

The Diversity and Heterogeneity of TCR and BCR in Pan-Cancer

Bo Li

Wuhan Institute of Technology School of Chemistry and Environmental Engineering

Jing Zhang

College of Life Science and Technology, Beijing University of Chemical Technology

Jin-dan Guo

Chinese Academy of Medical Sciences Institute of Medical Laboratory Animal Science, Beijing

Hui-xue TANG

College of Life Science and Technology, Beijing University of Chemical Technology

Wen-xiang Hu

Wuhan Institute of Technology School of Chemistry and Environmental Engineering

Li-da XU

College of Life Science and Technology, Beijing University of Chemical Technology

Yu-bao Chen (✉ chenyubao@cnilas.org)

Chinese Academy of Medical Sciences Institute of Medical Laboratory Animal Science, Beijing

Research Article

Keywords: Immune repertoire, Pancarcinoma, Immune checkpoints, Heterogeneity

Posted Date: October 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-926802/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: With the development of the sequencing technology, RNA-seq is becoming more and more popular. The method of using RNA-seq technology to analyze the immune repertoire has the advantages of low sample demand, simple operation, and comprehensive information.

Result: The RNA-seq data from public database were used to analyze and compare the TCR and BCR sequences of cancer and normal tissues in four types of cancers, including colon adenocarcinoma, lung adenocarcinoma, pancreatic adenocarcinoma and stomach adenocarcinoma. It was found that the clonal diversity of TCR among various types of cancer is different, and the variety of BCR between cancer tissues and healthy tissues is obvious. Based on the analysis of non-shared CDR3 sequences, the degree of the heterogeneity of TRB is: different cancer types > different individuals with same cancer > cancer tissues and paracancerous tissues of the same individual. Also, the correlation between the expression levels of the five main immune checkpoints genes and the count of TCR clones displayed various among different types of cancers.

Conclusion: The results verify the advantage of RNA-seq analyzing in acquiring individual immune repertoire. Moreover, the heterogeneity of TCR in different cancers and the diversity of BCR in tumour and adjacent tissues shown in the results, provide hints at developing the strategies of tumour immunotherapy.

Introduction

The adaptive immune system is an essential part of the vertebrate immune system. It is a necessary system to produce specific antibodies against antigen invasion and has a memory function. The system includes T cells and B cells, which provide different antibodies based on different antigenic structures to prevent pathogens and antigens from entering basal cells. The recognition of an antigen depends on the receptors on the surface of the immune cells, i.e. T cell receptors (TCR) or B cell receptors (BCR) [1]. When the antigen is recognized, T cells will clone in large numbers and promote B cells to secrete the specific antibodies to neutralize the antigen molecules.

TCR is a heterodimer composed of two chains TRA and TRB [2, 3]. Depending on the diversity caused by the different constituent chains of TCR, T cells can be divided into $\alpha\beta$ T cells and $\gamma\delta$ T cells [4]. Among them, $\alpha\beta$ T cells mainly involve in cellular immunity [5]. $\gamma\delta$ T cells mainly distribute in the mucosa and skin immune system, which can directly recognize specific antigens and kill target cells [6]. Here we only discuss the TCR produced by $\alpha\beta$ T cells. BCR, which is distributed on the surface of B cell membrane, is a tetramer with two heavy chains (H chain) and two light chains (L chain) connected by two disulfide bonds [7]. The epitope on the surface of the antigen molecule can be accurately recognized by and bound explicitly to BCR, thereby the humoral immunity arise [8].

The structures of TCR and BCR chains include V, J and CDR3 regions. The V region is the most diverse species relative to other regions and is one of the primary sources of receptor diversity. It including two

complete complementarity-determining areas (CDR1 and CDR2) and a portion of CDR3, is variable in TCR and BCR. The J region is a junction region in TCR and BCR. The type of J region is also an essential factor in the variety of TCR and BCR. The CDR3 area is the region connecting the V gene and the J gene. CDR3 integrates all nucleotide insertions or deletions during gene recombination and is the region with the most mutations. Therefore, analyzing the sequence of the CDR3 region is very important for studying the characteristics of TCR and BCR.

The more types of TCR and BCR clones and the higher diversity of T cells and B cells, the more active the immune system becomes[9]. The diversity mainly stems from four mechanisms[10, 11]: combinations of different VDJ gene fragments; the random insertion and deletion of different gene segments; combinations of different heavy and light chains; and unique random high-frequency mutations to the receptor of B cells. It has been estimated that these mechanisms could cause thousands of different B-cell and T-cell receptors[12]. The diversity of TCR and BCR can be stimulated by different antigens and has specificity for antigen. For a particular infection or disease, T cells and B cells in the immune system respond and produce specific TCR and BCR sequences for the antigen[13]. Thus a particular receptor sequence has the potential to be a marker of a specific virus, bacterium, or fungus[14, 15].

Common methods for detecting receptor sequences include flow cytometry[16], PCR[17, 18], and immune repertoires sequencing[19, 20]. In recent years, RNA-seq analysis gradually becomes a new technology to avoid the limitations of these methods. RNA-seq technology can provide important biological informations so as to reveal biomarkers for the diagnosis, monitoring, and treatment of diseases[21, 22]. In recent years, the studies on the diversity of TCR and the use of TCR-modified T cells for immunotherapy based on RNA-seq technology continue rising. The association of TCR diversity with MHC class II expression in tumour tissue and the public T cells enrichment in the tumour environment has been examined in the literature[23]. Some studies also reported the relationships between the diversity and clonality of TCR and certain diseases, such as DCI[24] and some cancer types[25], and proposed several identification methods based on TCR specific analysis[26]. In addition, the use of TCR to treat diseases has also made great progress. Jakobsen used high-affinity TCR modified T cells to treat patients with melanoma, in which 55% patients showed clinical responses[27]. Therefore, discussing the TCR and BCR sequences in pan-cancer in a more comprehensive manner is meaningful for developing tumour immunotherapy. Based on the aforementioned advantages of RNA-seq technology and the importance of TCR and BCR diversity, we collected the RNA-seq data of four types of cancers and fully described their immune repertoires, including the clone diversity and the sequences features, in order to find out the TCR and BCR characters with tumour specificity.

Results

Diversity and heterogeneity of TCR in different cancer types

Diversity of TCR

The number of TCR clones, clone types, and CDR3 sequences in normal and tumour tissues of different cancers were found out and displayed in Table 1. Because the $\gamma\delta$ T cells only account for 1%-5% of T cells, our analysis of TCR mainly considers $\alpha\beta$ T cells. As presented in Figure 1a-1d, individuals with different cancers are not consistent. For most of COAD individuals, the numbers of clones in normal samples is relatively higher than that in cancer tissues. But no consistent phenomenon can be observed in the other three cancers. For the overall comparison between the cancer and the adjacent noncancer samples, only the number of TRB in the cancer tissue of COAD is significantly lower than that of healthy tissue ($P=0.038$). For different types of tumour tissues, the numbers of TRA and TRB clones of STAD and PAAD are relatively higher than those of the other two types of cancers (Figure 1e-1f).

Table 1
Statistics of TCR clones discovered in each sample groups

Sample type	count	average ^{*g}	Diversity [*]	average	CDR3	average	
Normal	COAD	820	103	549	69	544	68
	LUAD	4929	183	3816	141	3775	140
	PAAD	3702	370	2684	268	2654	265
	STAD	21291	626	10600	312	10405	306
Tumour	COAD	393	49	313	39	310	39
	LUAD	5067	188	4257	158	4219	156
	PAAD	2515	252	1944	194	1928	193
	STAD	17441	513	9151	269	8991	264

Diversity: the diversity of clone types

average: the number of counts or diversity or CDR3/the number of samples of each type. The numbers of TRA and TRB clones were measured in each 10 million reads.

The TCR diversities between tumour tissues and normal tissues were compared in each type of cancers and across four types of tumour tissues. Shannon entropies was adopted to represent the TCR clonal diversity (Figure 1g-1h). There was no significant difference in Shannon entropies between tumour-normal samples for the four types of cancers ($P>0.05$). The diversity of TCR in COAD and LUAD did not reach the conclusion of the significant difference reported in the previous literatures[28, 29]. The reason is probably that the sample size of COAD is too small. But the Shannon entropies of TRA and TRB in COAD were significantly lower than those of the other three types of cancers ($P<0.05$). Interestingly, the significant differences in TCR diversity were mainly concentrated in different cancer types, rather than cancer and adjacent noncancers. It may be due to various immune responses in different organs.

heterogeneity of TCR

The shared TRB CDR3 sequences were analysed among all the samples to look into the heterogeneities of TCR for each cancer type. The share sequences of CDR3 are mainly concentrated between the cancer tissue and healthy tissue of the same person (Figure 2a, **Figure S1**). At the same time, for the totality of all the individuals of the four cancers, it can also be found that there were few TRB shared between different cancer tissues (Figure 2b). Taking STAD as an example, the number of shared sequences between the cancer tissue and adjacent cancer tissue shown in Figure 2c is greater than that of the cancer tissue of STAD and the other three cancers in Figure 2b. In a word, the TCR heterogeneity between different tumour types is very high, and the TCR heterogeneity of different individuals with the same cancer is higher than that of cancer and adjacent cancers in the same individual. This conclusion supported the ideas of differentiation of various tumours based on the TCR sequences.

The diversity of BCR in carcinoma and adjacent noncancerous tissues

Meanwhile, B cell receptor sequences (IGH, IGK, IGL) of all the samples were also analyzed. The number of BCR clones and BCR CDR3 sequences found in carcinoma and adjacent noncancerous tissues of each cancer type were shown in Table 2. The figures showed the number of BCR clones in four types of cancers (Figure 3a-3d). For individuals in the four types of cancers, patients with LUAD and PAAD had more BCR clones in tumor tissues, but patients with STAD and COAD had more BCR clones in normal tissues. The differences in BCR clone numbers across the sample groups were demonstrated in Figure 3e-3g. The number of BCR clones in LUAD cancer samples was significantly higher than that in adjacent noncancerous samples ($P < 0.05$). In comparison, the number of BCR clones in STAD cancer samples was considerably lower than that in noncancerous samples ($P < 0.05$). Besides, the number of BCR clones in STAD samples was significantly higher than those of the other three cancers ($P < 0.05$).

Table 2
Statistics of BCR clones

Sample type	count	average	diversity	average	CDR3	average	
Normal	COAD	315988	39499	103184	12898	83208	10401
	LUAD	289424	10719	88937	3294	69408	2571
	PAAD	351798	35179	74372	7437	62808	6281
	STAD	13732779	403905	1358124	39945	1038997	30559
Tumour	COAD	82500	10313	24713	3089	22002	2750
	LUAD	1562443	57868	283427	10497	208906	7737
	PAAD	701283	70128	94940	9494	79637	7964
	STAD	4867153	143152	638065	18767	508122	149445

Consistent with the analysis of TCR, we calculated the Shannon entropies of BCR in different samples (Figure 3h-3j). In the comparison between tumour and normal tissues of the same cancer, Shannon entropies of BCR (IGH, IGK and IGL) in COAD and STAD cancer samples were significantly lower than their adjacent noncancerous samples ($p < 0.05$). In contrast, BCR Shannon entropies of LUAD cancer samples was substantially higher than their adjacent noncancerous samples. Unlike TCR, the diversity of BCR between cancer and adjacent noncancerous cancers within the same cancer type was very significant ($p < 0.05$), except in PAAD. Furthermore, the BCR diversity between tumour and adjacent noncancerous tissues of the same cancer type was more significant than that among different cancers.

The frequency distribution of V and J genes in different sample groups

The frequencies of the V genes and the J genes in TCR and BCR were also presented in Figure 4. Based on the preferences expressed by the height of the bars, the frequency distribution of the V genes were similar in tumour and normal samples in each cancer type except for COAD, and so the frequency distributions of the J genes. This preference was only observed in COAD samples, probably because the number of COAD samples is less than one-third of STAD and LUAD, which makes the statistical results deviate the real world.

PDCD1 expression positively correlated with TCR counts

The immunotherapy targeting to many immune checkpoint inhibitors such as PD-L1 have been widely used. However, anti-immunotherapy for tumours was not always effective. The anti-PD-1 antibody had a therapeutic effect in approximately one in four to one in five patients with non-small-cell lung cancer, melanoma, or renal-cell cancers[30]. Therefore, novel prognostic markers are needed to predict which patients would respond. A study has demonstrated the potential for TCR repertoire profiling to serve as a biomarker of clinical response in pancreatic cancer patients receiving immunotherapy[31]. That study also provides the rationale to consider anti-CTLA-4 as an initial therapy for T cell priming and anti-PD-1 as a requirement for T cell expansion and maintenance. On the basis of the article, we introduced the associations between more immune checkpoints and TCR clones. The correlation between the expression levels (FPKM) of the immune checkpoint related genes, such as PDCD1, TIM-3, LAG3, BTLA and 4-1BB, and the number of TCR and BCR clones were displayed in Figure 5. The results showed that the TCR counts were correlated to the expression levels of different immune checkpoint genes in different types of cancers. For example, the number of TCR clones was correlated with the expression of PDCD1 and BTLA in LUAD, while the number of TCR clones is related to the expression of PDCD1 and TIM-3 in PAAD. Commonly, the TCR expression was positively correlated with the expression level of the PDCD1 gene in all the four cancer types.

Discussion

Through the analysis of the immune repertoires in cancer tissues and adjacent noncancer tissues of COAD, LUAD, PAAD and STAD, there were no significant differences found in TCR count and Shannon entropies within the same cancer type. But there are significant differences in TCR numbers and Shannon entropies among different cancer types. Based on the analysis results of non-shared CDR3 sequences, TRB heterogeneity between various cancers>heterogeneity between different individuals of the same disease>heterogeneity of cancer tissues and healthy tissues of the same individual. Our results show the diversity and heterogeneity of TCR among different cancers. In addition to the same TRB sequence as that in normal tissues, the TRB sequence unique to the cancer types predicts the presence of specific antigens for the disease and may be a molecular target for disease prediction and immunotherapy.

For BCR, the Shannon entropies of cancer tissues and adjacent tissues in COAD, LUAD, STAD are significantly different. It shows that the diversity of B cells infiltrated inside the tumours of these three cancers has changed significantly compared with that in healthy tissues. Compared with the other three cancers, the diversity of BCR in cancer and noncancer tissue of PAAD is not very obvious. It can be inferred that the degree of B cell infiltration of different cancers is also different. The diversity of BCR may become an indicator of cancerous, and provide new ways for clinical monitoring of diseases.

Through analyzing of the usage of the V genes and J genes of TCR and BCR sequence, it seems that the usage preference and frequency distribution has no obvious differences between normal and tumour tissues in LUAD, PAAD and STAD. The different frequency distribution of TRAV and TRAJ genes in COAD standard samples and cancer samples may be due to the small samples size. Besides, similar to the TCR genes, the frequency distribution of the V and J genes of the BCR heavy and light chains are also identical.

Immune checkpoints are important factors for suppressing the function of T lymphocytes in the body, thus triggering the immune escape of tumour cell antigens. We selected five immune checkpoints PDCD1, TIM-3, LAG3, BTLA, and 4-1BB, and analyzed the correlation between their expression levels and the number of TCR clones. It is found that the number of TCR clones related to the expression number of different immune checkpoints for different cancer types. Therefore, based on the number of TCR clones, the therapeutic effect of the immune checkpoint inhibitor can be inferred.

In conclusion, our results suggest possible complementary methods or companion diagnostic methods of tumour immunotherapy. At present, the impacts of only using TCR to modify T cells or using anti-immune checkpoint drugs are not very satisfactory. Based on the heterogeneity of TCR in different cancers and the diversity of BCR in tumour and adjacent tissues, the strategy of tumour immunotherapy can be probably more precise.

Materials And Methods

Data collection and preprocessing

RNA-sequencing data were downloaded from SRA in NCBI, including 79 pairs of cancerous and adjacent noncancerous tissues samples of four representative cancer types, based on the literature[32–34] (Table 3). CutAdapt[35] software was adopted to remove adapters, barcode bases, and low-quality bases.

Table 3
Basic information of the sample

Accession code	Cancer type	Pairs of samples
PRJDB4239	COAD	3
PRJNA288518	COAD	5
PRJNA344450	LUAD	27
PRJNA435914	STAD	34
PRJNA490335	PAAD	10

Extraction of immune repertoire

The MiXCR[36] software package was employed to extract immune repertoire receptor sequences. In the process of extracting TCR or BCR by MiXCR, to obtain more assembled reads containing full CDR3 sequence, two iterations of reads assembling and one germline sequences extend step were added to avoid the false positives, which may be caused by the randomly interruption of RNA reads during the subsequent comparison analysis.

Gene expression analysis

The HISTA2[37] software was used to map the reads to the human reference genome hg38 to obtain the SAM files. Samtools software[38] was used to convert the SAM file into a BAM file and sort reads according to the location of the chromosome. Finally, StringTie[37] software was used to process the upstream data to get the FPKM values of gene expression.

Statistical analysis and visualization

After using MiXCR to obtain the TCR and BCR sequences, VDJTools[39] and tcR were used to analyze the sequences. To eliminate the batch effect caused by the depth of sequencing, the number of clones were measured in each 10 million reads. Subsequently, VDJTools software package was adopted to analyze and count the obtained clones. The tcR[40] software package was used to count the number of shared TRB sequences among the samples in the group and calculate the Shannon entropies. The pheatmap package was used to draw heat maps. The ggpubr package was used to draw a box plot with rank-sum tests and to perform the correlation analysis.

Abbreviations

TCR T cell receptors BCR B cell receptors.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

All the sequencing data used in this manuscript were downloaded from NCBI SRA database as listed in table 3 in Materials and methods section.

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the State Key Laboratory of Pathogen and Biosecurity (Academy of Military Medical Science) SKLPBS2138 and Shenzhen Science and Technology Project (JCYJ20180507183842516)

Authors' contributions

Bo Li, Jing Zhang, Huixue Tang collected the clinical samples and performed the data analysis. Wenxiang Hu verified the analytical methods. Lida Xu and Yu-bao Chen designed the study. Jin-dan Guo drafted the manuscript with the help of Lida Xu. All authors read and approved the final manuscript.

Acknowledgments

This work was supported by the National Science and Technology Major Project during the 13th 5-Year Plan Period (project number: 2019ZX09721001-007-002) and Shenzhen Science and Technology Project (project number: JCYJ20180507183842516).

References

1. Robins H: **Immunosequencing: applications of immune repertoire deep sequencing.** *Curr Opin Immunol* 2013, **25**(5):646–652.
2. Samelson LE: **Signal transduction mediated by the T cell antigen receptor: the role of adapter proteins.** *Annu Rev Immunol* 2002, **20**:371–394.

3. Zou HY, Yu WZ, Zhang Q, Yang HC, Huang HY, Jiao M: **Preliminary study of the clonal characteristics of the TCR BV subfamilies in T cells in the peripheral blood from patients with uveitis.** *Genet Mol Res* 2014, **13**(1):1296–1303.
4. Vantourout P, Hayday A: **Six-of-the-best: unique contributions of gammadelta T cells to immunology.** *Nat Rev Immunol* 2013, **13**(2):88–100.
5. Vanhanen R, Heikkila N, Aggarwal K, Hamm D, Tarkkila H, Patila T, Jokiranta TS, Saramaki J, Arstila TP: **T cell receptor diversity in the human thymus.** *Mol Immunol* 2016, **76**:116–122.
6. Hayday AC: **[gamma][delta] cells: a right time and a right place for a conserved third way of protection.** *Annu Rev Immunol* 2000, **18**:975–1026.
7. Niiro H, Clark EA: **Regulation of B-cell fate by antigen-receptor signals.** *Nat Rev Immunol* 2002, **2**(12):945–956.
8. Cadera EJ, Wan F, Amin RH, Nolla H, Lenardo MJ, Schlissel MS: **NF-kappaB activity marks cells engaged in receptor editing.** *J Exp Med* 2009, **206**(8):1803–1816.
9. Wucherpfennig KW, Gagnon E, Call MJ, Huseby ES, Call ME: **Structural biology of the T-cell receptor: insights into receptor assembly, ligand recognition, and initiation of signaling.** *Cold Spring Harb Perspect Biol* 2010, **2**(4):a005140.
10. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR: **The promise and challenge of high-throughput sequencing of the antibody repertoire.** *Nat Biotechnol* 2014, **32**(2):158–168.
11. Meffre E, Wardemann H: **B-cell tolerance checkpoints in health and autoimmunity.** *Curr Opin Immunol* 2008, **20**(6):632–638.
12. Schwartz RS: **Shattuck lecture: Diversity of the immune repertoire and immunoregulation.** *N Engl J Med* 2003, **348**(11):1017–1026.
13. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P: **A direct estimate of the human alphabeta T cell receptor diversity.** *Science* 1999, **286**(5441):958–961.
14. Chitgopeker P, Sahni D: **T-cell receptor gene rearrangement detection in suspected cases of cutaneous T-cell lymphoma.** *J Invest Dermatol* 2014, **134**(4):1–5.
15. Kedzierska K, La Gruta NL, Stambas J, Turner SJ, Doherty PC: **Tracking phenotypically and functionally distinct T cell subsets via T cell repertoire diversity.** *Mol Immunol* 2008, **45**(3):607–618.
16. McLean-Tookey A, Barge D, Spickett GP, Gennery AR: **Flow cytometric analysis of TCR Vbeta repertoire in patients with 22q11.2 deletion syndrome.** *Scand J Immunol* 2011, **73**(6):577–585.
17. Wang CY, Fang YX, Chen GH, Jia HJ, Zeng S, He XB, Feng Y, Li SJ, Jin QW, Cheng WY *et al*: **Analysis of the CDR3 length repertoire and the diversity of T cell receptor alpha and beta chains in swine CD4+ and CD8+ T lymphocytes.** *Mol Med Rep* 2017, **16**(1):75–86.
18. Lim A, Baron V, Ferradini L, Bonneville M, Kourilsky P, Pannetier C: **Combination of MHC-peptide multimer-based T cell sorting with the Immunoscope permits sensitive ex vivo quantitation and follow-up of human CD8+ T cell immune responses.** *J Immunol Methods* 2002, **261**(1-2):177–194.

19. Robins HS, Campregher PV, Srivastava SK, Wachter A, Turtle CJ, Kahsai O, Riddell SR, Warren EH, Carlson CS: **Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells.** *Blood* 2009, **114**(19):4099–4107.
20. Weinstein JA, Jiang N, White RA, 3rd, Fisher DS, Quake SR: **High-throughput sequencing of the zebrafish antibody repertoire.** *Science* 2009, **324**(5928):807–810.
21. Luo W, Liao WJ, Ma L, Huang YT, Shi M, Wen Q, Wang XN: **Dynamic monitoring the TCR CDR3 spectratypes in patients with metastatic CRC treated with a combination of bevacizumab, irinotecan, fluorouracil, and leucovorin.** *Cancer Immunol Immunother* 2010, **59**(2):247–256.
22. Woodsworth DJ, Castellarin M, Holt RA: **Sequence analysis of T-cell repertoires in health and disease.** *Genome Med* 2013, **5**(10):98.
23. Brown SD, Raeburn LA, Holt RA: **Profiling tissue-resident T cell repertoires by RNA sequencing.** *Genome Med* 2015, **7**:125.
24. Kim BJ, Youn DH, Kim Y, Jeon JP: **Characterization of the TCR beta Chain CDR3 Repertoire in Subarachnoid Hemorrhage Patients with Delayed Cerebral Ischemia.** *Int J Mol Sci* 2020, **21**(9).
25. Li B, Li T, Pignon JC, Wang B, Wang J, Shukla SA, Dou R, Chen Q, Hodi FS, Choueiri TK *et al*: **Landscape of tumor-infiltrating T cell repertoire of human cancers.** *Nat Genet* 2016, **48**(7):725–732.
26. Zvyagin IV, Tsvetkov VO, Chudakov DM, Shugay M: **An overview of immunoinformatics approaches and databases linking T cell receptor repertoires to their antigen specificity.** *Immunogenetics* 2020, **72**(1-2):77–84.
27. Robbins PF, Kassim SH, Tran TL, Crystal JS, Morgan RA, Feldman SA, Yang JC, Dudley ME, Wunderlich JR, Sherry RM *et al*: **A pilot trial using lymphocytes genetically engineered with an NY-ESO-1-reactive T-cell receptor: long-term follow-up and correlates with response.** *Clin Cancer Res* 2015, **21**(5):1019–1027.
28. Sherwood AM, Emerson RO, Scherer D, Habermann N, Buck K, Staffa J, Desmarais C, Halama N, Jaeger D, Schirmacher P *et al*: **Tumor-infiltrating lymphocytes in colorectal tumors display a diversity of T cell receptor sequences that differ from the T cells in adjacent mucosal tissue.** *Cancer Immunol Immunother* 2013, **62**(9):1453–1461.
29. Wang X, Zhang B, Yang Y, Zhu J, Cheng S, Mao Y, Feng L, Xiao T: **Characterization of Distinct T Cell Receptor Repertoires in Tumor and Distant Non-tumor Tissues from Lung Cancer Patients.** *Genomics Proteomics Bioinformatics* 2019, **17**(3):287–296.
30. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, Powderly JD, Carvajal RD, Sosman JA, Atkins MB *et al*: **Safety, activity, and immune correlates of anti-PD-1 antibody in cancer.** *N Engl J Med* 2012, **366**(26):2443–2454.
31. Hopkins AC, Yarchoan M, Durham JN, Yusko EC, Rytlewski JA, Robins HS, Laheru DA, Le DT, Lutz ER, Jaffee EM: **T cell receptor repertoire features associated with survival in immunotherapy-treated pancreatic ductal adenocarcinoma.** *JCI Insight* 2018, **3**(13).
32. Lee JR, Kwon CH, Choi Y, Park HJ, Kim HS, Jo HJ, Oh N, Park do Y: **Transcriptome analysis of paired primary colorectal carcinoma and liver metastases reveals fusion transcripts and similar gene**

- expression profiles in primary carcinoma and liver metastases. *BMC Cancer* 2016, **16**:539.
33. Lin J, Wu YJ, Liang X, Ji M, Ying HM, Wang XY, Sun X, Shao CH, Zhan LX, Zhang Y: **Network-based integration of mRNA and miRNA profiles reveals new target genes involved in pancreatic cancer.** *Mol Carcinog* 2019, **58**(2):206–218.
34. Sun Z, Wang L, Eckloff BW, Deng B, Wang Y, Wampfler JA, Jang J, Wieben ED, Jen J, You M *et al*: **Conserved recurrent gene mutations correlate with pathway deregulation and clinical outcomes of lung adenocarcinoma in never-smokers.** *BMC Med Genomics* 2014, **7**:32.
35. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *Embnet Journal* 2011, **17**(1).
36. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, Chudakov DM: **MiXCR: software for comprehensive adaptive immunity profiling.** *Nat Methods* 2015, **12**(5):380–381.
37. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL: **Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown.** *Nat Protoc* 2016, **11**(9):1650–1667.
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**(16):2078–2079.
39. Shugay M, Bagaev DV, Turchaninova MA, Bolotin DA, Britanova OV, Putintseva EV, Pogorelyy MV, Nazarov VI, Zvyagin IV, Kirgizova VI *et al*: **VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires.** *PLoS Comput Biol* 2015, **11**(11):e1004503.
40. Nazarov VI, Pogorelyy MV, Komech EA, Zvyagin IV, Bolotin DA, Shugay M, Chudakov DM, Lebedev YB, Mamedov IZ: **tcR: an R package for T cell receptor repertoire advanced data analysis.** *BMC Bioinformatics* 2015, **16**:175.

Figures

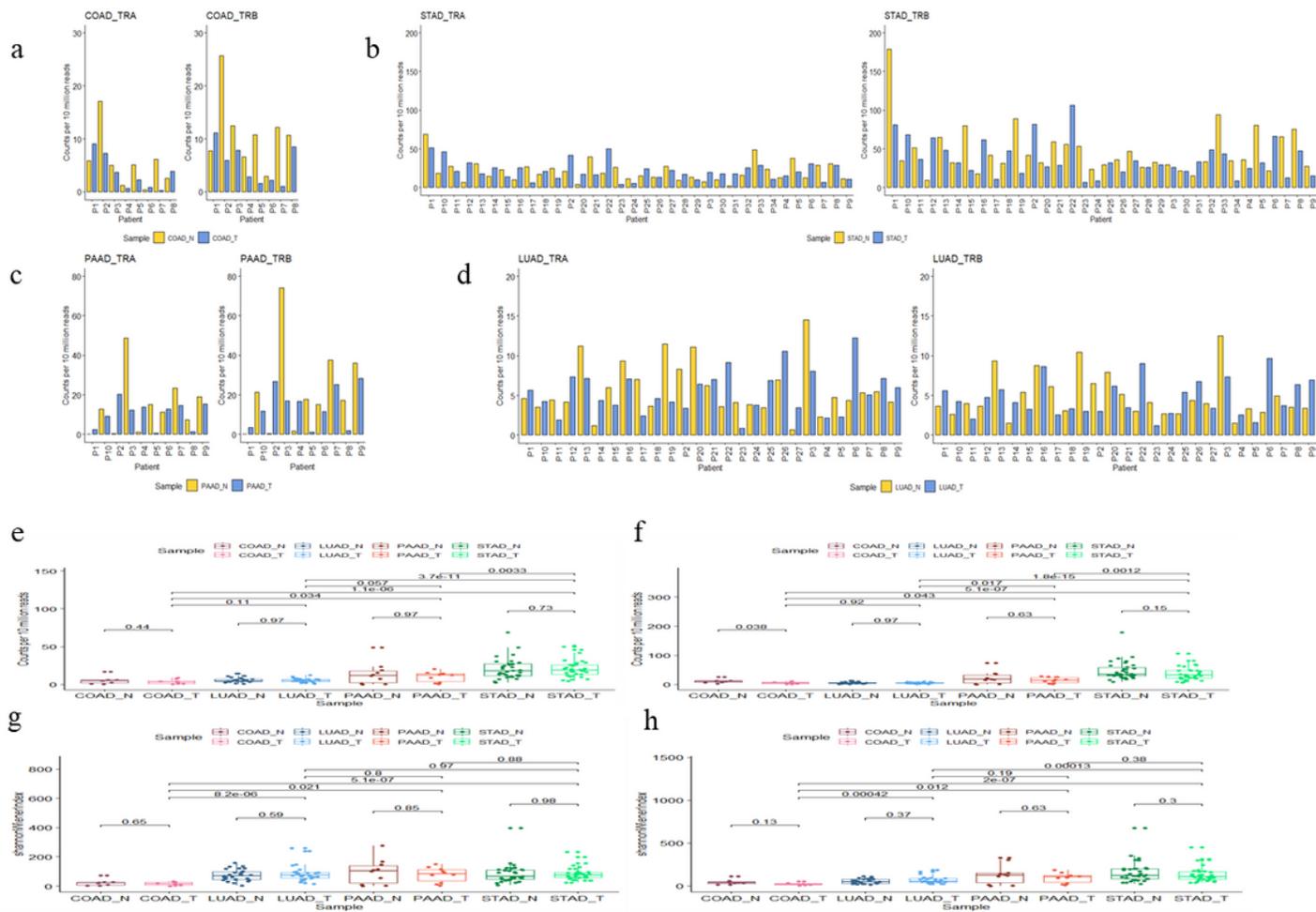


Figure 1

Overall description of TCR. a-d. Statistics of the number of TRA and TRB in each sample. e-f. Comparison of the number of TRA (e) and TRB (f) clones. g-h. Comparison of Shannon entropies of TRA (g) and TRB (h).

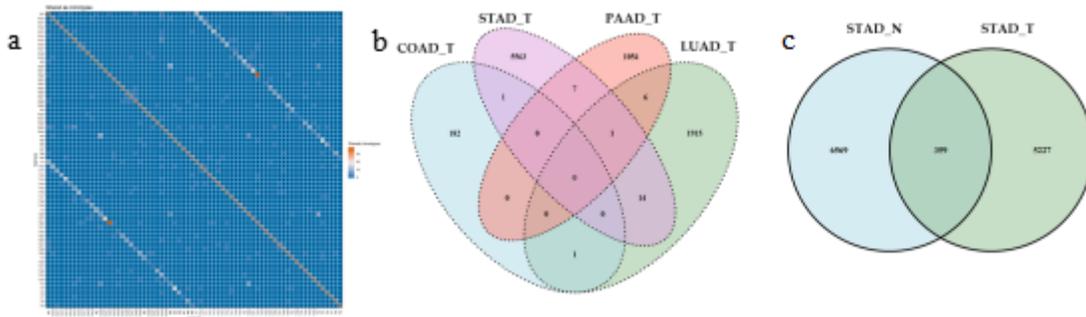


Figure 2

Analysis of shared TRB sequence of different samples. a. The distribution of the shared TRB CDR3 sequences between every two samples in STAD. b. The number of TRB sequences shared among different cancer tissues. c. The number of shared TRB between STAD cancer and adjacent tissues.

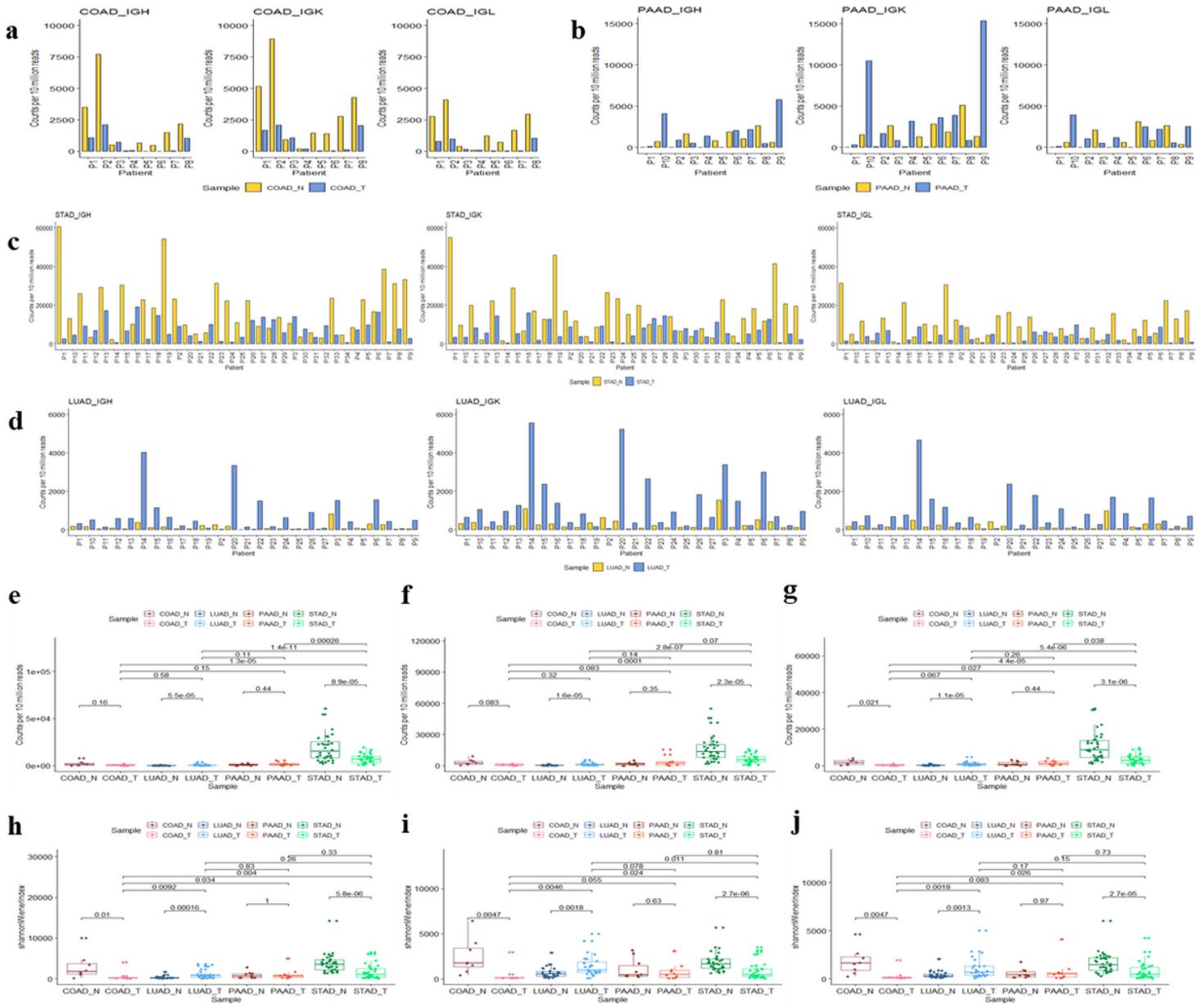
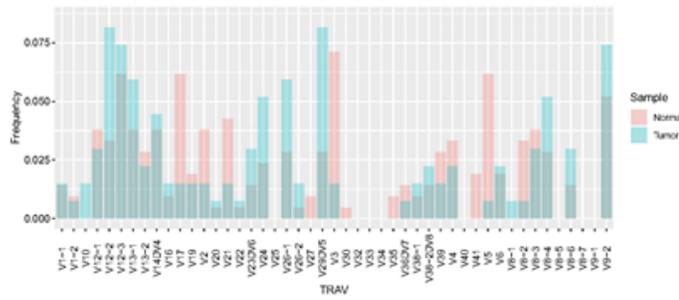


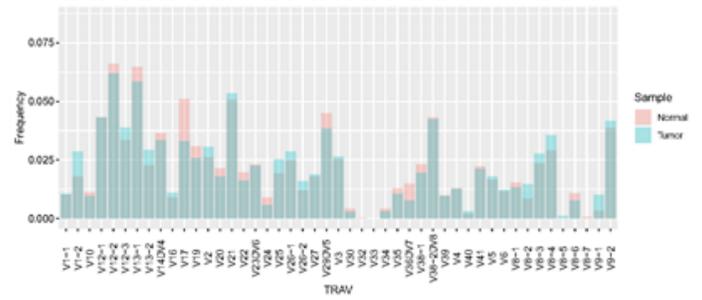
Figure 3

Overall description of BCR. a-d. Statistics of the number of light and heavy chains in each sample. e-g. Comparison of the number of IGH (e), IGK (f), and IGL(g). h-j. Comparison of Shannon entropies of IGH (h), IGK (i), and IGL (j).

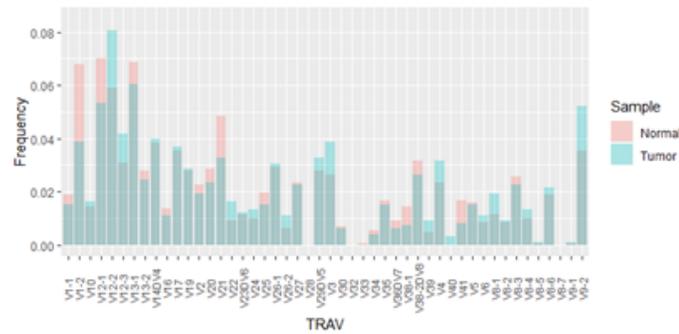
COAD



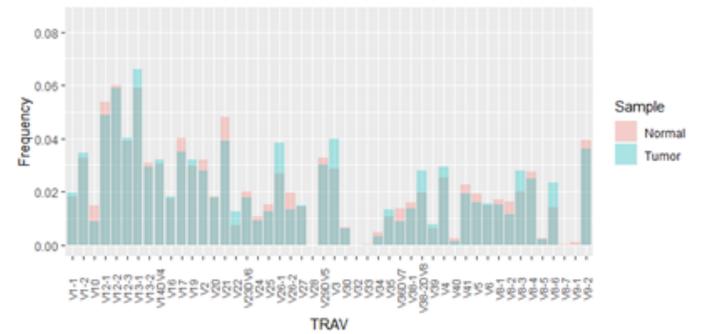
LUAD



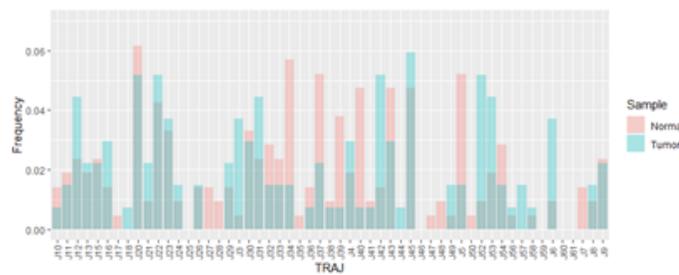
PAAD



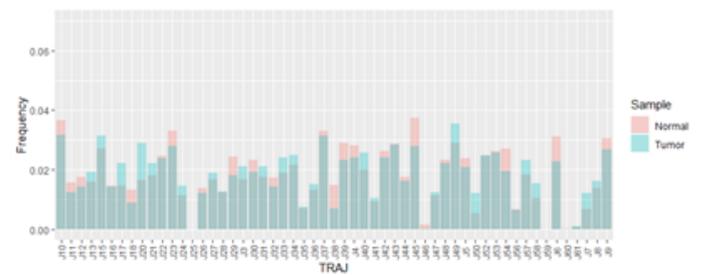
STAD



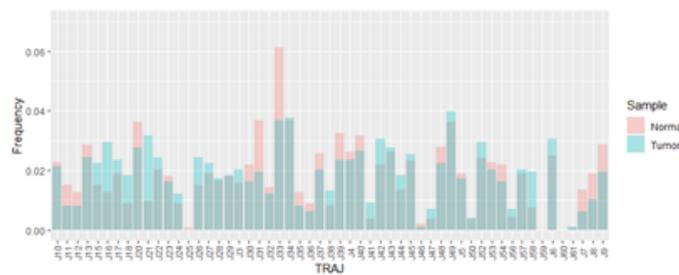
COAD



LUAD



PAAD



STAD

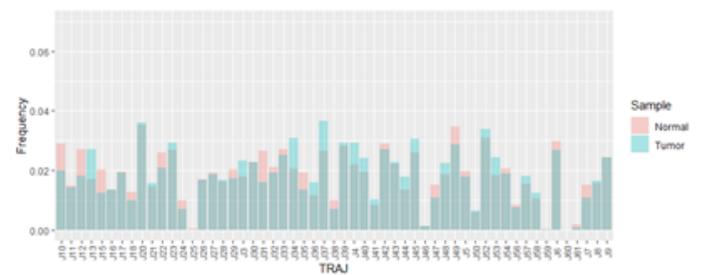


Figure 4

The usage of TRAV and TRAJ. The gray-green parts represent the overlap of the frequency of normal tissue and tumour tissue. The pink(or the blue) bars represent a certain V or J region was preferential to be used in normal(or tumour) tissues.

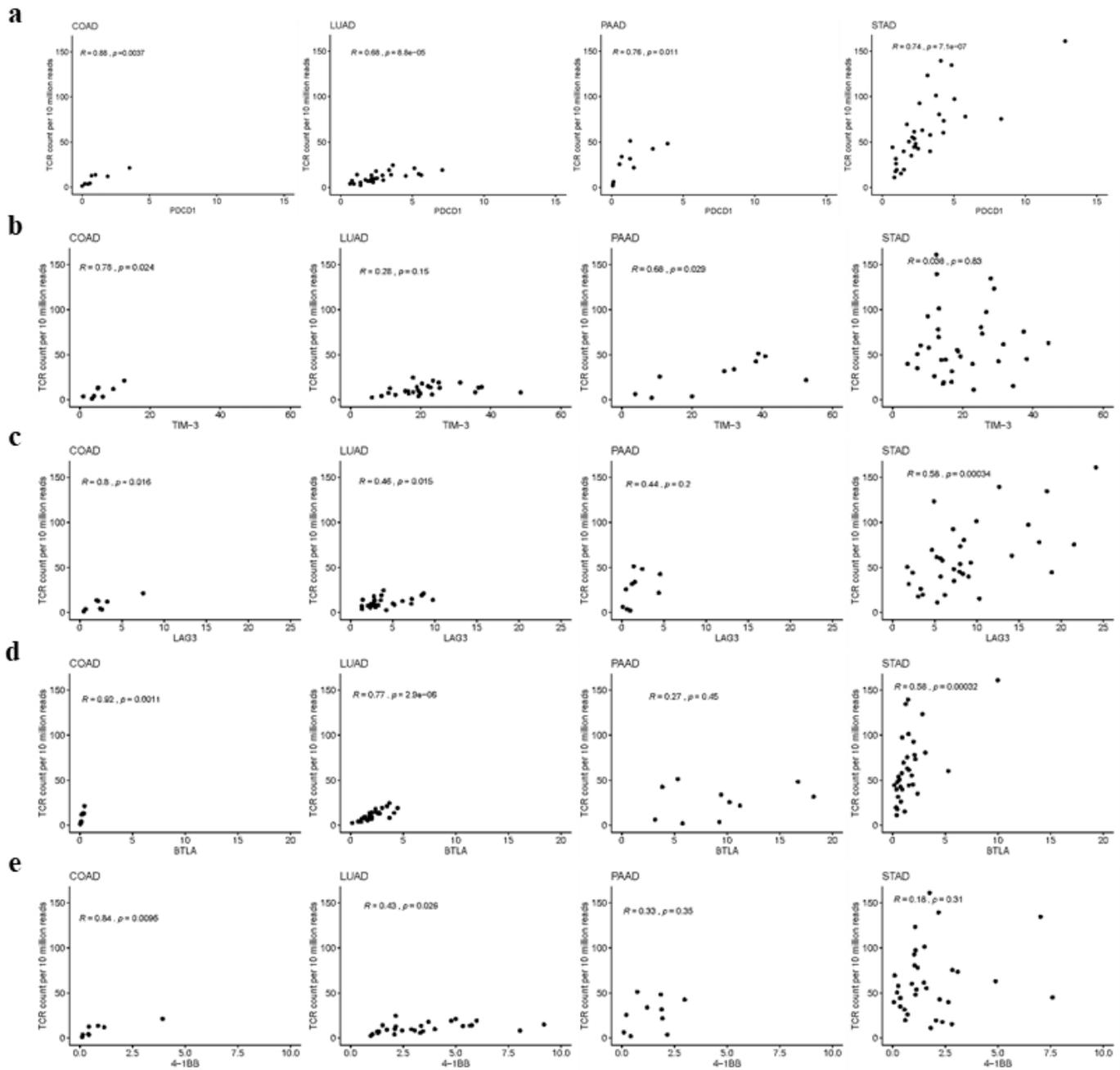


Figure 5

Correlation between the expression levels of immune checkpoints and the clone number of TCR

Supplementary Files

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

- [Supplementarymaterial.docx](#)