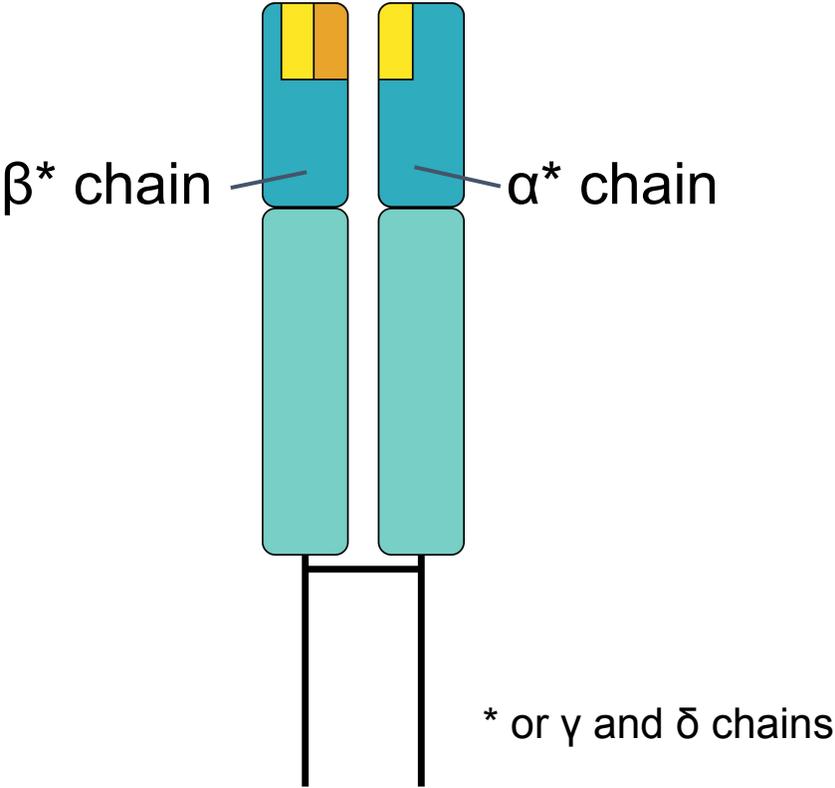# Receptor Data in the IEDB

iedb.org

Presented by: Lonneke Scheffer, Bioinformatics Postdoctoral Fellow

Adaptive Immune Receptors Bind Antigen Epitopes
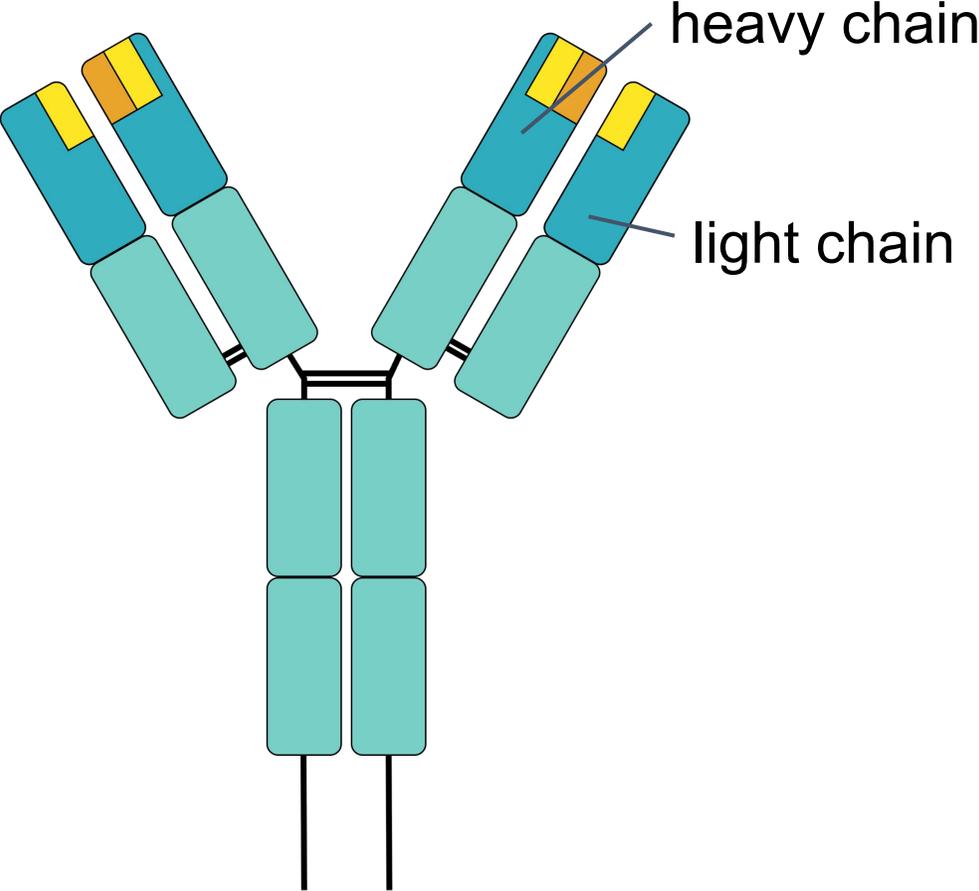

Immune receptor

Epitope

Antigen
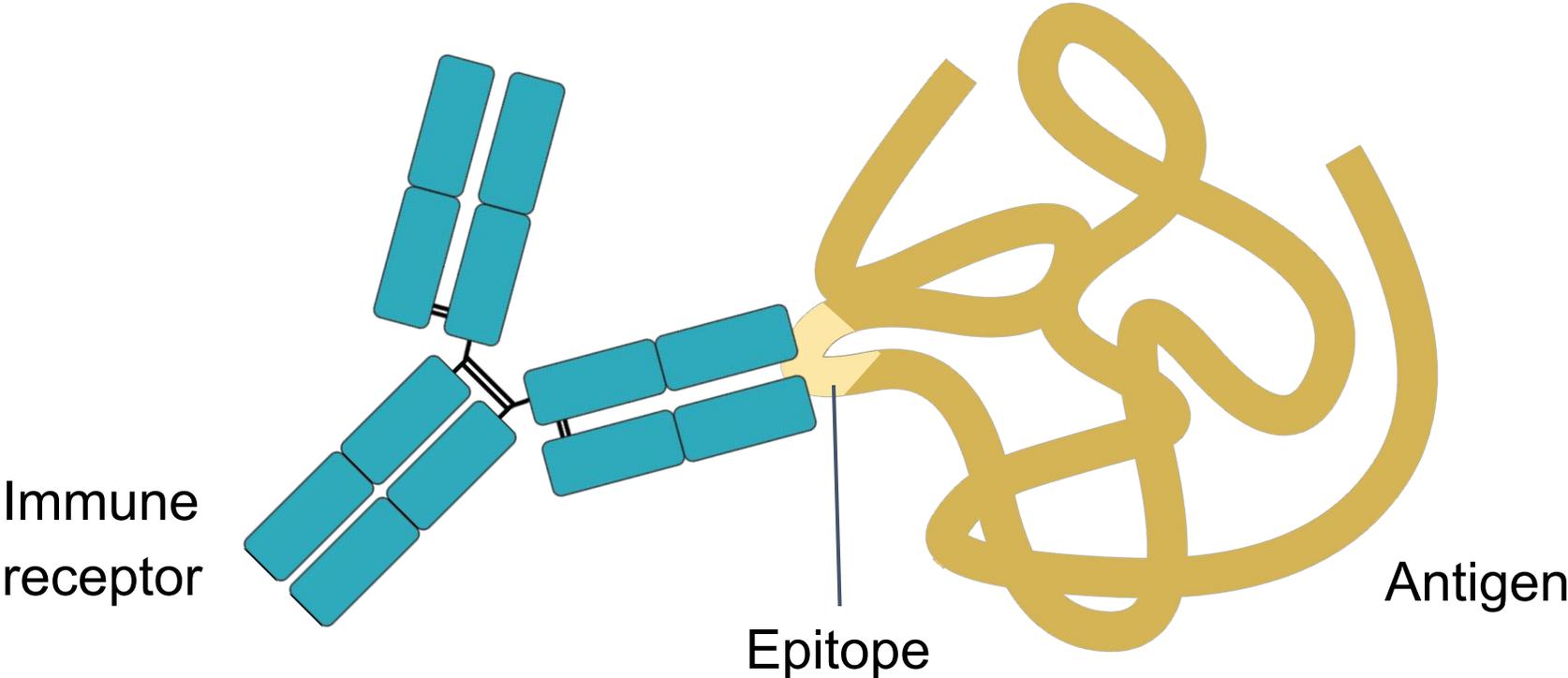
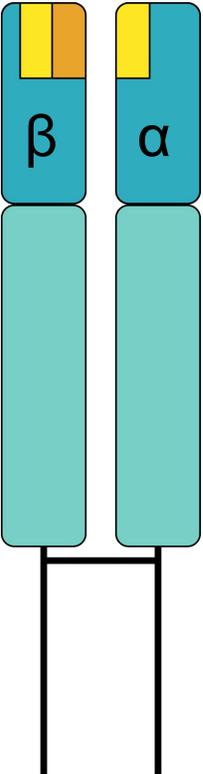# Adaptive Immune Receptors

T cell receptor (TCR)

B cell receptor (BCR)

β* chain

α* chain

heavy chain

light chain

* or γ and δ chains

Adaptive Immune Receptors

Adaptive Immune Receptors

Complementarity determining regions (**CDRs**), especially **CDR3β/heavy**, are most important for antigen binding

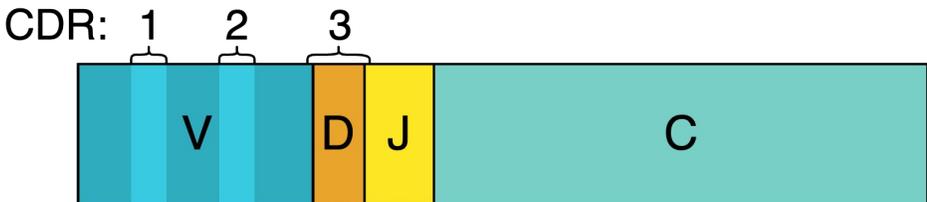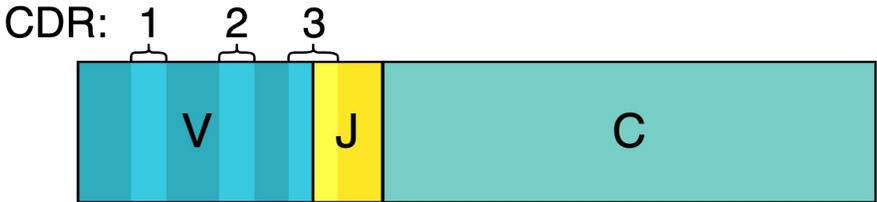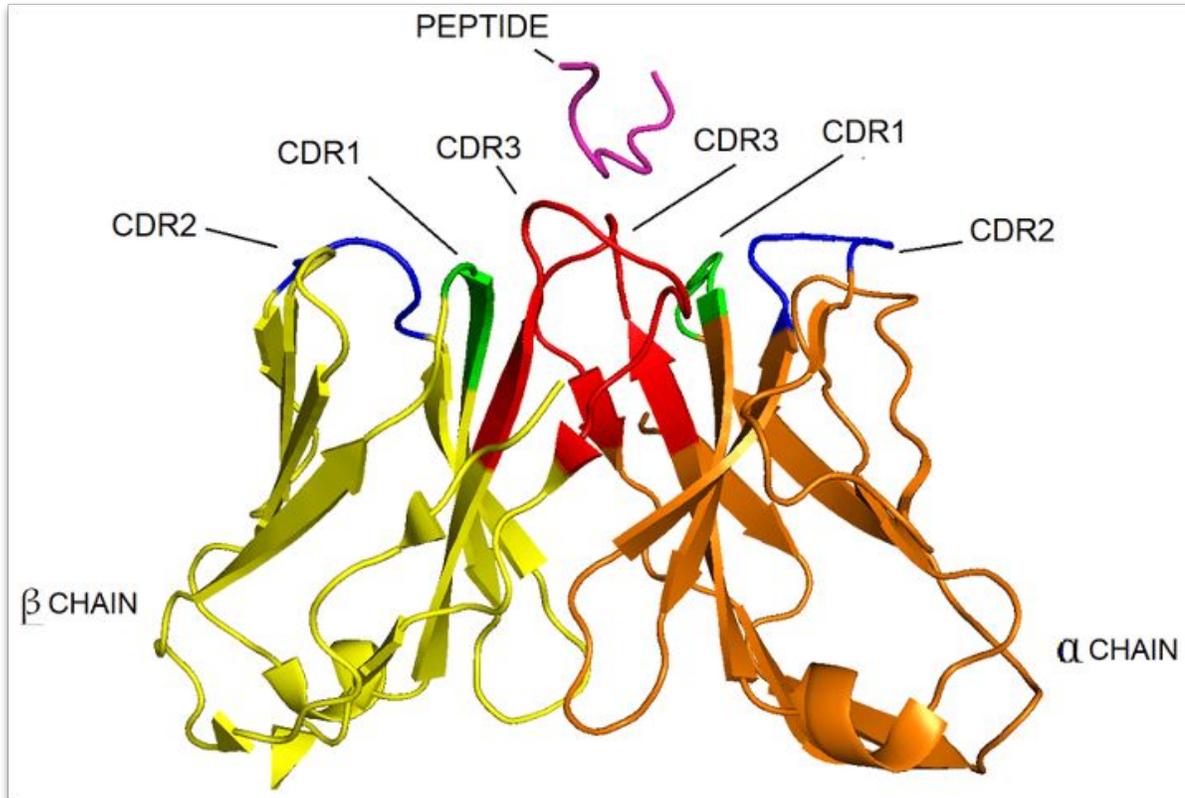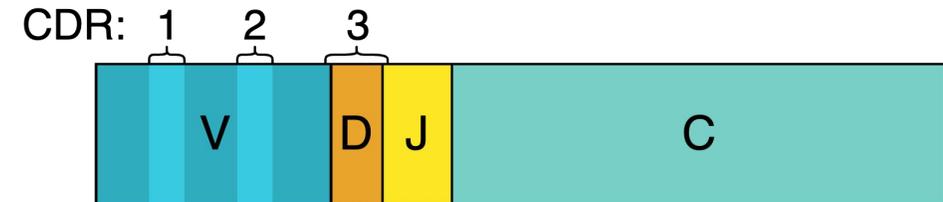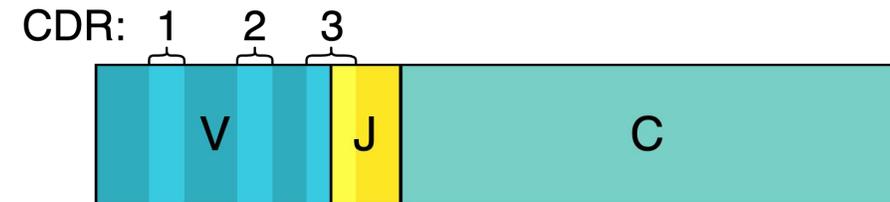Cinelli, Mattia. (2018). Analysis of murine CDR3β repertoires using machine learning techniques.
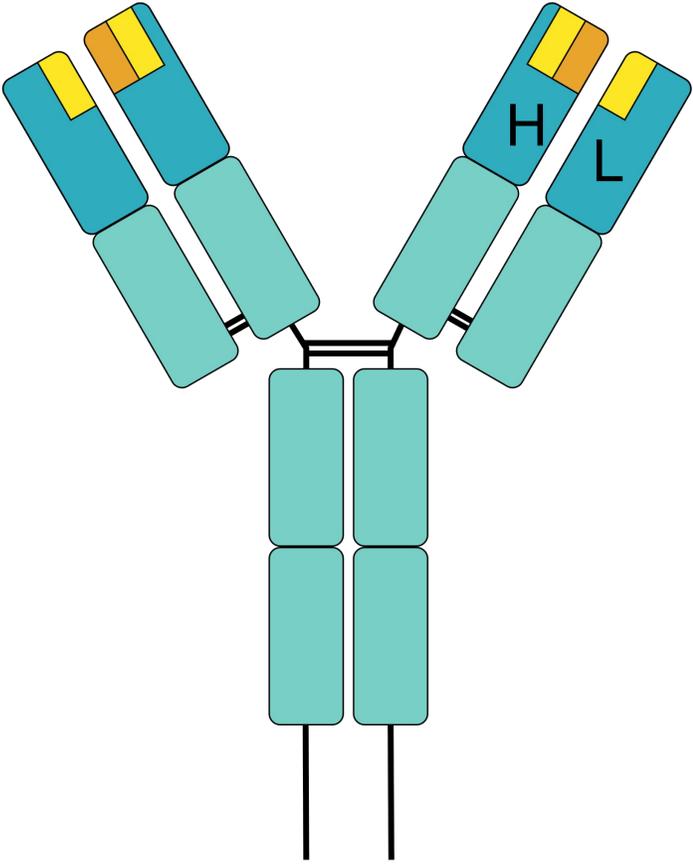
BCR heavy / TCR β

BCR light / TCR α

Receptor Data Currently in the IEDB

Receptor Data Currently in the IEDB

189,978 TCRs
(+915 negative)

4,790 BCRs
(+77 negative)

Receptor Data Currently in the IEDB

Receptor Data Currently in the IEDB

Receptor Data Currently in the IEDB

Receptor group: these receptors share the same CDR3s

T cell receptor (receptor group ID 94)

Alpha beta TCR with alpha chain CDR3 of AVRPLLDGTYIPT and beta chain CDR3 of ASSYLGNTGELF was reported in Homo sapiens (human). This TCR has accessions for alpha chain 2P5E_D ⬈, 2P5W_D ⬈, 2PYE_D ⬈ and beta chain 2P5E_E ⬈, 2P5W_E ⬈, 2PYE_E ⬈ and was shown in 3D Structures 2P5E, 2P5W and 2PYE.

alpha		beta		Epitopes (# assays)
Gene usage	CDR sequences	Gene usage	CDR sequences	
V:TRAV21*01 D: J:TRAJ6*01	CDR1:DSAIYN CDR2:ITPWQRE CDR3:AVRPLLDGTYIPT	V:TRBV6-5*01 D: J:TRBJ2-2*01	CDR1:MNHEY CDR2:SVAIQT CDR3:ASSYLGNTGELF	SLLMWITQC (2)
V Domain: KQEVTQIPAALSVPEGENLVLNCS FTDSAI...		V Domain: GVTQTPKFQVLKTGQSMTLQCAQD MNHEYM...		
V:TRAV21*01 D: J:TRAJ6*01	CDR1:DSAIYN CDR2:ITPWQRE CDR3:AVRPLLDGTYIPT	V:TRBV6-5*01 D: J:TRBJ2-2*01	CDR1:MNHEY CDR2:SVSVGM CDR3:ASSYLGNTGELF	SLLMWITQC (2)
V Domain: KQQVTQIPAALSVPEGENLVLNCS FTDSAI...		V Domain: GVTQTPKFQVLKTGQSMTLQCAQD MNHEYM...		
V:TRAV21*01 D: J:TRAJ6*01	CDR1:DSAIYN CDR2:IQSSQRE CDR3:AVRPLLDGTYIPT	V:TRBV6-5*01 D: J:TRBJ2-2*01	CDR1:MNHEY CDR2:SVGAGT CDR3:ASSYLGNTGELF	SLLMWITQC (2)
V Domain: KQEVTQIPAALSVPEGENLVLNCS FTDSAI...		V Domain: GVTQTPKFQVLKTGQSMTLQCAQD MNHEYM...		

Epitope summary

This TCR was studied for the following epitope SLLMWITQC studied as part of Cancer/testis antigen 1 (UniProt:P78358) from Homo sapiens (human) (epitope ID 59278, 1 publication, 6 assays).

IEDB Receptor Data Updates

Web Interface Updates	Data Standardization	Collaborative Projects
• Home page • Receptor results tab • AIRR-format data export	• New standardization pipeline • Old vs new receptor data • Re-curation	• IMMREP2025 • AIRR Knowledge Commons

Web Interface Updates: Receptor Search Panel

Web Interface Updates: Receptor Search Panel

Detailed receptor search side panel: <u>already available</u> on CEDAR & IEDB

Web Interface Updates: Epitope Column

Web Interface Updates: Epitope Column

Note: multiple rows per "Receptor group"

The AIRR Data Standard for Immune Receptors

- We are adding AIRR export format to IEDB/CEDAR

- The Adaptive Immune Receptor Repertoire (AIRR) Community has developed a standard for immune receptor and repertoire data

AIRR Community Standardized Representations for Annotated Immune Repertoires

Jason Anthony Vander Heiden [1,†], Susanna Marquez [2], Nishanth Marthandan [3], Syed Ahmad Chan Bukhari [2], Christian E Busse [4], Brian Corrie [5], Uri Hershberg [6,7,8], Steven H Kleinstein [2,9], Frederick A Matsen IV [10], Duncan K Ralph [10], Aaron M Rosenfeld [6], Chaim A Schramm [11]; The AIRR Community[‡], Scott Christley [12,*,†], Uri Laserson [13,*]

▸ Author information   ▸ Article notes   ▸ Copyright and License information

PMCID: PMC6173121   PMID: 30323809

doi: 10.3389/fimmu.2018.02206

More Info: docs.airr-community.org

Web Interface Updates: AIRR Export

Web Interface Updates: AIRR Export

Web Interface Updates: AIRR Export

Web Interface Updates: AIRR Export

Web Interface Updates: AIRR Export

AIRR format download:

cell_id	reactivity_ref	iedb_epitope_id	iedb_epitope_name	iedb_host_specie	locus	v_call	d_call	j_call	cdr1_aa	cdr2_aa	cdr3_aa	junction_aa	sequence_aa
1023_530809	IEDB_RECEPTOR:802	IEDB_EPITOPE:530809	ALEPGPVTA	NCBITAXON:9606	TRA	TRAV21*01		TRAJ53*01	DSAIYN	IQSSQRE	AVLSSGGSNYKLT	CAVLSSGGSNYKLTF	MKQEVTQIPAALSVPEGE
1023_530809	IEDB_RECEPTOR:802	IEDB_EPITOPE:530809	ALEPGPVTA	NCBITAXON:9606	TRB	TRBV7-3*01		TRBJ2-3*01	SGHTA	FQGTGA	ASSFIGGTDTQY	CASSFIGGTDTQYF	GAGVSQTPSNKVTEKGK
9144_16878	IEDB_RECEPTOR:1950	IEDB_EPITOPE:16878	FLRGRAYGL	NCBITAXON:9606	TRA	TRAV14/DV4*01		TRAJ49*01			AMREDTGNQFY	CAMREDTGNQFYF	KITQTQPGMFVQEKEAVT
9144_16878	IEDB_RECEPTOR:1950	IEDB_EPITOPE:16878	FLRGRAYGL	NCBITAXON:9606	TRB	TRBV11-2*01		TRBJ2-3*01	SGHAT	FQNNGV	ASSFTWTSGGATDTQY	CASSFTWTSGGATDTQY	MGVAQSPRYKIIEKRQSV
9152_44920	IEDB_RECEPTOR:824	IEDB_EPITOPE:44920	NLVPMVATV	NCBITAXON:9606	TRA	TRAV26-2*01		TRAJ43*01	TISGTDY	GLTSN	ILDNNNDMR	CILDNNNDMRF	MDAKTTQPNSMESNEEE
9152_44920	IEDB_RECEPTOR:824	IEDB_EPITOPE:44920	NLVPMVATV	NCBITAXON:9606	TRB	TRBV7-6*01		TRBJ1-4*01	SGHVS	FNYEAQ	ASSLAPGTTNEKLF	CASSLAPGTTNEKLFF	MGAGVSQSPRYKVTKRG
9160_16878	IEDB_RECEPTOR:1412	IEDB_EPITOPE:16878	FLRGRAYGL	NCBITAXON:9606	TRA	TRAV26-2*01		TRAJ52*01					
9160_16878	IEDB_RECEPTOR:1412	IEDB_EPITOPE:16878	FLRGRAYGL	NCBITAXON:9606	TRB	TRBV7-8*01	TRBD1*01	TRBJ2-7*01	SGHVS	FQNEAQ	ASSLGQAYEQY	CASSLGQAYEQYF	MGVSQSPRYKVAKRGQI

Curated versus Calculated IEDB Receptor Data

Chain 1	Chain 1	Chain 1	Chain 1	Chain 1	Chain 1	Chain 1
Protein Sequence	CDR3 **Curated**	CDR3 **Calculated**	CDR3 Start **Curated**	CDR3 End **Curated**	CDR3 Start **Calculated**	CDR3 End **Calculated**
KTTQPPSMDCAEGRAAN		IVVRSSNTGKLI			87	98
KTTQPPSMDCAEGRAAN		IVVRSSNTGKLI			87	98

Chain 1	Chain 1	Chain 1	Chain 1	Chain 1	Chain 1
Curated V Gene	**Calculated** V Gene	**Curated** D Gene	**Calculated** D Gene	**Curated** J Gene	**Calculated** J Gene
	TRAV26-1*01				TRAJ37*01
TRAV26-1*01	TRAV26-1*01			TRAJ37*01	TRAJ37*01

Curated versus Calculated IEDB Receptor Data
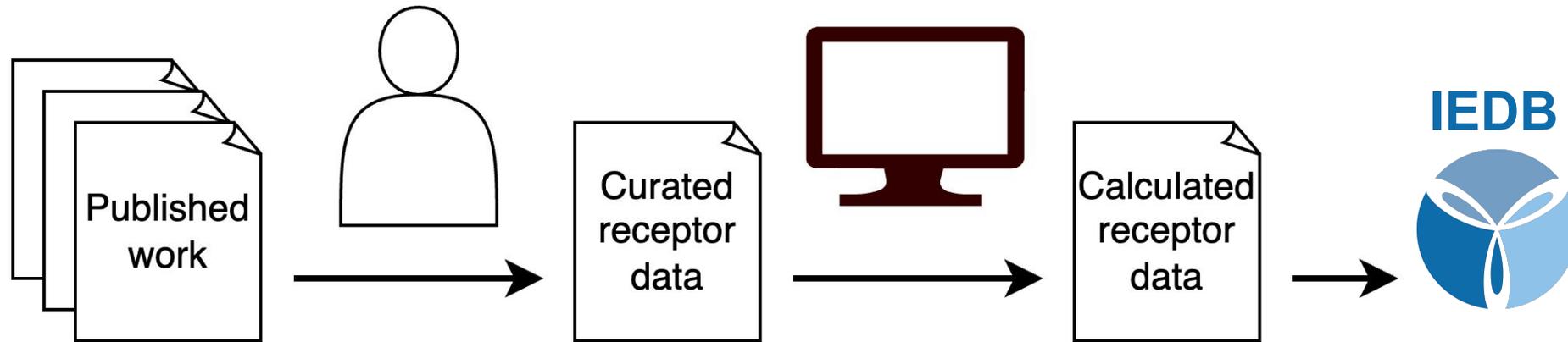
- **Curated** fields are heterogeneous, reported <u>as published</u>
- **Calculated** fields provide data in standardized, computer-readable format

Curated versus Calculated IEDB Receptor Data
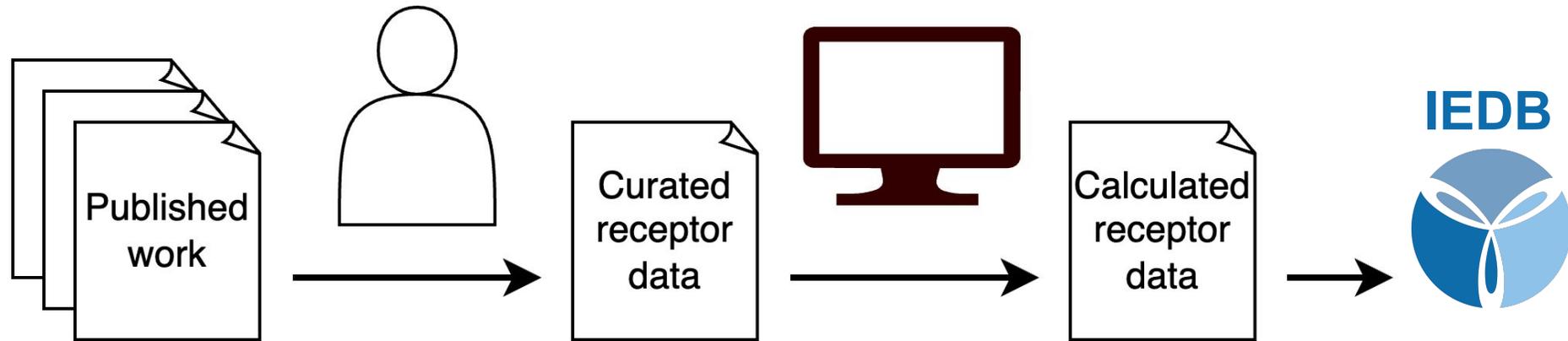
- **Curated** fields are heterogeneous, reported <u>as published</u>
- **Calculated** fields provide data in standardized, computer-readable format

We are revising the data to ensure all *calculated* fields are complete and ready-to-use

Available *Curated* Data for all Receptors in the IEDB

	TCR	BCR
Total receptors (receptor groups expanded)	225,983	8,639
Chain 1 has full nt	97	87
Chain 2 has full nt	717	89
Chain 1 has full aa	1,557	6,208
Chain 2 has full aa	1,661	5,300
Chain 1 has curated CDR3 and/or V/J (no full nt/aa)	62,894	2,291
Chain 2 has curated CDR3 and/or V/J (no full nt/aa)	192,326	1,366

Available *Curated* Data for all Receptors in the IEDB

	TCR	BCR	
Total receptors (receptor groups expanded)	225,983	8,639	
Chain 1 has full nt	97	87	
Chain 2 has full nt	717	89	
Chain 1 has full aa	1,557	6,208	mostly full aa sequence (often from PDB 3D structure)
Chain 2 has full aa	1,661	5,300	
Chain 1 has curated CDR3 and/or V/J (no full nt/aa)	62,894	2,291	
Chain 2 has curated CDR3 and/or V/J (no full nt/aa)	192,326	1,366	

Available *Curated* Data for all Receptors in the IEDB

	TCR	BCR
Total receptors (receptor groups expanded)	225,983	8,639
Chain 1 has full nt	97	87
Chain 2 has full nt	717	89
Chain 1 has full aa	1,557	6,208
Chain 2 has full aa	1,661	5,300
Chain 1 has curated CDR3 and/or V/J (no full nt/aa)	62,894	2,291
Chain 2 has curated CDR3 and/or V/J (no full nt/aa)	192,326	1,366

mostly CDR3 + gene names

Updated *Calculated* Data Standardization Pipeline

Curated data

Standardization pipeline

Calculated data

Receptor/chain type, species +
- full nt sequence or
- full aa sequence or
- V/D/J genes, CDRs

IgBLAST nt

IgBLAST aa + ANARCII

tidytcells

custom scripts

Standardized and validated:
- receptor/chain type
- V/D/J genes
- CDRs + junction

Anomalies

Flag for re-curation
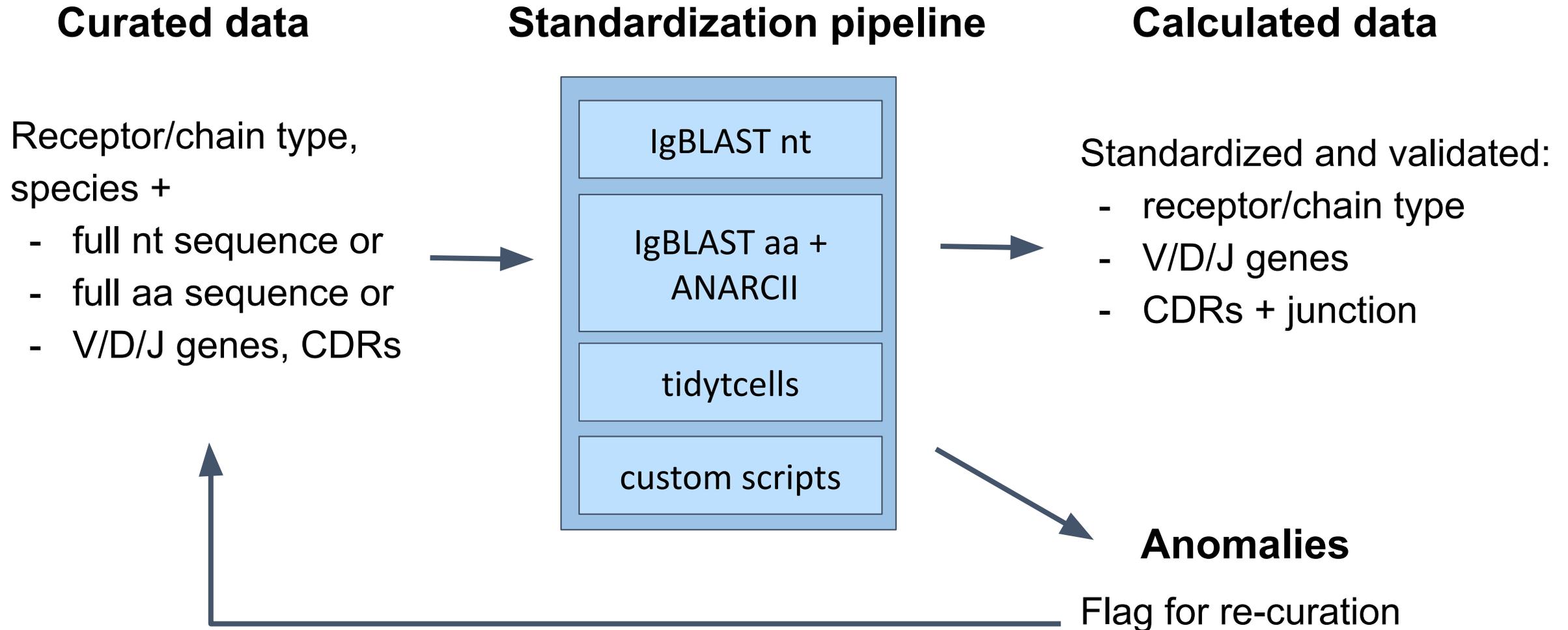
V/D/J Gene Name Standardization Examples

Full sequence data → recalculate (IgBLAST nt/aa)

Automated standardization (tidytcells*):

TCRAV1-2	→	TRAV1-2
TCRAJ52-01*01	→	TRAJ52*01
TRAV29-01	→	TRAV29/DV5

> *Support for BCR genes was added as part of this project, see github.com/yutanagano/tidytcells

Unable to standardize, human re-curation:

1-1	→	TRAV1-1
TRVB12	→	TRBV12
05-07*01	→	TRBV5-7*01
nonexistent gene name/wrong species		

TCR Gene Updates Per Chain

TCRV

- 1,247
- 36,764
- 19,161
- 179,271

TCRD

- 554
- 31,235
- 204,616

TCRJ

- 42,424
- 18,380
- 175,538
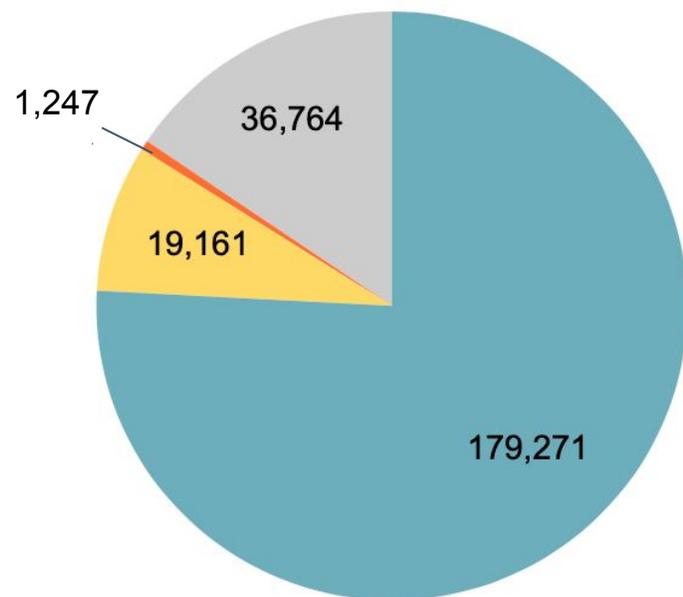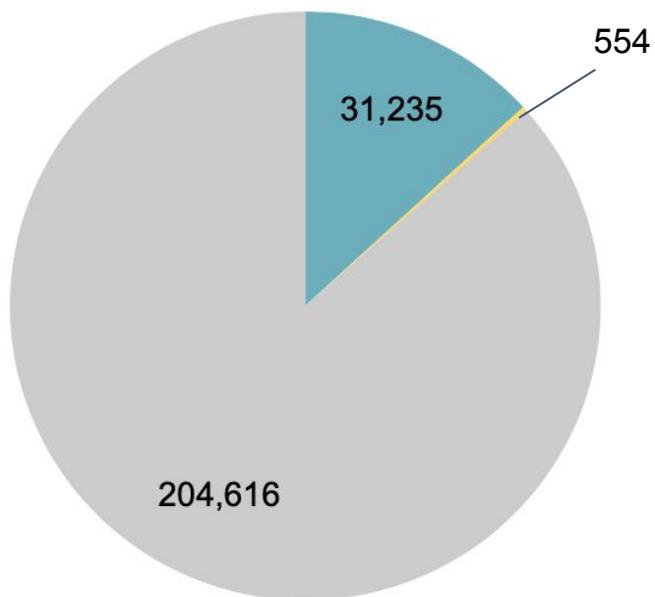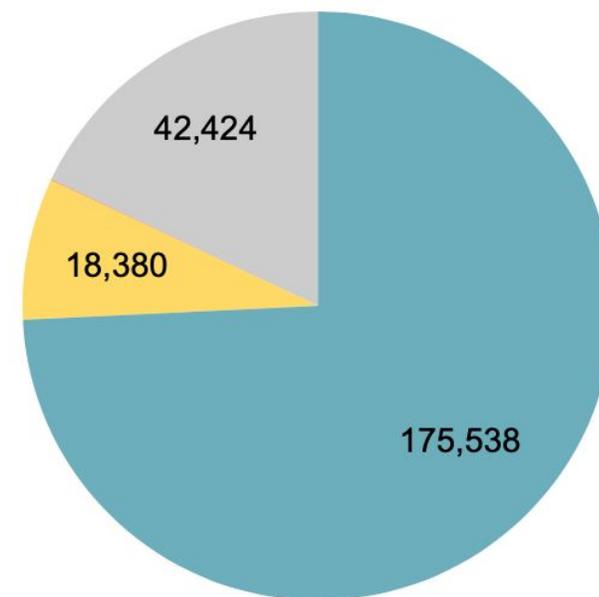
Legend:
- Gene name stayed the same (tidytcells validated)
- Gene name formatting was corrected (tidytcells)
- Gene was (re-)calculated from full sequence (IgBLAST nt/aa)
- No gene available

*human data only

BCR Gene Updates Per Chain

BCRV
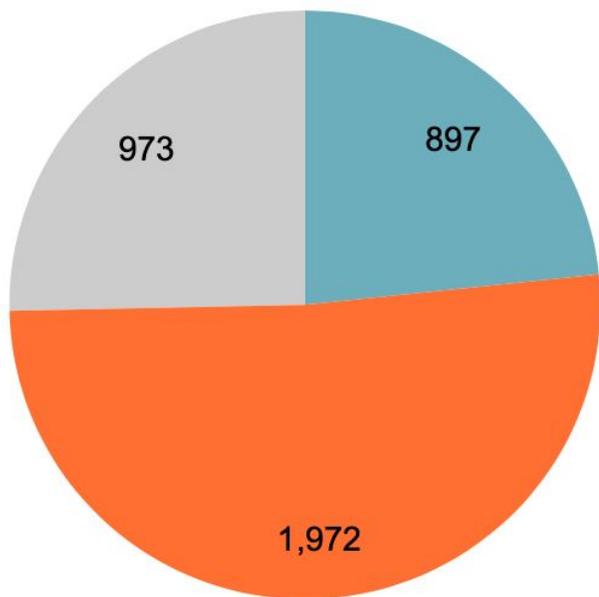
BCRD
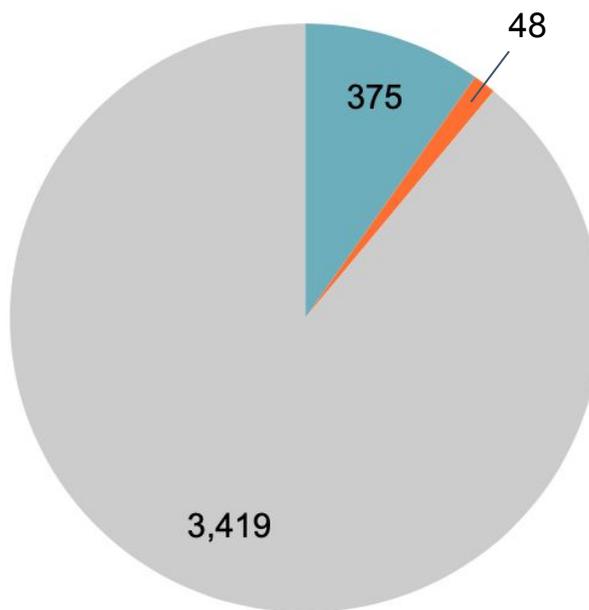
BCRJ
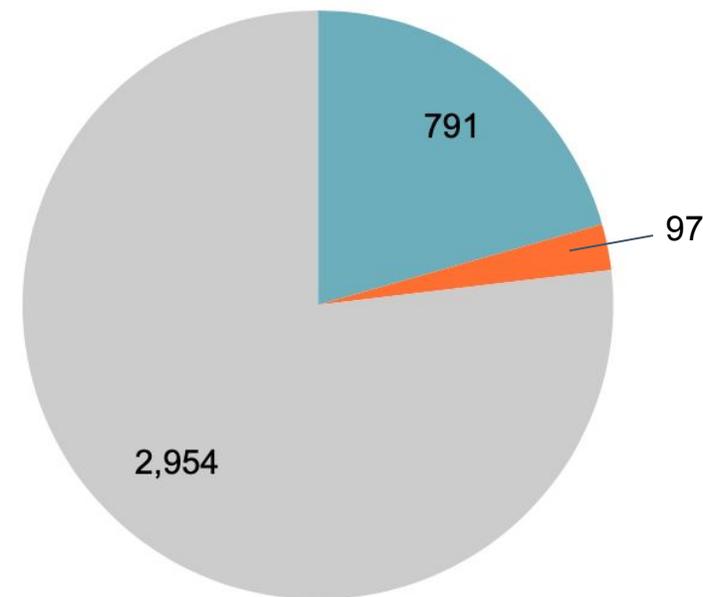
- ● Gene name stayed the same (tidytcells validated)
- ● Gene was (re-)calculated from full sequence (IgBLAST nt/aa)
- ● No gene available

*human data only

CDR3/Junction Standardization Examples

Full sequence data → recalculate (IgBLAST nt / ANARCII)

Curated CDR3		Calculated junction*	Calculated CDR3
CAVLNTGGFKTIF	→	CAVLNTGGFKTIF	
AVLNTGGFKTI	→	**C**AVLNTGGFKTI**F**	
AVLNTGG	→	**C**AVLNTGG**FKTIF**	→ AVLNTGGFKTI
ASYLCAVLNTGGFKTIF**G**	→	CAVLNTGGFKTIF	
SAVLNTGGFKTIF	→	**C**AVLNTGGFKTIF	
DASTGA	→	re-curate/omit	

*V/J gene names inform junction standardization, soon available in tidytcells v3

CDR3/Junction Updates per Chain

TCR

BCR

7,262 1,290 2,943

119

21,544

352

839

509

203,238

1,961
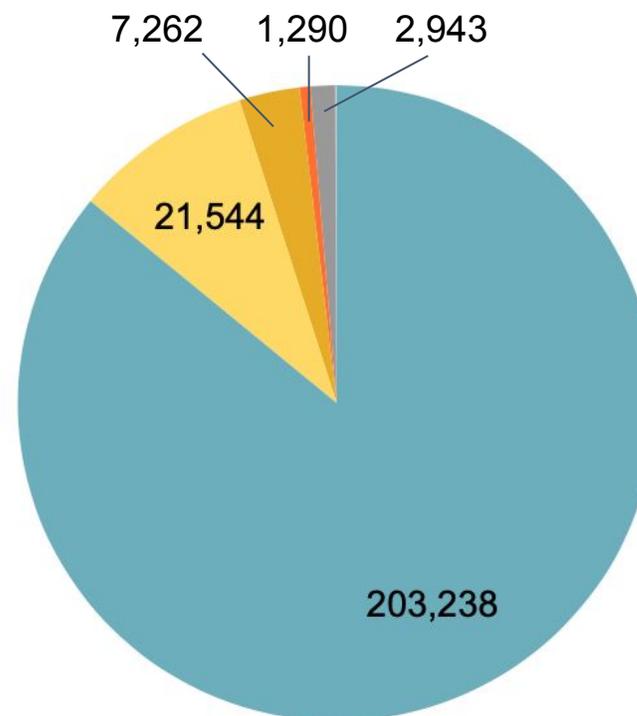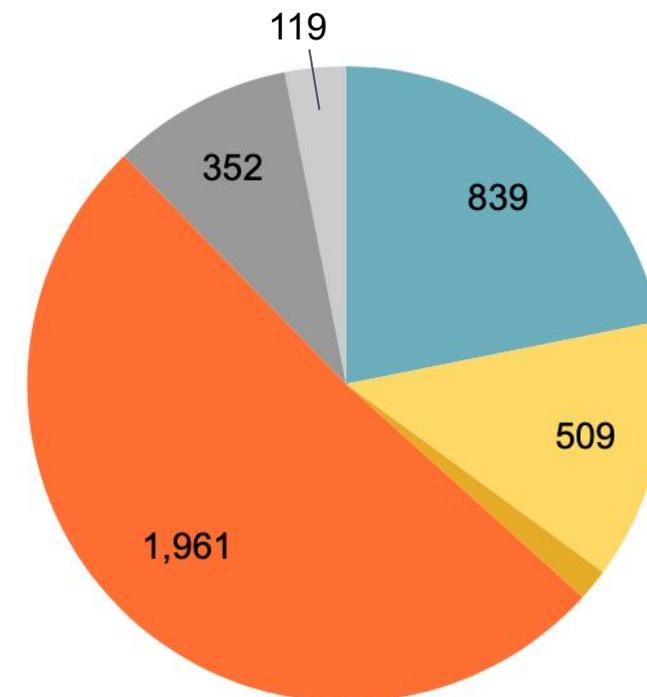
- Curated CDR3 is junction (tidytcells validated)
- Curated CDR3 is CDR3, recover junction (tidytcells)
- Non-trivial correction of curated CDR3 (tidytcells)
- (Re)calculated from full sequence (IgBLAST nt/ANARCII)
- Invalid CDR3, could not be standardized
- No CDR3 information available

*human data only

Anomalous Receptors Flagged for Re-Curation

Receptor re-curation by Dr. Laura Zarebski

Number of receptors (number of studies)

TCR

Total: 1,195 (170)

152 (64)
151 (47)
403 (35)
489 (24)

BCR

Total: 275 (89)

78 (25)
128 (11)
61 (45)
8 (8)

- V/D/J gene name errors
- CDR3 sequence errors
- Receptor/chain type errors
- Full sequence/accession errors
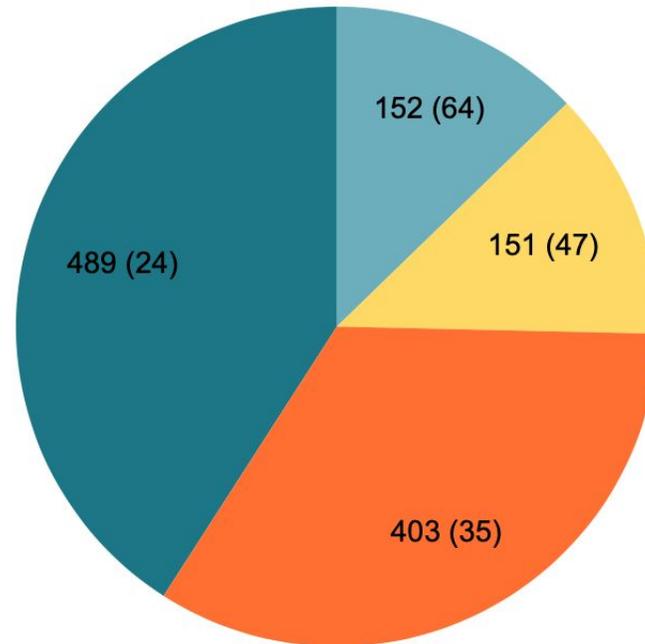
IMMREP25: TCR Specificity Prediction Challenge
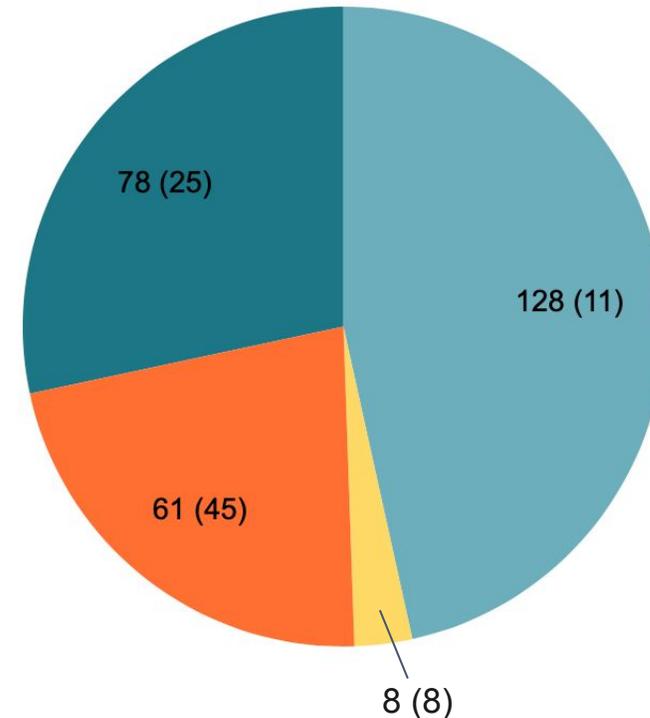

Training data provided by IEDB and VDJdb

IEDB data preprocessing by Dr. Eve Richardson

Integration into the AIRR Knowledge Commons (AKC)
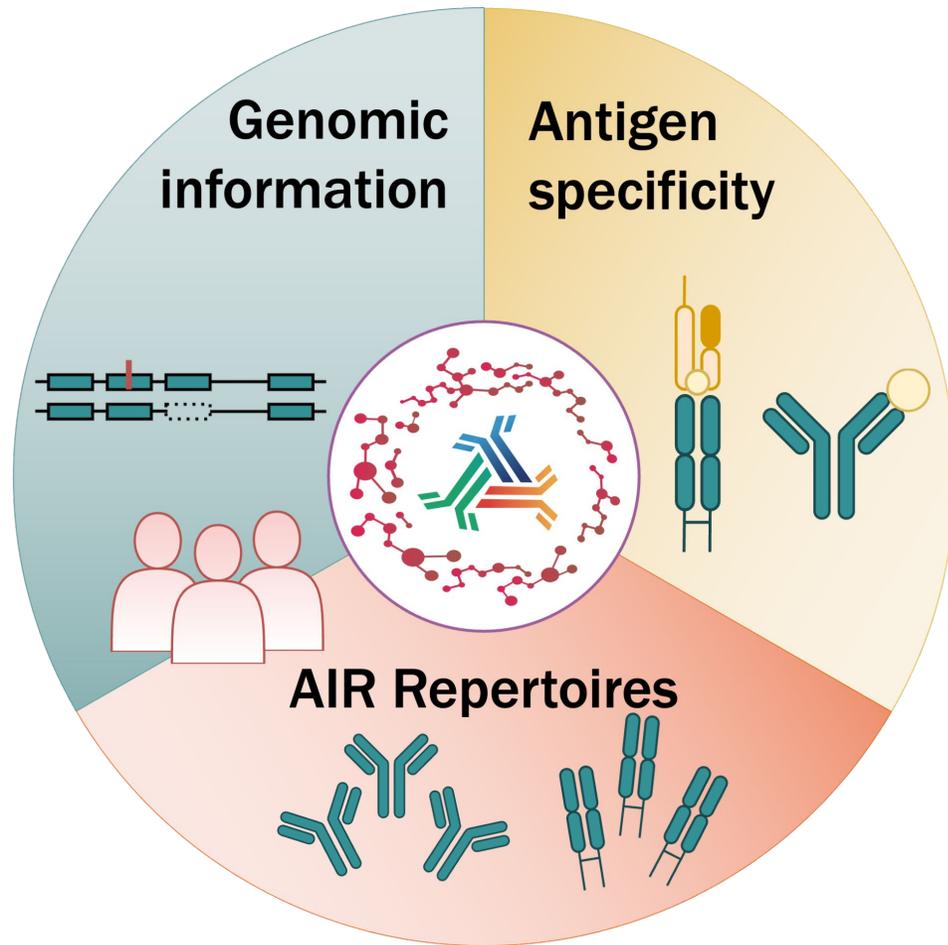
Component repositories:

Publication:

⬇ Download Full Issue

The Adaptive Immune Receptor Repertoire Knowledge Commons: An invitation to the community

Lindsay G. Cowell [1,2] ✉ · Scott Christley [1] · Felix Breden [3] · Kevin A. Burns [4,5] · Brian D. Corrie [3] · William D. Lees [6] · James A. Overton [7] · Bjoern Peters [8] · Eve Richardson [8] · Krishna M. Roskin [4,5] · Lonneke Scheffer [8] · Randi Vita [8] · Corey T. Watson [9] · Gur Yaari [10] Show less

more info: airr-knowledge.org

Acknowledgements

IEDB

Bjoern Peters

Eve Richardson

Randi Vita

Laura Zarebski

Nina Blazeska

Swapnil Mahajan

Daniel Wheeler (LEIDOS)

Jason Cantrell (LEIDOS)

AKC

Scott Christley (UTSW)

Lindsay Cowell (UTSW)

Brian Corrie (SFU)

William Lees (Clareo Biosciences)

James Overton (Knocean)

tidytcells

Yuta Nagano (UCL)

Thank you!