as alternative biomarkers for the screening and the diagnosis of IgAN.

Disclosure

All the authors declare no competing financial interests.

Author contributions

C.H., S.S., P.Z., G.X., H.W., H.L., and Y.L. provided patients samples and clinical information; L.L., X.W. and H.L performed the experiments; X.L. and X.W. performed the statistical analysis; X. L., J.W., W.Z., D.C., L.Z. and W.L. analyzed and interpreted the data; X.L. and L.W. designed and supervised the study; J.W. wrote most of the content of the manuscript, X. L. did an extensive revision and all authors participated in the writing and approval of the manuscript.

Acknowledgement

We would like to thank Lennart Hammarström and Liyong Wang for providing language help. This work was supported by the National Natural Science Foundation of China (No. 81672593 and No. 81670655), the Major Nature Science Foundation of Shaanxi Province (No. 2017ZDXM-SF-045), Beijing Natural Science Foundation (No. 7152127) and Shenzhen Municipal Government of China (JCYJ20170817145404433 and JCYJ20170817145428361).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jaut.2018.10.018.

References

[1] F.P. Schena, A retrospective analysis of the natural history of primary IgA nephropathy worldwide, Am. J. Med. 89 (1990) 209–215 http://www.ncbi.nlm.nih.

gov/pubmed/2200265.
[2] K.T. Woo, C.M. Chan, Y.M. Chin, H.L. Choong, H.K. Tan, M. Foo, V. Anantharaman, G.S.L. Lee, G.S.C. Chiang, P.H. Tan, C.H. Lim, C.C. Tan, E. Lee, H.B. Tan, S. Fook-Chong, Y.K. Lau, K.S. Wong, Global evolutionary trend of the prevalence of primary glomerulonephritis over the past three decades, Nephron Clin. Pract. 116 (2010), https://doi.org/10.1159/000319594.
[3] F. De Zhou, M.H. Zhao, W.Z. Zou, G. Liu, H. Wang, The changing spectrum of primary glomerular diseases within 15 years: a survey of 3331 patients in a single Chinese centre, Nephrol. Dial. Transplant. 24 (2009) 870–876, https://doi.org/10.1093/ndt/gfn554.
[4] Jingyuan Xie, N. Chen, Primary glomerulonephritis in mainland China: an overview, New Insights into Glomerulonephritis, 2012, pp. 1–11, , https://doi.org/10.1159/000348642.
[5] K.N. Lai, Pathogenesis of IgA nephropathy, Nat. Rev. Nephrol. 8 (2012) 275–283, https://doi.org/10.1038/nrneph.2012.58.
[6] H. Suzuki, K. Kiryluk, J. Novak, Z. Moldoveanu, A.B. Herr, M.B. Renfrow, R.J. Wyatt, F. Scolari, J. Mestecky, A.G. Gharavi, B.A. Julian, The pathophysiology of IgA nephropathy, J. Am. Soc. Nephrol. 22 (2011) 1795–1803, https://doi.org/10.1681/ASN.2011050464.
[7] H.-H. Yu, B.-L. Chiang, Diagnosis and classification of IgA nephropathy, Autoimmun. Rev. 13 (2014) 556–559, https://doi.org/10.1016/j.autrev.2014.01.030.
[8] R.N. Moresco, M.M. Speeckaert, J.R. Delanghe, Diagnosis and monitoring of IgA nephropathy: the role of biomarkers as an alternative to renal biopsy, Autoimmun. Rev. 14 (2015) 847–853, https://doi.org/10.1016/j.autrev.2015.05.009.
[9] Y. Suzuki, H. Suzuki, Y. Makita, A. Takahata, K. Takahashi, M. Muto, Y. Sasaki, A. Kelimu, K. Matsuzaki, H. Yanagawa, K. Okazaki, Y. Tomino, Diagnosis and activity assessment of immunoglobulin A nephropathy: current perspectives on non-invasive testing with aberrantly glycosylated immunoglobulin A-related biomarkers, Int. J. Nephrol. Renovasc. Dis. 7 (2014) 409–414, https://doi.org/10.2147/IJNRD.S50513.
[10] Z. Moldoveanu, R.J. Wyatt, J.Y. Lee, M. Tomana, B.A. Julian, J. Mestecky, W.Q. Huang, S.R. Anreddy, S. Hall, M.C. Hastings, K.K. Lau, W.J. Cook, J. Novak, Patients with IgA nephropathy have increased serum galactose-deficient IgA1 levels, Kidney Int. 71 (2007) 1148–1154, https://doi.org/10.1038/sj.ki.5002185.
[11] H. Yanagawa, H. Suzuki, Y. Suzuki, K. Kiryluk, A.G. Gharavi, K. Matsuoka, Y. Makita, B.A. Julian, J. Novak, Y. Tomino, A panel of serum biomarkers differentiates IgA nephropathy from other renal diseases, PloS One 9 (2014), https://doi.org/10.1371/journal.pone.0098081.
[12] T. Obara, S. Mizoguchi, Y. Shimozuru, T. Sato, O. Hotta, The complex of immunoglobulin A and uromodulin as a diagnostic marker for immunoglobulin A nephropathy, Clin. Exp. Nephrol. 16 (2012) 713–721, https://doi.org/10.1007/s10157-012-0617-3.
[13] X. Liu, J. Wu, History, applications, and challenges of immune repertoire research, Cell Biol. Toxicol. (2018) 1–17, https://doi.org/10.1007/s10565-018-9426-0.
[14] G. Georgiou, G.C. Ippolito, J. Beausang, C.E. Busse, H. Wardemann, S.R. Quake, The promise and challenge of high-throughput sequencing of the antibody repertoire, Nat. Biotechnol. 32 (2014) 158–168, https://doi.org/10.1038/nbt.2782.
[15] D. Hou, C. Chen, E.J. Seely, S. Chen, Y. Song, High-throughput sequencing-based immune repertoire study during infectious disease, Front. Immunol. 7 (2016), https://doi.org/10.3389/fimmu.2016.00336.
[16] C.M. Tipton, C.F. Fucile, J. Darce, A. Chida, T. Ichikawa, I. Gregoretti, S. Schieferl, J. Hom, S. Jenks, R.J. Feldman, R. Mehr, C. Wei, F.E.H. Lee, W.C. Cheung, A.F. Rosenberg, I. Sanz, Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus, Nat. Immunol. 16 (2015) 755–765, https://doi.org/10.1038/ni.3175.
[17] I. Gomez-Tourino, Y. Kamra, R. Baptista, A. Lorenc, M. Peakman, T cell receptor β-chains display abnormal shortening and repertoire sharing in type 1 diabetes, Nat. Commun. 8 (2017), https://doi.org/10.1038/s41467-017-01925-2.
[18] T. Wang, C. Wang, J. Wu, C. He, W. Zhang, J. Liu, R. Zhang, Y. Lv, Y. Li, X. Zeng, H. Cao, X. Zhang, X. Xu, C. Huang, L. Wang, X. Liu, The different T-cell receptor repertoires in breast cancer tumors, draining lymph nodes, and adjacent tissues, Cancer Immunol. Res. 5 (2017) 148–156, https://doi.org/10.1158/2326-6066.CIR-16-0107.
[19] T.J. Looney, J.Y. Lee, K.M. Roskin, R.A. Hoh, J. King, J. Glanville, Y. Liu, T.D. Pham, C.L. Dekker, M.M. Davis, S.D. Boyd, Human B-cell isotype switching origins of IgE, J. Allergy Clin. Immunol. 137 (2016) 579–586, https://doi.org/10.1016/j.jaci.2015.07.014.
[20] E. Cha, M. Klinger, Y. Hou, C. Cummings, A. Ribas, M. Faham, L. Fong, Improved survival with T cell clonotype stability after anti-CTLA-4 treatment in cancer patients, Sci. Transl. Med. 6 (2014), https://doi.org/10.1126/scitranslmed.3008211.
[21] D. Wu, A. Sherwood, J.R. Fromm, S.S. Winter, K.P. Dunsmore, M.L. Loh, H.A. Greisman, D.E. Sabath, B.L. Wood, H. Robins, High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia, Sci. Transl. Med. 4 (2012), https://doi.org/10.1126/scitranslmed.3003656 134ra63-134ra63.
[22] J. Wu, S. Jia, C. Wang, W. Zhang, S. Liu, X. Zeng, H. Mai, X. Yuan, Y. Du, X. Wang, X. Hong, X. Li, F. Wen, X. Xu, J. Pan, C. Li, X. Liu, Minimal residual disease detection and evolved IGH clones analysis in acute b lymphoblastic leukemia using IGH deep sequencing, Front. Immunol. 7 (2016), https://doi.org/10.3389/fimmu.2016.00403.
[23] R.O. Emerson, W.S. DeWitt, M. Vignali, J. Gravley, J.K. Hu, E.J. Osborne, C. Desmarais, M. Klinger, C.S. Carlson, J.A. Hansen, M. Rieder, H.S. Robins, Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire, Nat. Genet. 49 (2017) 659–665, https://doi.org/10.1038/ng.3822.
[24] W. Zhang, Q. Feng, C. Wang, X. Zeng, Y. Du, L. Lin, J. Wu, L. Fu, K. Yang, X. Xu, H. Xu, Y. Zhao, X. Li, U.H. Schoenauer, A. Stadlmayr, N.K. Saksena, H. Tilg, C. Datz, X. Liu, Characterization of the B cell receptor repertoire in the intestinal mucosa and of tumor-infiltrating lymphocytes in colorectal adenoma and carcinoma, J. Immunol. 198 (2017) 3719–3728, https://doi.org/10.4049/jimmunol.1602039.
[25] X. Liu, W. Zhang, X. Zeng, R. Zhang, Y. Du, X. Hong, H. Cao, Z. Su, C. Wang, J. Wu, C. Nie, X. Xu, K. Kristiansen, Systematic comparative evaluation of methods for investigating the TCRβ repertoire, PloS One 11 (2016), https://doi.org/10.1371/journal.pone.0152464.
[26] W. Zhang, Y. Du, Z. Su, C. Wang, X. Zeng, R. Zhang, X. Hong, C. Nie, J. Wu, H. Cao, X. Xu, X. Liu, Imonitor: a robust pipeline for TCR and BCR repertoire analysis, Genetics 201 (2015) 459–472, https://doi.org/10.1534/genetics.115.176735.
[27] U. Laserson, F. Vigneault, D. Gadala-Maria, G. Yaari, M. Uduman, J.A. Vander Heiden, W. Kelton, S. Taek Jung, Y. Liu, J. Laserson, R. Chari, J.-H. Lee, I. Bachelet, B. Hickey, E. Lieberman-Aiden, B. Hanczaruk, B.B. Simen, M. Egholm, D. Koller, G. Georgiou, S.H. Kleinstein, G.M. Church, High-resolution antibody dynamics of vaccine-induced immune responses, Proc. Natl. Acad. Sci. 111 (2014) 4928–4933, https://doi.org/10.1073/pnas.1323862111.
[28] I.V. Zvyagin, M.V. Pogorelyy, M.E. Ivanova, E.A. Komech, M. Shugay, D.A. Bolotin, A.A. Shelenkov, A.A. Kurnosov, D.B. Staroverov, D.M. Chudakov, Y.B. Lebedev, I.Z. Mamedov, Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing, Proc. Natl. Acad. Sci. 111 (2014) 5980–5985, https://doi.org/10.1073/pnas.1319389111.
[29] K.M. Roskin, N. Simchoni, Y. Liu, J.Y. Lee, K. Seo, R.A. Hoh, T. Pham, J.H. Park, D. Furman, C.L. Dekker, M.M. Davis, J.A. James, K.C. Nadeau, C. Cunningham-Rundles, S.D. Boyd, IgH sequences in common variable immune deficiency reveal altered B cell development and selection, Sci. Transl. Med. 7 (2015), https://doi.org/10.1126/scitranslmed.aab1216.
[30] T. Schneider-Hohendorf, H. Mohan, C.G. Bien, J. Breuer, A. Becker, D. Görlich, T. Kuhlmann, G. Widman, S. Herich, C. Elpers, N. Melzer, K. Dornmair, G. Kurlemann, H. Wiendl, N. Schwab, CD8 + T-cell pathogenicity in Rasmussen encephalitis elucidated by large-scale T-cell receptor sequencing, Nat. Commun. 7 (2016), https://doi.org/10.1038/ncomms11153.
[31] A.W.L. Van den Wall Bake, M.R. Daha, J. Radl, J.J. Haaijman, A. Van der Ark, R.M. Valentijn, L.A. Van Es, The bone marrow as production site of the IgA deposited in the kidneys of patients with IgA nephropathy, Clin. Exp. Immunol. 72 (1988).
[32] H. Suzuki, Z. Moldoveanu, S. Hall, R. Brown, H.L. Vu, L. Novak, B.A. Julian, M. Tomana, R.J. Wyatt, J.C. Edberg, G.S. Alarcón, R.P. Kimberly, Y. Tomino, J. Mestecky, J. Novak, IgA1-secreting cell lines from patients with IgA nephropathy produce aberrantly glycosylated IgA1, J. Clin. Invest. 118 (2008) 629–639, https://doi.org/10.1172/JCI33189.
[33] A. Maeda, T. Gohda, K. Funabiki, S. Horikoshi, I. Shirato, Y. Tomino, Significance of serum IgA levels and serum IgA/C3 ratio in diagnostic analysis of patients with IgA nephropathy, J. Clin. Lab. Anal. 17 (2003) 73–76, https://doi.org/10.1002/jcla.10071.
[34] J.J. Zhang, L.X. Xu, G. Liu, M.H. Zhao, H.Y. Wang, The level of serum secretory IgA of patients with IgA nephropathy is elevated and associated with pathological phenotypes, Nephrol. Dial. Transplant. 23 (2008) 207–212, https://doi.org/10.1093/ndt/gfm492.
[35] N. Zhao, P. Hou, J. Lv, Z. Moldoveanu, Y. Li, K. Kiryluk, A.G. Gharavi, J. Novak, H. Zhang, The level of galactose-deficient IgA1 in the sera of patients with IgA nephropathy is associated with disease progression, Kidney Int. 82 (2012) 790–796, https://doi.org/10.1038/ki.2012.197.
[36] J. Feehally, T.J. Beattie, P.E.C. Brenchley, B.M. Coupes, N.P. Mallick, R.J. Postlethwaite, Sequential study of the IgA system in relapsing IgA nephropathy, Kidney Int. 30 (1986) 924–931, https://doi.org/10.1038/ki.1986.274.
[37] K.N. Lai, R.T. Ho, C.K. Lai, C.H. Chan, P.K. Li, Increase of both circulating Th1 and Th2 T lymphocyte subsets in IgA nephropathy, Clin. Exp. Immunol. 96 (1994) 116–121 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1534550&tool=pmcentrez&rendertype=abstract.
[38] S.I. Toyabe, W. Harada, M. Uchiyama, Oligoclonally expanding γδ T lymphocytes induce IgA switching in IgA nephropathy, Clin. Exp. Immunol. 124 (2001) 110–117, https://doi.org/10.1046/j.1365-2249.2001.01494.x.
[39] M.C. Falk, G. Ng, G.Y. Zhang, G.C. Fanning, L.P. Roy, K.M. Bannister, A.C. Thomas, A.R. Clarkson, A.J. Woodroffe, J.F. Knight, Infiltration of the kidney by alpha beta and gamma delta T cells: effect on progression in IgA nephropathy, Kidney Int. 47 (1995) 177–185 http://www.ncbi.nlm.nih.gov/pubmed/7731144.
[40] H. Wu, G.Y. Zhang, A.R. Clarkson, J.F. Knight, Conserved T-cell receptor β chain CDR3 sequences in IgA nephropathy biopsies, Kidney Int. 55 (1999) 109–119, https://doi.org/10.1046/j.1523-1755.1999.00243.x.