Databases and Ontologies

CoV-AbDab: the Coronavirus Antibody Database

Matthew I. J. Raybould [1], Aleksandr Kovaltsuk [1], Claire Marks [1] and Charlotte M. Deane [1,*]

[1] Oxford Protein Informatics Group, Department of Statistics, University of Oxford, 24-29 St. Giles', Oxford, OX1 3LB, United Kingdom.

* To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: The emergence of a novel strain of betacoronavirus, SARS-CoV-2, has led to a pandemic that has been associated with over 700,000 deaths as of 5th August 2020. Research is ongoing around the world to create vaccines and therapies to minimise rates of disease spread and mortality. Crucial to these efforts are molecular characterisations of neutralising antibodies to SARS-CoV-2. Such antibodies would be valuable for measuring vaccine efficacy, diagnosing exposure, and developing effective biotherapeutics. Here, we describe our new database, CoV-AbDab, which already contains data on over 1400 published/patented antibodies and nanobodies known to bind to at least one betacoronavirus. This database is the first consolidation of antibodies known to bind SARS-CoV-2 as well as other betacoronaviruses such as SARS-CoV-1 and MERS-CoV. It contains relevant metadata including evidence of cross-neutralisation, antibody/nanobody origin, full variable domain sequence (where available) and germline assignments, epitope region, links to relevant PDB entries, homology models, and source literature.

Results: On 5th August 2020, CoV-AbDab referenced sequence information on 1402 anti-coronavirus antibodies and nanobodies, spanning 66 papers and 21 patents. Of these, 1131 bind to SARS-CoV-2.

Availability: CoV-AbDab is free to access and download without registration at *http://opig.stats.ox.ac.uk/webapps/coronavirus*. Community submissions are encouraged.

Contact: deane@stats.ox.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

To respond effectively to the recent Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) pandemic, it is essential to understand the molecular basis for a successful immune response to coronavirus infection (Iwasaki and Yang (2020)). In particular, characterising the B-cell response is important as the identification of potent neutralising antibodies could pave the way for effective treatments, aid in prior exposure diagnosis, or assist in predicting vaccine efficacy (e.g. Tay *et al.* (2020); Liao *et al.* (2020); Wen *et al.* (2020); Zhao *et al.* (2020); Nielsen *et al.* (2020); Galson *et al.* (2020)). Such molecular characterisations of binding/neutralising antibodies to SARS-CoV-2 antigens are now beginning to emerge.

Early studies have shown a large proportion of the SARS-CoV-2 neutralising response is SARS-CoV-2 specific, although some SARS-CoV-2/SARS-CoV-1 (the virus responsible for the 2003 epidemic) cross-neutralising antibodies do exist (Pinto *et al.* (2020); Yuan *et al.* (2020); Robbiani *et al.* (2020)). This is possible because many of the SARS-CoV-1 and SARS-CoV-2 spike protein domains, including the Receptor Binding Domain (RBD), share high sequence and structural similarity (Tay *et al.* (2020)). Other SARS-CoV-2 surface proteins also display high sequence similarity to more distantly related betacoronaviruses, such as the Middle East Respiratory Syndrome coronavirus (MERS-CoV). Therefore, molecular knowledge of any antibody able to bind a betacoronavirus antigen could be relevant in treating SARS-CoV-2 infection.

In addition to the large existing body of work studying SARS-CoV-1 and MERS-CoV, the number of investigations into SARS-CoV-2 is extremely high. As an indication, 1,134 SARS-CoV-2 preprints were

uploaded to BioRxiv and MedRxiv between January-March 2020, while 3,462 were uploaded in April and May. We have therefore built CoV-AbDab, a database that documents molecular information (*i.e*. sequence and structure) and metadata on all preprinted, published, and patented anti-coronavirus antibodies. It will save valuable time in the fight against COVID-19 and act as a central hub to consolidate knowledge and coordinate efforts to identify novel antibodies that neutralise SARS-CoV-2. Researchers can use CoV-AbDab to yield new insights, including deriving crucial sequence/structural patterns that distinguish neutralising from non-neutralising SARS-CoV-2 binders (Iwasaki and Yang (2020)), or deducing independent neutralising epitopes exploitable by combination therapies (Hansen *et al.* (2020)).

2 Data Sources

Academic papers and patents containing coronavirus-binding antibodies were sourced by querying PubMed, BioRxiv, MedRxiv, GenBank, and Google Patents with relevant search terms. A full list of references is given in the database and in the Supporting Information. ANARCI (Dunbar and Deane (2016)) was used to number amino acid sequences in the IMGT (Lefranc *et al.* (2003)) numbering scheme, and to assign V and J gene origins. Our Structural Antibody Database (Dunbar *et al.* (2013)), which tracks all antibody structures submitted to the Protein Data Bank (PDB, Berman *et al.* (2000)), was mined to identify relevant solved structures. Our antibody/nanobody homology modelling tool, ABodyBuilder (Leem *et al.* (2016)), was used to generate full Fv region structural models.

3 Contents

As of 5th August 2020, CoV-AbDab contains 1402 entries from 66 publications and 21 patents. Of these entries, 147 bind to MERS-CoV, 483 bind to SARS-CoV-1, and 1131 bind to SARS-CoV-2 (each entry may be tested against multiple coronaviruses). All entries contain a minimum of germline information, the CDR3 sequence, or Material Transfer Agreement contact information (a complete list of metadata is provided in the Supporting Information). Currently, 1303/1402 entries (92.9%) contain full variable domain antibody (Fv) or nanobody (VHH) sequences.

A comparative analysis of SARS-CoV-2, SARS-CoV-1, and MERS-CoV binder biological/synthetic origins (Fig. S1), binder targets (Fig. S2), and human antibody binder heavy V-gene germlines (Fig. S3) is available in the Supporting Information. The origin plots show how technological advances have impacted the nature of isolated binders and this should be considered when comparing binders to different coronaviruses.

We are reaching out to authors of new studies characterising coronavirus binding antibodies to send us their data in Excel or CSV format (*opig@stats.ox.ac.uk*). We require a minimum of the antibody/nanobody clonotype (closest heavy and light V and J gene transcripts plus the CDR3 amino acid sequences) but ideally seek the full antibody variable domain [Fv] or full nanobody [VHH] sequence. We also require binding data against at least one specified coronavirus protein reported in a preprint, publication, or patent. Through these submissions and our own efforts to track the scientific literature, we hope to provide a central community resource for coronavirus antibody sequence and structural information.

4 Usage

CoV-AbDab is located at *http://opig.stats.ox.ac.uk/webapps/coronavirus* (Fig. S4A). Users can download the entire database or search-filtered results as a CSV file and bulk download all ANARCI numberings, IMGT-numbered antibody- or nanobody-CoV structure files, and IMGT-numbered antibody/nanobody homology models (Fig. S4B). A summary of tracked (but not included) antibodies and coronavirus studies is also provided.

The database can be queried by variable domain sequence (Fig. S4C) or by attribute (Fig. S4D). Attribute search results can be further filtered by a search term and ordered by any metadata field for maximum interpretability (Fig. S4E). Searching by sequence returns the top-10 same-length whole chain and/or CDR3 sequence identities to the query. Query-target alignments are displayed (Fig. S4F). Any entry with a homology model can be viewed in-browser using our native molecular viewer (Fig. S4G).

Acknowledgements and Funding

We would like to thank the many authors who have responded in a timely and helpful manner to our requests for data. This work was supported by an EPSRC and MRC grant [EP/L016044/1] awarded to MIJR, a BBSRC grant [BB/M011224/1] awarded to AK, and funding from GlaxoSmithKline plc, UCB Pharma Ltd., AstraZeneca plc, and F. Hoffmann-La Roche AG.

References

Berman, H. M. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.*, **28**(1), 235–242.

Dunbar, J. and Deane, C. M. (2016). ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics*, **32**(2), 298–300.

Dunbar, J. *et al.* (2013). SAbDab: the structural antibody database. *Nucleic Acids Res*, **42**(D1), D1140–D1146.

Galson, J. D. *et al.* (2020). Deep sequencing of B cell receptor repertoires from COVID-19 patients reveals strong convergent immune signatures. *bioRxiv*.

Hansen, J. *et al.* (2020). Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. *Science*.

Iwasaki, A. and Yang, Y. (2020). The potential danger of suboptimal antibody responses in COVID-19. *Nat. Rev.*, **20**, 339–341.

Leem, J. *et al.* (2016). ABodyBuilder: Automated antibody structure prediction with data–driven accuracy estimation. *mAbs*, **8**(7), 1259–1268.

Lefranc, M.-P. *et al.* (2003). IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.*, **27**(1), 55–77.

Liao, M. *et al.* (2020). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med*, **26**, 842–844.

Nielsen, S. C. A. *et al.* (2020). B cell clonal expansion and convergent antibody responses to SARS-CoV-2. *Research Square (Nature Preprint)*.

Pinto, D. *et al.* (2020). Structural and functional analysis of a potent sarbecovirus neutralizing antibody. *bioRxiv*.

Robbiani, D. F. *et al.* (2020). Convergent Antibody Responses to SARS-CoV-2 Infection in Convalescent Individuals. *Nature*.

Tay, M. Z. *et al.* (2020). The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.*, **20**, 363–374.

Wen, W. *et al.* (2020). Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov.*, **6**, 31.

Yuan, M. *et al.* (2020). A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science*, **368**, 630–633.

Zhao, J. *et al.* (2020). Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clin. Infect. Dis.*