

### 1 Overview

Next generation sequencing (NGS) of B-cell repertoires has provided a new lens through which to view the adaptive immune response. As a consequence, a rich source of antibody candidates with strong affinity to an immunogen is now accessible for antibody discovery. However, recent studies highlight the challenge of selecting candidates from B-cell repertoires.

Proteogenomics, which combines NGS with mass spectrometry-based proteomics, has received considerable attention for antibody discovery. At Digital Proteomics, we have developed Alicanto, a proteogenomic platform for antibody discovery that harnesses decades of technology development across genomics and proteomics<sup>1,2,3,4,5,6</sup>. An overview of the Alicanto pipeline is shown in Figure 1.

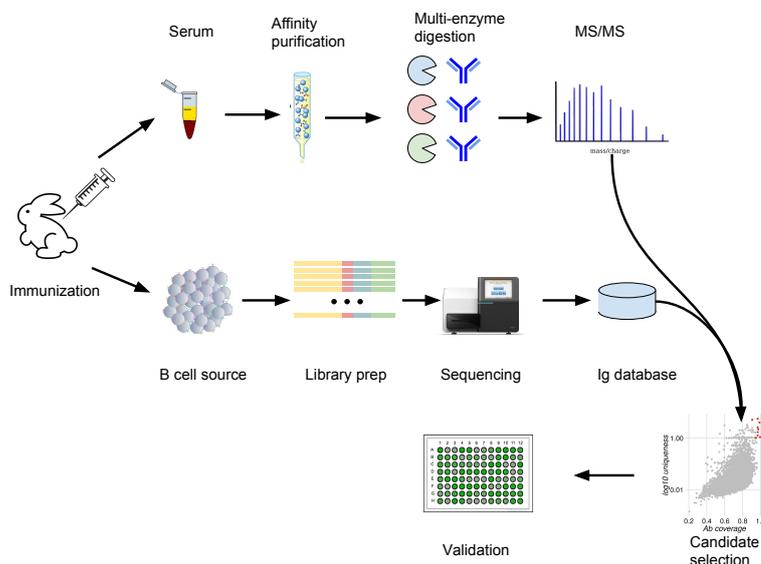


Figure 1: Alicanto workflow

### 2 Experimental Set-up

Alicanto can be applied to a variety of host organisms, including mouse, rat, rabbit, and human. However, minimum material requirements constrain which host animals may be practical. Larger mammals produce acceptable quantities of purified material, however obtaining sufficient antibody quantity from mice, for example, may be challenging. To ensure full coverage of high-abundance antibodies, and distinguish closely-related members of the same clonal lineage, we digest the antibodies with 4-6 different proteases. We require at least 200µg of affinity purified polyclonal antibodies for the mass-spectrometry analysis. If multiple hosts are used, it is preferable to keep the affinity purified antibodies separate for each host.

The choice of B-cell source(s) for NGS sequencing of antibody transcripts depends on the experimental set-up. Ideally the transcripts sequenced should match the isotype that is targeted for proteomics analysis; but it is not strictly necessary.

### 3 Deliverable

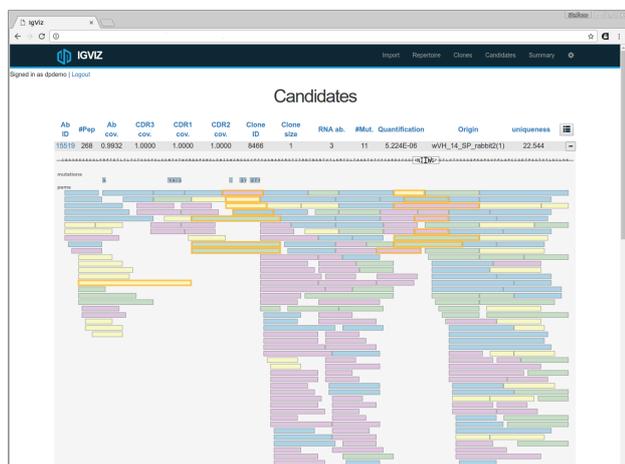


Figure 2: Screenshot of Alicanto’s interactive candidate reporting.

Alicanto provides a short list of antibody candidates. In addition to the candidate sequences, we provide an interactive cloud-backed interface to browse the list of the candidates, as well as antibodies that were not selected, shown in Figure 2.

In summary, the input requirements can be tailored to the specific experiment, or started at different stages. However, we accept these starting inputs:

Input	Immunization	NGS	MS	Analysis	Turnaround
Antigen	10-14 weeks	4 weeks	2 weeks	1 week	17-21 weeks
Tissue+serum	⇒	4 weeks	2 weeks	1 week	7 weeks
NGS+serum	⇒	⇒	2 weeks	1 week	3 weeks

## References

- [1] N. E. Castellana, S. H. Payne, Z. Shen, M. Stanke, V. Bafna, and S. P. Briggs. Discovery and revision of Arabidopsis genes by proteogenomics. *Proc. Natl. Acad. Sci. U.S.A.*, 105:21034–21038, Dec 2008. PMID: PMC2605632.
- [2] Y. Safonova, S. Bonissone, E. Kurpilyansky, E. Starostina, A. Lapidus, J. Stinson, L. DePalatis, W. Sandoval, J. Lill, and P. A. Pevzner. IgRepertoireConstructor: a novel algorithm for antibody repertoire construction and immunoproteogenomics analysis. *Bioinformatics*, 31(12):53–61, Jun 2015.
- [3] N. E. Castellana, V. Pham, D. Arnott, J. R. Lill, and V. Bafna. Template proteogenomics: sequencing whole proteins using an imperfect database. *Mol. Cell Proteomics*, 9:1260–1270, Jun 2010. PMID: PMC2877985.
- [4] Stefano R Bonissone and Pavel A Pevzner. Immunoglobulin classification using the colored antibody graph. *Research in Computational Molecular Biology*, pages 44–59, 2015.
- [5] Sunghee Woo, Seong Won Cha, Stefano Bonissone, Seungjin Na, David L Tabb, Pavel A Pevzner, and Vineet Bafna. Advanced proteogenomic analysis reveals multiple peptide mutations and complex immunoglobulin peptides in colon cancer. *Journal of proteome research*, 2015.
- [6] S. Woo, S. W. Cha, S. Na, C. Guest, T. Liu, R. D. Smith, K. D. Rodland, S. Payne, and V. Bafna. Proteogenic strategies for identification of aberrant cancer peptides using large-scale next-generation sequencing data. *Proteomics*, 14(23-24):2719–2730, Dec 2014.