Journal of Human Genetics
https://doi.org/10.1038/s10038-017-0364-0

ARTICLE

Characterization of the B-cell receptor repertoires in peanut allergic subjects undergoing oral immunotherapy

Kazuma Kiyotani[1] · Tu H Mai[1,2] · Rui Yamaguchi[3] · Poh Yin Yew[1,4] · Mike Kulis[5] · Kelly Orgel[5] · Seiya Imoto[6] · Satoru Miyano[3] · A Wesley Burks[5] · Yusuke Nakamura[1,7]

Received: 24 July 2017 / Revised: 9 September 2017 / Accepted: 9 September 2017
© The Author(s) 2017. This article is published with open access

Abstract

B-cell receptors (BCRs) play a critical role in adaptive immunity as they generate highly diverse immunoglobulin repertoires to recognize a wide variety of antigens. To better understand immune responses, it is critically important to establish a quantitative and rapid method to analyze BCR repertoire comprehensively. Here, we developed "Bcrip", a novel approach to characterize BCR repertoire by sequencing millions of BCR cDNA using next-generation sequencer. Using this method and quantitative real-time PCR, we analyzed expression levels and repertoires of BCRs in a total of 17 peanut allergic subjects' peripheral blood samples before and after receiving oral immunotherapy (OIT) or placebo. By our methods, we successfully identified all of variable (V), joining (J), and constant (C) regions, in an average of 79.1% of total reads and 99.6% of these VJC-mapped reads contained the C region corresponding to the isotypes that we aimed to analyze. In the 17 peanut allergic subjects' peripheral blood samples, we observed an oligoclonal enrichment of certain immunoglobulin heavy chain alpha (IGHA) and IGH gamma (IGHG) clones ($P = 0.034$ and $P = 0.027$, respectively) in peanut allergic subjects after OIT. This newly developed BCR sequencing and analysis method can be applied to investigate B-cell repertoires in various research areas, including food allergies as well as autoimmune and infectious diseases.

Introduction

The adaptive immune system is a complex network of cells, including B and T cells, and multiple organs defending against pathogens. Antibodies (immunoglobulin; IG), the soluble form of B-cell receptors (BCRs), are essential in adaptive immunity. Healthy humans have ~3×10^9 B cells in their peripheral blood, and this population consists of a repertoire of distinct B cells expressing different BCRs to recognize a wide variety of antigens and play critical roles in an effective humoral immune response. BCRs are a heterodimeric protein complex, composed of two identical heavy-chain (IGH) and two identical light-chain proteins [1, 2]. *BCR* genes consist of multiple distinct gene segments corresponding to the variable (V), diversity (D, only for *IGH* genes), and joining (J) regions, and undergo site-specific V(D)J recombination [3, 4]. During the recombination, random deletion and/or non-templated insertion at the junction site occurred and this process significantly increases the BCR diversity; within this rearranged protein, a complementarity-determining region 3 (CDR3) is considered to be critical for antigen recognition. Rearranged *BCR* genes obtain further diversity by helper T-cell-mediated somatic

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s10038-017-0364-0) contains supplementary material, which is available to authorized users.

✉ Yusuke Nakamura
ynakamura@bsd.uchicago.edu

1 Department of Medicine, The University of Chicago, Chicago, IL 60637, USA

2 Committee on Clinical Pharmacology and Pharmacogenomics, The University of Chicago, Chicago, IL 60637, USA

3 Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan

4 OncoTherapy Science, Inc., Kawasaki, Kanagawa 213-0012, Japan

5 Department of Pediatrics, University of North Carolina, Chapel Hill, NC 27599, USA

6 Health Intelligence Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan

7 Department of Surgery, The University of Chicago, Chicago, IL 60637, USA

Published online: 30 November 2017

hypermutation induced by activation-induced cytosine deaminase [5]. Through clonal affinity selection for enhanced antigen binding, somatic hypermutation-mediated variation further increases diversity of the mature B-cell repertoire [5, 6]. The recent advancement in next-generation sequencing (NGS) technologies has allowed us to generate millions of BCR cDNA sequence reads and characterize BCR repertoires comprehensively and quantitatively in a single experiment. In our previous study, we identified several abnormally spliced or rearranged, and potentially novel T-cell receptor (TCR) transcripts by NGS [7]. Hence, a method to comprehensively analyze BCR repertoires including potential novel sequences is important. These techniques are being applied in several research areas, including characterization and/or monitoring of B-cell repertoires in various types of disease, such as food allergy as well as autoimmune diseases, infectious diseases, and cancer.

Food allergy affects ~2–10% of children in the Unites States [8–10]. Peanut allergy is one of the most common and serious types of food allergy, and its prevalence has been increasing [11, 12]. Allergic reactions are characterized by various kind of symptoms, such as skin (rashes, angioedema), gastrointestinal (vomiting, diarrhea), respiratory (coughing, sneezing), and circulatory symptoms (cardiovascular collapse). Sometimes peanut allergy can lead to anaphylaxis, a life-threatening condition. Although the adaptive and innate immune systems were reported to be responsible for peanut allergic reactions, the detailed molecular mechanisms of peanut allergic reactions largely remain unclear. Current treatment options are limited to strict peanut avoidance and preparation to treat an accidental exposure with epinephrine. Recently, oral immunotherapy (OIT) has become a promising experimental treatment for peanut allergy where gradually increasing the amount of peanut intake for many months is expected to induce desensitization.

In this study, we have developed a novel method and successfully characterized the B-cell repertoires by sequencing BCR cDNAs, and obtained the comprehensive information of BCR repertoire using blood samples obtained before and after receiving peanut OIT or placebo. The data generated in this manuscript demonstrates this kind of approach is a very promising tool to better understanding the characteristics or changes of immune repertoires during the course of various immune-related diseases as well as in the course of the treatments by drugs or desensitization.

Materials and methods

Patient samples

This study was approved by the Institutional Review Board at the University of North Carolina-Chapel Hills and the University of Chicago, and written informed consent was obtained. In this study, we analyzed the samples obtained in a randomized controlled study of peanut OIT reported previously [13]. Among them, we studied 12 subjects who received peanut OIT and 5 subjects received a placebo due to the sample limitation. Detail information of study subjects is summarized in Supplementary Table 1. Briefly, subjects (age 3–11 years) with a clinical history of reaction to peanut within 60 min of ingestion, a peanut CAP-FEIA >15 kU/L (Phadia AB; Pharmacia, Inc, Uppsala, Sweden) or >7 kU/L if a significant reaction occurred within 6 months of enrollment, and a positive skin prick test (>3 mm of negative control) were enrolled, and received OIT in the form of peanut flour or placebo. The initial day escalation began with subjects receiving 0.1 mg of peanut protein, then doses were approximately doubled every 30 min to reach the target 6 mg dose. Subjects ingested the dose tolerated on initial escalation for 2 weeks then returned to the research unit for build-up dosing. Every 2 weeks, subjects ingested larger doses until the 4000 mg maintenance dose was reached. The maintenance dose of peanut flour was ingested for the remainder of the study until the final oral food challenges, in which increasing doses of peanut flour every 10–20 min up to a cumulative dose of 5000 mg were given to assess desensitization. Peripheral blood samples were obtained before initial treatment (pre-OIT) and during OIT dosing, typically at the time of final oral food challenge, unless subjects withdrew from the study then it was collected at the last visit). Peripheral blood mononuclear cells (PBMCs) were isolated using Ficoll-based density separation (LymphoH; Atlanta Biologicals, Lawrenceville, GA), and 1 million PBMCs were cultured in RPMI-1640 supplemented with 10% autologous plasma at 37 °C in a 5% CO_2 humidified atmosphere for 72 h in the presence of 200 μg/mL crude peanut extract or media alone. We further analyzed the samples cultured with media alone in this study.

Gene expression assay

To quantify the mRNA expression levels of each *BCR* isotype, we designed specific primers corresponding to parts of constant (C) exon of each *BCR* as shown in Supplementary Table 2. RNeasy mini kit (Qiagen, Valencia, CA, USA) was used to extract total RNA from the cultured PBMCs. After checking the quality of RNA using TapeStation 2100 (Agilent Technologies, Santa Clara, CA, USA), cDNA was first synthesized from 1 μg of total RNA using SMART cDNA library construction kit (Clontech Laboratories, Mountain View, CA, USA). A common 5′ rapid amplification of cDNA end (5′ RACE) adapter was added at 5′ ends of cDNAs during the reverse transcription reaction. Real-time quantitative PCR (qPCR) was performed using

SYBR Select Master Mix (Thermo Fisher Scientific, Carlsbad, CA, USA) on an ABI ViiA7 system (Thermo Fisher Scientific) to quantify the expression of *IGHA*, *IGHD*, *IGHE*, *IGHG*, *IGHM*, *IGK*, and *IGL* transcripts. To adjust the PCR efficiency bias of each assay, we used concentration-known standards, which were prepared by subcloning of each PCR amplicon into pCR2.1/TOPO vector (Thermo Fisher Scientific). All transcript expression levels were normalized to the expression of *GAPDH* (Hs02758991_g1) as the housekeeping gene. We estimated the expression levels of subclasses of IGHs, such as *IGHG4*, based on the expression levels of IGHs and the proportion of each subclass obtained from the BCR sequencing data described as below.

Library preparation for BCR sequencing

The library preparation method was basically same as the TCR sequencing method we previously reported [7, 14]. First-strand cDNA described above were used for BCR sequencing analysis. PCR was designed to separately amplify each BCR isotype using a forward primer for the 5′ RACE adapter and the reverse primers for the C regions of each BCR isotype as shown in Supplementary Table 3. The PCR reaction was performed as follows: 94 °C for 3 min, followed by 20 cycles of 94 °C for 30 s, 65 °C for 30 s, and 72 °C for 1 min. Nextera XT Index kit (Illumina, San Diego, CA, USA) was used to add Illumina sequence adapters with barcode sequences to generate multiplexed sequencing libraries, allowing the sequencing of multiple samples in a single experiment. The PCR condition was as follows: 95 °C for 3 min; 8 cycles of 95 °C for 30 s, 55 °C for 30 s, and 72 °C for 30 s; and a final extension at 72 °C for 5 min. Multiple dual indexed samples were pooled in a single library, and were sequenced by 300-bp paired-end reads on the Illumina MiSeq platform, using MiSeq Reagent v3 600-cycels kit (Illumina).

BCR sequence analysis

To analyze BCR sequencing data, we developed an algorithm of V(D)J decomposition with soft-clipping, Bcrip, similar to Tcrip [7], which we previously developed for TCR repertoire analysis. Each sequencing read was mapped to the V, J, and C reference sequences of each BCR isotype obtained from IMGT/GENE-DB [15, 16] (www.imgt.org) using the Bowtie2 aligner (Version 2.1.0) with -local option [17], which allows the soft-clipped (trimmed) mapping at the one or both ends of reads. Reads in which V, J, and C segments were found in a proper order are characterized by identifiers of the reference sequences, deletion of V, (D), and J segments and inserted nucleotides. For IGHs, a D segment was searched by scoring similarities of subsequences of the junction sequences to either of the D segment references with a sliding window method [18]. After decomposition of sequencing reads into V, (D), and J segments, CDR3 defined between second conserved cysteine in the 3′ portion of the V segment and the conserved phenylalanine (or tryptophan) in the 5′ portion of the J segment were searched using six nucleotides corresponding to the conserved amino acid and a followed amino acid in the reference sequence. When nucleotide sequences of both conserved amino acids were identified within the same reading frame, the amino acid sequence of the CDR3 was identified. If the V, J, and C segments were not properly identified in a read, especially in which J and C exons were identified but not a V exon, we remapped the unmapped parts of those reads to the IGHs, IGK, and IGL genomic sequences in GRCh38 reference.

Statistical analysis

The diversity index (DI) of unique V(D)J combinations with CDR3 sequences was calculated using inverse Simpson's diversity index formula: $\mathrm{DI} = \left[\frac{\sum_{i=1}^{K} n_i(n_i-1)}{N(N-1)}\right]^{-1}$, where K is the total number of clonotypes, n_i is the number of i-th clonotype sequences, and N is the total number of sequences for which each clonotype is determined [19]. When DIs were not able to be calculated by a very small number of the VJC mapped reads in IGHE sequencing, we assigned DI to be 0. The Wilcoxon signed-rank test was used to compare the DI or gene expression before and after either OIT or placebo treatments. These statistical tests were conducted using Prism software, version 6.0 (GraphPad, La Jolla, CA, USA). A P value of <0.05 was considered statistically significant.

Results

Establishment of BCR sequencing platform using Illumina sequencer

To develop the BCR sequencing platform, we first used a peripheral blood sample (8 mL) from a healthy donor. We basically applied the similar concept to our TCR sequencing method [7, 14], and modified it for BCR sequencing. Addition of 5′ RACE adapter at the 5′ end during cDNA synthesis from total RNA allows us to amplify each BCR isotype using a single forward common primer. The reverse primers were designed on the C regions corresponding to each of BCR isotypes (Supplementary Table 3). Fusion PCR were further applied to prepare sequencing amplicons by adding sequencing adapters and barcodes for read 1 and read 2 of Illumina sequencing, respectively, to the 5′ end of

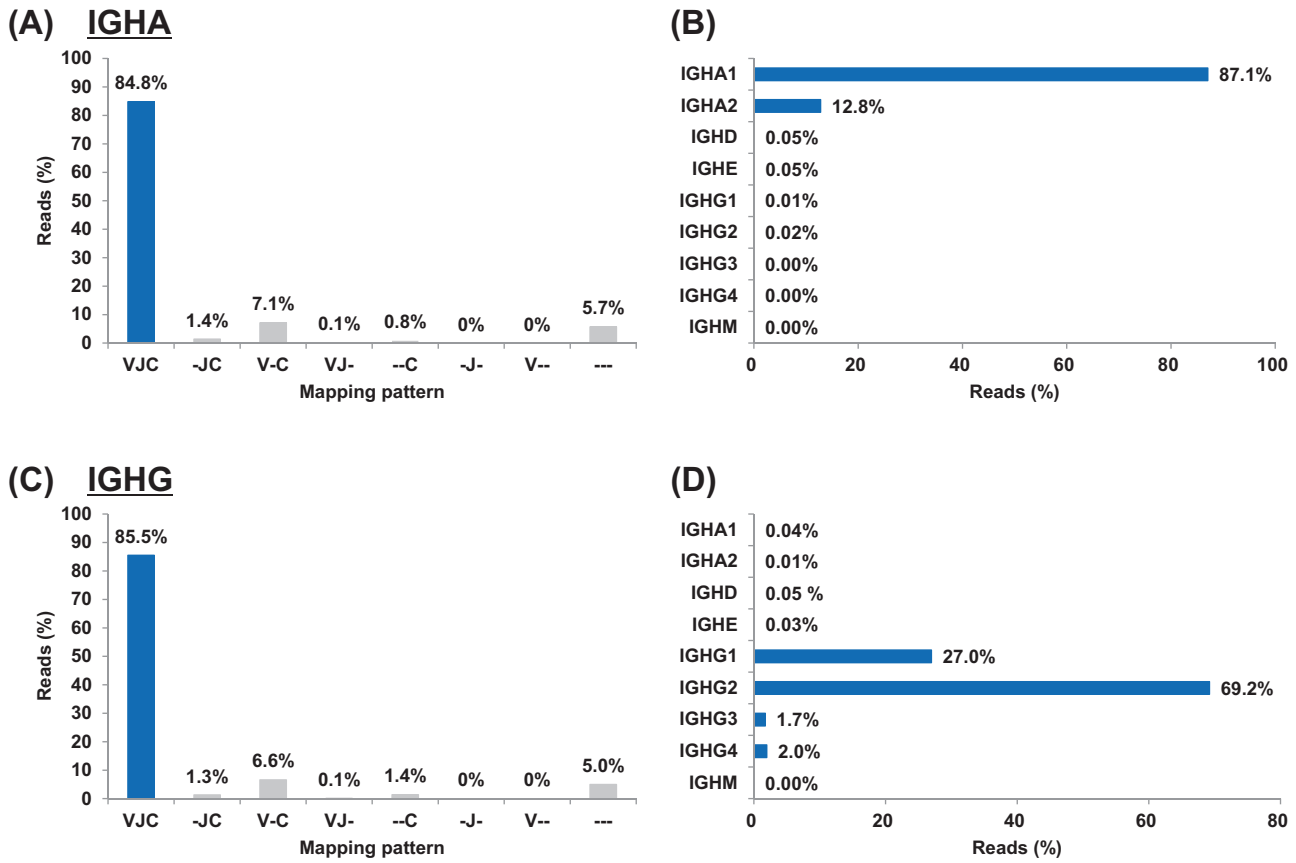


Fig. 1 Representative mapping results of BCR cDNA sequences from PBMCs of a healthy donor. Deep sequencing of BCR cDNA was performed using Illumina MiSeq with the paired-end 300-bp sequencing, and the sequencing reads were analyzed by Bcrip software. **a**, **c** Percentages of the reads mapped to variable (V), joining (J), and constant (C) regions in IGHA (**a**) and IGHG (**c**) reads. Detailed sequencing information is summarized in Table 1. **b**, **d** Mapping percentages of C regions of BCR in IGHA (**b**) and IGHG (**d**) reads. Detailed mapping information is summarized in Table 2

C regions and 3′ end of V regions of BCRs. After pooling these libraries, deep sequencing by 300-bp paired-end reads was then performed using Illumina MiSeq.

We obtained an average of 1,569,568 (782,935–2,609,687) sequencing reads for each isotype of BCRs and analyzed each of the sequencing reads by newly developed Bcrip software. Since *IGH* genes are located within the same gene locus on chromosome 14 and share the V and J exons, we analyzed the IGH reads using a reference set of IGHC, including C exons of five IGHs, whereas the IGK and IGL reads were separately analyzed using each reference of these two isotypes. We first examined the sequencing quality by calculating the mapping percentages of BCR reads. 84.8, 84.1, 42.3, 85.5, 85.6, 90.1, and 81.4% of sequence reads using specific primers for IGHA, IGHD, IGHE, IGHG, IGHM, IGK, and IGL, respectively, were identified to contain all of V, J, and C regions (Fig. 1 and Table 1). These mapping percentages are almost comparable with or slightly lower than the results, 91.5, 90.2, 43.9, 91.9, 87.6, 91.1, and 85.3%, obtained using another BCR analysis software, MiXCR [20]. Among the VJC-mapped reads of these IGHs, 98.3% or more of the sequencing reads were confirmed to contain the corresponding C regions (Fig. 1 and Table 2), which is comparable to MiXCR results (more than 98.3%). Moreover, compared to MiXCR, our algorithm provided highly concordant clonotype (V–J–C and CDR3) assignments although there were some discordance in the clones with <0.001% frequencies ($R \geq 0.89$ for ≥0.001% clones; Supplementary Fig. 1).

We also developed a qPCR method to quantify mRNA expression of each of IGHs, IGK, and IGL by designing the specific primers to their C regions (Supplementary Table 2). We confirmed the PCR efficiency bias of these assays was within only 1.8 times, which is corresponding to 0.87 PCR cycle, using concentration-known standards. The expression levels of IGHA, IGHD, IGHE, IGHG, IGHM, IGK, and IGL were calculated to be 1.5, 0.10, 0.0073, 2.5, 1.0, 5.5, and 3.8 amol/μL, respectively. The IGHE expression was confirmed to be significantly correlated with serum IgE levels ($R = 0.51$, $P = 0.038$; Supplementary Fig. 2).

Table 1 Mapping pattern of BCR reads in a healthy donor

Isotype	Number of reads (%)														
	Total	VJC		–JC		V–C		VJ–		–C		–J–		V–		—	
IGHA	1,202,818	1,020,233	(84.8%)	16,993	(1.4%)	85,904	(7.1%)	733	(0.061%)	9,734	(0.81%)	41	(0.0034%)	261	(0.022%)	68,919	(5.7%)
IGHD	1,228,405	1,032,486	(84.1%)	41,995	(3.4%)	48,516	(3.9%)	7,607	(0.62%)	12,513	(1.0%)	396	(0.032%)	860	(0.070%)	84,032	(6.8%)
IGHE	1,873,358	791,790	(42.3%)	5,481	(0.29%)	48,505	(2.6%)	835	(0.045%)	337,116	(18.0%)	32	(0.0017%)	2,617	(0.14%)	686,982	(36.7%)
IGHG	782,935	669,764	(85.5%)	10,192	(1.3%)	51,582	(6.6%)	416	(0.053%)	11,279	(1.4%)	24	(0.0031%)	261	(0.033%)	39,417	(5.0%)
IGHM	1,335,980	1,143,350	(85.6%)	33,765	(2.5%)	14,740	(1.1%)	1,068	(0.080%)	38,492	(2.9%)	9	(0.00067%)	212	(0.016%)	104,344	(7.8%)
IGK	2,609,687	2,350,637	(90.1%)	11,954	(0.46%)	93,484	(3.6%)	302	(0.012%)	15,600	(0.60%)	68	(0.0026%)	1,047	(0.040%)	136,595	(5.2%)
IGL	1,953,796	1,590,087	(81.4%)	5,196	(0.27%)	209,993	(10.7%)	998	(0.051%)	13,972	(0.72%)	14	(0.00072%)	1,436	(0.073%)	132,100	(6.8%)

V variable region, *J* joining region, *C* constant region, – not mapped

Unmapped sequence read analysis

Since we previously identified several abnormally spliced or rearranged TCR transcripts in the reads containing J and C segments but not V segments (called as "unmapped-JC reads") [7], we explored the unmapped reads in IGH sequences. In total, 1.7% of the IGH reads were classified into "unmapped-JC reads" (Table 1). As similar to unmapped-JC reads of TCRβ [7], we found those reads mapped to intronic sequences in the IGH locus. Some of the unmapped parts corresponded to either an intronic region adjacent to the J segment ("J Intron"; 25.4% of the unmapped-JC reads) or an intronic region adjacent to the D segment ("D Intron"; 13.2% of the unmapped-JC reads) (Fig. 2a). In addition, 5.0% of the "unmapped-JC" reads contained the sequences that were classified as "V Intron," which was defined to contain only exon 1 sequences corresponding to L-part of IGH V genes or sequences between two known V exons. The remaining reads were classified as "Unmapped" reads, which were mapped to the non-IGH locus or which were not mapped to any regions in the human genome reference sequence. Interestingly, there was a strong bias in the usage of J segments in those unmapped-JC reads; for example, IGHJ3 or IGHJ5 were frequently observed in the J intron reads, and IGHJ4 or IGHJ6 were frequently observed in the other unmapped-JC reads (Fig. 2b). We further investigated one possible novel exon candidate (an exonic sequence that has not been deposited in the database) showing 84.4% sequence identity to IGHV3-7 sequence as shown in Fig. 2c. When we analyzed nucleotide sequences of read 2 (from the other end of cDNA starting from the V region), these reads showed the highest sequence identity of 99.2% to IGHV3-21 exon 1, but the sequence identity of exon 2 of IGHV3-21 was as low as 61.0%. Since we confirmed that 110-bp overlapped sequences between read 1 and read 2 were 100% concordant, we can exclude a possibility of sequencing errors. Although further extensive analysis is required, our cDNA approach using primers in the constant region would be able to identify any novel exons that have not been deposited in the public databases.

BCR expression and sequencing of OIT samples

We then obtained PBMCs before and after OIT treatments from a total of 17 peanut allergic subjects, and analyzed expression levels and repertoires of each isotype of BCR. Among them, 12 patients received peanut flour and the remaining 5 patients received a placebo until the final oral food challenges. In this analysis, we focused on IGHA, IGHE, and IGHG, which have been reported to associate with food allergy [13, 21–25]. In the expression analysis using real-time qPCR, we observed the tendency of

Table 2 Percentage of reads correctly mapped to IGH constant region

PCR-amplified isotype	IGH constant region									Percentage of reads mapped to correct constant region
	IGHA1	IGHA2	IGHD	IGHE	IGHG1	IGHG2	IGHG3	IGHG4	IGHM	
IGHA	87.1%	12.8%	0.00%	0.05%	0.01%	0.02%	0.00%	0.00%	0.00%	99.9%
IGHD	0.02%	0.01%	99.9%	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	99.9%
IGHE	0.61%	0.07%	0.00%	98.3%	0.01%	0.91%	0.00%	0.01%	0.00%	98.3%
IGHG	0.04%	0.01%	0.00%	0.03%	27.0%	69.2%	1.72%	1.95%	0.00%	99.9%
IGHM	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	100.0%	100.0%

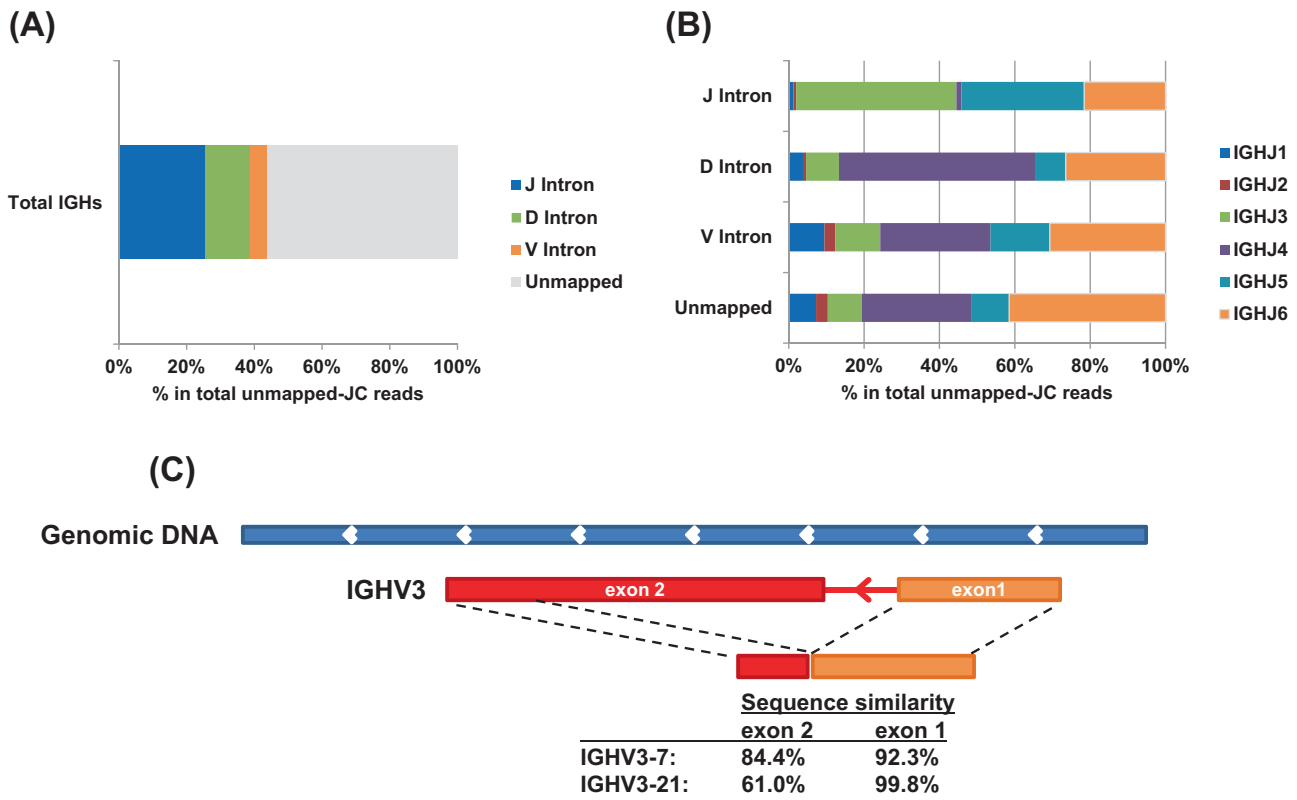


Fig. 2 The analysis of unmapped BCR reads. **a** The proportion of the patterns among the unmapped-JC reads. J Intron: the reads containing an intronic region adjacent to the J segment, D Intron: the reads containing an intronic region adjacent to the D segment, V Intron: the reads containing the sequences within IGH V locus including the reads containing a part of V exons or the sequences between two known V exons, Unmapped: the reads which did not map to IGH locus. **b** The J segment usage in the unmapped-JC reads of IGH. **c** The structure of potentially novel IGHV3 exon similar to IGHV3-7 an IGHV3-21

decrease of *IGHE* expression after OIT compared to before OIT ($P = 0.052$) but not in placebo group ($P = 0.63$) (Fig. 3), implying the decrease in the numbers of IgE-producing cells during OIT desensitization. The mRNA expression levels of *IGHA* and *IGHG* did not differ between samples before and after OIT or placebo treatments. We also estimate the expression levels of subclasses of IGHs using the results of qPCR and mapping data, and compared them between before and after OIT. IGHG3 showed significant decrease after OIT but not placebo (median, 0.014–0.0056; $P = 0.021$), while IGHG4 showed an increase tendency after OIT compared to the baseline (median, 0.00018–0.0041; $P = 0.27$; Supplementary Fig. 3).

Through BCR deep sequencing, we obtained total sequence reads of 594,633 ± 301,523 for IGHA, 455,989 ± 420,091 for IGHE, and 579,983 ± 295,685 for IGHG, and identified average 39,228, 770, and 48,927 unique CDR3 clonotypes for each IGH, respectively (Table 1). DIs for IGHA and IGHG were significantly decreased after peanut OIT, compared to that before the treatment in the OIT group ($P = 0.034$ and $P = 0.027$ for IGHA and IGHG, respectively), but this tendency was not observed in the placebo

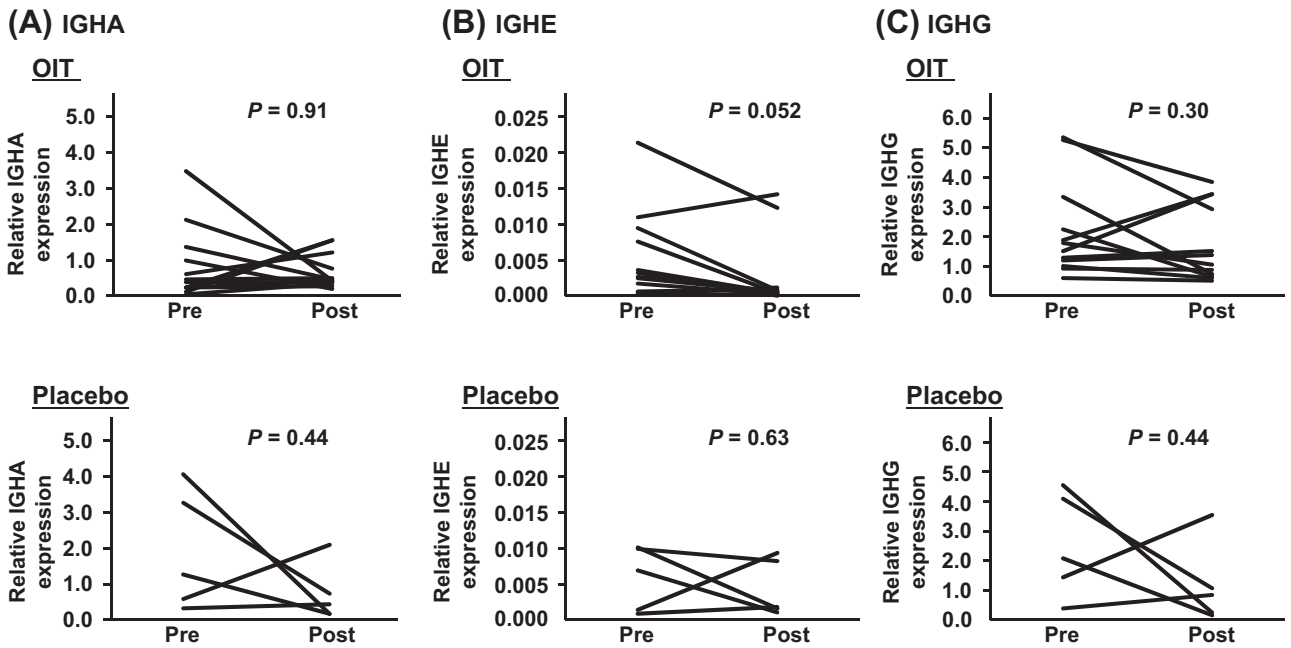


Fig. 3 Changes of the expression levels of *BCR* mRNA during the oral immunotherapy for peanuts allergy patients. **a** The expression levels of *IGHA* mRNA in the oral immunotherapy group (upper) and placebo group (lower). **b** The expression levels of *IGHE* mRNA in the oral immunotherapy group (upper) and placebo group (lower). **c** The expression levels of *IGHG* mRNA in the oral immunotherapy group (upper) and placebo group (lower). Statistical significance was examined using the Wilcoxon signed-rank test

group ($P = 0.81$ and $P = 0.31$ for IGHA and IGHG, respectively) (Fig. 4). DIs for IGHE were not statistically different between before and after OIT ($P = 0.68$) probably because the numbers of cells expressing IgE were very low and in some cases we did not have sufficient information to calculate DIs. We summarized the clonotypes of IGHA, IGHE, and IGHG in Supplementary Tables 4–6, and further analyzed common clones which observed in multiple subjects. While there was no commonly observed clone with frequency above 0.01% among at least three subjects in IGHA and IGHG, we identified one IGHE clone (IGHV2-5, IGHJ5, CAHGQVHQWLGDVHWFGPW) which was commonly detected in 2 out of 12 pre-treated OIT and all of 5 pre-treated placebo samples (Supplementary Table 5). However, this clone was also detected in post-treated samples in both OIT and placebo groups.

Discussion

In this study, we reported the establishment of a novel and effective method using NGS to comprehensively characterize the human BCR repertoire, including the information of V(D)J combination and CDR3 sequences, and investigated the application of this platform to monitor B-cell repertoire changes in peanut allergy patients receiving OIT. This platform can characterize all of seven isotypes (including five IGHs and two IG light chains) of the BCR simultaneously and should be applied for a wide range of research areas that require deep full-length antibody profiling, such as monitoring of immune responses during autoimmune diseases, in which B cells are known to play crucial roles, leading to disease-causing immunoglobulins, as well as examining host–immune responses during infectious diseases.

In our system, we have applied 5′ RACE-PCR method to prepare the sequence library of each BCR isotype using a single primer set, which can minimize PCR amplification bias when compared with multiplex PCR amplification method in which the annealing efficiency of different PCR primers may cause uneven amplification [26], but we are also unable to completely exclude the possibility of PCR amplification bias due to the differences in the amplicon size of each BCR. We confirmed our algorithm provided very highly concordant clonotype (V–J–C and CDR3) assignments compared to MiXCR [20], although there were some discordance, when the frequencies clones are lower than 0.001% (Supplementary Fig. 1). We also found some discordance due to assignment errors of V exons for which V exons had very high-sequence similarities (particularly in the cases that the sequence read lengths are not sufficient enough for accurate mapping), in both Bcrip and MiXCR [20]. Further optimization of the parameters as well as further update of sequencing technology, including sequence read length and quality, will be required to improve accurate clonotype assignments. In addition, our

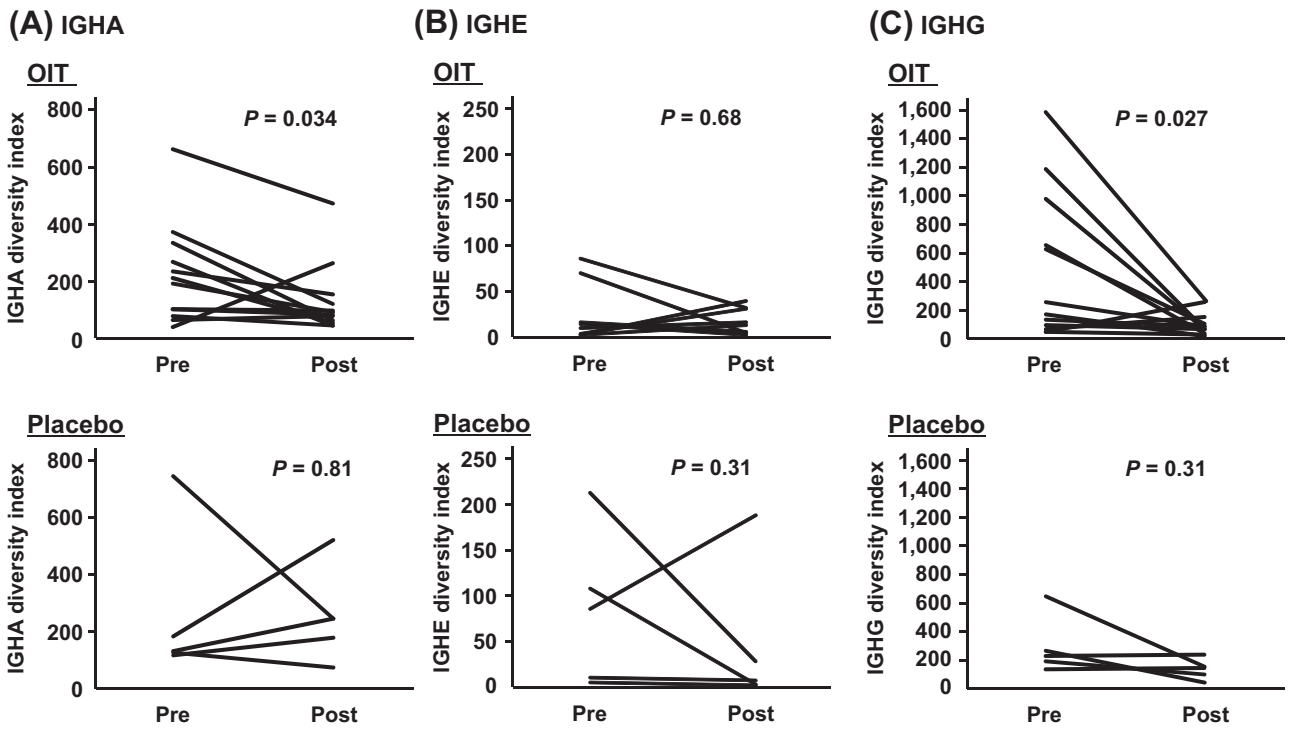


Fig. 4 Changes of BCR repertoires during the oral immunotherapy for peanuts allergy patients. **a** IGHA diversity index in the oral immunotherapy group (upper) and placebo group (lower). **b** IGHE diversity index in the oral immunotherapy group (upper) and placebo group (lower). **c** IGHG diversity index in the oral immunotherapy group (upper) and placebo group (lower). Statistical significance was examined using the Wilcoxon signed-rank test

method has potential to obtain the full-length sequences of BCR transcripts, including those resulted from unexpectedly aberrant splicing or rearrangement, and those containing potentially novel exons that are not reported in the human genome reference database, whereas other previously reported algorithms such as IMGT/HighV-QUEST [27], IgBLAST [28], and MiXCR [20], use the information of only V(D)J regions. Indeed, we identified the reads that unexpectedly mapped to IGHV region with a deletion in a known V segment as shown in Fig. 2, although further experiment is required to confirm whether this transcript was produced by aberrant splicing or resulted from a novel sequence. One of the potential limitations arises from difference in BCR mRNA expression levels in individual B cells, and this makes it challenging to deduce actual B-cell clonotype, compared to the sequencing methods using genomic DNA as a template. However, our method is likely to more precisely reflect the amount of individual immunoglobulin transcripts. Another potential limitation was that Bcrip did not implement somatic hypermutation analysis, because it is still challenging to accurately identify somatic hypermutations when considering the sequence quality and error rate of the current NGS. To achieve this, adding unique molecular barcodes, which are introduced as random oligonucleotides at the cDNA synthesis step or a very early amplification step of library preparation, would be beneficial solution [29], but further optimization would be necessary.

In the analysis of expression and repertoire of IGHA, IGHE, and IGHG in the peripheral blood samples from peanut allergic patients who underwent either OIT or placebo, we found that *IGHE* expression was decreased after OIT compared to before OIT ($P = 0.052$, Fig. 3), which likely corresponds to the reduction of the numbers of IgE-producing cells. This result was concordant with the results based on serum peanut-specific IgE levels ($P = 0.034$ and 0.62, respectively, in OIT and placebo groups), and also several reports demonstrating that serum peanut-specific IgE levels were significantly decreased by the treatment of OIT [13, 21, 24, 25]. We observed one common IGHE clone among two pre-treated OIT and five pre-treated placebo samples (Supplementary Table 5), but this clone was still retained in the samples after OIT and placebo treatment. In this study, we used PBMCs after 72-h culture obtained in our previous study [13] and it might affect BCR repertoire. Further analysis will be needed to clarify whether this clone is associated with peanut allergy. We also observed that mRNA expression levels of total IGHA and IGHG were not significantly different before and after OIT (Fig. 3) while the DIs of BCR repertoire were significantly decreased after OIT ($P = 0.034$ and $P = 0.027$, Fig. 4), indicating that oligoclonal expansion of B cells with certain BCR sequences.

The expression levels of IGHG4 estimated from the results of expression and mapping data showed a tendency of increase after OIT compared to the baseline (median, 0.00018–0.0041; $P = 0.27$; Supplementary Fig. 3). These results were supported by the reports that Ara h 2- or egg-specific IgA and IgG in serum were increased at as early as 7 days or 3 months after starting the OIT [21, 30]. In addition, the CDR3 sequences reported as Ara h 2-specific IgA and IgG were not observed in our analysis (Supplementary Tables 4–6) [21]. As small samples size is a limitation of this study, further experiments with lager sample size are guaranteed for validation of these results. In this study, since it was also difficult to figure out which BCR sequences are critically and clinically important in development of peanut allergy due to the limited number of samples, further experiments with more time points, including especially early time points after starting OIT, will be required to analyze the dynamics of B-cell repertoire changes during OIT.

In summary, we have developed a new high-throughput BCR analysis system using the NGS platform with a novel algorithm, Bcrip, which has allowed us to obtain the comprehensive B-cell repertoire data. This methodology enables us to provide a detailed antibody profiling in patients with various types of diseases including food allergy, autoimmune diseases, and infectious diseases.

Acknowledgements The authors thank Makiko Harada for excellent technical assistance. The super-computing resource was provided by Human Genome Center, the Institute of Medical Science, the University of Tokyo (http://sc.hgc.jp/shirokane.html). This work was supported by the funding from OncoTherapy Science, Inc, and the National Institutes for Health (Grant no. 5R01-AI068074-09).

Author contributions KK designed and performed experiments, analyzed the data, and wrote the manuscript. THM and PYY performed experiments. RY, SI, and SM developed algorithm and analyzed the data. MK, KO, and AWB collected and prepared the samples, and performed experiments. YN planned and supervised the entire project, and edited the manuscript.

Compliance with Ethical Standards

Conflict of interest YN is a stockholder and an adviser of OncoTherapy Science Inc. PYY is employee of OncoTherapy Science Inc. The remaining authors declare no conflict of interest.

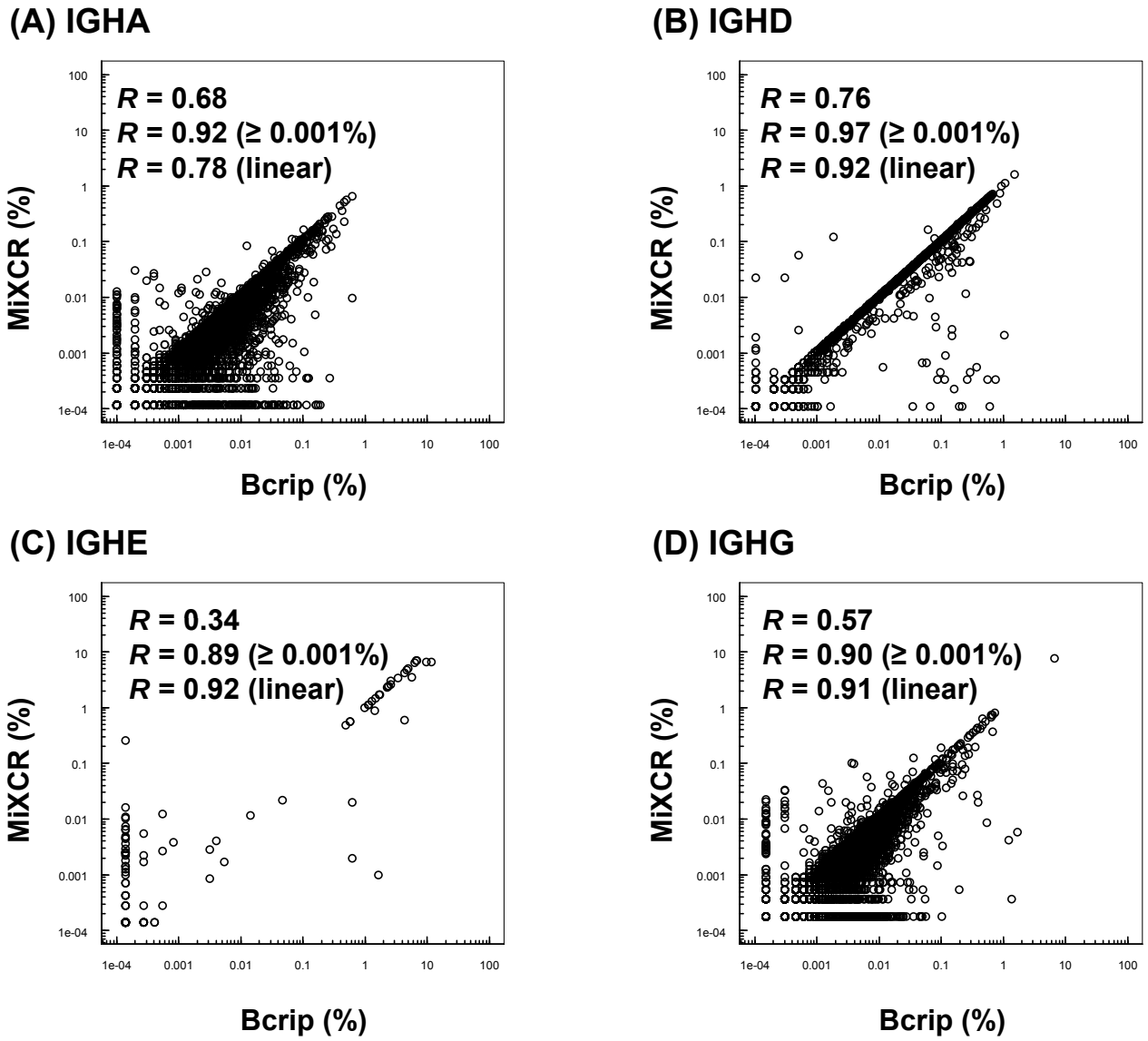
Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. If you remix, transform, or build upon this article or a part thereof, you must distribute your contributions under the same license as the original. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/.

References

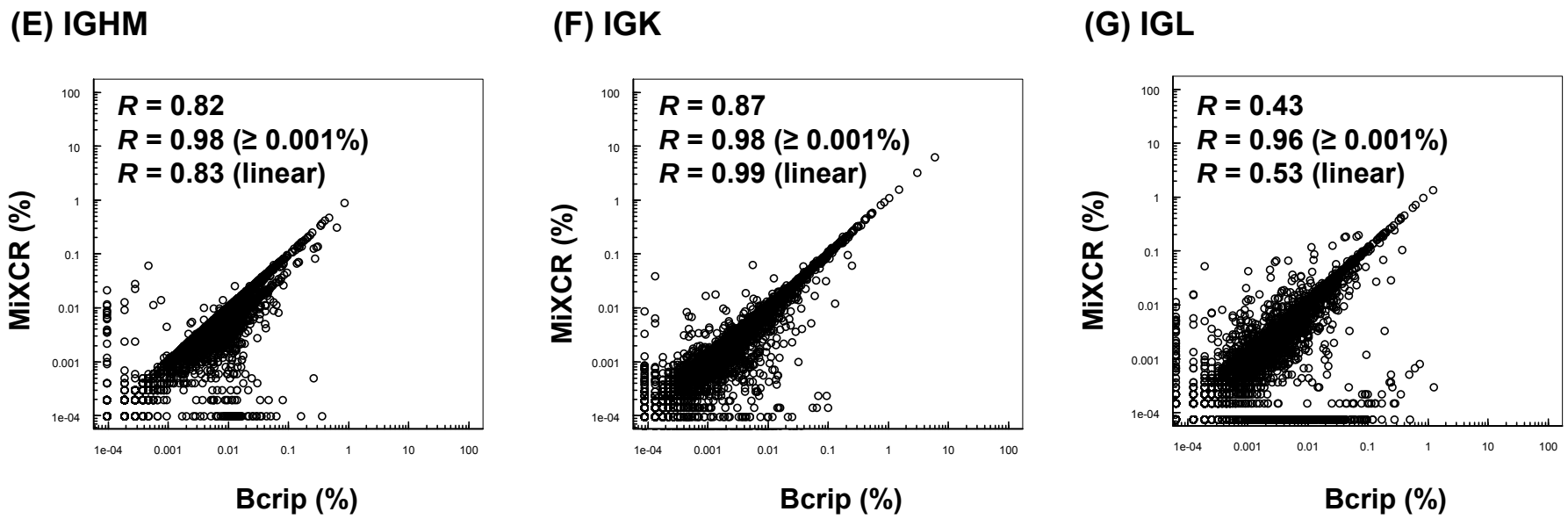
1. Tonegawa S. Somatic generation of antibody diversity. Nature. 1983;302:575–81.
2. Woof JM, Burton DR. Human antibody-Fc receptor interactions illuminated by crystal structures. Nat Rev Immunol. 2004;4:89–99.
3. Jung D, Giallourakis C, Mostoslavsky R, Alt FW. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. Annu Rev Immunol. 2006;24:541–70.
4. Schatz DG, Swanson PC. V(D)J recombination: mechanisms of initiation. Annu Rev Genet. 2011;45:167–202.
5. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell. 2000;102:553–63.
6. Batrak V, Blagodatski A, Buerstedde JM. Understanding the immunoglobulin locus specificity of hypermutation. Methods Mol Biol. 2011;745:311–26.
7. Fang H, Yamaguchi R, Liu X, Daigo Y, Yew PY, Tanikawa C, et al. Quantitative T cell repertoire analysis by deep cDNA sequencing of T cell receptor alpha and beta chains using next-generation sequencing (NGS). Oncoimmunology. 2014;3:e968467.
8. Boyce JA, Assa'ad A, Burks AW, Jones SM, Sampson HA, Wood RA, et al. Guidelines for the diagnosis and management of food allergy in the United States: report of the NIAID-sponsored expert panel. J Allergy Clin Immunol. 2010;126:S1–58.
9. Gupta RS, Springston EE, Warrier MR, Smith B, Kumar R, Pongracic J, et al. The prevalence, severity, and distribution of childhood food allergy in the United States. Pediatrics. 2011;128:e9–17.
10. Sicherer SH, Sampson HA. Food allergy: epidemiology, pathogenesis, diagnosis, and treatment. J Allergy Clin Immunol. 2014;133:291–307.
11. Sicherer SH, Munoz-Furlong A, Godbold JH, Sampson HA. US prevalence of self-reported peanut, tree nut, and sesame allergy: 11-year follow-up. J Allergy Clin Immunol. 2010;125:1322–6.
12. Osborne NJ, Koplin JJ, Martin PE, Gurrin LC, Lowe AJ, Matheson MC, et al. Prevalence of challenge-proven IgE-mediated food allergy using population-based sampling and predetermined challenge criteria in infants. J Allergy Clin Immunol. 2011;127:668–76.
13. Varshney P, Jones SM, Scurlock AM, Perry TT, Kemper A, Steele P, et al. A randomized controlled study of peanut oral immunotherapy: clinical desensitization and modulation of the allergic response. J Allergy Clin Immunol. 2011;127:654–60.
14. Choudhury N, Kiyotani K, Yap KL, Campanile A, Antic T, Yew PY, et al. Low T-cell receptor diversity, high somatic mutation burden, and high neoantigen load as predictors of clinical outcome in muscle-invasive bladder cancer. Eur Urol Focus. 2015;2:445–52.
15. Giudicelli V, Chaume D, Lefranc MP. IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res. 2005;33:D256–D261.
16. Lefranc MP, Giudicelli V, Kaas Q, Duprat E, Jabado-Michaloud J, Scaviner D, et al. IMGT, the international ImMunoGeneTics information system. Nucleic Acids Res. 2005;33:D593–D597.

17. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
18. Yousfi Monod M, Giudicelli V, Chaume D, Lefranc MP. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. Bioinformatics. 2004;20(Suppl 1):i379–i385.
19. Venturi V, Kedzierska K, Turner SJ, Doherty PC, Davenport MP. Methods for comparing the diversity of samples of the T cell receptor repertoire. J Immunol Methods. 2007;321:182–95.
20. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods. 2015;12:380–1.
21. Patil SU, Ogunniyi AO, Calatroni A, Tadigotla VR, Ruiter B, Ma A, et al. Peanut oral immunotherapy transiently expands circulating Ara h 2-specific B cells with a homologous repertoire in unrelated subjects. J Allergy Clin Immunol. 2015;136:125–34. e112
22. Jones SM, Pons L, Roberts JL, Scurlock AM, Perry TT, Kulis M, et al. Clinical efficacy and immune regulation with peanut oral immunotherapy. J Allergy Clin Immunol. 2009;124:292–300. 300 e291-297
23. Hoh RA, Joshi SA, Liu Y, Wang C, Roskin KM, Lee JY, et al. Single B-cell deconvolution of peanut-specific antibody responses in allergic patients. J Allergy Clin Immunol. 2016;137:157–67.
24. Vickery BP, Berglund JP, Burk CM, Fine JP, Kim EH, Kim JI, et al. Early oral immunotherapy in peanut-allergic preschool children is safe and highly effective. J Allergy Clin Immunol. 2017;139:173–81.e178
25. Tang ML, Ponsonby AL, Orsini F, Tey D, Robinson M, Su EL, et al. Administration of a probiotic with peanut oral immunotherapy: a randomized trial. J Allergy Clin Immunol. 2015;135:737–44.e738
26. Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. Nat Commun. 2013;4:2680.
27. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res. 2008;36:W503–508.
28. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. 2013;41:W34–40.
29. Kivioja T, Vaharautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods. 2011;9:72–74.
30. Wright BL, Kulis M, Orgel KA, Burks AW, Dawson P, Henning AK, et al. Component-resolved analysis of IgA, IgE, and IgG4 during egg OIT identifies markers associated with sustained unresponsiveness. Allergy. 2016;71:1552–60.

Supplementary Figure 1, Kiyotani K, *et al.*

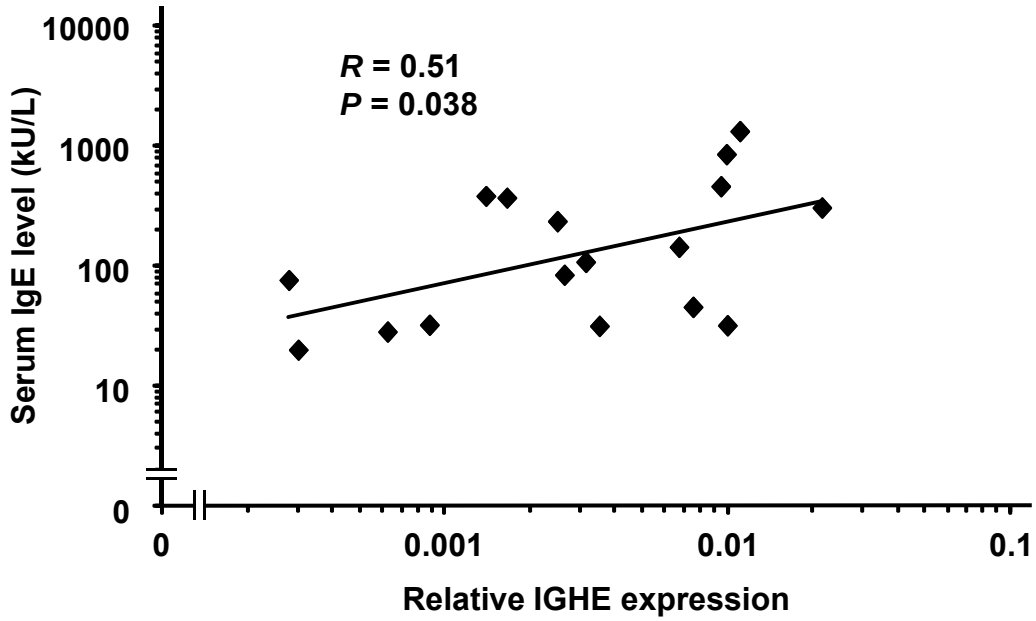


Supplementary Figure 1, Kiyotani K, *et al. (continued)*



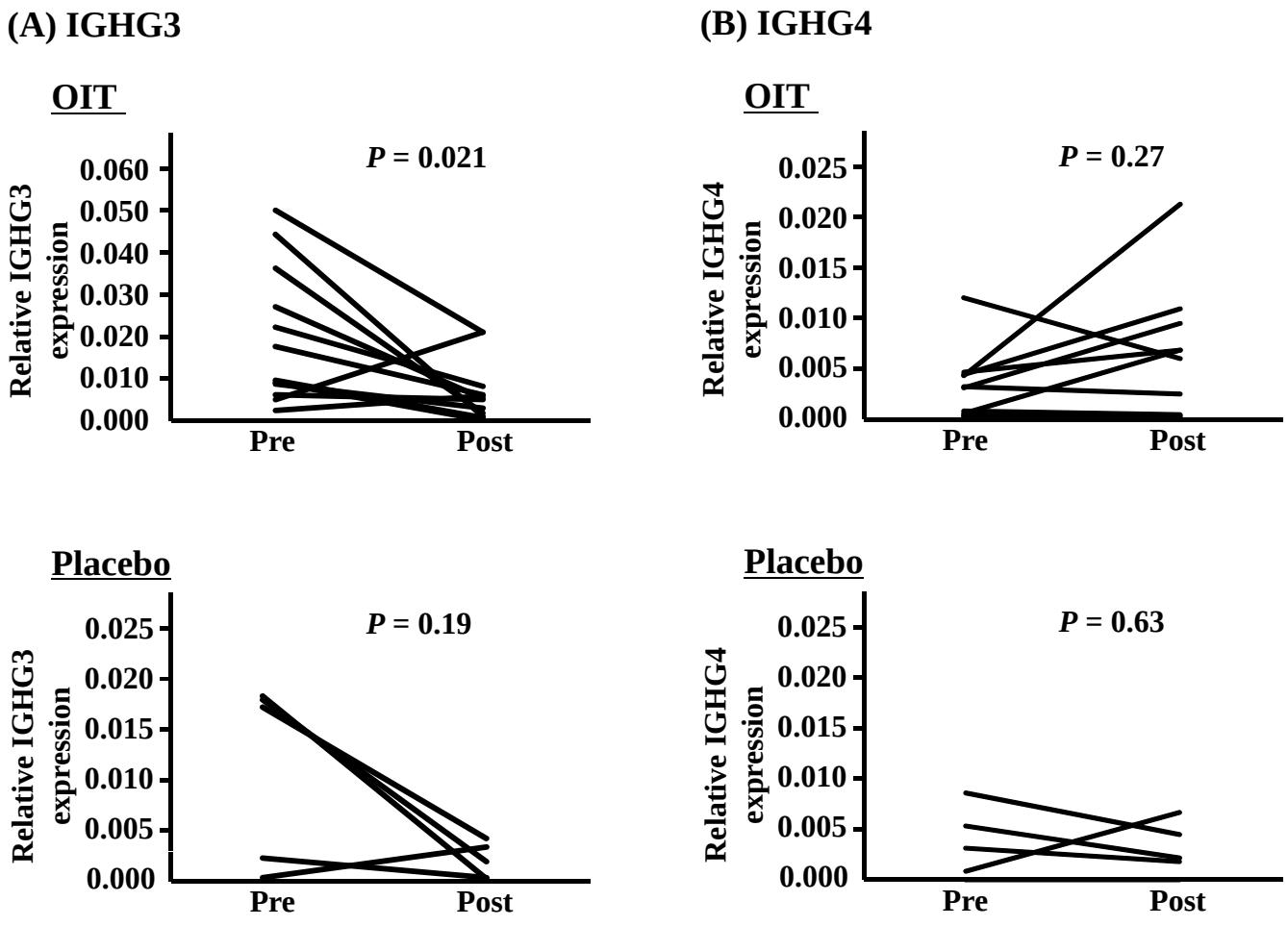
Supplementary Figure 1.
Comparisons of clonotype frequencies between Bcrip and MiXCR.
Pearson's correlation coefficients (*R*) were calculated at log-scale, at log-scale for only clones with ≥0.001%, and at linear-scale for IGHA (**A**), IGHD (**B**), IGHE (**C**), IGHG (**D**), IGHM (**E**), IGK (**F**), and IGL (**G**).

Supplementary Figure 2, Kiyotani K, *et al.*



Supplementary Figure 2.
Significant correlation between expression levels of IGHE mRNA and serum IgE levels.
Pearson's correlation coefficients (*R*) were calculated.

Supplementary Figure 3, Kiyotani K, *et al.*



Supplementary Figure 3.
Changes of the expression levels of *IGHG3* and *IGHG4* mRNA during the oral immunotherapy for peanuts allergy patients.
The expression levels of *IGHG3* (A) and *IGHG4* (B) mRNA in the oral immunotherapy group (upper) and placebo group (lower). Statistical significance was examined using the Wilcoxon signed-rank test.

Supplementary Table 1. Characteristics of study participants

Subject ID	Treatment	Age (years)	Gender	Duration of OIT (months)	Outcome of OIT
PMIT19	OIT	4	Male	47	Withdrew before challenge
PMIT20	OIT	5	Female	55	Passed 5 g challenge
PMIT21	OIT	3	Male	33	Passed 5 g challenge
PMIT22	OIT	6	Female	41	Passed 5 g challenge
PMIT24	OIT	4	Male	28	Withdrew before challenge
PMIT25	OIT	4	Female	25	Withdrew before challenge
PMIT14	OIT	6	Male	26	Withdrew before challenge
PMIT15	OIT	4	Female	63	Passed 5 g challenge
PMIT18	OIT	4	Female	48	Passed 5 g challenge
PMIT12	OIT	5	Male	64	Passed 5 g challenge
PMIT11	OIT	4	Female	34	Withdrew before challenge
PMIT26	OIT	5	Female	61	Passed 5 g challenge
SLB201	Placebo	11	Female	NA	Failed
SLB204	Placebo	7	Male	NA	Failed
SLB205	Placebo	11	Male	NA	Failed
SLB210	Placebo	9	Female	NA	Failed
SLB212	Placebo	8	Female	NA	Failed

OIT: oral immunotherapy. NA: not applicable.

Supplementary Table 2. Primer sequences for quantitative RT-PCR

Primer	Sequence (5' to 3')
qIGHA-F	AGTACCTGACTTGGGCATCC
qIGHA-R	CCCACCATGCAGGAGAAGG
qIGHD-F	GGAACAAGAAGAGAGAGAGAC
qIGHD-R	AGCAGGTGAAGGTGGCTTTG
qIGHE-F2	AGATGTTCACCTGCCGTGTG
qIGHE-R2	CGCAGGACGACTGTAAGATC
qIGHG-F3	GCCTGGTCAAGGACTACTTC
qIGHG-R3	CACGCTGCTGAGGGAGTAG
qIGHM-F	ACCATCAAAGAGAGCGACTG
qIGHM-R	CGGATGGCTGTGTCTTGATC
qIGKC-F	AACTGCCTCTGTTGTGTGCC
qIGKC-R	TTGCTGTCCTGCTCTGTGAC
qIGLC-F	CTCCAAACARAGCAACAACAAG
qIGLC-R	TGTCTTCTCCACGGTGCTC

Supplementary Table 3. Primer sequences for *BCR* amplification

Primer	Sequence (5' to 3')
5'-F	[overhang adapter 2] - TATCAACGCAGAGTGGCCAT
Heavy chains	
IGHA-R	[overhang adapter 1] - $(N_{-1})_{0-5}$ - GCGAYGACCACGTTCCCATCT
IGHD-R	[overhang adapter 1] - $(N_{-1})_{0-5}$ - TGTCTGCACCCTGATATGATG
IGHE-R	[overhang adapter 1] - $(N_{-1})_{0-5}$ - CCCAGAGTCACGGAGGTG
IGHG-R	[overhang adapter 1] - $(N_{-1})_{0-5}$ - GAAGTAGTCCTTGACCAGGCA
IGHM-R	[overhang adapter 1] - $(N_{-1})_{0-5}$ - GTGATGGAGTCGGGAAGGAAG
Light chains	
IGKC-R	[overhang adapter 1] - $(N_{-1})_{0-5}$ - GAGGCAGTTCCAGATTTCAA
IGLC-R	[overhang adapter 1] - $(N_{-1})_{0-5}$ - GCTTGRAGCTCCTCAGAGG

Illumina overhang adapter 2 (for Index 2; i5):
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
Illumina overhang adapter 1 (for Index 1; i7):
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG
N_{-1}: random nucleotides to increase the sequence diversity for cluster differentiation in the Illumina sequencer