NRC·CNRC

NRC Publications Archive
Archives des publications du CNRC

Oxford nanopore sequencing enables rapid discovery of single-domain antibodies from phage display libraries

Lowden, Michael J.; Henry, Kevin A.

This publication could be one of several versions: author's original, accepted manuscript or the publisher's version. / La version de cette publication peut être l'une des suivantes : la version prépublication de l'auteur, la version acceptée du manuscrit ou la version de l'éditeur.
For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

Publisher's version / Version de l'éditeur:
https://doi.org/10.2144/btn-2018-0123
BioTechniques, 65, 6, pp. 351-356, 2018-11-27

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=aa89e335-a35a-4c1d-bc1f-ce803f6d6c02
https://publications-cnrc.canada.ca/fra/voir/objet/?id=aa89e335-a35a-4c1d-bc1f-ce803f6d6c02

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.

National Research Council Canada | Conseil national de recherches Canada

Canada

Benchmark

BioTechniques

For reprint orders, please contact: reprints@futuremedicine.com

Oxford nanopore sequencing enables rapid discovery of single-domain antibodies from phage display libraries

Michael J Lowden[1] & Kevin A Henry*[,1]

ABSTRACT

Antibody (Ab) repertoire sequencing using high-throughput massively parallel technologies has contributed substantially to the understanding of Ab responses following infection, vaccination and autoimmunity. Because individual B-cell receptors are recombined and diversified somatically, genomic comparisons are limited, and distinguishing rare variants from sequencing errors is a major challenge. Oxford Nanopore Technologies' MinION is a highly portable and cost-effective third-generation sequencing instrument, but has not been used for Ab repertoire sequencing due to its high error rate (approximately 1/10 bases). Here, we applied nanopore sequencing to single-domain Ab (sdAb) repertoires and phage-displayed sdAb libraries. We show that despite low overall data fidelity, sdAb sequences could be reconstructed above a frequency threshold (~100 copies); however, distinguishing clonal sdAb variants was not always possible. The data quality was sufficient to enable rapid identification of antigen-specific sdAb sequences enriched during panning of phage display libraries, obviating the need for screening single clones.

KEYWORDS:

antibody • nanopore sequencing • next-generation DNA sequencing • phage display • single-domain antibody • V_HH

[1]*Human Health Therapeutics Research Centre, National Research Council Canada, 100 Sussex Drive, Ottawa, Ontario, Canada, K1A 0R6; *Author for correspondence: kevin.henry@nrc-cnrc.gc.ca*

METHOD SUMMARY

We demonstrate that Oxford Nanopore Technologies' MinION instrument can be used to sequence single-domain antibody (sdAb) repertoires of limited diversity, and that the resulting data quality is sufficient to enable sdAb discovery from phage display libraries. The method relies on (i) minimizing error by limiting the unit of analysis to the sdAb CDR3, and (ii) taking the consensus after aligning all full-length sdAb sequences associated with each CDR3 clonotype.

INTRODUCTION

Antibodies (Abs) and their cell-surface counterparts, the B-cell receptors (BCRs) are critical mediators of humoral adaptive immunity. The development and commercialization of massively parallel high-throughput DNA sequencing technologies starting in the mid-2000s dramatically accelerated studies of adaptive immune repertoires and their involvement in infectious diseases, cancer and autoimmunity [1]. In the early days of Ab repertoire sequencing, 454 pyrosequencing was the platform of choice due to its longer read lengths (400–500 bp, sufficient to span the entire length of a rearranged variable domain exon). Subsequently, improvements in sequencing chemistry led to the increased read lengths of the Illumina and Ion Torrent platforms, enabling single or merged paired-end reads to cover a complete variable domain exon at lower cost. Since each BCR is rearranged somatically during B-cell development, distinguishing rare variants from sequencing errors is a major challenge for which a variety of tools have been developed [2].

Oxford Nanopore Technologies' MinION is a highly portable and cost-effective third-generation sequencing instrument, but has not been used for Ab repertoire sequencing due to its high error rate (approximately 10% for single reads). The single-read error rate can be mitigated by sequencing the complementary strands of a double-stranded DNA molecule individually and taking the consensus (so-called '2D' and '1D^2' sequencing). Nanopore sequencing has been applied to genomic [3], metagenomic [4], transcriptomic [5] and HLA typing [6] investigations. Here, we applied it to sequencing of single-domain Ab (sdAb) repertoires and phage-displayed libraries thereof, theorizing that the small size of the rearranged V_HH exon (~300–400 bp, with sequence diversity concentrated in only three complementarity-determining regions [CDRs]) and the absence of heavy:light chain pairing would sufficiently reduce the complexity of the analysis so as to permit identification of sdAb sequences.

We first conducted a preliminary assessment of the data quality derived from nanopore sequencing of a phage-displayed V_HH library prepared from the lymphocytes of llamas (*Lama glama*) immunized with human CD73 (Cat. No. 10904-H08H; Sino Biological, Beijing, China) and human HER2 (Cat. No. HE2-H5225; ACROBiosystems, Beijing, China) as described previously [7,8]. All animal procedures were conducted using protocols approved by the National Research Council Canada Animal Care Committee and in accordance with the guidelines set out in the OMAFRA Animals for Research Act, R.S.O. 1990, c. A.22. Briefly, phagemid DNA was extracted from library-containing *Escherichia coli* TG1 cells using a QIAprep® Spin Miniprep Kit (QIAGEN, CA, USA) and PCR-amplified using primers MJ7 and MJ8 (Table 1) as previously described [8,9]. The resulting amplicon was purified using a PureLink® PCR Purification Kit (Life Technologies, CA, USA), 3′ dA-tailed using the NEBNext® Ultra™ End Repair/dA-Tailing Module (New England Biolabs, MA, USA), ligated to the Nanopore sequencing adaptors supplied in the Ligation Sequencing Kit 1D (SQK-LSK108; Oxford Nanopore Technologies, Oxford, UK) using Blunt/TA Ligase Master Mix (New England Biolabs), and purified by ▶

Table 1. Oligonucleotide primers used in this study.

Primer	Sequence (5′-3′)	Specificity
MJ7	GCCCAGCCGGCCATGGCC	pMED1 phagemid pelB leader
MJ8	TGAGGAGACGGTGACCTGG	V_HH FR4
nano-MJ7	TTTCTGTTGGTGCTGATATTGCNNNNNGCCCAGCCGGCCATGGCC	pMED1 phagemid pelB leader
nano-MJ8	ACTTGCCTGTCGCTCTATCTTCNNNNNTGAGGAGACGGTGACCTGG	V_HH FR4

▸ solid-phase reversible immobilization using AMPure XP beads (Beckman Coulter, CA, USA). Approximately 500 ng of DNA was mixed with library loading beads and sequenced on a MinION instrument (Oxford Nanopore Technologies) using the 48-h sequencing script in MinKNOW v1.13.1, yielding a total of 1,550,579 basecalled reads. Sequence quality was assessed using FastQC [10] and showed a relatively uniform distribution of quality scores across reads (Figure 1A–C). We tested the effect of filtering the data using the FASTX toolkit [11]; as expected, based on the quality scores, filtering stringency had only a minor impact on the proportion of functional, in-frame sdAb sequences identified by IMGT/HighV-QUEST [12] after alignment to alpaca germline IGHV, IGHD and IGHJ genes and removal of insertions and deletions (Figure 1D–F).

Next, we panned the phage-displayed V_HH libraries for four rounds against CD73 and HER2 directly adsorbed in wells of microtiter plates and eluted the bound sdAb-phage using triethylamine as previously described [7,8]. Library phage and phage particles eluted from each round of panning were used as PCR templates to amplify rearranged V_HH genes using primers nano-MJ7 and nano-MJ8 (Table 1), as previously described [8,9]. The resulting amplicons were purified using a PureLink® PCR Purification Kit and 30 ng was used as template for a second PCR in which each sample was barcoded with primers LWB01–LWB10 from the Low Input by PCR Barcoding Kit (SQK-LWB001; Oxford Nanopore Technologies). The amplicons were pooled and purified using a QIAquick® Gel Extraction Kit (QIAGEN) followed by solid-phase reversible immobilization using AMPure XP beads. Approximately 100 fmol of the pooled DNA was ligated with 1 μl of rapid 1D sequencing adaptor in 10 μl ▸

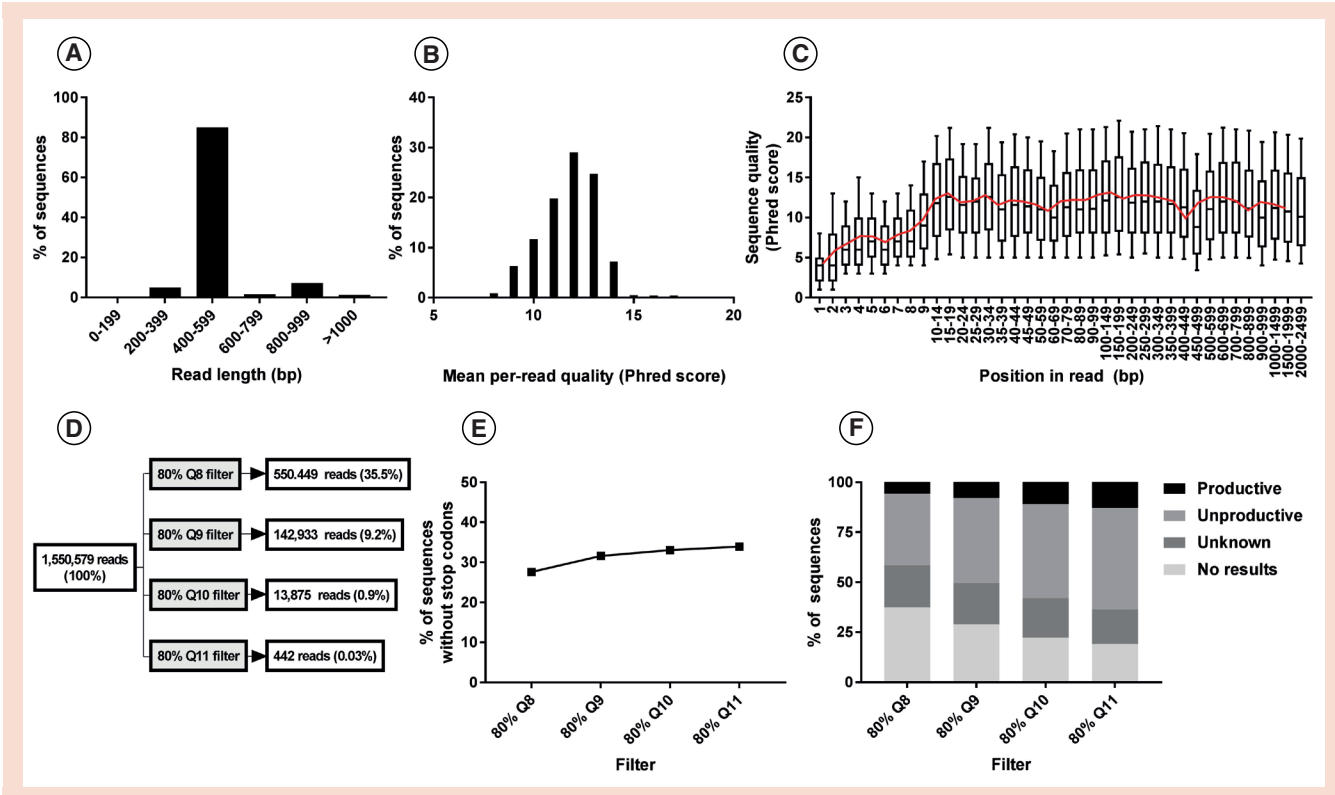

Figure 1. Quality metrics for Oxford Nanopore sequencing of single-domain antibody repertoires and phage-displayed libraries. (A) Read length distribution, (B) mean per-read quality score and (C) distribution of quality scores according to read position. In the boxplot in C, black lines represent medians, the box shows the lower and upper boundaries of the first and third quartiles, respectively, the whiskers represent the 10th and 90th percentiles of the data, and the red line shows the mean quality score. Analyses in A–C were conducted using FastQC [10]. (D) Output after quality filtering with the FASTX toolkit [11] using various stringencies. (E) Frequency of single-domain antibody sequences lacking stop codons and (F) functional annotation using IMGT/HighV-QUEST [12] after filtering.

Table 2. Putative CD73- and HER2-specific single-domain antibody clonotypes identified by nanopore sequencing.

Clonotype	Frequency (%)		Enrichment (fold)		No. of sequences	Outcome
	Round 3	Round 4	Round 3	Round 4		
CD73–1	0.36	0.30	111	93	448	Multiple clonal variants
CD73–2	0.20	0.16			257	Multiple clonal variants
CD73–3	0.09	0.15			166	Out of frame
CD73–4	0.08	0.08	25	24	118	Multiple clonal variants
CD73–5	0.06	0.04	19	14	69	Multiple clonal variants
CD73–6	0.03	0.03	11	10	47	Out of frame
CD73–7	0.23	0.45	73	141	452	Synthesized and expressed
CD73–8	0.14	0.19			211	Out of frame
CD73–9	0.07	0.14			133	Multiple clonal variants
CD73–10	0.07	0.07	20	23	91	Synthesized and expressed
CD73–11	0.49	0.04	15	13	63	No consensus
CD73–12	0.22	0.24			309	Multiple clonal variants
CD73–13	0.10	0.08			104	Out of frame
CD73–14	0.06	0.09			100	Synthesized and expressed
CD73–15	0.06	0.06	19	17	78	Out of frame
CD73–16	0.04	0.04	13	12	53	Out of frame
HER2–1	3.5	2.6	76	56	2,784	Synthesized and expressed
HER2–2	2.3	1.6	50	34	1,651	Out of frame
HER2–3	1.1	1.9			898	Synthesized and expressed
HER2–4	1.0	0.5			650	Out of frame
HER2–5	0.67	1.0			477	Multiple clonal variants
HER2–6	0.57	0.76			381	Out of frame
HER2–7	0.46	0.65			329	Out of frame
HER2–8	0.40	0.37			371	Synthesized and expressed
HER2–9	0.34	0.43			335	Out of frame
HER2–10	0.230	0.44			213	Out of frame

▶ of 10 mM Tris-HCl, pH 8.0, containing 50 mM NaCl for 5 min at room temperature and sequenced on the MinION instrument as described above, yielding a total of 1,225,772 basecalled reads. The run data were demultiplexed using EPI2ME v2.52, aligned without quality filtering with alpaca germline IGHV, IGHD and IGHJ genes using IMGT/HighV-QUEST, and the resulting tabular output was analyzed using R v3.4.3. Between 75 and 80% of sequences were eliminated from the analysis because the sdAb CDR3 was either not detected or was shorter than three residues (Supplementary material). We then identified potential sdAb CDR3s ('clonotypes') of interest using the following criteria: (i) among the ten highest-frequency clonotypes in either of rounds three or four of panning; (ii) among the ten most highly enriched clonotypes in either of rounds three or four of panning with respect to the library; and (iii) free of stop codons. Applying these criteria yielded 16 unique clonotypes of interest for CD73 and ten for HER2 (Table 2). Enrichment of these clonotypes over the course of panning is depicted in Figure 2.

Finally, we collected all copies of the full-length sdAb coding sequences (gapped according to IMGT numbering) corresponding to each clonotype from the original library and all four rounds of panning. The full-length sequences were aligned with Clustal Omega [13] using default parameters (Figure 3A) and the consensus was determined with the Bio3d package [14] using a cutoff of 0.3 (30% sequence identity threshold). We assessed the presence of somatic clonal variation by: (i) examining the alignment for positions where the consensus was ▶

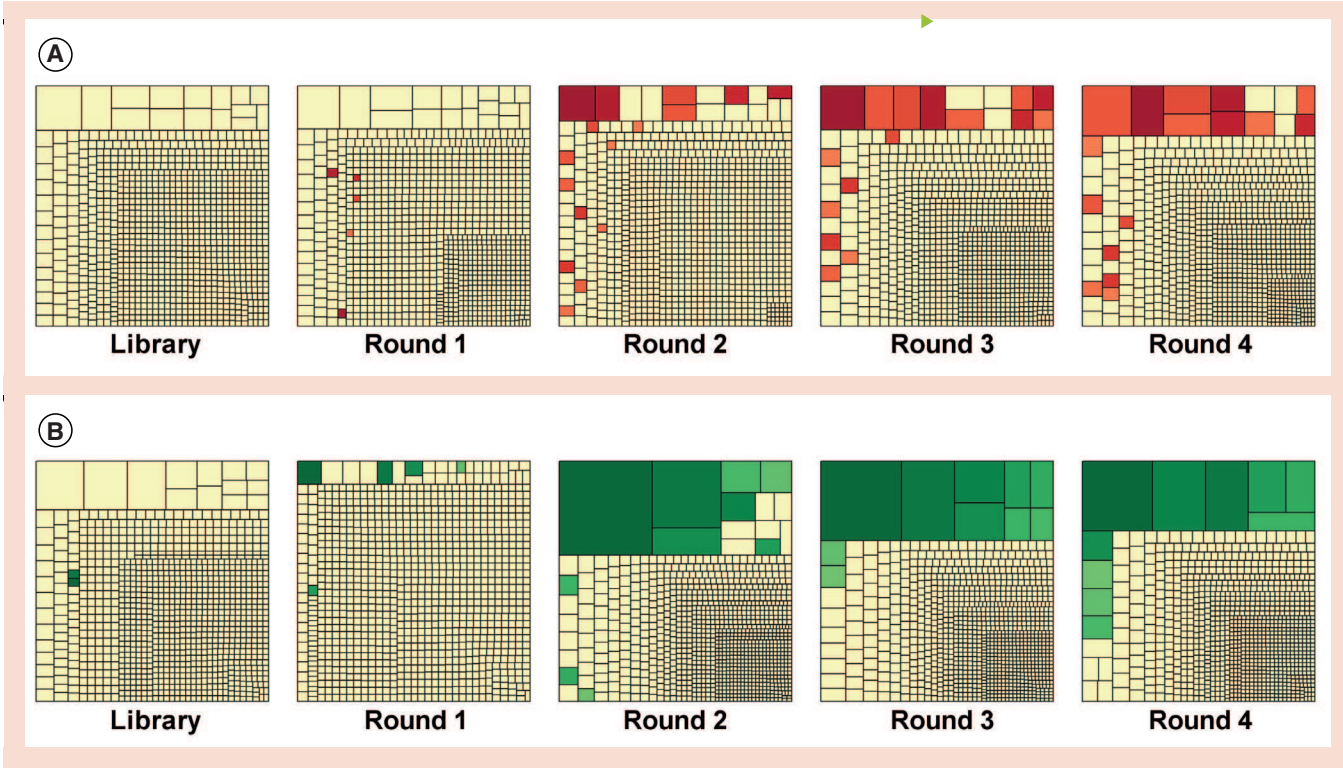

Figure 2. Enrichment of putative antigen-specific single-domain antibody CDR3 clonotypes over the course of four rounds of panning against CD73 (A) and HER2 (B). The most frequent 1000 unique CDR3 sequences are shown for each dataset, with lead clonotypes listed in Table 3 colored in green (CD73) and red (HER2). The size of each square is proportional to the clonotype's frequency. Figure created using R package 'treemap'.
CDR: Complementarity determining region.

dramatically lower (Figure 3A), (ii) determining whether single (Figure 3B & C) or multiple (Figure 3D & E) unique CDR1 and CDR2 sequences were associated with each CDR3; and (iii) visual inspection of the alignment. Only clonotypes for which alignments showed strong consensus throughout and in which singular unique CDR1 and CDR2 sequences could be associated with the index CDR3 sequences were carried forward. A significant number of enriched clonotypes represented sdAb sequences that were frameshifted in CDR3 and could not be corrected by IMGT (6/16 for CD73 and 6/10 for HER2; Table 2); many of these represented frameshifted variants of in-frame enriched clonotypes (data not shown). For clonotypes comprising multiple clonal lineages (as indicated by low alignment consensus and the presence of multiple common CDR1/2 sequences; 6/16 for CD73 and 1/10 for HER2), variants could not be consistently distinguished from one another (e.g., using clustering algorithms). For one clonotype (CD73–11), we could not confidently determine a consensus sequence based on alignment of only 63 full-length sequences. However, three putative anti-CD73 sdAbs (CD73–7, CD73–10 and CD73–14) and three putative anti-HER2 sdAbs (HER2–1, HER2–3 and HER2–8) showed strong consensus sequences and no evidence of clonality; genes encoding these sdAbs were synthesized, cloned into the expression vector pSJF2H and expressed in *E. coli* as previously described [7,8]. All six of the sdAbs bound their cognate antigens with low nanomolar affinities (Figure 3F & G).

In summary, we showed here that Oxford Nanopore sequencing can be used to correctly identify antigen-specific sdAb sequences derived from panning of phage-displayed V_HH libraries by tracking enrichment of CDR3 sequences and then aligning full-length sdAb coding sequences associated with each CDR3. Although we panned the phage-displayed V_HH libraries for four rounds in this study to ensure that the sequenced material was primarily composed of antigen-specific V_HHs, this turned out not to be necessary, as enrichment of antigen-specific V_HHs primarily occurred during the first two rounds of panning (Figure 2). However, minimizing the number of panning rounds would also restrict the available criteria for identifying lead antigen-specific sdAb clonotypes (e.g., to abundance after two rounds of panning and enrichment during the first and second round of panning), and the impact this might have on recovery of antigen-specific sdAbs is unclear. We used this strategy to identify three high-affinity sdAbs against each of two model antigens (CD73 and HER2). This method is currently limited to identification of sdAb sequences that are not members of extended clonal lineages, and requires a minimum of approximately 100 sequences to produce alignments of reasonable quality (Figure 4). Nonetheless, we note that: (i) separation of somatic variants by clustering was partially successful for larger V_HH families, and thus simply increasing throughput represents a viable strategy for delineation of such sequences; and (ii) in V_HH lineages comprising limited clonal variation but with high coverage, many polymorphisms were identifiable by visual inspection (and/or by matching high-frequency CDR1 and CDR2 sequences). We have ignored these possibilities here as they are unlikely to represent general solutions to sdAb

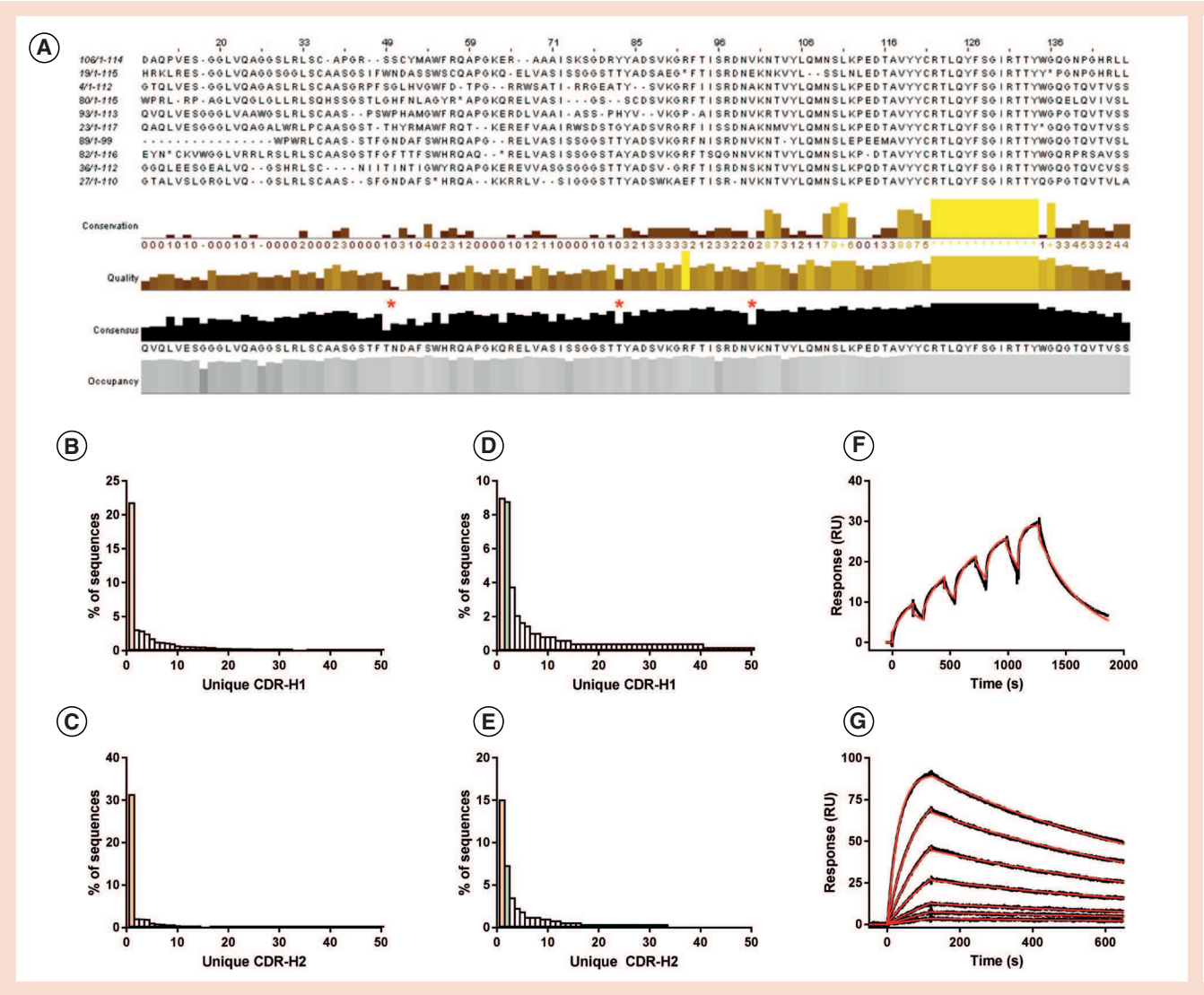

Figure 3. Identification of full-length single-domain antibody sequences by alignment consensus. (A) An example alignment of full-length single-domain antibody sequences derived from a single CDR3 clonotype after Oxford nanopore sequencing and indel removal by IMGT/HighV-QUEST. Red asterisks denote low consensus positions suggestive of clonal variation. Figure produced using Jalview [17]. (B) Distribution of unique CDR1 and (C) distribution of unique CDR2 sequences within a clonotype (HER2–1) comprising a single clonal lineage. (D) Distribution of unique CDR1 and (E) distribution of CDR2 sequences within a clonotype (HER2–5) comprising multiple somatic variants. (F) Binding of recombinant V_HHs CD73–10 and (G) HER2–3 to their cognate antigens using surface plasmon resonance. Each antigen (713–1345 RUs) was immobilized on a sensor chip CM5 or a CM5 Series 5 sensor chip (GE Healthcare, NJ, USA) by amine coupling, and the relevant V_HH was flowed over the surface at concentrations ranging from 12.5 to 200 nM (CD73–10) and 0.25 to 25 nM (HER2–3). Data for CD73–10 and HER2–3 V_HHs were collected on a Biacore T200 instrument (GE Healthcare) and a Biacore 3000 instrument (GE Healthcare), respectively, and analyzed using BIAevaluation 4.1 software. Single-cycle kinetic analysis (for CD73–10) and multi-cycle kinetic analysis (for HER2–3) revealed K_Ds of 15 nM and 1.4 nM, respectively. The affinities of V_HHs CD73–7 (0.6 nM), CD73–14 (35 nM), HER2–1 (6 nM) and HER2–8 (10 nM) were determined in the same manner. CDR: Complementarity determining region; RU: Response unit.

▶ identification. However, using these strategies, we were able to identify several additional unique antigen-specific sdAbs derived from clonotypes comprising multiple clonal lineages and subsequently confirmed them at the single-clone level using conventional Sanger sequencing (Supplementary material). Thus, the overall yield of sdAbs recovered using nanopore sequencing here (~5–10) was similar to that of conventional screens performed in our lab involving ~50–100 single clones, but unlikely to accomplish the ultra-deep repertoire profiling accomplished using other sequencing platforms [15,16]. However, we note that the results shown here represent the worst-case scenario, and we anticipate that sequence quality could be substantially improved using 2D or $1D^2$ reads to achieve Illumina MiSeq-like data at much lower cost.

AUTHOR CONTRIBUTIONS

MJL and KAH designed the study, MJL performed the experiments, KAH analyzed the data and KAH wrote the manuscript.

FINANCIAL & COMPETING INTERESTS DISCLOSURE

This work was supported by funding from the National Research Council Canada. The authors have no other relevant affiliations or financial involvement with any ▶

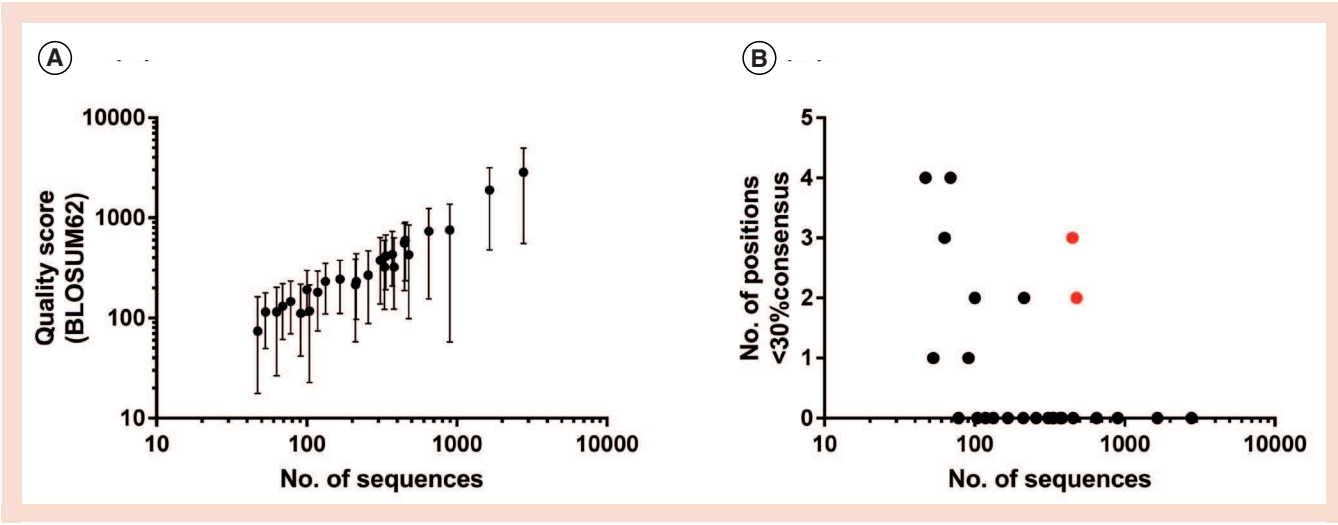

Figure 4. Dependence of alignment quality on sequence coverage. (A) Distribution of median alignment quality score (BLOSUM62) across all positions in the full-length sequences of the clonotypes shown in Table 3. Error bars represent 95% confidence intervals. (B) Number of ambiguous positions falling below 30% consensus (sequence identity). Red circles indicate clonotypes comprising multiple somatic variants, and in these cases low consensus was due to polymorphism rather than poor alignment quality.

organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

OPEN ACCESS

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/

SUPPLEMENTARY DATA

To view the supplementary data that accompany this paper please visit the journal website at: www.future-science.com/doi/suppl/10.2144/btn-2018-0123

REFERENCES

1. Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nat. Biotechnol.* 32(2), 158–168 (2014).
2. Robinson WH. Sequencing the functional antibody repertoire – diagnostic and therapeutic discovery. *Nat. Rev. Rheumatol.* 11(3), 171–182 (2015).
3. Jain M, Koren S, Miga KH *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36(4), 338–345 (2018).
4. Goordial J, Altshuler I, Hindson K, Chan-Yam K, Marcolefas E, Whyte LG. *In situ* field sequencing and life detection in remote (79 degrees 26′ N) Canadian high arctic permafrost ice wedge microbial communities. *Front. Microbiol.* 8, 2594 (2017).
5. Byrne A, Beaudin AE, Olsen HE *et al.* Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat. Commun.* 8, 16027 (2017).
6. Liu C, Xiao F, Hoisington-Lopez J *et al.* Accurate typing of human leukocyte antigen class I genes by Oxford Nanopore sequencing. *J. Mol. Diagn.* 20(4), 428–435 (2018).
7. Baral TN, Mackenzie R, Arbabi Ghahroudi M. Single-domain antibodies and their utility. *Curr. Protoc. Immunol.* 103, Unit 2.17 (2013).
8. Henry KA, Hussack G, Collins C, Zwaagstra JC, Tanha J, Mackenzie CR. Isolation of TGF-β-neutralizing single-domain antibodies of predetermined epitope specificity using next-generation DNA sequencing. *Protein Eng. Des. Sel.* 29(10), 439–443 (2016).
9. Henry KA. Next-generation DNA sequencing of V_H/V_L repertoires: a primer and guide to applications in single-domain antibody discovery. *Methods Mol. Biol.* 1701, 425–446 (2018).
10. Andrews S. FastQC: a quality control tool for high throughput sequence data. www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).
11. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6), 863–864 (2011).
12. Alamyar E, Duroux P, Lefranc MP, Giudicelli V. IMGT® tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods Mol. Biol.* 882, 569–604 (2012).
13. Sievers F, Wilm A, Dineen D *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539 (2011).
14. Grant BJ, Rodrigues AP, Elsawy KM, Mccammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 22(21), 2695–2696 (2006).
15. Deschaght P, Vintem AP, Logghe M *et al.* Large diversity of functional nanobodies from a camelid immune library revealed by an alternative analysis of next-generation sequencing data. *Front. Immunol.* 8, 420 (2017).
16. Turner KB, Naciri J, Liu JL, Anderson GP, Goldman ER, Zabetakis D. Next-generation sequencing of a single domain antibody repertoire reveals quality of phage display selected candidates. *PLoS One* 11(2), e0149393 (2016).
17. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9), 1189–1191 (2009).

First draft submitted: 17 August 2018; Accepted for publication: 18 September 2018