Cell Reports Methods

CellPress
OPEN ACCESS

Preview

Learning what not to select for in antibody drug discovery

Beichen Gao,[1] Jiami Han,[1] and Sai T. Reddy[1,*]
[1]Department of Biosystems Science and Engineering, ETH Zurich, Basel 4058, Switzerland
*Correspondence: sai.reddy@bsse.ethz.ch
https://doi.org/10.1016/j.crmeth.2022.100258

Identifying antibodies with high affinity and target specificity is crucial for drug discovery and development; however, filtering out antibody candidates with nonspecific or polyspecific binding profiles is also important. In this issue of *Cell Reports Methods*, Saksena et al. report a computational counterselection method combining deep sequencing and machine learning for identifying nonspecific antibody candidates and demonstrate that it has advantages over more established molecular counterselection methods.

Traditional antibody discovery processes make heavy use of screening platforms, such as phage, yeast, and mammalian display technologies (Parola et al., 2018). These methods link the phenotype and genotype of proteins displayed and allow for rapid selection of large libraries of proteins including antibodies and antibody fragments. For phage display, the platform used in Saksena et al. (2022), the most common type of antibody fragments displayed are single-chain variable fragments (scFvs) and antigen-binding fragments (Fabs). Till now, phage display libraries with diversities up to 10^{11} sequences have been constructed and used extensively for antibody drug discovery. Traditional discovery by display platforms consist of performing several rounds of selection against target antigens, resulting in a pool of antigen-specific clones that then undergo additional experimental characterization and possibly follow-up engineering (e.g., affinity maturation) to select for therapeutic lead candidates.

In addition to high affinity to the target, a critical property for any antibody drug candidate is having minimal off-target binding. However, antibodies with off-target binding may also become enriched and selected through a screening process such as phage display. These polyspecific or nonspecific antibodies have the capacity to bind unrelated antigens and are often linked to non-ideal pharmacokinetic profiles (Hötzel et al., 2012). There has been evidence that suggests nonspecific antibody sequences can be identified through shared features. For example, the VH6 germline family was identified as a source of nonspecific clones derived from a yeast-displayed naive human antibody library (Kelly et al., 2017). Additionally, certain physicochemical properties of amino acids in antibody variable regions have been found to be associated with nonspecific binding (Zhang et al., 2020). Therefore, to identify and remove nonspecific antibodies during discovery campaigns, molecular counterselections can be performed such as screening against heterogeneous antigen panels (e.g., cell membrane extracts or other purified, unrelated target antigens) (Xu et al., 2013).

In recent years, deep sequencing has become a valuable tool for antibody screening and discovery (Parola et al., 2018), including the sequencing of display libraries to augment selection of candidates with high affinity to target antigens (Hu et al., 2015). In particular, deep sequencing enables quantitative analysis of enrichment and binding profiles of antibody sequences during various screening steps (e.g., rounds of selection). Recently, with the rapid advancement of machine learning tools for biological sequence analysis and the leveraging of deep sequencing data, researchers have also started to apply machine learning for antibody discovery and engineering (Pertseva et al., 2021). For example, in a recent study by our group, we screened by mammalian display mutagenesis libraries of the therapeutic antibody trastuzumab for binding to the HER2 antigen, deep sequencing data was then used to train supervised deep learning models (e.g., convolutional neural networks, CNNs) for classification of binding and non-binding antibodies (Mason et al., 2021). The deep learning models were then used to screen *in silico* a large library of trastuzumab sequence variants and identify possible clones that possess better developability properties while maintaining target binding. In another work, phage display and deep sequencing of antibody libraries was used to construct ensemble machine learning models, which were deployed for *in silico* affinity maturation and to identify novel antibody sequences with specificity to selected target antigens (Liu et al., 2020). These studies have established that deep sequencing and machine learning offer powerful tools for antibody discovery and engineering, but in nearly all cases thus far, they have focused on antibody specificity to target antigens.

Saksena et al. (2022) now report an application of deep sequencing and machine learning for antibody discovery, which identifies and removes nonspecific antibodies by computational counterselection (Figure 1). Specifically, the authors develop a machine learning approach based on multi-task ensemble models with the aim to identify and remove off-target, nonspecific sequences following phage display screening. For their experiments, the authors used a single-framework, randomized phage-displayed Fab library with a diversified heavy-chain complementarity determining region 3 (CDRH3) and selected two monoclonal antibodies as their target antigens: trastuzumab and omalizumab. The authors

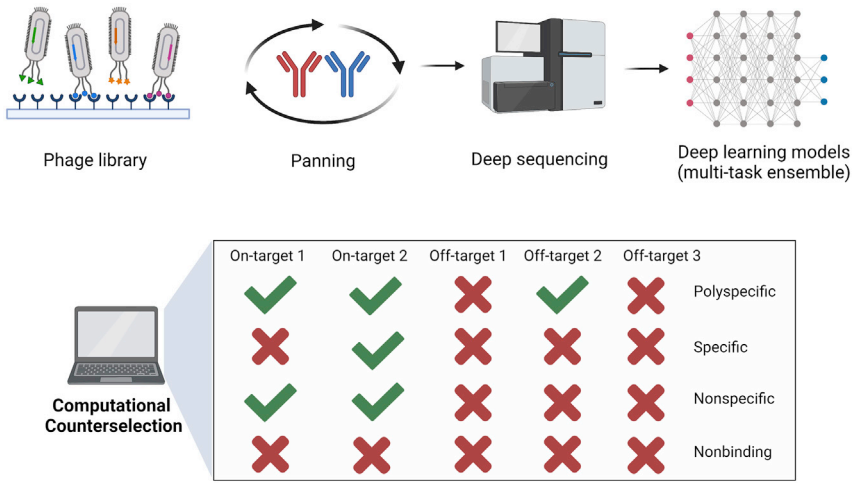

Figure 1. Computational counterselection

A phage display antibody library was selected against five different targets (two on targets, three off targets). Deep sequencing data across panning rounds was used to train a set of deep learning multi-task ensemble models to perform computational counterselection by identifying antibody sequences that are nonspecific or polyspecific. Figure created with Biorender (https://biorender.com/)

preferred monoclonal antibodies as targets given their public availability and the presence of defined shared epitopes (e.g., Fc domain) that provides a natural source of nonspecific binding and serves as a control in counterselections. Machine learning models were trained on deep sequencing data obtained through three different sets of phage display selections: on-target selection, off-target counterselection, and unrelated antigen counterselection. Five different CNNs were trained, which allows for the extraction and learning of different "perspectives" of the antigen-binding landscape, in addition to one dense neural network, to create an ensemble model for computational counterselection. Ensemble models improve the average prediction output through increasing robustness and variance reduction—which are very useful qualities when dealing with highly noisy data, such as those generated through phage display.

First, by performing two rounds of selections on their target antigens, trastuzumab and omalizumab, and then one round of counterselection on the opposing off-target antibody, Saksena et al. (2022) generated three sets of labeled deep sequencing data for model training: trastuzumab-specific, omalizumab-specific, and cross-reactive antibody sequences that were enriched in the counterselections. Using these combined datasets, multi-task ensemble models were trained to identify if a specific antibody sequence would be enriched in one, the other, or both selection conditions. The model achieved high performance (area under curve, AUC > 0.9) on the held-out test sets and, when experimentally evaluated, demonstrated that model predictions were highly accurate in predicting on-/off-target binding of sequences and with greater sensitivity than molecular counterselection. In particular, the authors found that the machine learning models are capable of correctly identifying nonspecific sequences that were unenriched during cross-panning counterselection, as well as specific sequences that were enriched during counterselection, both cases which may have been misclassified using conventional molecular counterselection.

The study also describes the training of machine learning models on deep sequencing data from selections on three unrelated targets (baculovirus extract, BSA, and TGF-b) in order to predict polyspecific antibody sequences. This machine learning approach was subsequently shown to also achieve similar performance in identifying polyspecific sequences from within the libraries. Importantly, this suggests it may be possible that such polyspecific machine learning models could be used in place of experimental molecular off-target counterselections. The finding of common motifs shared between nonspecific and polyspecific sequences also supports previous findings that there are common physicochemical properties shared by polyspecific antibody sequences (Zhang et al., 2020).

One of the major utilities of computational counterselection described in Saksena et al. (2022) is that it incorporates data from antibody binding to unrelated targets to improve off-target sequence identification. This is highly compatible with antibody discovery practices in industry such as screening multiple libraries in parallel against panels of target antigens and constructing large databases with antibody sequences with defined binding specificities. However, it is possible that instead of the ensemble deep learning models described in Saksena et al. (2022), other machine learning approaches may be used to identify polyspecific sequences; for example, recently, K-mer-embedded logistic regression models were trained using sequencing data from yeast display antibody libraries and demonstrated relatively high performance (AUC > 0.8) for predicting polyspecificity (Harvey et al., 2022). Finally, it is important to note that such computational counterselection methods may be adapted for other applications such as engineering of therapeutic TCRs, where off-target binding or cross-reactivity to peptide-major histocompatibility complex targets can result in serious safety concerns.

DECLARATIONS OF INTERESTS

The authors declare no competing interests.

REFERENCES

Harvey, E.P., Shin, J.-E., Skiba, M.A., Nemeth, G.R., Hurley, J.D., Wellner, A., Shaw, A.Y., Miranda, V.G., Min, J.K., Liu, C.C., et al. (2022). An in Silico Method to Assess Antibody Fragment Polyreactivity. Preprint at bioRxiv. https://doi.org/10.1101/2022.01.12.476085.

Hötzel, I., Theil, F.-P., Bernstein, L.J., Prabhu, S., Deng, R., Quintana, L., Lutman, J., Sibia, R., Chan, P., Bumbaca, D., et al. (2012). A strategy for risk mitigation of antibodies with fast clearance.

mAbs *4*, 753–760. https://doi.org/10.4161/mabs.22189.

Hu, D., Hu, S., Wan, W., Xu, M., Du, R., Zhao, W., Gao, X., Liu, J., Liu, H., and Hong, J. (2015). Effective Optimization of antibody affinity by phage display Integrated with high-throughput DNA Synthesis and sequencing technologies. PLoS One *10*, e0129125. https://doi.org/10.1371/journal.pone.0129125.

Kelly, R.L., Zhao, J., Le, D., and Wittrup, K.D. (2017). Nonspecificity in a nonimmune human scFv repertoire. mAbs *9*, 1029–1035. https://doi.org/10.1080/19420862.2017.1356528.

Liu, G., Zeng, H., Mueller, J., Carter, B., Wang, Z., Schilz, J., Horny, G., Birnbaum, M.E., Ewert, S., and Gifford, D.K. (2020). Antibody complementarity determining region design using high-capacity machine learning. Bioinformatics *36*, 2126–2133. https://doi.org/10.1093/bioinformatics/btz895.

Mason, D.M., Friedensohn, S., Weber, C.R., Jordi, C., Wagner, B., Meng, S.M., Ehling, R.A., Bonati, L., Dahinden, J., Gainza, P., et al. (2021). Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. Nat Biomed Eng *5*, 600–612. https://doi.org/10.1038/s41551-021-00699-9.

Parola, C., Neumeier, D., and Reddy, S.T. (2018). Integrating high-throughput screening and sequencing for monoclonal antibody discovery and engineering. Immunology *153*, 31–41. https://doi.org/10.1111/imm.12838.

Pertseva, M., Gao, B., Neumeier, D., Yermanos, A., and Reddy, S.T. (2021). Applications of machine and deep learning in adaptive Immunity. Annu. Rev. Chem. Biomol. Eng. *12*, 39–62. https://doi.org/10.1146/annurev-chembioeng-101420-125021.

Saksena, S.D., Liu, G., Banholzer, C., Horny, G., Ewert, S., and Gifford, D.K. (2022). Computational counterselection identifies nonspecific therapeutic biologic candidates. Cell Reports Methods *2*, 100254-1–100254-8.e5. https://doi.org/10.1016/j.crmeth.2022.100254.

Xu, Y., Roach, W., Sun, T., Jain, T., Prinz, B., Yu, T.-Y., Torrey, J., Thomas, J., Bobrowicz, P., Vásquez, M., et al. (2013). Addressing polyspecificity of antibodies selected from an in vitro yeast presentation system: a FACS-based, high-throughput selection and analytical tool. Protein Eng. Des. Sel. *26*, 663–670. https://doi.org/10.1093/protein/gzt047.

Zhang, Y., Wu, L., Gupta, P., Desai, A.A., Smith, M.D., Rabia, L.A., Ludwig, S.D., and Tessier, P.M. (2020). Physicochemical Rules for identifying monoclonal antibodies with drug-like specificity. Mol. Pharm. *17*, 2555–2569. https://doi.org/10.1021/acs.molpharmaceut.0c00257.