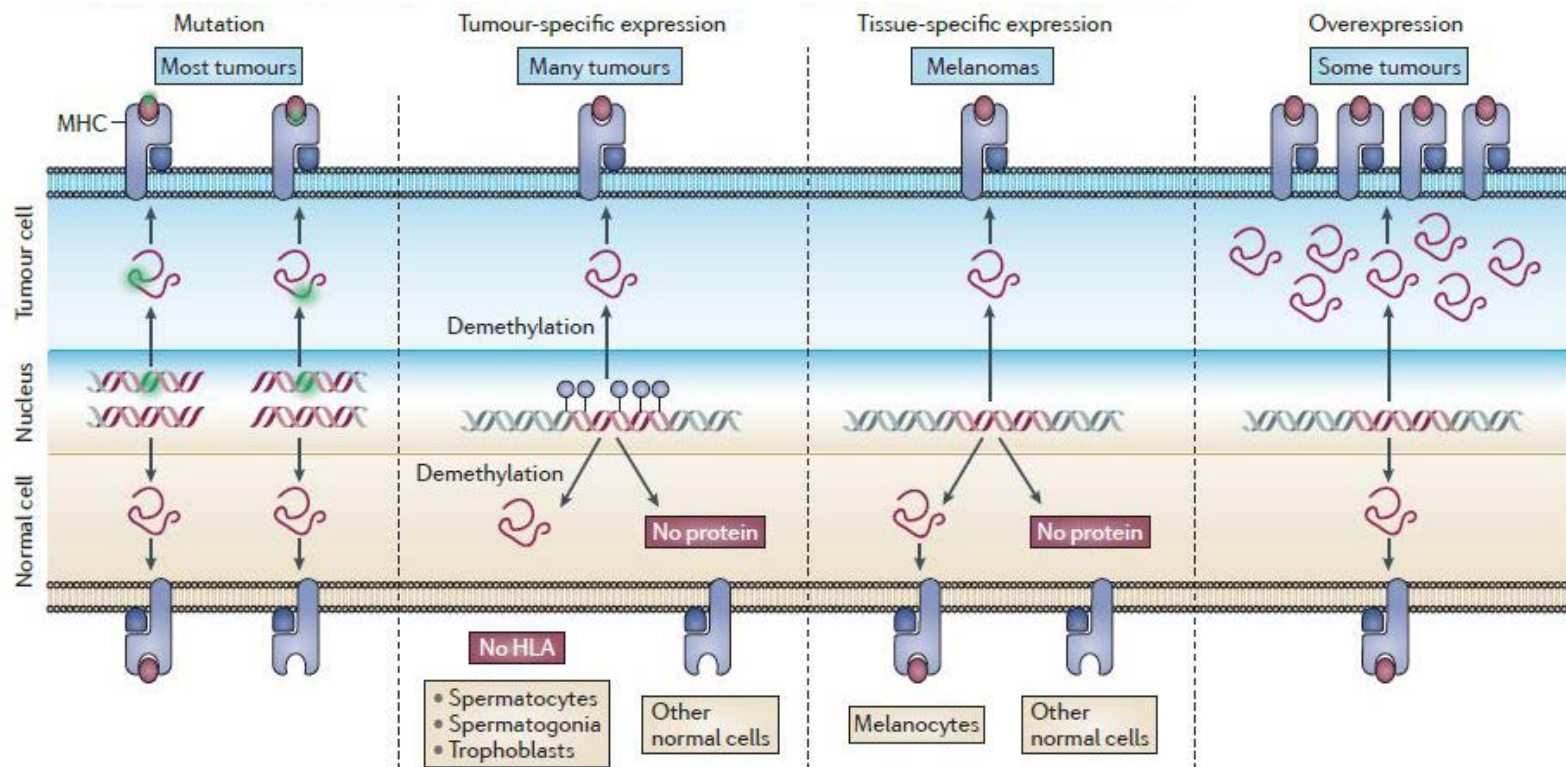# Cancer Epitope Database and Analysis Resource (CEDAR)

Presented by: Alessandro Sette, IEDB Principal Investigator

Cancer Epitopes are derived from Cancer Antigens

Coulie et al, Nat Rev Cancer. 2014 Feb

Motivation for the CEDAR Project

- **IEDB hosts epitope data for**
  - Allergy
  - Infectious diseases
  - Autoimmune diseases
  - Transplantation / Alloantigens
  - But **NOT** Cancer

We received funding from the **National Cancer Institute** to develop a resource for cancer epitopes

NIH ❭ NATIONAL CANCER INSTITUTE

Informatics Technology for Cancer Research (ITCR)

Motivation for the CEDAR Project (II)

1 **IMPORTANCE**
- Cancer epitopes play a key role in cancer immunology and immunotherapy
- They are important in understanding the biological mechanisms associated with treatment efficacy and developing more effective therapeutic approaches

2 **COMMUNITY NEED**
- Several resources attempted to catalog cancer epitopes (e.g. TANTIGEN, CAPED, NEPdb, dbPepNEO, etc.)
  - o Existing resources do not capture all epitope data in a granular fashion linked to the biological, immunological, and clinical contexts
  - o All resources only provide limited computational prediction and analysis tools

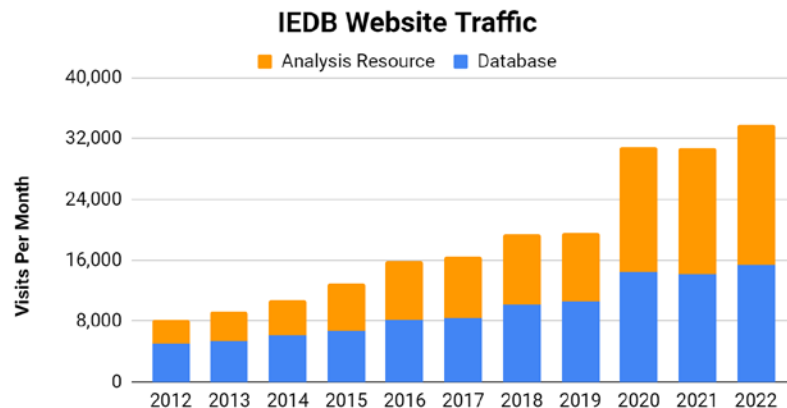*We developed The Cancer Epitope Database and Analysis Resource to fill these gaps*

Building on IEDB's Success
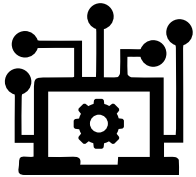**(Immune Epitope Database, www.iedb.org)**

1 **19 years of experience and learnings**

2 **Tried and tested IT and curation processes**

3 **Engaged user base with continuous growth over the years**

IEDB Website Traffic

Annual Citations to the IEDB in Published Literature
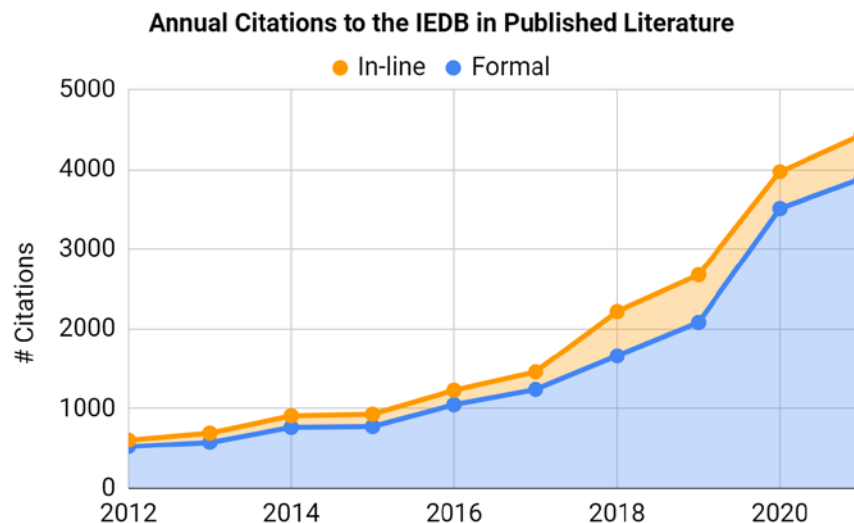
Cancer Epitope Database and Analysis Resource (CEDAR)

Database

Comprehensively cataloging all cancer epitope-related data linked to the biological, immunological, and clinical contexts

Analysis Resource

Computational epitope prediction and analysis tools providing researchers access to predictive strategies and objective evaluations of their performance

Specific Aims of the CEDAR Project
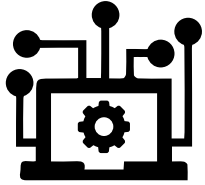
1) Establish the CEDAR database, ontology, and query and reporting functionality
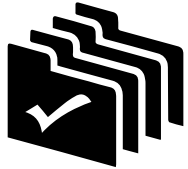
2) Curate literature epitope data, relevant to cancer immunology

3) Provide a validated set of cancer epitope prediction and analysis tools

4) Implement a multifaceted outreach program to engage the cancer research community

1
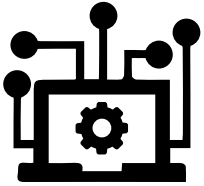
Establish the CEDAR database, ontology, and query and reporting functionality

1 CEDAR Database Design Process

- **Reached Out to Experts:** Interviewed cancer experts to identify important search fields for cancer research

- **Prototyped User Interfaces:** Designed database query interface wireframes

- **Introduced Cancer Curation Rules**, for example:

  i. Captured cancer types in more detail

  ii. Distinguished vaccination types (prophylactic / therapeutic)

  iii. Distinguished between allo- and xeno-adoptive transfer.

The CEDAR database base is available at cedar.iedb.org

Cancer-specific search parameters

Example Query: Prostate-specific Antigen

Example Query: Prostate-specific Antigen

Positive and negative data for T cell assay outcomes with MHC Class I restriction in human host

Example Results: Prostate-specific Antigen

Positive and negative data for T cell assays outcomes with MHC Class I restriction in human host

Query can be refined further on the 'Results' page via our 'Filter Options'

2
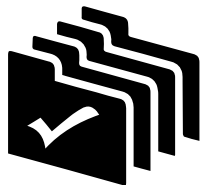
Curate literature epitope data, relevant to cancer immunology

CEDAR
Cancer Epitope Database and Analysis Resource

PubMed / PDB	Classifier	Abstract Review	Manual Curation
• Complex query • Bi-weekly	• Content based categories • Retrained annually	• Manual scan • Confirmation of classification	• Assigned to curators • Peer review
240K retrieved	**151K epitope related**	**44K likely curatable**	**23K curated**

Breakdown of Classified and Curatable References

Others 12.8%
Cancer 10.6%
Transplant 3.5%
Autoimmunity 20.3%
Allergy 6.7%
Infectious Disease 46.1%
3,543

Curated Cancer References by Quarter
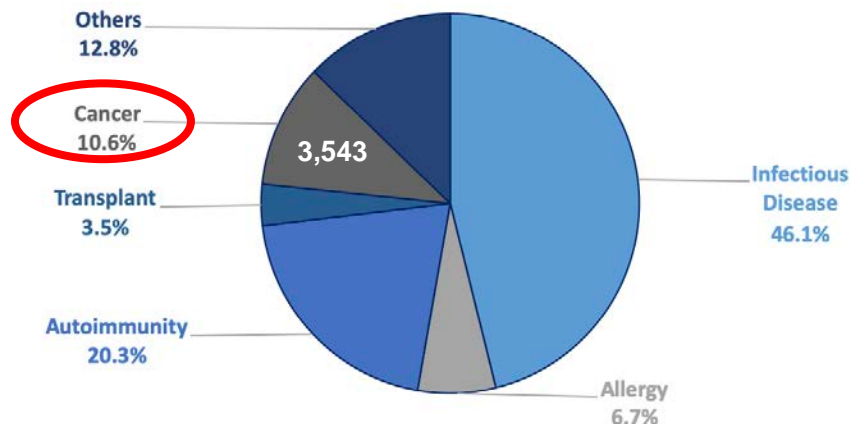
3,543 papers identified as cancer relevant

- Internal queue for curation: started with neoepitopes and prostate antigens
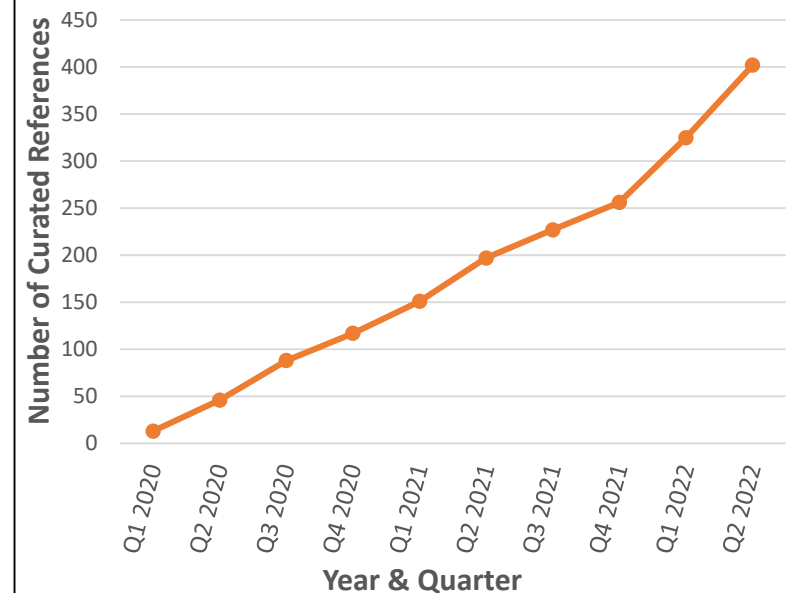
Prostate Cancer Curation

2

- 209 references finalized (as at October 24)

- ~97% complete with this category

- Category maintenance with 7 references in progress
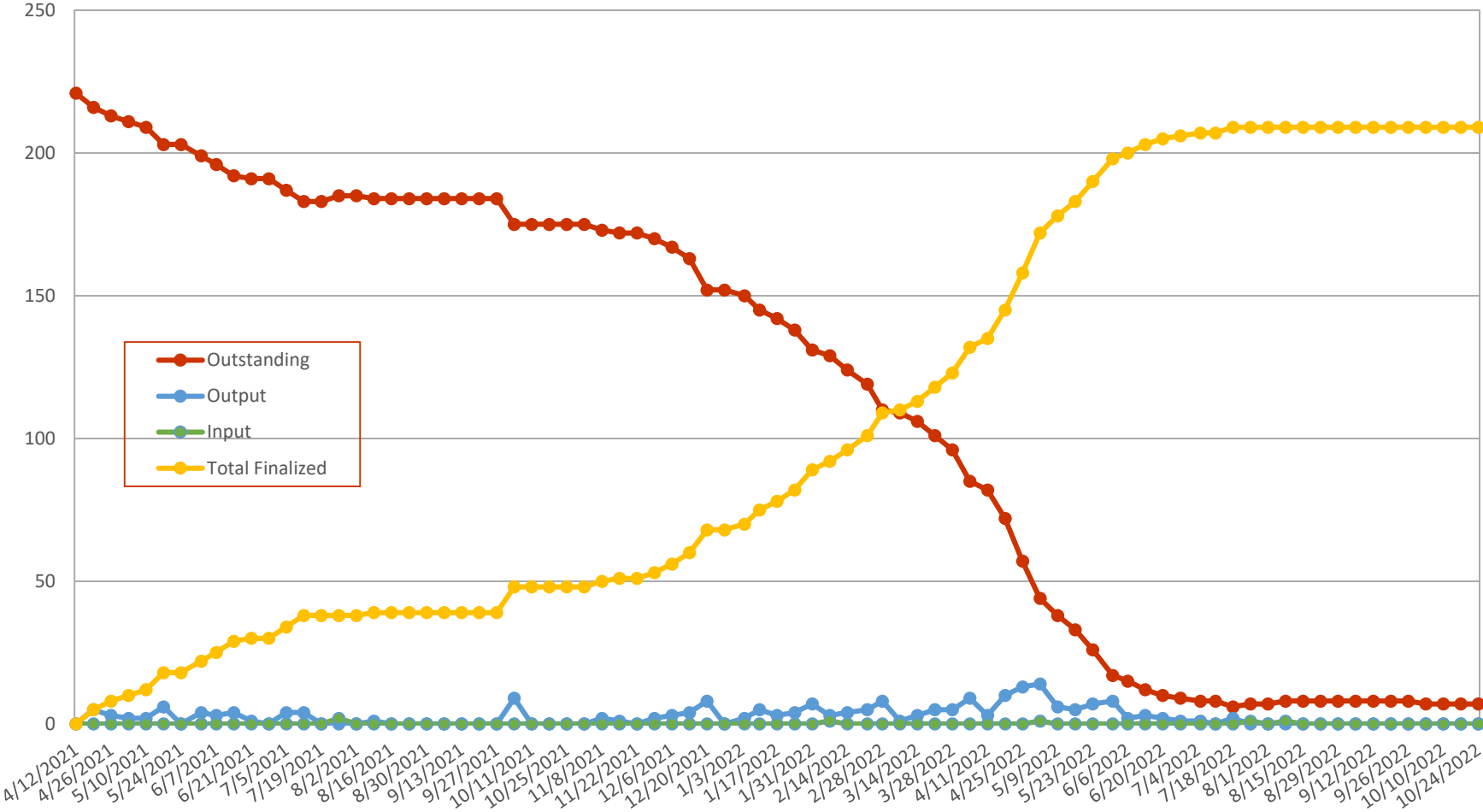
3

Provide a validated set of cancer epitope prediction and analysis tools

CEDAR
Cancer Epitope Database and Analysis Resource

- **Provide prediction tools <u>tailored to the needs of cancer immunologists</u>**

  - what neoepitopes are generated by a given mutation?

  - side-by-side predictions for mutant and wild-type peptides

- **Develop <u>novel prediction tools</u> for cancer epitopes**

  - combined assessment of expression and binding

  - include additional features when predicting epitopes
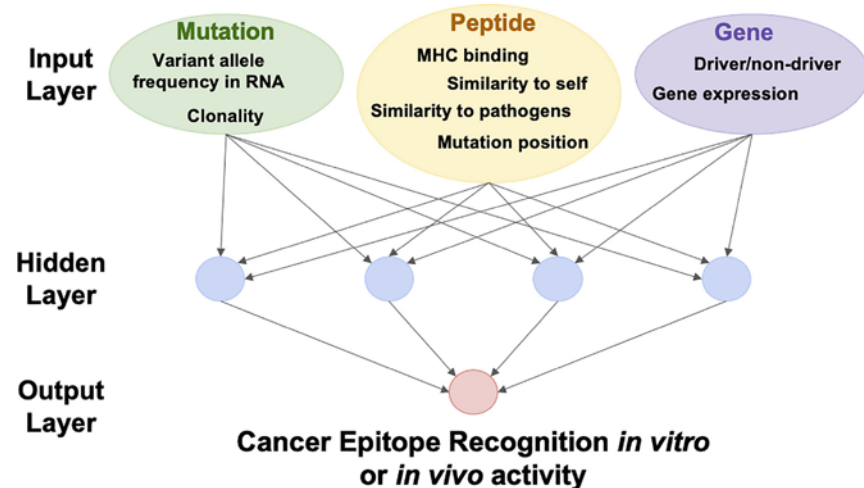
Overarching Analysis Resource Goals

- Provide web-implementations for published but hard to access cancer-epitope related tools in CEDAR

- Prioritize tools to implement in CEDAR based on cost-benefit analysis

Resource

Cell

Key Parameters of Tumor Epitope Immunogenicity Revealed Through a Consortium Approach Improve Neoantigen Prediction

Graphical Abstract

1. Global consortium to improve neoantigen prediction

Authors

Daniel K. Wells, Marit M. van Buuren, Kristen K. Dang, ..., Ton N. Schumacher, Pia Kvistborg, Nadine A. Defranoux

nature biotechnology

ARTICLES
https://doi.org/10.1038/s41587-019-0322-9

A large peptidome dataset improves HLA class I epitope prediction across most of the human population

Siranush Sarkizova[1,2,13], Susan Klaeger[2,13], Phuong M. Le[3], Letitia W. Li[3], Giacomo Oliveira[3], Hasmik Keshishian[2], Christina R. Hartigan[2], Wandi Zhang[3], David A. Braun[2,3,4,5], Keith L. Ligon[2,4,6,7], Pavan Bachireddy[2,3,5], Ioannis K. Zervantonakis[8], Jennifer M. Rosenbluth[8], Tamara Ouspenskaia[2], Travis Law[2], Sune Justesen[9], Jonathan Stevens[10], William J. Lane[4,10], Thomas Eisenhaure[2], Guang Lan Zhang[3,4,11], Karl R. Clauser[2], Nir Hacohen[2,3,12*], Steven A. Carr[2*], Catherine J. Wu[2,3,4,5*] and Derin B. Keskin[2,3,4,5,11*]

LETTER

doi:10.1038/nature24473

A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy

Marta Łuksza[1], Nadeem Riaz[2,3], Vladimir Makarov[3,4], Vinod P. Balachandran[5,6,7], Matthew D. Hellmann[7,8,9], Alexander Solovyov[10,11,12,13], Naiyer A. Rizvi[14], Taha Merghoub[7,15,16], Arnold J. Levine[1], Timothy A. Chan[2,3,4,7], Jedd D. Wolchok[7,8,15,16] & Benjamin D. Greenbaum[10,11,12,13]

3 Use curated cancer epitope datasets to benchmark epitope prediction tools

- Assemble comprehensive sets of cancer epitope data

- Make available in simple format for bioinformaticians for tool training and testing

- Conduct benchmarks of prediction tools on cancer epitope datasets

  - Manual compile and run benchmarks (initially)

  - Automated benchmarks of all tools implemented in CEDAR, using newly curated data

---

Examples of benchmark targets for prediction tools
- What peptides in a tumor sample are processed and presented on MHC
- What neo-epitopes are recognized by T cells from a cancer patient?

4
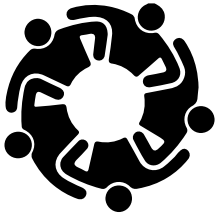
Implement a multifaceted outreach program to engage the cancer research community

4 **Community Outreach**

2 CEDAR-Specific Manuscripts Published To-Date

ORIGINAL RESEARCH article
Front. Immunol., 24 August 2021 | https://doi.org/10.3389/fimmu.2021.735609

The Cancer Epitope Database and Analysis Resource: A Blueprint for the Establishment of a New Bioinformatics Resource for Use by the Cancer Immunology Community

Zeynep Koşaloğlu-Yalçın[1], Nina Blazeska[1], Hannah Carter[2,3], Morten Nielsen[4,5], Ezra Cohen[3], Donald Kufe[6], Jose Conejo-Garcia[7,8], Paul Robbins[9], Stephen P. Schoenberger[10], Bjoern Peters[1,2] and Alessandro Sette[1,2*]

1 The Cancer Epitope Database and Analysis Resource (CEDAR).

Koşaloğlu-Yalçın Z[1], Blazeska N[1], Vita R[1], Carter H[2], Nielsen M[3], Schoenberger S[4], Sette A[1], Peters B[1]

Author information ▸

Nucleic Acids Research, 17 Oct 2022, :gkac902   **2**
DOI: 10.1093/nar/gkac902   PMID: 36250634

2 Conference Booths Held

1. **AACR Annual Meeting:** April 8-13, 2022

2. **FOCIS 2022:** June 21-24, 2022

9 Presentations or Posters Completed To-Date

| #	Presentation Title	Presenter	Event	Date
1	Talk: The Cancer Epitope Database & Analysis Resource	Sette & Peters	NCI Cancer Moonshot IOTN WG Meeting	July 15, 2021
2	Talk: The Cancer Epitope Database & Analysis Resource	Sette & Peters	Bioinformatics and Computational Biology WG	July 23, 2021
3	Talk: The Cancer Epitope Database & Analysis Resource	Sette & Peters	ITCR Annual Conference	September 21, 2021
4	Talk: The Cancer Epitope Database & Analysis Resource	Kosaloglu-Yalcin	2021 IEDB Virtual User Workshop	October 28, 2021
5	Talk: Recent Progress Within Immunopeptidome Prediction	Nielsen	5th Neoantigen Summit Europe	April 28, 2022
6	Poster: The Cancer Epitope Database & Analysis Resource	Kosaloglu-Yalcin	RECOMB-CCB 2022	May 20, 2022
7	Talk: The Cancer Epitope Database & Analysis Resource	Kosaloglu-Yalcin	Dynamics of Immune Repertoires	July 14, 2022
8	Poster: The Cancer Epitope Database & Analysis Resource	Kosaloglu-Yalcin	SIXTH CRI-ENCI-AACR INTERNATIONAL CANCER IMMUNOTHERAPY CONFERENCE	September 28-October 1, 2022
9	Talk: The Cancer Epitope Database & Analysis Resource	Kosaloglu-Yalcin	2022 ITCR Investigators Meeting	September 12-14

Summary

CEDAR is an extension of the IEDB, containing cancer-related epitope data and tools

1 **cedar.iedb.org** is a working prototype cancer epitope database
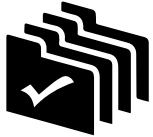
2 Literature curation is moving rapidly, with prostate cancer category completed and neoepitope category 74% complete

3 Exiting tools will be adapted to the needs of cancer researchers and novel cancer-specific tools will be developed. Cancer benchmark datasets are being assembled.

4 Community engagement has been initiated and will increase as we have more to show

Looking forward…

Improved style and functionality!