# An introduction of TBAdb database

TBAdb is a manually curated database of T-cell receptor (TCR) and B-cell receptor (BCR) targeting specific antigen or diseases.

The database contains three parts, namely TCR-AB, TCR-GD and BCR. These three parts are aimed at collecting sequences and specificity information of TCRA and TCRB, TCR- gamma and TCR-delta and BCR separately.

Our mission is to present the accurate relationship between TCR/BCR and antigens of diseases in a humanized manner. Therefore, in addition to a broad collection and irregularities correction of articles (early since 1990), many other processes also contribute to the goal:

* The detailed information, including disease ICD code, sequencing platform, the method used to identify the specificities and so on, is collected.
* The exporting of V/J fragments and CDR3 sequences have been double-checked manually to ensure them stand and correct.
* In order to help users to obtain specificities of sequence, the experiments in article are grading as showed in the tutorial document.
* The code of V/J fragments are standardized following the IMGT to our best.
* The format of CDR3 sequences are ruled with starting with (YX)**C** in V gene region and ending with **W or F**(GXG) in the J gene region.

The website provides a series of convenient tools to browse, query, stat and visualize the information. A tutorial document is open to download; it will help you to unitize the database.

1. **ICDname:** the disease ID of International Classification of Disease (ICD) 10th, which the sequence is correlated to.
2. **Disease.name:** disease which sequences are special to
3. **Category**: the category of the disease
4. **Antigen**: the antigen which the TCR sequence is specific to.
5. **Antigen.sequence**: the amino acid sequence of antigen
6. **HLA:** the type of leukocyte antigen (HLA) of the sample.
7. **Locus:** the single chain, such as TRA, TRB, TRD, TRG, IGH, IGK, IGL.
8. **CDR3.alpha.aa:** the amino acid sequence of complementarity determining region 3 (CDR3) for the TCR sequence alpha chain
9. **CDR3.alpha.nt:** the nucleotide sequence of complementarity determining region 3 (CDR3) for the TCR sequence alpha chain
10. **CDR3.****beta.aa:** the amino acid sequence of complementarity determining region 3 (CDR3) for the TCR sequence beta chain
11. **CDR3.beta.nt:** the nucleotide sequence of complementarity determining region 3 (CDR3) for the TCR sequence beta chain
12. **CDR3.gamma.aa:** the amino acid sequence of complementarity determining region 3 (CDR3) for the TCR sequence gamma chain
13. **CDR3.gamma.nt:** the nucleotide sequence of complementarity determining region 3 (CDR3) for the TCR sequence gamma chain
14. **CDR3.delta.aa:** the amino acid sequence of complementarity determining region 3 (CDR3) for the TCR sequence delta chain
15. **CDR3.delta.nt:** the nucleotide sequence of complementarity determining region 3 (CDR3) for the TCR sequence delta chain
16. **CDR3.light.aa:** the amino acid sequence of complementarity determining region 3 (CDR3) for the BCR sequence light chain
17. **CDR3.light.nt:** the nucleotide sequence of complementarity determining region 3 (CDR3) for the BCR sequence light chain
18. **CDR3.heavy.aa:** the amino acid sequence of complementarity determining region 3 (CDR3) for the BCR sequence heavy chain
19. **CDR3.heavy.nt:** the nucleotide sequence of complementarity determining region 3 (CDR3) for the BCR sequence heavy chain
20. **Species:** the specie in which the TCR sequence is found (ig. Human, mice)
21. **Valpha:** identified V(variable) gene segment locus on alpha chain. The reference genes are derived from IMGT database
22. **Jalpha:** the id of identified V(joining) gene segment locus on alpha chain. The reference genes are derived from IMGT database
23. **Vbeta:** identified V(variable) gene segment for the sequence. The reference genes are derived from IMGT database
24. **Vbeta:** identified V(variable) gene segment locus on beta chain. The reference genes are derived from IMGT database
25. **Dbeta:** identified D(diversity) gene segment locus on beta chain. The reference genes are derived from IMGT database
26. **Jbeta:** the id of identified V(joining) gene segment locus on beta chain. The reference genes are derived from IMGT database
27. **Seq.platform:** the sequencing platform (eg. Sanger, ABI3100, ….)
28. **Species:** the species of samples
29. **Origin:** the tissue from which the sample obtained
30. **Nucleotide.type:** DNA or RNA, the material used for sequencing
31. **Cell\_subtype:** the cell type used for sequencing, CD8+ T cells, CD4+T cells or total T cells.
32. **Prepare.method:** the method used for librarying
33. **Evaluate.method:** the assays are used for identifying the specificity of TCR to the antigen or disease
34. **Case.num:** how many samples from patients are studied
35. **Control.type:** the control sample used in the study
36. **Control.num:** how many samples from control individuals are used in this study.
37. **Filteration:** the threshold for filting
38. **Article.name:** the article from which the data get
39. **Pubmed\_id:** the id categorized by pubmed
40. **Grade:** the reliability of specificity of the TCR sequence.

**Evaluate.method and grade**

There are four classes of methods been used to evaluate the specificity between sequences and antigens; These methods graded by their accuracy.

1. **Selection of antigen-specific T cells using peptide-MHC multimers**. peptide-MHC multimers are useful to identify the antigen-specific directly. If specificity of sequences in an article were evaluated by this method, **4 points** will be given.
2. **Antigen-specific ex vivo proliferation** AND **Antigen-specific stimulation ex vivo**. They are classical methods, that stimulate cells ex vivo with antigen to proliferate them or just detect the proteins expressed by antigen-specific cells. **5 points**.
3. Binding assay. These methods, containing SPR, ITC, MST and so on, are used to quantify the affinity between TCR/BCR and antigen. **4 points**.
4. Laser microdissection. V/J specific antibodies are used to stain each type V and/or J motifs, then Laser microdissection is employed to detect them on tissue or cells. **1 point.**
5. Statistical analysis. Actually, high frequency sequences or public clones are often considered as disease-specific clones. **2 points.**
6. **If more than two methods are employed, the points will be added up.**