

The composition and variation of the BCR CDR3s in gastric cancer

SONG LIU^{1*}, YING ZHU^{1*}, LIE-WEN LIN², SHUN-KAI DING², XIAO-CONG LIN¹,
KE-LI ZHONG², KAI PAN^{2*} and YONG DAI¹

¹Clinical Medical Research Center; ²Department of Gastrointestinal Surgery,
The Second Clinical Medical College of Jinan University, Shenzhen People's Hospital,
Shenzhen, Guangdong 518020, P.R. China

Received November 21, 2017; Accepted April 26, 2018

DOI: 10.3892/ol.2018.8677

Abstract. Gastric cancer (GC) is the fourth most common type of cancer and the second most common cause of cancer-associated mortality worldwide. B cell-associated autoantibodies against tumor-associated antigens are attractive biomarkers for the development of noninvasive serological tests for the early detection of cancer. This is due to their specificity and stability in the sera. In the present study multiplex polymerase chain reaction and Illumina high-throughput sequencing (HTS) was used to study the composition and variation of the B cell receptor (BCR) complementary-determining region 3 (CDR3) in GC. The peripheral blood, cancer tissues and peri-cancer tissues were included from 7 patients with GC. On average there was a total of 403,959 CDR3 sequences, with 72,367 unique CDR3 nt sequences and 61,709 unique CDR3 aa sequences per sample identified, which are critical for further understanding the BCR repertoire in GC. The details of GC CDR3s may accelerate the screening process for possible new autoantigens and may provide additional information necessary for generating effective B cell targeted diagnosis and therapeutic strategies.

Introduction

Gastric cancer (GC) is the second most common cause of cancer-related death in the world, with the estimated 5-year survival rate ranges from 4 to 20% (1). High mortality rate in GC is mostly due to its detection at advanced stage (IIIA-IV), which based on endoscopic examination followed by the histological analysis of gastric biopsy, which is an invasive technique not applicable for the screening of asymptomatic population (2,3). Hence, development of noninvasive or minimally invasive tests for GC, body fluids such as plasma, serum or urine is really necessary. For the development of serological tests for early cancer detection, Autoantibodies against tumor-associated antigens (TAAs) are very attractive biomarkers for the development of noninvasive serological tests for the early detection of cancer because of their specificity and stability in the sera (4). By using the classical serological identification of antigens by recombinant expression cloning (SEREX) and proteomics techniques, a variety of TAAs of GC were identified, while the antibody repertoire in GC has not been comprehensively characterized and the diagnostics significance of the autoantibodies has not been evaluated so far (4).

In recent years, major efforts have been made to develop sophisticated experimental and bioinformatic workflows for sequencing adaptive immune repertoires (5). B cells are selectively activated by the specific recognition of an antigen via the variable region of surface B cell receptor (BCR). The BCRs undergo sequential mechanisms to maximize diversity, mainly including rearrangement of various V, D and J gene segments, insertion and deletion of nucleotide at V-D or D-J joint and somatic hyper-mutation of the V region to generate BCRs with high-affinity antigen binding sites. The hypervariable regions in BCRs are also called CDR. Within the variable domain, CDR1 and CDR2 are found in the V region of a polypeptide chain, and CDR3 includes some of V, all of D (heavy chains only) and J regions (6). We and other researchers focused on CDR3 because of its contribution to the generation of antibody diversity. The diversity of distinct BCRs were estimated about 3×10^9 in peripheral blood for healthy people, making the repertoire particularly difficult to analyze (6). Recently, high-throughput sequencing (HTS)

Correspondence to: Professor Yong Dai, Clinical Medical Research Center, The Second Clinical Medical College of Jinan University, Shenzhen People's Hospital, 1017 DongmenBei Road, LuoHu, Shenzhen, Guangdong 518020, P.R. China
E-mail: daiyong22@aliyun.com

Professor Kai Pan, Department of Gastrointestinal Surgery, The Second Clinical Medical College of Jinan University, Shenzhen People's Hospital, 1017 DongmenBei Road, LuoHu, Shenzhen, Guangdong 518020, P.R. China
E-mail: szpankai@126.com

*Contributed equally

Key words: gastric cancer, B cell receptor, immune repertoire, next-generation, complementary-determining region 3

Table I. Human immunoglobulin heavy chain sequence statistics.

Data analysis	Blood	Cancer	Peri-cancer
Total reads number	519954	544159	503171
Immune sequences number	515136	541462	500994
Unknown sequences number	4818	2697	2177
Productive sequences number	407885	461434	413926
Non_productive sequences number	107251	80028	87068
In-frame sequences number	438135	494654	443395
Out-of_frame sequences number	76828	46654	57415
Total CDR3 sequences number	382253	437438	392186
Unique CDR3 nt sequences number	57658	76166	83278
Unique CDR3 aa sequences number	48547	64121	72460

CDR3, complimentary-determining region 3.

technologies have transformed our ability to examine antigen receptor repertoires at single nucleotide resolution, which makes possible the study of BCR repertoire diversity and selection mechanisms at a greater depth than in the past (7-9).

In the present study, we studied the composition and variation of the BCR CDR3s in peripheral blood, cancer tissues, and peri-cancer tissues included from GC individuals by multiplex (polymerase chain reaction) PCR and HTS. Millions of BCR reads will be obtained for each sample, such large-scale sequencing of the BCR repertoire in peripheral blood, cancer tissues, and peri-cancer tissues from GC individuals may reform our perception of the immune system. Moreover, a deep understanding of CDR3s from GC may provide more information for the development of serological tests for early cancer detection and promising treatment target.

Materials and methods

Clinical samples. 5 ml peripheral blood samples, cancer and peri-cancer tissue from 4 males' and 3 females' GC were collected at Shenzhen People's Hospital (Shenzhen, China). GC patients enrolled in the group were required to comply with the pathologic diagnosis, and the patients had a mean age of 62.4 years, ranging from 26 to 85 years. All patients gave written informed consent and the present study was approved by the Medical Ethics Committee of Shenzhen People's Hospital.

DNA extraction and mix. 5 ml peripheral blood samples, 10 mg cancer and 10 mg peri-cancer tissue were obtained from each patient and DNA was extracted by standard methods. Briefly, dewaxing was done by xylene and followed by overnight proteinase K digestion for tissues. QIAamp DNA Mini kit (Qiagen GmbH, Hilden, Germany) was further used for DNA extraction following the manufacturer's instructions. DNA quality was evaluated by loading on a 0.8% agarose gel electrophoresis and DNA concentration was quantified by Qubit fluorometer. DNA from 7 patients' peripheral blood samples were mixed together by 1:1:1:1:1:1:1 according to

Table II. Blood sample's HEC aa sequences for ratios $\geq 0.1\%$.

HEC aa sequence	HEC ratio (%)
ARDLSSWYRDMDV	0.52
ARDQRYYYMDV	0.48
ARDRGYWYFDL	0.45
ARLLTMIHYMDV	0.38
ARTTDSGQHWYFDL	0.37
ARDVIAIYYYYMDV	0.36
ARGYNPDYGMVDV	0.35
ARLHRSWNEVYYGMDV	0.33
ARFGTRSNFQH	0.31
ARVGC GGGRCSLGMHV	0.31
ARGMYGDYVYYGMDV	0.30
ARDRAAPYYYYYMDV	0.28
ARQQLNLHMDV	0.28
AREDYSYYYYMDV	0.27
ARSSYYYGMDV	0.27
ARIPWDYDWYFDL	0.27
ARRRSIAAGPLDV	0.27
ARRDTLGV	0.25
ARAPYVPMVDV	0.25
ARDSLHSWYDDFQN	0.25
ARDPLYGMDV	0.25
ARSSIGYYYYMDV	0.25
ARHWS DQHLHYYYGMDV	0.25
ARGRGYVTYGMVDV	0.25
ATGRYYYYGMDV	0.24
ARLPWLGGMDV	0.24
ARLPLAYSNYYYYYGMVDV	0.23
ARDSGGYGMDV	0.23
ARDGGDYYYYYMDV	0.22
AMGGGYYYYYGMDV	0.22
ARHPRPVTDYYYYYYMDV	0.21
ARDTYGDYASYYYYGMDV	0.21
ARFWTNSSSWYYYYGMDV	0.21
ARDWSYGLDV	0.21
ARDHWN YDGGYMDV	0.21
ARTSSGWYLG SYYYGMDV	0.21
ARDNDYSDYYGMDV	0.20
ARDGAAGTQYGMVDV	0.20
ARDTAMVGRGRGT YGMVDV	0.20
ARDPGSSNWYFDL	0.20
ATGRRDY YYYGMDV	0.20
ARVAYCGGDCYRNLDV	0.19
ARQGF DYYMDV	0.19
ARSLSSTYYYYYMDV	0.18
ARAVVPA AIDGWYFDL	0.18
AKCGPRGARGTMDV	0.18
AKAGRAAAGTGYFQH	0.18
ARSQMATIDYY YGMDV	0.18
ARDFGDYMDV	0.18
ARGGWSWYFDL	0.18
ARSPGKHIVVVT VVFDL	0.18
ARDEATGVAGGMDV	0.18

Table II. Continued.

HEC aa sequence	HEC ratio (%)
ARLSPTQYYYYGMDV	0.17
ARDGIAGLDY	0.17
ARDMELGFDYYMDV	0.17
ARVYSGYTITRWYYMDV	0.17
AKQRHTRMNYMDV	0.17
ARFKNYYYYGMDV	0.17
ARGWNTDSYFYMDV	0.17
ARVGHMVRIYYYYGMDV	0.17
AREFPAVTTPGGMDV	0.17
ARDAEGMDV	0.17
ARDTRLLIDYYYYYMDV	0.17
ARDTGASHFYYYYGMDV	0.16
ARDTRYDYYYYMDV	0.16
ARTDGLGTYYYGMDV	0.16
GRVDTLIMYGMDV	0.16
AREDEYYGMDV	0.16
ARDAF	0.16
ARVTVAADREGMDV	0.15
AKADFRSAGDYYYYMDV	0.15
ARVRFDGSGSRHYYYGMDV	0.15
ATYTSSWYFYGMDV	0.15
ARDSPWNYYYGMDV	0.15
ARHTKVYGDYDDSYYYYMDV	0.15
ARDDEAYRRSSYYYYGMDV	0.15
ARGQRTYGRGYYYGMDV	0.15
ARGGSSSWYMDV	0.15
ARHPYGYNWNELGTEPNYYYMDV	0.15
ARDRGSYFDL	0.15
ARDQASIAAQNYGMDV	0.15
ARLSQGYGDRDDIFQFYWYFDL	0.15
AKDGMSYSSSWHYWYFDL	0.15
TRDAEGMDV	0.15
ARDRGIAAAGTPYYYYGMDV	0.15
ARDGTHCSGSRGYFYFDL	0.15
AKAGPRWYYYYYMDV	0.15
AREGLSGSYYYYGMDV	0.15
AREISYYYYMDV	0.15
AKHGGDIAGLRYFHY	0.15
ARHSMTTSYYYGMDV	0.14
ARDRTAVVVAASLLYGMDV	0.14
ARGEWEPIGYYYYGMDV	0.14
ARDIAARPEQSVQH	0.14
ARDFEQQLLYYYYMDV	0.14
ARIAAAGHYYYYGMDV	0.14
AKSATGPRPYWYFDL	0.14
ARPPEGSYYAFDI	0.14
ARGGAANDYYYYYMDV	0.14
ARDGQTTWLYYYYMDV	0.14
ARDKVPAANYYYGMDV	0.14
ARDLITGTTGMDV	0.14
APKPGRRLLVDV	0.14
ARDLRGGSSYGMDV	0.14

Table II. Continued.

HEC aa sequence	HEC ratio (%)
ARHDDTSGQDLHEH	0.13
ARDTLSGGYYYYYYMDV	0.13
ARDFPGGMDV	0.13
ARDVDP	0.13
ARDRPYSSGWSWGYGMDV	0.13
AKSAYCTPKCNALDV	0.13
AKVVVFVVPVPAASDYMDV	0.13
ARQHRRVRGVMNYYGMDV	0.13
ARDRRGGDYYYYGMDV	0.13
AKDSWYSRPYYGMDV	0.13
ARHGWFGEGLGVYYYGMDV	0.13
AKVYRDYYYYGMDV	0.13
ARRRYSGYRDDAFDI	0.12
ARDLRSGEMATPNYYYGMDV	0.12
ARGGVDTAMGYYYYMDV	0.12
ARVRGSSSWYRPRGMDV	0.12
ARVGVPIADEDYYYYMDV	0.12
ARDGRDSSGWYSYWYFDL	0.12
ARHLTGELFGMDV	0.12
ASVAATTNYYYVLDV	0.12
ARGNWGTSWYFDL	0.12
ARHKL RDGSGSYFLYMDV	0.12
ARDLAVLGSGSLGGL	0.12
ARDRGPYYYYGMDV	0.12
ARDPGGGTTGMTYYYYGMDV	0.12
ARVVLPFGELEAMDV	0.12
ARGVTIFGVVTTYYYMDV	0.12
ARFRVSYGYVDYYYYGMDV	0.12
ASAYGDYGAAFDI	0.12
AKDEKGVIIYYGMDV	0.12
AREGLRNYYYYYMDV	0.12
ARGKLLSPMDV	0.12
ARDLGTTVTTERSYYYGMDV	0.12
ARDIRREWEPFYYGMDV	0.12
ARDFGHIYDEYDFDF	0.12
ATIRCSGGSCPYYYYGMDV	0.12
ARSGYYDRYYYMDV	0.12
ARDFTDQNTVYYGMDV	0.12
ARGGPYYYYYMDV	0.12
AREYSSNHYYYMDV	0.11
ARGRVPAAKGYYYYMDV	0.11
ARDRVADDAFDI	0.11
ARDLGGYFLYRYGMDV	0.11
ARDRNYYGSGSYYSTYMDV	0.11
AKDDGPYYYGMDV	0.11
AKDPHGSSWYYYYYYYMDV	0.11
ARHSYDILTGYAYYYGMDV	0.11
AKGPRSGIRFMDV	0.11
ARQDYDSSGYYYDYYYYGMDV	0.11
ARRVSSSWYGWSDY	0.11
ARAHGPSSWGMDV	0.11
ASPNSSTAYTFDY	0.11

Table II. Continued.

HEC aa sequence	HEC ratio (%)
AREAMVRGALYYYYYGMDV	0.11
ARECFSSWYCYYYYYGMDV	0.11
AKDSIAVAGDGMDV	0.11
ARHVGHHYGGSGSYNTYMDV	0.11
ARALRRYVVVPAAYYYYYYMDV	0.11
ATASGRYFDYYYYYMDV	0.11
ARPGHVVPAAAFDI	0.11
ARHGLRGCSDRCYTSFYNGMDD	0.10
ARMSMTTPVK	0.10
ARGHMTDPDYMDV	0.10
ARESDYYYGMDV	0.10
AGPDTAMVSRDYYYYYMDV	0.10
ARHEAGDLGYDAFDI	0.10
ARDQGLVVVIKDYGGYMDV	0.10
ARDKEDPWNALDL	0.10
ALEEVGY	0.10
AKVRRRAQGYAMDV	0.10
ARDRGGIQLWRTRYRYYYYYGMDV	0.10
ARVPTPFDYGGYMDV	0.10
ARGRQVDSSGWYDYGGYMDV	0.10
ARDHYSSDAFDI	0.10
ATLYYDFFVRPGGMDV	0.10
GRDCVRAGDYGVVDV	0.10
AREPLYDYGGYMDV	0.10
HEC, highly expanded clones.	

Qubit value, renamed one blood sample. Meanwhile, DNA from 7 patients' cancer tissues and peri-cancer tissues were mixed by the same way.

Multiplex-PCR amplification of the IGH CDR3 region. The utilized 12 forward primers and 4 reverse primers were used for multiplex PCR to amplify rearranged IGH CDR3 region (10). The PCR condition was set as 95°C for 15 min, followed by 25 cycles of 94°C for 15 sec, 60°C for 3 min, with a final extension for 10 min at 72°C. And the PCR products were purified by AMPure XP beads to remove primer sequences (Beckman Coulter, Inc., Brea, CA, USA). A second round PCR was performed to add a sequencing index to each sample. The PCR condition was set as 98°C for 1 min, followed by 25 cycles of 98°C for 20 sec, 65°C for 30 sec and 72°C for 30 sec, with a final extension for 5 min at 72°C. The library was separated on agarose gel, and the target region was isolated and cleaned by QIAquick Gel Extraction kits (Qiagen GmbH).

HTS and data analysis. The library was quantitated by Agilent 2100 bioanalyzer instrument (Agilent DNA 1000 Reagents; Agilent Technologies, Inc., Santa Clara, CA, USA) and real-time quantitative PCR (TaqMan Probe), sequenced by Illumina Miseq. Briefly, the adaptor reads and the low quality reads were filtered from raw data, and the clean data

Table III. Cancer sample's HEC aa sequences for ratio >0.1%.

HEC aa sequence	HEC ratio (%)
ARWRAAAGTRRNYSYHYMDV	1.23
AKDFGAGLGGYYMDV	1.01
ARLQGGAVFQH	0.79
ARGSMSIRAGWYFDL	0.75
AADRGDYGDYEDYYYYIDV	0.64
AREVHQRRQSEDAFDV	0.57
ARDLCSGVCWYFFMDV	0.52
ARDGRTWHYESRGFHWFDA	0.51
AKEGIPAAGMSEGYGGYMDV	0.38
ARELRGGSWAGGMDV	0.29
ARTRITILGDMDV	0.27
ARDLRGSDDY	0.27
TIGHYST	0.26
ARTRITIFGDMDV	0.23
AGSSGFDPFDC	0.23
AQEIRPND	0.22
ATDAVNNWNSHY	0.21
EAGGAGFDC	0.21
ARDASSGLRYYGMDV	0.21
ARRNKGSLGWDFDY	0.21
AKGGNTGGTNQFLSYYYHYMDV	0.20
ARDRAGDYAADD	0.20
ARQRGYYNMDV	0.20
AADGFQLEDFRYGMDV	0.20
ARTQWEYWYFDL	0.18
ARLWTVTPTDYGMDV	0.18
TIGHYSS	0.18
ARTVTWGYMDV	0.17
AKDDTTYCGGDCYFDL	0.17
AKGGHMGPNQFLSYYYHYMDV	0.17
TRGGSRGDSISWYTMHSSYYGMDV	0.17
ARVLHGGGGHFHH	0.17
ARSPYDSSGSKYYGMDV	0.16
TRDERGRAAQTNYYYYYMDV	0.16
ARGRQDYFDF	0.15
AKDARYCTPTSCNTPLSYSSFYMDA	0.15
ATEDSNGGSYRRH	0.15
ARTSGSSGDAFDI	0.15
AKDRGPNIPDWYFDL	0.15
ARPSGAVTTMRMDV	0.14
ARVLHGGGGNFHH	0.14
ATDSLVD	0.14
ARGGSRGDSTSWYTMHSSYYGMDV	0.14
AKDDTTYCGGDCYFDI	0.14
ATLDTHPAHYGGYMDV	0.13
AKGGEGRTTYWYLDL	0.13
ARGRQDYFDY	0.13
ARFHLKDGTSSEY	0.13
ARLSGSYYYYYYMDV	0.13
AKDQQWLIRLVHDY	0.12
ARDPYYGSGSRYYYMDV	0.12
ARGPGAAAGTYYYYSMDV	0.12

Table III. Continued.

HEC aa sequence	HEC ratio (%)
ARDYGEN	0.12
ARGPRQIGARSSSGLEDWHFEL	0.12
ARDISPAYRGNHAFDI	0.12
AKARTPGTPYYYHMDV	0.12
ARSSGRAAGVDS	0.11
ARETQGFDP	0.11
VRGTITFDY	0.11
AKTYGDYGGETAFTDM	0.11
ASEVHLQSDRLGMDV	0.11
ARGGLPSSYYYMDV	0.11
ARTSQLGRYLDL	0.11
ASEIHLGADRLGMGV	0.11
ARESGGYRYGHIDHYNAMDV	0.11
ARLDCSSTSCYPDLGLGYFDL	0.11
ARDIGSSSWYEHLEQ	0.11
AKLNSALGVTGRRGRPVYFED	0.11
ARSRDYYSYMDV	0.11
ARTTYSYDV	0.11
ARRRYGEGFQH	0.11
ASGTASWYDYYYGMDV	0.11
ARDPGGYSFDH	0.11
VRGEDDDGDYVDYYYGMDV	0.11
ARIRPVYTIDF	0.11
AKGGEGRTTYWYFDL	0.11
VTQGDGSLDF	0.10
SRARITIFGDMDV	0.10
ARETEGFDP	0.10
ARGRGNDYGDYSYYYMDV	0.10
GHIGTFDL	0.10
ARGGSRGDSTSWYTGMLVYYGLDV	0.10
HEC, highly expanded clones.	

was used in further alignment. Subsequently, the clean data was aligned to human IGH database and analyzed by online IMGT/HighV-QUSET tool. The data were including V, D, J assignment, CDR3 length distribution, clustering and other analyses (11). Statistical significance was determined using Student's t-test and $P < 0.05$ was considered to indicate a statistically significant difference.

Results

Summary of sequencing. We obtained 519,954 reads number of blood sample, 544,159 reads number of cancer sample and 503,171 reads number of peri-cancer sample from 7 GC patients. After filtering, including the removal of contamination, adaptor sequences and low-quality reads, the data were aligned to human IGH database. On average, the mapping IGH sequences reads number was 519,197 per sample. The mean unknown sequences number (reads=3,031), productive sequences number (reads=427,748), Non-productive

Table IV. Peri-cancer sample's HEC aa sequences for ratio $>0.1\%$.

HEC aa sequence	HEC ratio (%)
VRHSSGDYRNWYFDL	1.16
ARAGYTYGEDMDV	0.72
AREMDV	0.59
CAYHDY	0.52
ARDIGSSSWYEHLEQ	0.39
ASQRAIFRPMMDV	0.30
AALLQHNGRGTFDF	0.26
AKDDEYHDSFGLDV	0.21
ARHEVASHDSYMDV	0.19
ARLPEGLDWHLDL	0.17
ARGRRPDHTYFYMDV	0.16
GVYV	0.16
ARGRTPGRMDV	0.15
AVAVHGTIFYWYFDV	0.15
ATSDNYMDV	0.14
ASEIHLGADRLGMGV	0.14
ARVPWGWFFDY	0.12
AREGI	0.12
A	0.12
AKDVGTYIYNYMDV	0.12
ARSSDRAEFGGNYYSMDV	0.11
ARPGQQRGGWYFDL	0.11
VRRGFWSEAAIGKDGNYYSMDV	0.11
ARSATTAANWYFNL	0.11
AASSDY	0.11
AKDRIRWSLYDFCSGFDV	0.10
AKAGGAFLYMDV	0.10
ARGRRGYSGYENRPLLDQ	0.10
ATLQQGHSRSLTNPSFDYITMDV	0.10
ARHRRYYGSGYYMDV	0.10
ARDGMRTGNMDV	0.10
HEC, highly expanded clones.	

sequences number (reads=91,449), In_frame sequences number (458,728), and Out-of_frame sequences number (reads=60,299) per sample were listed on Table I. The CDR3 sequences are identified by the conserved motif. The abundance of each CDR3 clone and the number of distinct CDR3 clone species are calculated. On average, there were 403,959 total CDR3 sequences, 72,367 unique cdr3 nt sequences number, 61,709 Unique cdr3 aa sequences number per sample (Table I).

Distribution characteristics of CDR3 length. The length of the BCR CDR3 loop is an important determinant of B cell repertoire diversity. In our study, we first assessed length distribution of BCR CDR3 sequences (aa) in the blood, cancer and peri-cancer group of GC (Fig. 1). Each group displayed good Gaussian distribution of CDR3 length. Then we fitted Gaussian distribution curve for each sample by excel. And the

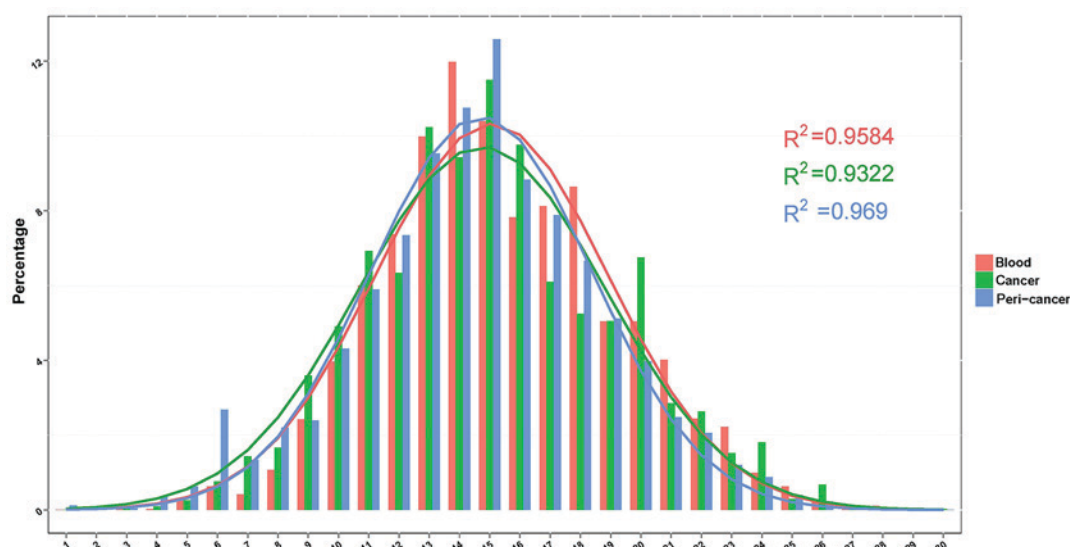


Figure 1. Length distribution of B cell receptor complementary-determining region 3 sequences (aa) in the blood, cancer and peri-cancer group of gastric cancer.

goodness of fit was quantified by R^2 , which ranges from 0 to 1 (that is, from worst fitness to best fitness). R^2 were calculated for each sample, and for the R^2 value peri-cancer ($R^2=0.969$) > blood ($R^2=0.9584$) > cancer ($R^2=0.9322$).

Highly expanded clones (HEC). The expression level of each clone was calculated according to the identity of each sequence after alignment. And the degree of expansion of each sample clone was based on the unique CDR3 sequence frequency. Here, we defined that BCR clones with frequency above 0.1% of total reads in a sample were HECs. In the blood sample, we observed 180 clones were HECs, with HEC ratio 29.80% (Table II), clone aa sequence 'ARDLSSWYRDMV' was the only sequence above 0.5% (0.52%). While in the cancer sample, 82 clones were defined HECs, with HEC ratio 17.24% (Table III), there were 8 aa sequences with ratio above 0.5% among them listed in Table III. And for peri-cancer sample, we observed 31 clones were HECs (Table IV), with HEC ratio 7.07%, and among them 4 aa sequences were HEC above 0.5%. Next, we investigated whether HECs overlap among blood, cancer and peri-cancer sample. Only two HEC sequences were found in both cancer and peri-cancer sample, with sequence 'ASEIHLGADRLGMGV' 0.11% in cancer vs. 0.14% in peri-cancer, 'ARDIGSSSWYEHLEQ' 0.11% in cancer vs. 0.39% in peri-cancer. While sequence 'ASEIHLGADRLGMGV' and sequence 'ARDIGSSSWYEHLEQ' were found 0.0007 and 0.001% in blood sample.

Cancer special CDR3nt and aa. We further to investigate the cancer special CDR3s with blood sample and peri-cancer sample no expression, which may be the perfect B cell treatment target. And the top 20 cancer special CDR3nt and aa were listed in Table V, with the sequence ratio from 0.04 to 0.11%. And the identified cancer special CDR3s need further research.

Discussion

GC is biologically and genetically heterogeneous with a poorly understood carcinogenesis at the molecular level. Despite there

were many prognostic, predictive, and therapeutic biomarkers: HER2, E-cadherin, fibroblast growth factor receptor, mammalian target of rapamycin, and hepatocyte growth factor receptor as well as sections on microRNAs, long noncoding RNAs, matrix metalloproteinases, PD-L1, Baniak *et al* investigated to date (12), GC continues to be detected at an advanced stage with resultant poor clinical outcomes (13,14).

Cancer immunoediting, which is the capacity of immunity to control and shape cancer, that is, the result of three processes: Elimination, equilibrium, and escape (15,16). While the influence of gastrointestinal tract tumor differentiation on the diversity of T-cell repertoire was investigated by Luo *et al*, and they found TCR repertoire diversity have a significant correlation with the degree of tumor differentiation (17). In this study, we used a novel HTS protocol to investigate the BCR CDR3s in blood, cancer tissues and peri-cancer tissues of GC individuals. On average, we obtained 403,959 total CDR3 sequences, 72,367 unique cdr3nt sequences number, 61,709 Unique cdr3 aa sequences number per sample. Among them, we defined that BCR clones with frequency above 0.1% of total reads in a sample were HECs. The overlap HECs and cancer special HECs were counted and listed in the data. Furthermore, although this study provided some insight into the pathological processes associated with the cancer-specific BCR CDR3s, the function role of these sequences in the development and/or progression of cancer and antitumor immune response is still elusive, and it would be of great interest to explore their predictive and prognostic significance.

Acknowledgements

Not applicable.

Funding

The present study was supported by funds received from Science and Technology Plan of Shenzhen, Guangdong (grant nos. JCYJ20160422150329190 and JCYJ20140416122811914)

Table V. Top 20 cancer special CDR3 aa and nt sequence.

aa sequence	nt sequence	Sequence ratio (%)
GCGAGAGATCCCGGTGGATACTCCTTTGACAC	ARDPGYSFDH	0.11
GCGAGACAGGGTGTGAAGCAGACGTGATTCCTACTACTACGGTATGGACGTC	ARQVEAAADSYYYGMDV	0.08
GCGAGAGTATACATGGCGGTGGCGGATCCTCCATCAC	ARVLHGGGGLHH	0.07
GCGAGAGATGGTTCGGGGAGCCCCATGGACGTC	ARDGSGSPMDV	0.07
GCGAGAGATCCTTATTACTATGGTTCGGGAAGTCTTCGTACTACTACTTTTGACTCC	ARDPYYYGSGSSSYYYFDS	0.07
GCGAGAACGGCAATTGCTACGAGGGGCTACTACTACATGGACGTC	ARTAIATRGYYMDV	0.07
GCAAGAGGTACTATGGTTCGGGAGTCCCTCAGAGGCTACATGGACGTC	ARGTMVRGVLRGYMDV	0.06
GCGAGAGATTTGACTGCAGCAGCTGGCACCCCTTCTACTACTACAAAGTTTGGACGTC	ARDLTAAGTPFYYKGLDV	0.06
GCGAAATGGGAGGGCACTTACCCTAAGTATTACATGGACATC	AKWEGTYPKYMDI	0.06
GCGAGAGCTCGTGCTTTTGATATC	ARARAFDI	0.06
GCGAGAGGGGTTTGGTAGCTGGAGCTATGGACGTC	ARGGLVAGAMDV	0.06
GCGAGAGATCCTTCTTACGGTATGGACGTC	ARDPSYGMDV	0.05
GCGAGAGGTTTCGGGAGTTCCTTACTACTACTACGGTATGGACGTC	ARGSGSSYYYGMDV	0.05
GCGAGAGTATACATGGCGGTGGCGGATCTCCACCAC	ARVLHGGGGIFHH	0.05
GCGAGACTTGCCCGTGACCTGCTATGGAGACTTTAGCCGGGTGCTTCCGATCTC	ARLARDPAMETLAGWYFDL	0.05
GCGAGACCGCGGTATAACTGGAACTACATGGGGGACTACTACTACGGTATGGACGTC	ARPRYNWNYMGDYYYGMDV	0.05
GCGAGAACAGGCAAGAAATGATGGACTACATGGTCGTC	ARTGKEMMDYMVV	0.05
GTGAGAGGGGAGGGCGGCTACTACTACGGTATGGAAGTC	VRGEGGYYYGMEV	0.05
GCGAGATCGTCTGGAAGGCGAGCGGGCGTTGATTC	ARSSGRAAGVDS	0.05
GCGAAATATAGTAGTCTGCTAAGTACTATTATATATGGACGTC	AKYSSSSKYYYMDV	0.04

CDR3, complimentary-determining region 3.

and China Postdoctoral Science Foundation Grant (grant no. 2017M610575).

Availability of data and materials

The aa sequence raw data is available from the National Center for Biotechnology Information database at ncbi.nlm.nih.gov/bioproject/438006.

Authors' contributions

SL and YZ conceived and designed the study. LWL, SKD, XCL and KLZ performed the experiments. SL wrote the paper. KP and YD interpreted the data and gave final approval of the version to be published. All authors read and approved the manuscript.

Ethics approval and consent to participate

All patients provided written informed consent and the study was approved by the Medical Ethics Committee of Shenzhen People's Hospital.

Consent for publication

All patients provided written informed consent for the publication of their data.

Competing interests

The authors declare that they have no competing interests.

References

1. Park JM and Kim YH: Current approaches to gastric cancer in Korea. *Gastrointest Cancer Res* 2: 137-144, 2008.
2. Rosati G, Ferrara D and Manzione L: New perspectives in the treatment of advanced or metastatic gastric cancer. *World J Gastroenterol* 15: 2689-2692, 2009.
3. Choi IS and Tsungteh TT: Epigenetic alterations in gastric carcinogenesis. *Cell Res* 15: 247-254, 2005.
4. Zayakin P, Ancāns G, Siliņa K, Meistere I, Kalniņa Z, Andrejeva D, Endzeliņš E, Ivanova L, Pismennaja A, Ruskule A, *et al*: Tumor-associated autoantibody signature for the early detection of gastric cancer. *Int J Cancer* 132: 137-147, 2013.
5. Friedensohn S, Khan TA and Reddy ST: Advanced methodologies in high-throughput sequencing of immune repertoires. *Trends Biotechnol* 35: 203-214, 2017.
6. Jung D, Giallourakis C, Mostoslavsky R and Alt FW: Mechanism and control of V (D)J recombination at the immunoglobulin heavy chain locus. *Annu Rev Immunol* 24: 541-570, 2006.
7. Calis JJA and Rosenberg BR: Characterizing immune repertoires by high throughput sequencing: Strategies and applications. *Trends Immunol* 35: 581-590, 2014.
8. Mori A, Deola S, Xumerle L, Mijatovic V, Malerba G and Monsurro V: Next generation sequencing: New tools in immunology and hematology. *Blood Res* 48: 242-249, 2013.
9. Weinstein JA, Jiang N, White RA III, Fisher DS and Quake SR: High-throughput sequencing of the zebrafish antibody repertoire. *Science* 324: 807-810, 2009.
10. Liu S, Hou XL, Sui WG, Lu QJ, Hu YL and Dai Y: Direct measurement of B-cell receptor repertoire's composition and variation in systemic lupus erythematosus. *Genes Immun* 18: 22-27, 2017.
11. Jacobi AM and Diamond B: Balancing diversity and tolerance: Lessons from patients with systemic lupus erythematosus. *J Exp Med* 202: 341-344, 2005.
12. Baniak N, Senger J, Ahmed S, Kanthan SC and Kanthan R: Gastric biomarkers: A global review. *World J Surg Oncol* 14: 212, 2016.
13. Kodera Y: The current state of stomach cancer surgery in the world. *Jpn J Clin Oncol*: 30 Aug, 2016 (Epub ahead of print).
14. Lazăr DC, Tăban S, Cornianu M, Faur A and Goldis A: New advances in targeted gastric cancer treatment. *World J Gastroenterol* 22: 6776-6799, 2016.
15. Koebel CM, Vermi W, Swann JB, Zerafa N, Rodig SJ, Old LJ, Smyth MJ and Schreiber RD: Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* 450: 903-907, 2007.
16. Lee K, Hwang H and Nam KT: Immune response and the tumor microenvironment: How they communicate to regulate gastric cancer. *Gut Liver* 8: 131-139, 2014.
17. Luo W, Liao WJ, Huang YT, Shi M, Zhang Y, Wen Q, Zhou MQ and Ma L: Cancer of the gastrointestinal tract results in a restricted T-cell repertoire dependent upon tumor differentiation. *Cell Immunol* 270: 47-52, 2011.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.