CaAtlas: an immunopeptidome atlas of human cancer

Xinpei Yi, Yuxing Liao, Kai Li, Bo Wen, Bing Zhang

Department of Molecular and Human Genetics, Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX, USA

Abstract

- **Human cancer antigen atlas (caAtlas)** is a comprehensive resource developed through extensive collection and analysis of publicly available **immunopeptidome** datasets from human cancer cell lines and tumor tissues from 9 different cancer types.
- To allow identification of putative tumor antigens inlcuding antigens with modifications, we used an **open search** tool.
- In total, we identified almost **370,000 antigens**, which is **60%** more of HLA class I peptides than those in SysteMHC Atlas and **tripled** the number of HLA class II peptides.
- To identify neoantigens from the cancer immunopeptidomes, we developed **NeoQuery** on the basis of **PepQuery**.
- The antigens we identified included a substantial number of • **neoantigens** • **C/T antigens** • **cancer-associated antigens**.
- Peptides from these verified antigen genes by caAtlas are potential therapeutic targets for **cancer immunotherapy**.
- We created a web resource named caAtlas (**`http://www.zhang-lab.org/caatlas/`**) to make all these data easily available and accessible to the broad cancer research community.

Construction of a comprehensive immunopeptidome atlas of human cancer

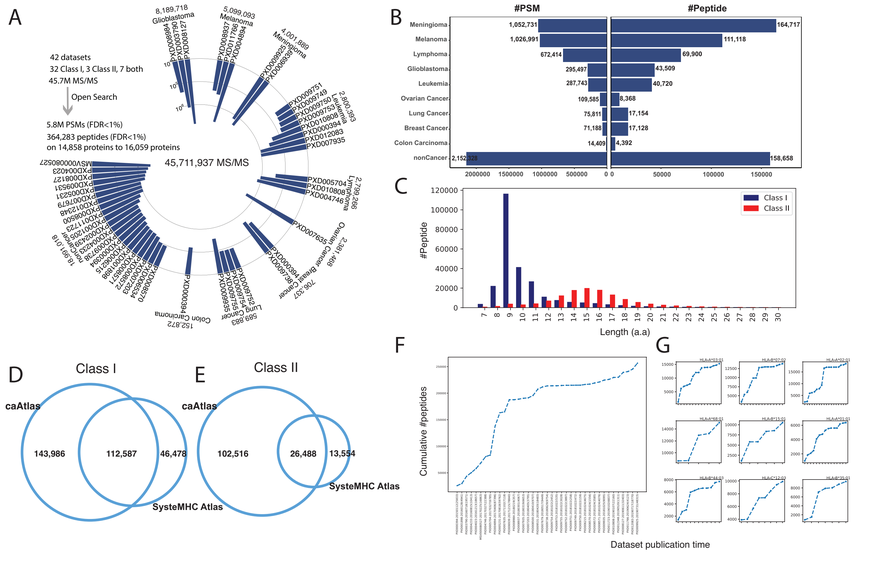

Figure 1:(A). MS/MS spectra numbers of the immunopeptidomics datasets included in this study and summary peptide identification results.(B). The distributions of PSM and peptide corresponding to 9 different cancer types and non-cancerous samples, respectively. (C). Typical length distribution of HLA class I and HLA class II peptides. (D) Comparison of the HLA class I peptide numbers between this study and SysteMHC Atlas. (E) Comparison of the HLA class II unique peptide numbers between this study and SysteMHC Atlas. (F) Cumulative number of all distinct HLA class I peptides as a function of dataset publication time. Each dataset is denoted as dataset ID: dataset public time (#peptides).(G) Cumulative number of distinct peptides for each of the top nine HLA class I alleles as a function of dataset publication time.

Neoantigens & C/T antigens & Cancer-associated antigens verified by caAtlas

Neoantigens

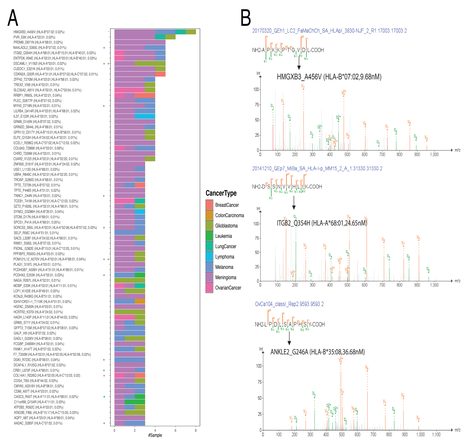

Figure 2:Neoantigens verified by HLA class I immunopeptidome of cancer samples.(A) Somatic mutations supported by immunopeptidomics data from three or more tumor samples. The identified HLA Alleles and the frequencies of these mutations in ICGC are list on the left. (B) Annotated spectra of the three neoantigen peptides in (A).

C/T antigens

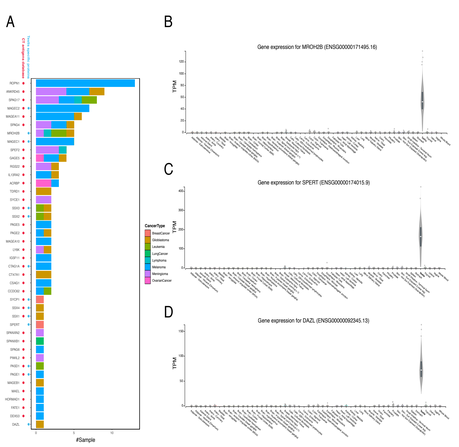

Figure 3:42 CT antigens from the CT antigens database and the GTEx-supported testis-specific proteins with the antigen source genes in caAtlas identified in tumor samples but not normal samples. The gene expression distribution among all the 54 normal tissues from GTEx for the 3 previously unrecognized CT antigens.

Cancer-associated antigens

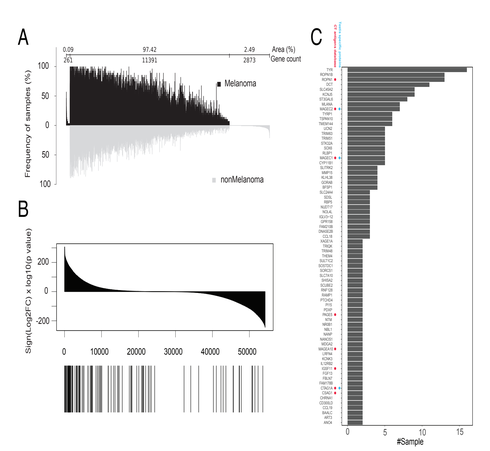

Figure 4:Differential analysis of antigens between melanoma samples and non-melanoma samples. Gene names for the 73 melanoma-associated genes which are significantly higher expression in the TCGA melanoma samples compared to GTEx normal skin tissues and sorted by sample numbers.

Open-pFind showed relatively higher sensitivity and specificity

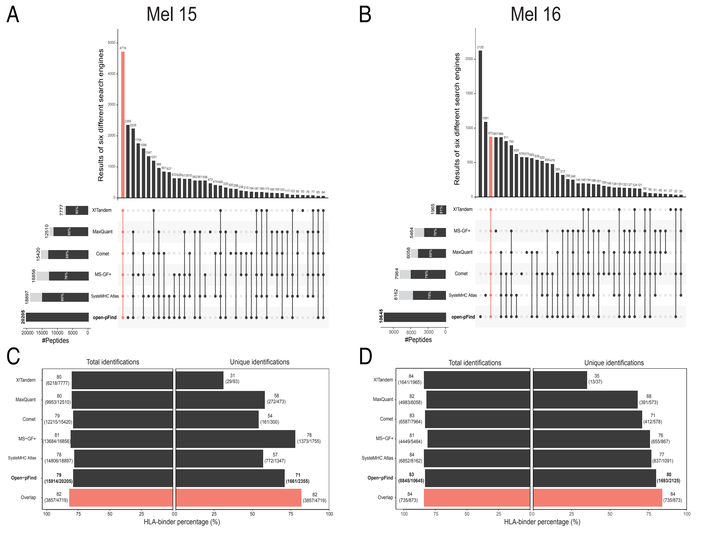

Figure 5:Open-pFind produces relatively higher sensitivity and specificity compared to four other search engines and SysteMHC Atlas in peptide identification from HLA class I data, and thus was selected for further analyses in this study.

Data dissemination through caAtlas

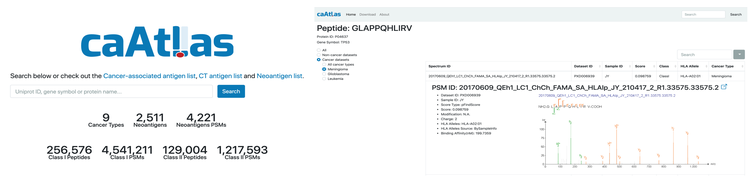

Figure 6:CaAtlas web resource: `http://www.zhang-lab.org/caatlas/`. All antigens are searchable by gene symbol, protein name, or protein ID. Moreover, neoantigens, CT antigens, and cancer-associated antigens, are browsable in three separate lists, in which the antigen source genes can be sorted by sample number or associated cancer type

Conclusion

- CaAtlas expanded immunopeptidomics validated neoantigens from a few dozens in the existing literature to more than 3,000.
- CaAtlas provided direct evidence to support tumor specific-presentation of some known and putative novel CT antigens and also revealed non-tumor-specific presentation of some previously annotated CT antigens.
- CaAtlas provided direct mass spectrum evidence to support known and putative novel tumor-associated differentiation or overexpressed antigen genes.
- CaAtlas website allows users to manually check and download the mass spectrum matching results for all antigens identified in this study.

Acknowledgements: This study was supported by the Cancer Prevention and Research Institute of Texas (CPRIT) award RR160027, and funding from the McNair Medical Institute at The Robert and Janice McNair Foundation.