Diversity of Immunoglobulin Heavy Chain Repertoire in Patients With Rheumatoid Arthritis

Hung-Cheng Tsai
Taipei Veterans General Hospital
Chik-On Choy
SP BioMED Co., Ltd. Taiwan
Tsai-Hung Wu
Taipei Veterans General Hospital
Chih-Wei Liu
Taipei Veterans General Hospital
Yu-Jen Pan
Buddhist Tzu Chi General Hospital
Wei-Sheng Chen
Taipei Veterans General Hospital
Chia-Sheng Pai
Buddhist Tzu Chi General Hospital
Ning Hsu
SP BioMED Co., Ltd. Taiwan
Shih-Feng Tsai
National Health Research Institutes
Shih-Tzu Tsai
Cheng Hsin General Hospital
Kuei-Ying Su
Buddhist Tzu Chi General Hospital
Chang-Youh Tsai
Taipei Veterans General Hospital
Hsien-Tzung Liao (htliao@vghtpe.gov.tw)
Taipei Veterans General Hospital

Research Article

Keywords: B cell, immune repertoire, complementarity determining region (CDR)3, rheumatoid arthritis, clonal diversity, next generation sequencing

Posted Date: December 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1125791/v1

Abstract

Objectives Rheumatoid Arthritis (RA) is associated with polymorphism in major histocompatibility complex class II genes and dysregulations of $CD4^+$ T cells which cause abnormalities in immune repertoire (iR) expression and intracellular signaling. We monitored nucleotide sequence changes in iR of immunoglobulin heavy chain (*IGH*), particularly complementarity determining region 3 (CDR3) during the course of treatments in RA patients using massively parallel sequencing technology.

Methods CDR3 sequencing was carried out on clinical blood samples from RA patients for disease progress monitoring. The iR of each sample was measured using next generation sequencing (NGS) pipeline. Data analysis was done with a web-based iRweb server. Principal components analysis (PCA) was completed with commercial statistical pipeline.

Results Datasets from 14 patients covered VDJ regions of *IGH* gene. D50 stayed low for all cases (mean D50 = 6.5). A pattern of shared CDR3 sequences was confirmed by a clustering pattern using PCA. Shared profile of 608 CDR3 sequences unique to the disease baseline was identified. D50 analyses revealed clonal diversity would remain low throughout the disease course even after treatment (mean D50 = 11.7 & 8.2 for csDMARD & bDMARD groups respectively) regardless of fluctuated disease activity. PCA has provided a correlation of change in immune diversity along the whole course of RA.

Conclusion We have successfully constructed the experimental design, data acquisition, processing, and analysis pipeline of a high throughput massively parallel CDR3 sequences detection to be used to correlate RA disease activity and *IGH* CDR3 iR during disease progression with or without treatments.

Introduction

Rheumatoid arthritis (RA) is a chronic, destructive autoimmune inflammatory disease which not only causes joint deformity, variant extra-articular manifestations, and increased mortality in individual patients but also poses socioeconomic burdens[1]. Ever-increasing awareness of its pathogenesis has elicited developments of various novel therapeutic modalities for RA. Currently, conventional synthetic disease modifying ant-rheumatic drugs (csDMARDs), biologic (b)DMARDs and targeted synthetic (ts)DMARDs are 3 mainstays for the treatments in RA[2–6]. Traditionally, RA has been described as a T-cell mediated autoimmune disease, which develops via self-antigen presentation to T cell receptor (TCR) and subsequent amplification of the immune cell activation[7]. However, ever-increasing evidence has suggested that B cells also play essential roles in its pathogenesis[8].

The immune repertoire (iR) is the summation of T and B cells in a human body at any given moment[9]. It represents the very moment or the chronology of individual's immune functions. Application of the iR sequencing in clinical medicine is cumulative in recent years, including cancer characterization and human T cell subset study in organ transplantation[10,11]. Sequencing the expressed T or B cell genes in iR may enable an accurate evaluation of disease activity, treatment response, and disease prognosis for RA. Previous studies have demonstrated the effects of bDMARDs on different auto-reactive T cell subtypes and TCR repertoire in RA[11–16]. We have used high-throughput sequencing approach to identify TCR iRs in RA patients receiving different types of bDMARDs[12]. Although, no significant difference has been found among various biological treatment strategies, an inverse tendency between TCR repertoire diversity and RA disease activity was found, suggesting that TCR iR can be a potential response biomarker for bDMARDs-treated RA.

It is known that RA patients express a skewed repertoire of polyclonal, hypo-mutated B cell receptor (BCR)[17]. Additionally, the dominant B cell clones in blood can predict onset of arthritis in individuals at risk for RA[18]. However, less little studies have been focused on BCR iR in RA, either the relationship between RA disease *per se* and BCR iR or the changes upon its treatments. In this study, we monitored the changes of iR in immunoglobulin heavy chain (*IGH*) of BCR, particularly complementarity determining region 3 (CDR3) sequences using massively parallel sequencing technology. We aimed to identify the correlation between disease activity and changes in BCR iR and its diversity in RA patients before and after treatment with csDMARDs or bDMARDs.

Methods

Patient selection, treatment protocol, and traditional disease markers

The whole study protocol was verified by the institutional review board of Buddhist Hualian Tzu Chi Hospital (IRB-105-65-A). Seventeen patients were initially enrolled after they had signed the informed consent form, but three were lost followed up due to personal reasons. All procedures were performed after informed consent was signed, and all methods were performed in accordance with the relevant guidelines and regulations. The 16 samples for 14 patients were further categorized into those with newly diagnosed RA who were treatment naïve (Group 1, sample numbers: 8) and those with established RA who had been treated with csDMARDs but were bDMARD naïve (Group 2, sample numbers: 8). Two patients (1st patient designated as sample No. 4 and sample No. 14; 2nd patient designated as sample No. 8 and sample No. 16) participated in both groups (initially in Group 1 and then being extended into Group 2). The major demographics, and the treatment history including csDMARDs, and bDMARDs are listed in Table 1. Their additional laboratory information is included in the supplementary Table 1. Among them, 14 patients were scrutinized throughout the disease progression with blood samples being collected at a total of 4 time-points, day 0, day 30, day 90, and day 180 after the trial entry. Thus, a total of 4 specimens for serial detections were collected for each of these 14 patients (except patients No. 4 [or 14] and No. 8 [or 16] who were collected with blood samples for 7 times each). After treatments, a clinically meaningful effective treatment is defined as decrease in DAS28 score of more than 1.2 or outcome DAS28 score of less than 3.2.[19, 20]

Table 1

RA patients¶ and their therapeutic status at the respective sampling time point

Group	Patient No.	Sex	Age (year)	DMARD doses at sampling time***			
				Day 0	**Day 30**	**Day 90**	**Day 180**
cs-DMARD	1	M	58	M7.5+S0+H4+L0	M15+S0+H4+L0	M15+S0+H4+L0	M7.5+S0+H4+L0
	2	M	70	M0+S0+H4+L0	M0+S2+H4+L0	M0+S1+H2+L0	M0+S1+H2+L0
	3	F	50	M0+S0+H4+L0	M7.5+S0+H2+L0	M7.5+S0+H2+L0	M7.5+S0+H2+L0
	4	F	53	M10+S0+H2+L0	M15+S0+H4+L0	M15+S0+H2+L2	M15+S0+H0+L2
	5	F	73	M10+S0+H0+L0	M10+S0+H4+L0	M15+S0+H2+L0	M15+S0+H0+L1
	6	F	45	M0+S2+H0+L0	M0+S2+H0+L0	M0+S2+H0+L0	M0+S2+H0+L0
	7	F	63	M7.5+S0+H0+L0	M7.5+S0+H0+L0	M7.5+S0+H0+L0	M10+S2+H0+L0
	8	M	53	M10+S0+H0+L0	M10+S0+H0+L0	M15+S2+H0+L0	M17.5+S2+H0+L0
Add-on b-DMARD	9*	M	64	M15+S0+H4+L0	M15+S0+H4+L0	M15+S0+H4+L0	M12.5+S0+H2+L0
	10*	F	45	M15+S0+H4+L0	M15+S0+H2+L0	M15+S0+H2+L0	M7.5+S0+H2+L0
	11**	F	67	M15+S2+H0+L0	M15+S2+H0+L0	M15+S1+H0+L0	M15+S1+H0+L0
	12#	F	85	M10+S1+H2+L0	M10+S1+H2+L0	M10+S1+H2+L0	M10+S1+H2+L0
	13**	M	53	M15+S2+H2+L0	M15+S2+H2+L0	M15+S2+H2+L0	M17.5+S2+H2+L0
	14◊	F	53	M15+S0+H0+L2	M15+S0+H0+L0	M15+S0+H0+L0	M10+S0+H0+L0
	15#	F	77	M15+S1+H0+L0	M15+S2+H0+L0	M7.5+S1+H0+L0	M7.5+S1+H0+L0
	16ϕ	M	53	M17.5+S2+H0+L0	M17.5+S2+H0+L0	M17.5+S2+H0+L0	M0+S2+H0+L1

¶ A total of 14 patients with 16 enrollments (No. 4 and No. 14 are the same patient and No. 8 and No. 16 are the same patient. csDMARD= conventional synthetic disease modifying anti-rheumatic drug; bDMARD = biological disease modifying anti-rheumatic drug = biologics, which included * tocilizumab, ** abatacept, # adalimumab, ϕ certolizumab pegol and ◊ golbumab; *** csDMARD doses are represented with [M (methotrexate mg/wk)+ S (sulfasalazine gm/d) + H (hydroxychloroquine $X10^2$ mg/d)+ L (lefluonomide X 10 mg/d)].

Next generation sequencing (NGS)

In order to evaluate iR at the disease baseline by CDR3 sequencing in BCR heavy chain, all specimens are collected from each patient with minimal required treatments at the beginning of study. Disease progression after treatments was evaluated by CDR3 sequencing of BCR heavy chain using longitudinal specimen series of four samples from each patient as mentioned above. A methodology overview contained the following four items, i.e. experimental design, RNA extraction, iRepertoire library preparations, and MiSeq® (Illumina, San Diego, CA) sequencing.

Data analysis

Raw datasets fastq files were uploaded into the server maintained by iRepertoire® for initial data processing. Data analysis was performed by iRweb software pipeline (iRepertoire, Huntsville, AL) (https://irweb.irepertoire.com/nir/). Sequence diversity of the iR is represented by D50 which is calculated by the following formula:

$$\text{Assume that } \frac{r_1 \geq r_2 \cdots \geq r_i \geq_{i+1} \cdots \geq r_s}{s}, \sum_{i=1}^{s} r_i = J$$

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

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

where D50 is defined as the diversity of an iR of total number of CDR3s including S distinct CDR3s in a ranked dominance configuration, and "r" stands for the amount of most abundant CDR3s, r1 is the amount of the 20 most abundant CDR3s, r2 is the amount of the second most abundant CDR3, and so on (i.e., $r_1 > r_2 > r_3 > \ldots ri > r_{i+1} > \ldots r_n$), X is the number of distinct CDR3s of the specific sample, C is the minimum number of distinct CDR3s amounting to ≥50% of total sequencing reads[21]. The Diversity Index (DI) is calculated by the following formula, which is defined mathematically as follows:

assume that the frequency of the individual unique CDR3 are

$r_1 > r_2 > r_3 > \cdots r_k > r_{k+1} > \cdots > r_n$

where k stands for the k-th sample, r_i is the frequency of CDR3 in the i-th sample, r_k is the frequency of CDR3 in the k-th sample and n is the total number of unique CDR3s.

$$x_k = \frac{k}{n}, \qquad y_k = \frac{\sum_{i=1}^{k} r_i}{\sum_{i=1}^{n} r_i}$$

Principal component analysis (PCA) for baseline RA disease activity before treatment

A very convenient feature of the iRweb software is CDR3 algebra, which allows the comparison of the CDR3 sequences from one data set to another data sets to identify shared CDR3s profile. This allows for a comparison amongst samples of different time points during treatment. All CDR3 frequencies were artificially scaled to 10 million reads to account for differences in read depth among samples, making comparisons between samples easier with this normalization step. A shared profile of RA disease baseline was obtained using this function. We constructed observations/variables table based on this profile taking different samples as well as different observations and different CDR3 clone sequences as different variables. We then applied the PCA function of a commercial module of Excel (Microsoft), namely xlstat (Addinsoft), to compute the PCA results on this set of baseline samples. Finally, observation chart with PC1 & PC2 for dimension reduced CDR3 listings of all samples was plotted for evaluation and observations based on factor scores.

Principal component analysis (PCA) for RA disease progression

Raw data in the form of CDR3 listings from serial detection of each patient were processed with the same pipeline of iRweb and analyzed using the PCA function of xlstat module in a PC windows environment. Both groups of patients undergoing csDMARD/bDMARD treatments were processed separately.

Table 2

RA patients¶ and their disease activity as scored by DAS28 at the respective sampling time points

Group	Patient No.	Sex	Age (year)	DAS 28 (D50) score at sampling time			
				Day 0	**Day 30**	**Day 90**	**Day 180**
cs-DMARD	1	M	58	5.84	5.57	5.21	4.92
	2	M	70	4.51	4.39	4.60	4.34
	3	F	50	5.16	4.42	4.43	5.10
	4	F	53	5.54	5.42	6.43	6.21
	5	F	73	4.48	4.71	3.49	3.19
	6	F	45	5.91	3.18	4.04	4.16
	7	F	63	5.29	4.94	5.24	2.73
	8	M	53	8.56	4.34	7.01	7.00
Add-on b-DMARD	9*	M	64	4.69	3.25	2.88	1.5
	10*	F	45	4.59	3.29	3.69	2.32
	11**	F	67	8.16	5.43	5.16	4.98
	12#	F	85	4.80	4.78	4.79	4.64
	13**	M	53	5.50	5.43	4.65	4.06
	14◊	F	53	6.21	4.90	5.76	4.96
	15#	F	77	6.31	5.59	5.52	4.85
	16ϕ	M	53	7.00	4.39	3.95	3.91

¶ A total of 14 patients with 16 enrollments (No. 4 and No. 14 are the same patient and No. 8 and No. 16 are the same patient). csDMARD= conventional synthetic disease modifying anti-rheumatic drug; bDMARD = biological disease modifying anti-rheumatic drug = biologics, which included * tocilizumab, ** abatacept, # adalimumab, ϕ certolizumab pegol and ◊ golbumab.

Results

CDR3 detection, D50 and DI analysis

For the disease baseline study, the D50 values of most samples were below the value 10, suggesting low diversity of B cells in these samples. There were 1,035,476 average CDR3 reads (130,395 average unique CDR3) detected for all specimens with an average D50 value of 6.4 (DI 18.0 and Entropy 11.1). DI and entropy are alternate measurement values for diversity. In addition to the above definition," DI" could alternatively defined as 100 minus the area under the curve between percentage of total reads and percentage of unique CDR3s, when the frequencies of unique CDR3s are accumulated from most frequent to least frequent. On the other hand, "entropy" is defined as that of the Shannon Entropy[22]. There was a shared profile containing 608 BCR heavy chain CDR3 sequences which was exhibited across the samples as listed in supplementary Table 2. For the disease progression study, the results were presented in two groups, i.e. datasets obtained from 8 csDMARD-naïve patients undergoing csDMARD treatment and those from 8 bDMARD-naïve patients undergoing bDMARD treatment. The average D50 values from the 8 csDMARD-treated patients were 11.7, suggesting low diversity of B cells in them (Table 3). In average, there were 945,519 CDR3 reads (128,715 unique CDR3) detected for each sample. The average D50 values from the 8 bDMARD-treated patients were 8.2, suggesting also low diversity of B cells in them (Table 4). In average, there were

1,065,693 CDR3 reads (123,518 unique CDR3) detected for each sample. There was not steady trend of increase in D50 values within the same individual during the treatment process for most patients.

Table 3
CDR3 sequence detections and immune diversity of the patients along the csDMARD treatment course

Sample No.	CDR3	Unique CDR3	D50	Diversity Index	Entropy
1	970181	69775	5.7	13.7	11.5
	135446	2541	0.1	3.1	4.9
	1192380	82431	5.4	17.9	11.7
	907993	115562	7.7	19.1	12
2	773966	3042	0	1.8	0.7
	55691	320	0.3	2.2	1.1
	657603	2424	0	1.8	0.9
	714459	2620	0	1.8	0.7
3	925892	154922	6.4	19.2	11.8
	388511	7678	0.4	3.3	7.3
	171610	70426	26.3	34.9	12.8
	839957	94713	1.7	13.4	10.1
4	1129489	156706	5.9	17.6	11.5
	1215866	167172	24.6	33.9	12.6
	1108047	151344	35	40.5	13
	1307567	206063	37.3	42.1	13.1
5	1615613	147021	9.7	24	12.2
	1193499	91408	5.3	21.9	11.8
	1882124	81473	12.8	18.1	12.1
	961420	101366	35.6	41	13.2
6	856974	270047	7.2	23.1	11.9
	1201338	53715	7.2	13	11.4
	950539	200175	4.6	16.2	11.2
	1282392	238848	4.6	15.8	11.4
7	640059	33960	6.1	12.4	11.2
	986845	84375	3.1	11.8	10.9
	1459166	126362	4.5	14.3	11.4
	547755	74709	13.8	27.8	12.3
8	1061610	290379	18.5	29.8	12.5
	641132	339762	30.9	37.6	12.9
	1319762	446015	22.7	33.2	12.5
	1161736	251537	30.4	37.7	13

Table 4
CDR3 sequence detection and immune diversity of the patients receiving add-on bDMARD treatment with longitudinal follow-up

Sample No.	CDR3	Unique CDR3	D50	Diversity Index	Entropy
9	973776	67896	3.3	15.5	11.2
	1343145	75759	4.2	16.2	11.5
	1240304	60757	2.4	13.7	11
	1054288	41441	1.1	9.7	10
10	902657	157768	7.9	22.1	12.1
	138116	2084	0.4	3	5.4
	1329572	227500	6.2	17.3	11.6
	1052733	179702	1.3	14	8.9
11	1162217	36524	2.8	6.7	10.4
	762070	114477	2.2	20.4	10.8
	1248839	178649	8.5	26.2	11.9
	1277564	90165	4.6	12.1	11.1
12	1182785	133418	7.9	25	12.1
	703661	78337	2	13.1	10.5
	1331235	111798	0.1	9.7	7.6
	1048034	127518	5.7	17	11.3
13	644467	45600	5.3	14.2	11.4
	1596790	151582	8.9	20.6	12.1
	631714	90093	14.2	27.8	12.5
	1252213	128029	11.2	25.4	12.1
14	1307567	206063	37.3	42.1	13.1
	1134029	207189	36	41.1	13
	939318	155525	19.4	31.1	12.3
	1068616	213338	8.7	24.6	12.1
15	1360150	79116	3.8	17.8	11.5
	1242343	77484	8	17.9	11.7
	1263108	127132	12.4	23.6	12.3
	1120029	86748	5.6	14.2	11.5
16	1161736	251537	30.4	37.7	13
	1049932	232889	8.5	19.2	11.9
	1146453	307009	10.6	24.5	12.3
	528772	37468	2.8	9.8	10.5

The CDR3 region is of particularly interest to us as the antigen specificity is highly correlated with this region of the BCR. The D50 is a quantitative measure of the degree of diversity of B cells within a sample. It is the percentage of dominant and unique B cell clones that account for the cumulative 50% of the total CDR3s counted in the sample. The more diverse a library, the closer the value will be to 50. Low diversity values are associated with decreased diversity. BCR heavy chain datasets and CDR3 listing in patients were obtained for disease baseline evaluation and *IGH* datasets from patients were obtained for disease progression evaluation.

Principal component analysis (PCA) for baseline disease activity of RA

Apart from shared CDR3 analysis, we also performed a PCA on the disease baseline datasets and tried to analyze numerical data (CDR3 listings) structured in observations/ many variables table as described previously. We analyzed the CDR3 sequence listings as principal component 1 & 2 (PC1 & PC2) for a 2D visualization. The PC1 & PC2 of CDR3 listings from RA samples are presented as Figure 1. We found that the CDR3 listings from RA samples held a converging profile.

Principal component analysis (PCA) for RA disease progression

PCA results of IGH datasets on serial detections from 8 csDMARD-naïve patients undergoing DMARD treatment were plotted as principal component 1 and 2 (PC1 & PC2) for a 2D visualization (Figure 2). PC1 and PC2 data points within the same patient exhibit a converging pattern in patients. Patient No. 3 exhibited a slight consistent horizontal shift along the PC1 axis. PCA results of IGH datasets on serial detections of 8 bDMARD-naïve patients receiving biologic treatment were also plotted as principal component 1 and 2 (PC1 & PC2) for a 2D visualization (Figure 3). PC1 and PC2 data points within the same individual exhibited a converging pattern for patients. The correlation between this PCA pattern with respect to clinical phenotype has also yet to be examined with more clinical disease markers.

Discussion

There have been a lot of studies dealing with the iR of T cells in human RA, but only little studies have been dedicated to the iR of B cells in RA, especially in patients under bDMARD treatments. The present investigation might have demonstrated a consequential change in iR of BCR in RA patients for the first time. It is accepted that clonal diversity is generally getting low in disease state[8, 23]. We have shown that RA disease baseline condition is matching to what would be expected in CDR3 sequence diversity change.

The present investigation has also shown that RA disease activity was inversely correlated with D50 level, i.e., for the CD3 clonal diversity, patients with active RA demonstrated low D50. Both csDMARD and bDMARD treatment groups have shown steady growing trend of D50 values compared with their baseline samples along the treatment course, which is matching to what would be expected for disease improvement. The DAS28 values of two patients, No.13 and No.15 were decreasing accompanied with clonal expansion in the bDMARD group. The DAS28 values of Patient No. 13, which were 5.50, 5.43, 4.65, and 4.06 in sequence along the timeline, exhibited a decreasing trend in disease activity. The DAS28 values of Patient No. 15, which were 6.31, 5.59, 5.52, and 4.85 in sequence along the timeline, also exhibited a decreasing trend in disease activity. But in csDMARD group, one patient (No. 4) showed irrelevance of DAS28 to D50 level (Table 2 & Table 3), i.e., while improving D50 level, there was no improvement in terms of DAS28. Another patient (No.5) revealed the same trend as that in bDMARD group, which were 4.48, 4.71, 3.49 and 3.19 of D50 in sequence along the timeline, exhibiting a decreasing trend in disease activity. However, in contrary to increase in clonal diversity, the last DAS28 value on both cases was still above 3.2, indicating no significant improvement. Beside possible confounders such as undercurrent infection, this may conversely suggest D50 level is more sensitive than traditional DAS28 value to reflect disease activity. After all, pain or tenderness is so subjective that varies among patients, irrelevant to actual inflammatory status. Since tender joint count is one of the components of DAS28, the disease activity can also not be so reliably assessed by DAS28. Instead, D50 can be a better companion indicator of disease improvement trend, superior to DAS28 score. For a long time, lacking reliable biomarkers for prompt stratification of

individual patients to fit the most appropriate and effective medications has rendered current treatment consensus for RA so challenging[24]. D50 level may potentially serve as an immediate biomarker to overcome this predicament.

Regarding the PCA, both “specific treatment” naïve samples did show a clustering effect, implying that specific clonal expansion pattern is present for RA. We have performed shared profile analysis on patient datasets of disease baseline to identify shared dominant CDR3 sequences. We identified a shared profile of 608 CDR3 clone sequences, i.e. clone expansion sequences which were present in at least two patients. Sequence “ARDLDY” (A=alanine R=arginine D=aspartic acid L=leucine Y=tyrosine) is the most significant CDR3 expansion sequence in these RA patients. Although currently, we still could not identify specific peptide that contains this sequence, it is expected that it could eventually be clarified when more data are accumulated.

We have also performed shared profile analysis on patient datasets of disease progression to identify shared dominant CDR3 sequences in patients with csDMARD/ bDMARD treatment. PCA analysis on CDR3 shared profile revealed a divergent tendency is some patients both in csDMARD/ bDMARD groups (Sample No. 5 in DMARD group, No.13, No. 15 in biologic group) indicating changes in immune diversity. Besides, the D50 values of Patient No. 5, No. 13 and No. 15 also exhibited an increasing trend in immune diversity, which were 9.7, 5.3, 12.8, 35.6; 5.3, 8.9, 14.2, 11.2; and 3.8, 8.0, 12.4, 5.6 in sequence along the timeline respectively, exhibiting an increasing trend in immune diversity. Taken together, PCA-measured divergent profiles can be a good indicator for an increase in immune diversity expressed as D50 values. On the other hand, serial detections during treatment course also exhibit clustering effect, supporting our hypothesis that specific clonal expansion pattern in individual patient may occur in response to some unique exogenous stimuli such as csDMARD/ bDMARD treatments within a same kind of underlying disease such as RA, regardless of disease improvement or deterioration.

Oral and gut microbiota have been thought to be one of the environmental factors that may enhance the development of RA[25–28]. For example, periodontitis caused by microorganisms might alter post-transcriptional regulation and citrullinated self-protein, leading to autoantibody production and local inflammation, and finally triggering the onset of RA. The change of healthy microbiota and pathogen-host immune system interaction may integrate the reduced immune diversity and BCR CD3 sequences clustering effect. Other environmental risk factor, especially cigarette smoking may also result in same clustering effect, eventually precipitating RA development[29–31]. On the contrary, treatments such as csDMARD/bDMARD break down the vicious cycle and lead to increased immune diversity. Clinically, autoantibodies such as RF and anti-CCP, though with very high specificity for disease entity, can be detected up to a decade before the development of RA[32]. Hence, no one can predict the disease onset of RA. CDR3 sharing profile analysis may potentially predict the development of RA and guide physicians to a timely and appropriate intervention. Clustering effect may represent inchoate of disease and serves as a predictor biomarker; however, more data are needed to confirm this hypothesis.

In spite of the present promising findings, our investigation holds some limitations. Firstly, this study was restrained by a small sample size, which may limit the statistical power for detecting the variation of the B-cell repertoire in clinical assessments. Yet, we still observed a trend toward a decrease of BCR iR diversity in RA patients prior to treatment and an increase in it after either csDMARDs or biologics, and a larger cohort study is required for solid validation. Secondly, our study only focuses on peripheral B cell repertoire in RA patients without B-cell and/ or plasma-cell clones in synovial tissue, and we will plan to extend our study in the future. Finally, we measured and analyzed the BCR *IGH* CD3 iR sequencing to monitor the disease activity, and further functional investigation will be conducted to elucidate the potential roles in the pathogenesis of RA.

In conclusion, we have successfully established a pipeline of experimental design, data acquisition, processing, and analysis to be used in monitoring the disease progression of RA. PCA is used to perform dimension reduction to make sense of CDR3 sequences variations, and the BCR CDR3 sequences do have a clustering pattern upon PCA representation which is an indication of specific clonal expansion in RA. We demonstrated that BCR CD3 iR sequencing can be a good tool to follow up on the disease progression of individual RA patients. Low BCR repertoire diversity was noted in treatment naïve RA patients.

Conversely, immune diversity expands upon recovery or remission from active disease after csDMARDs and biologics treatment.

Declarations

Acknowledgements

We thank the patients and their caregivers as well as the investigators who contributed to this study.

Funding:

This study was supported by the Ministry of Science & Technology, Taiwan (MOST107-2314-B075-051-MY3), Taipei Veterans General Hospital (V107D37-002-MY3), National Health Research Institutes and Buddhist Tzu-Chi Foundation (TC-NHRI-105-03).

Disclosure statement:

the authors have declared no conflicts of interest.

References

1. McInnes, I. B. & Schett, G. The pathogenesis of rheumatoid arthritis. *N Engl J Med*, **365**, 2205–19 (2011).
2. Emery, P. *et al.* Efficacy of monotherapy with biologics and JAK inhibitors for the treatment of rheumatoid arthritis: a systematic review. *Adv Ther*, **35**, 1535–63 (2018).
3. Köhler, B. M., Günther, J., Kaudewitz, D. & Lorenz, H-M. Current therapeutic options in the treatment of rheumatoid arthritis. *J Clin Med*, **8**, 938 (2019).
4. Aletaha, D. & Smolen, J. S. Diagnosis and management of rheumatoid arthritis: a review. *JAMA*, **320**, 1360–72 (2018).
5. Davis, J. M. III & Matteson, E. L. My treatment approach to rheumatoid arthritis. *Mayo Clin Proc*, **87**, 659–73 (2012).
6. Perrone, V. *et al.* Real-world analysis of therapeutic patterns in patients affected by rheumatoid arthritis in Italy: a focus on baricitinib. *Rheumatol Ther*, **7**, 657–65 (2020).
7. Granato, A., Chen, Y. & Wesemann, D. R. Primary immunoglobulin repertoire development: time and space matter. *Curr Opin Immunol*, **33**, 126–31 (2015).
8. Panayi, G. S. B cells: a fundamental role in the pathogenesis of rheumatoid arthritis? Rheumatology 2005;44 (Suppl 2):ii3-ii7
9. Han, Y., Li, H., Guan, Y. & Huang, J. Immune repertoire: a potential biomarker and therapeutic for hepatocellular carcinoma. *Cancer Lett*, **379**, 206–12 (2016).
10. Jang, M. *et al.* Characterization of T cell repertoire of blood, tumor, and ascites in ovarian cancer patients using next generation sequencing. *Oncoimmunology*, **4**, e1030561 (2015).
11. Wang, C. *et al.* High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc Natl Acad Sci USA*, **107**, 1518–23 (2010).
12. Chang, C-M. *et al.* Characterization of T-cell receptor repertoire in patients with rheumatoid arthritis receiving biologic therapies. *Dis Markers*, **2019**, 2364943 (2019).
13. McGovern, J. L. *et al.* Th_{17} cells are restrained by Treg cells via the inhibition of interleukin-6 in patients with rheumatoid arthritis responding to anti-tumor necrosis factor antibody therapy. *Arthritis Rheum*, **64**, 3129–38 (2012).
14. Nguyen, D. X. & Ehrenstein, M. R. Anti-TNF drives regulatory T cell expansion by paradoxically promoting membrane TNF-TNF-RII binding in rheumatoid arthritis. *J Exp Med*, **213**, 1241–53 (2016).
15. Samson, M. *et al.* Brief report: inhibition of interleukin-6 function corrects Th_{17}/Treg cell imbalance in patients with rheumatoid arthritis. *Arthritis Rheum*, **64**, 2499–503 (2012).

16. Eggleton, P. & Bremer, E. Direct and indirect rituximab-induced T cell depletion: comment on the article by Mélet. *Arthritis Rheumatol*, **66**, 1053 (2014).
17. Cowan, G. J. M. *et al.* In human autoimmunity, a substantial component of the B cell repertoire consists of polyclonal, barely mutated $IgG^{(+ve)}$ B Cells. *Front Immunol*, **395**, https://doi.org/doi: 10.3389/fimmu.2020.00395 (2020).
18. Tak, P. P. *et al.* Dominant B cell receptor clones in peripheral blood predict onset of arthritis in individuals at risk for rheumatoid arthritis. *Ann Rheum Dis*, **76**, 1924–30 (2017).
19. Fransen, J. & van Riel, P. L. C. M. The disease activity score and the EULAR response criteria. *Clin Exp Rheumatol*, **23** (5 Suppl 39), S93–S99 (2005).
20. Wells, G. *et al.* Validation of the 28-joint disease activity score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. *Ann Rheum Dis*, **68**, 954–60 (2009).
21. Cao, X. *et al.* High throughput sequencing reveals the diversity of TRB-CDR3 repertoire in patients with psoriasis vulgaris. *Int Immunopharmacol*, **40**, 487–91 (2016).
22. Shannon, C. E. A mathematical theory of communication. Bell. *System Tech J*, **27**, 379–423 (1948).
23. Bashford-Rogers, R. J. M., Smith, K. G. C. & Thomas, D. C. Antibody repertoire analysis in polygenic autoimmune diseases., **155**, 3–17 (2018).
24. Smolen, J. S. *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. *Ann Rheum Dis*, **79**, 685–99 (2020).
25. Sun, Y. *et al.* Characteristics of gut microbiota in patients with rheumatoid arthritis in Shanghai, China. *Front Cell Infect Microbiol*, **9**, 369 (2019).
26. Corrêa, J. D. *et al.* Oral microbial dysbiosis linked to worsened periodontal condition in rheumatoid arthritis patients. *Sci Rep*, **9**, 8379 (2019).
27. Lorenzo, D. *et al.* Oral-gut microbiota and arthritis: Is there an evidence-based axis? *J Clin Med*, **8**, 1753 (2019).
28. Wells, P. M., Williams, F. M. K., Matey-Hernandez, M. L., Menni, C. & Steves, C. J. 'RA and the microbiome: do host genetic factors provide the link? *J Autoimmun*, **99**, 104–15 (2019).
29. Chang, K. *et al.* Smoking and rheumatoid arthritis. *Int J Mol Sci*, **15**, 22279–95 (2014).
30. Hutchinson, D., Shepstone, L., Moots, R., Lear, J. T. & Lynch, M. P. Heavy cigarette smoking is strongly associated with rheumatoid arthritis (RA), particularly in patients without a family history of RA. *Ann Rheum Dis*, **60**, 223–7 (2001).
31. Silman, A. J., Newman, J. & MacGregor, A. J. Cigarette smoking increases the risk of rheumatoid arthritis. Results from a nationwide study of disease-discordant twins. *Arthritis Rheum*, **39**, 732–5 (1996).
32. Nielen, M. M. J. *et al.* Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. *Arthritis Rheum*, **50**, 380–6 (2004).

Figures

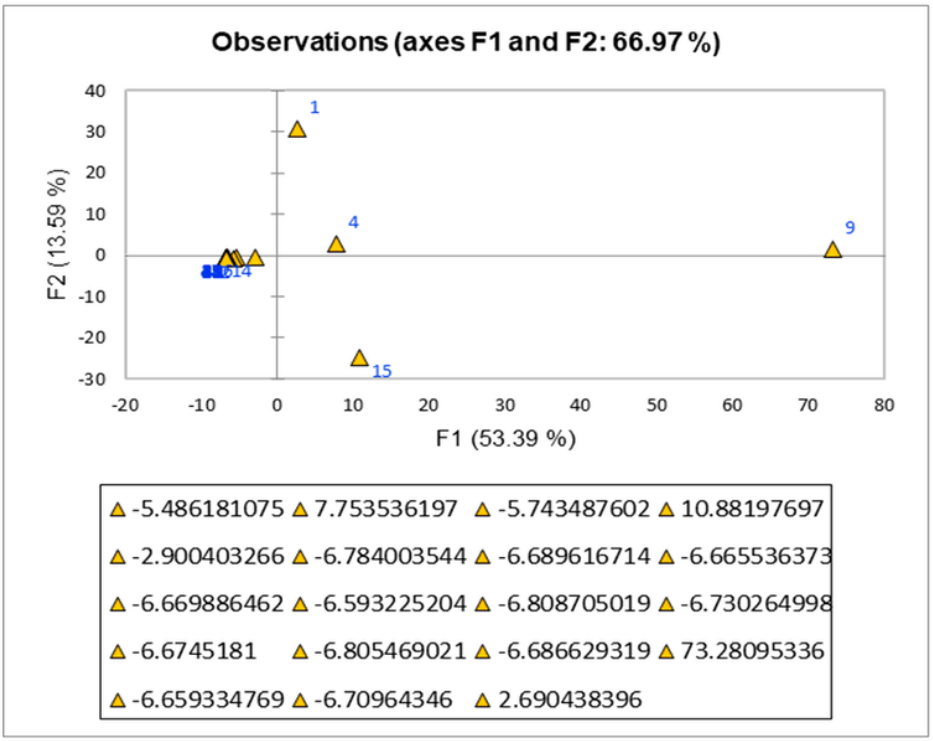

Figure 1

Principal component analysis was used to distinguish IGH repertoires among disease baseline datasets. The observations charts represented the observations in the PCA space. A new set of variables or principal components (PCs) was created to describe more complex variability in the datasets. The first PC (PC1) explains the maximum variance of the datasets, followed by PC2, and so on. X and Y axes show principal components 1 (F1) and 2 (F2), and the percent variation explained by each component is shown in parenthesis. The numbers under X axe are the factor scores calculated for each dataset under F1 (X coordinates). Factor scores are the observations coordinates on the PCA dimensions.

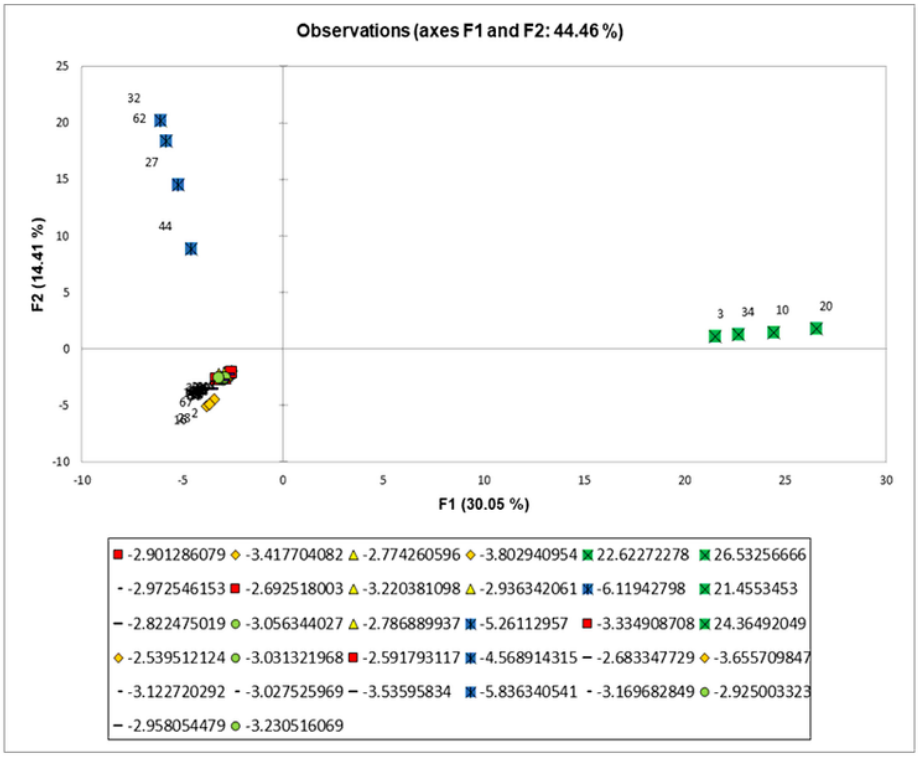

Figure 2

PCA of IGH datasets on serial detection of 8 RA patients undergoing DMARD treatment for disease progression. Principal component analysis was used to distinguish IGH repertoires on serial detection datasets. The observations charts represented the observations in the PCA space. A new set of variables or principal components (PCs) was created to describe more complex variability in the datasets. The first PC (PC1) explains the maximum variance of the datasets, followed by PC2, and so on. X and Y axes show principal components 1 (F1) and 2 (F2), and the percent variation explained by each component is shown in parenthesis. The numbers under X axe are the factor scores calculated for each dataset under F1 (X coordinates). Factor scores are the observations coordinates on the PCA dimensions. (X, Y) coordinates derived from consecutive samples of the same patient have the same color and shape labels. Time sequence of blood sample drawn is marked as separate sample numbers in an ascending order.

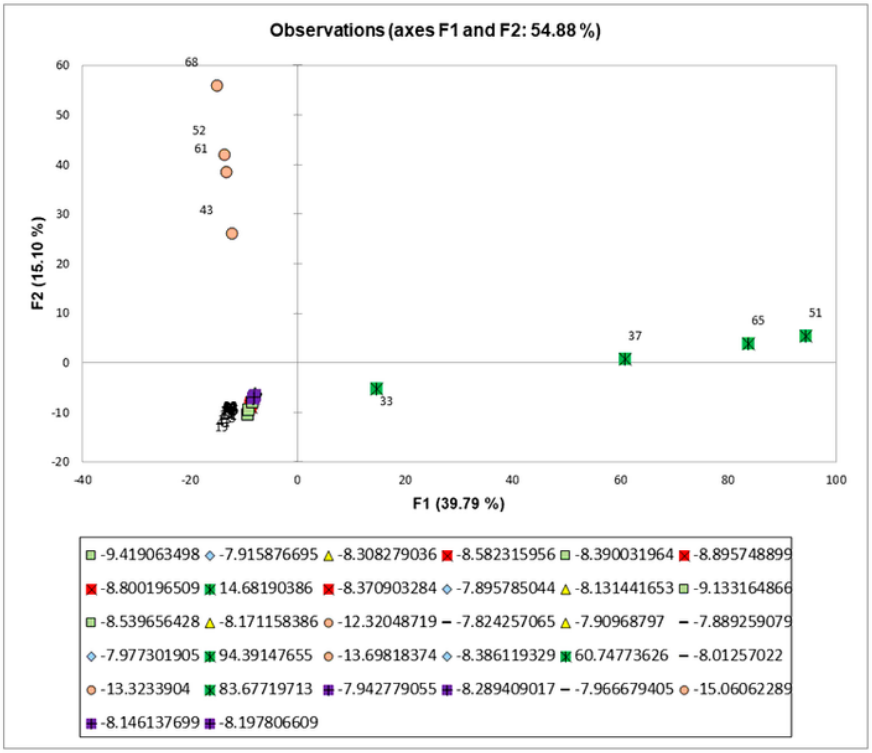

Figure 3

PCA of IGH datasets on serial detection of 8 RA patients undergone biologic treatment for disease progression. Principal component analysis was used to distinguish IGH repertoires on serial detection datasets. The observations charts represented the observations in the PCA space. A new set of variables or principal components (PCs) was created to describe more complex variability in the datasets. The first PC (PC1) explains the maximum variance of the datasets, followed by PC2, and so on. X and Y axes show principal components 1 (F1) and 2 (F2), and the percent variation explained by each component is shown in parenthesis. The numbers under X axe are the factor scores calculated for each dataset under F1 (X coordinates). Factor scores are the observations coordinates on the PCA dimensions. (X, Y) coordinates derived from consecutive samples of the same patient have the same color and shape labels. Time sequence of blood sample drawn is marked as separate sample numbers in an ascending order.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementtable1.docx
- Supplementtable2.docx