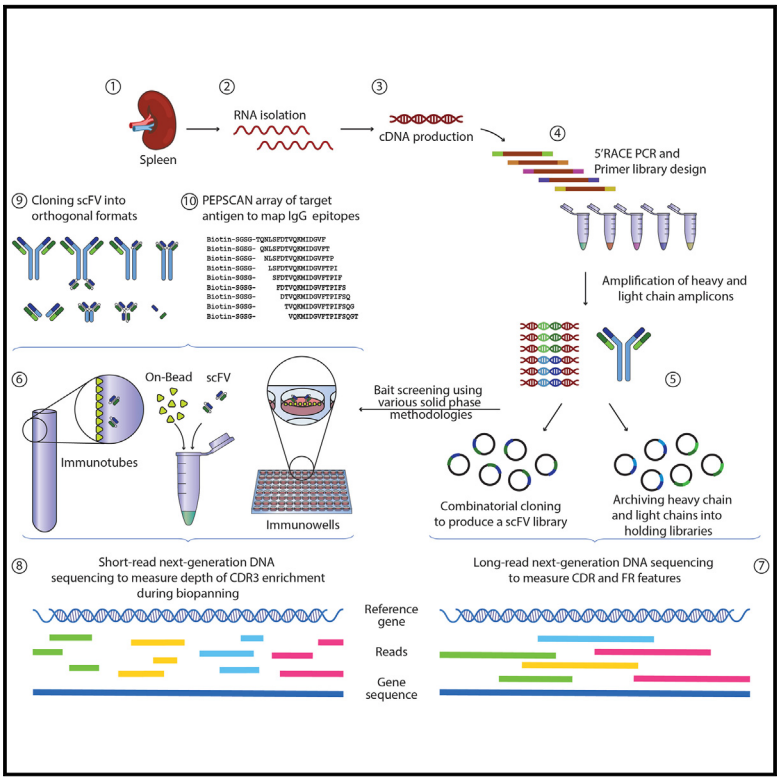
Cell Reports
Methods

Article

The development of a canine single-chain phage antibody library to isolate recombinant antibodies for use in translational cancer research

Graphical abstract



Highlights

- We generate a canine scFv monoclonal antibody library cloned into M13 phage
- We use long-read DNA sequencing to annotate features of antibody genes
- We use short-read DNA sequencing to measure CDR3 enrichment after biopanning
- A chimeric IgG is developed to the canine immune checkpoint receptor PD-1

Authors

Małgorzata Lisowska, Erin G. Worrall, Filip Zavadil-Kokas, ..., David J. Argyle, Maciej Parys, Ted R. Hupp

Correspondence

gosia.nus@gmail.com

In brief

Monoclonal antibodies form compelling medicines for use in cancer treatment. Spontaneous canine tumors can share clinical characteristics with corresponding human diseases. Lisowska et al. present development of a canine antibody library for isolating immunotolerant antibodies. Data from this approach may help to accelerate and diversify therapeutic antibody development.

Lisowska et al., 2025, Cell Reports Methods *5*, 101008
March 24, 2025
https://doi.org/10.1016/j.crmeth.2025.101008

CellPress

Cell Reports Methods

CellPress
OPEN ACCESS

Article

The development of a canine single-chain phage antibody library to isolate recombinant antibodies for use in translational cancer research

Małgorzata Lisowska,[1,10,11,12,*] Erin G. Worrall,[2,11] Filip Zavadil-Kokas,[3,11] Keith Charlton,[5] Euan Murray,[2,4] M. Aiman Mohtar,[2,9] Radovan Krejcir,[3] Vaclav Hrabal,[3] Jack Brydon,[2] Ainhoa Gonzalez Urionabarrenetxea,[2] David G. Saliba,[6] Mariana Grima,[6] Umesh Kalathiya,[1] Petr Muller,[3] Adam Krejci,[3] Borivoj Vojtesek,[3,8] Kathryn L. Ball,[2] Robin Fahraeus,[4] David J. Argyle,[7] Maciej Parys,[7] and Ted R. Hupp[1,2]

[1]International Centre for Cancer Vaccine Science, University of Gdansk, Gdansk, Poland
[2]University of Edinburgh, Institute of Genetics and Cancer, Edinburgh, Scotland, UK
[3]Research Centre for Applied Molecular Oncology, Masaryk Memorial Cancer Institute, Brno, Czech Republic
[4]Inserm UMRS1131, Institut de Génétique Moléculaire, Université Paris 7, Hôpital St Louis, Paris, France
[5]University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen, UK
[6]Faculty of Health Sciences/Department Applied Biomedical Science, Centre for Molecular Medicine and Biobanking, University of Malta, Msida, Malta
[7]The University of Edinburgh, Royal (Dick) School of Veterinary Studies and Roslin Institute, Edinburgh, UK
[8]Laboratory of Growth Regulators, Institute of Experimental Botany, The Czech Academy of Sciences, Olomouc, Czech Republic
[9]Present address: Universiti Kebangsaan Malaysia, UKM Medical Molecular Biology Institute (UMBI), Kuala Lumpur, Malaysia
[10]Present address: Viatris Healthcare Sp. z o.o., ul. Postępu 21B, 02–676 Warszawa
[11]These authors contributed equally
[12]Lead contact
*Correspondence: gosia.nus@gmail.com
https://doi.org/10.1016/j.crmeth.2025.101008

MOTIVATION Spontaneous tumors in canines have been shown to share important genetic, clinical, and immunologic characteristics with their corresponding human equivalents. Naturally occurring canine cancers can therefore form the foundation for disease models that take into account whole organism physiology, while tumor-immune repertoires in canines provide an expanded natural diversity of antibody sequences and targets. To help enable tumor-immune studies in canines, we develop a robust methodology to clone and archive a highly diverse naive canine scFv M13 phage display library. This library can be used for isolating canine immunotolerant monoclonal antibodies that could be tested in preclinical canine diseases and accelerate strategies for improving human health management.

SUMMARY

The development of canine immunotolerant monoclonal antibodies can accelerate the invention of new medicines for both canine and human diseases. We develop a methodology to clone the naive, somatically mutated variable domain repertoire from canine B cell mRNA using 5′RACE PCR. A set of degenerate primers were then designed and used to clone variable domain genes into archival "holding" plasmid libraries. These archived variable domain genes were then combinatorially ligated to produce a scFv M13 phage library. Next-generation long-read and short-read DNA sequencing methodologies were developed to annotate features of the cloned library including CDR diversity and IGHV/IGKV/IGLV subfamily distribution. A synthetic immunoglobulin G was developed from this scFv library to the canine immune checkpoint receptor PD-1. This synthetic platform can be used to clone and annotate archived antibody variable domain genes for use in perpetuity in order to develop improved preclinical models for the treatment of complex human diseases.

INTRODUCTION

Disease in mammals, including viral infections and cancer, is regulated by the genetic background, environmental agents, and the robustness of the host immune system.[1] Developing more physiologically relevant models to integrate human and veterinary medical sciences is emerging as "One World Health."[2,3] This program involves several fields of biological

sciences that focus on domestic livestock, wildlife, companion animals, and humans. The knowledge gained would aim to improve the health of living organisms with diseases, such as cancer or infections caused by viruses and parasites that use wildlife as vectors.[4]

Several animal species can serve as models to reflect a feature of a normal or diseased state in humans, including; yeast (*Saccharomyces cerevisiae*), nematodes (*Caenorhabditis elegans*), fruit fly (*Drosophila melanogaster*), zebrafish (*Danio rerio*), and mouse (*Mus musculus*). The main advantage of the murine model for disease is the ability to create transgenes with an intact mammalian immune system.[5,6] However, a deficiency of the murine model relates to husbandry in relatively pathogen-free environments and that models are often genetically driven rather than spontaneously arising. Immunodeficient xenograft models do not always predict clinical responses relevant for spontaneous diseases.

Millions of domestic dogs develop age-dependent spontaneous cancers with overlapping pathological and immunological features to human cancers.[7,8] The study of canine veterinary patients therefore has the potential to not only improve the health of companion animals, but to support novel therapeutic strategies for the treatment of human cancers.[3] For example, spontaneous canine glioma has molecular features similar to human pediatric, but not adult, glioma.[9] The anatomical similarities between the canine and human brain form a natural cross-species model for testing agents to be delivered *in vivo* into spontaneous mammalian brain tumors.[10] A study of hemangiosarcoma in canines might form another powerful model to improve outcomes in the rare human equivalent, angiosarcoma.[11] Other canine cancers that overlap with the human equivalent include osteosarcoma, urothelial carcinoma, breast, lung, lymphoma, and melanoma.[8] The study of such shared, spontaneously developing cancers can drive novel therapeutic approaches for improving both canine and human disease outcomes.

Immunotherapeutics are emerging as the preferred treatment choice for several human cancers,[12,13] and these platforms could be applied to spontaneous cancers in canine patients. However, immunotolerant ''dogized'' antibodies would be beneficial for modeling antibody responses in canine. Apart from the interchangeable use of mouse and human immunoglobulin (Ig)G scaffolds for using in creating chimeric or fully humanized therapeutic monoclonal antibodies, other animals, including sharks, camelids, and bovines, have provided unique structural variations on the mammalian CDR or IG fold.[14–16] Thus, the cloning of the IG genes of several species has provided opportunities for exploiting antibody diversity in nature.[17–19] Cloning the IG repertoire from dogs has the unique benefit of creating ''non-immunogenic'' species-specific antibodies that are less likely to induce immune toxicity in dogs. Developing monoclonal antibodies for use in veterinary animals is only in its infancy and production of antibodies that are tolerated by the canine immune system is a key step toward canine immunotherapeutics. The first set of ''dogized'' antibodies came from re-engineering rodent hybridoma-derived IgG.[20–26] The next generation of canine antibodies are ''dogized'' at inception (i.e., fully canine, not ''dogized'') and have emerged from phage display libraries.[27,28] The pharmaceutical industry has also invented synthetic canine antibody phage display libraries as well as engineered mice expressing canine immunoglobulin sequences.

Naive antibody libraries have been reported to be a robust source of biologics that mirror immunized libraries.[29] We report here on a detailed methodology to develop a naive canine scFv phage library that introduces three notable advances over contemporary synthetic libraries: (1) The incorporation of 5′ RACE-PCR in the methodology allows for the precise capture of the expressed variable domain repertoire. 5′RACE PCR has been reported to capture the entire repertoire of variable domain sequences from expressed TCR genes.[30,31] (2) The use of heavy and light chain gene ''holding plasmid libraries'' immortalizes the naive, expressed antibody repertoire for use in perpetuity. (3) Leveraging next-generation long-read and short-read DNA sequencing represents a significant advancement in quantifying cloned diversity of the naive mRNA repertoire, assessing cloning library efficiency, defining IG subfamily representation, and measuring CDR3 amplification during biopanning. These advances allow the implementation of a toolbox that can exploit the untapped potential of veterinary medicine to create spontaneous disease-specific antibody libraries[32] that is not possible with the current mouse, rat, camelid, shark, and other animals used to create antibody libraries.

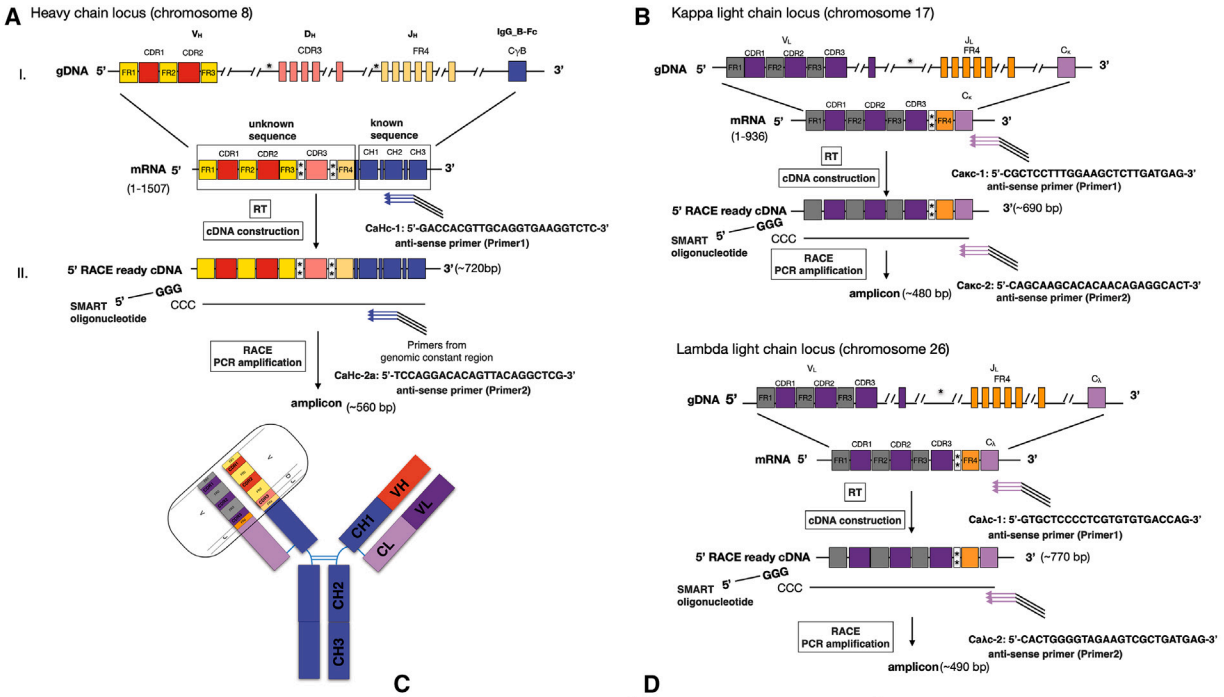
RESULTS

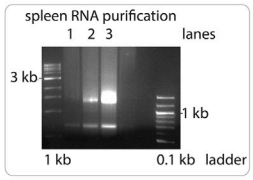
Development of primers for the amplification of canine heavy and light chain variable domain genes using 5′RACE PCR

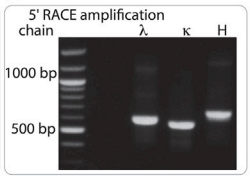
The first-generation canine scFv library[27] utilized primers designed from known genomic DNA sequences to amplify the heavy and light chains, and then these products were PCR amplified into scFvs. We took an alternative approach to clone separately the heavy and light chain genes, using primers designed from 5′RACE RNA-derived sequences, into holding vectors prior to cloning into scFv libraries. Both methods have their advantages and disadvantages; the first method[27] has the advantage of being rapid, while the main disadvantage is that affinity matured sequences with mutations near the 5′ end or in FR4 (J-region) of the mRNA will not necessarily be fully captured. By contrast, while the construction of holding libraries is more time-consuming, the advantage is that this immortalizes the naive antibody repertoire into vectors that can be accessed in perpetuity. The use of holding libraries also allows for future chain-shuffling to increase diversity, construction of alternate conformers (such as FAB libraries), developing novel fusion proteins (such as chromobodies that would allow for fluorescent antibody imaging),[33] and directed evolution to increase diversity, which is an advantage of any *in vitro* display system where buffer conditions and bait conformation can be controlled.[34]

To obtain heavy and light chain variable genes for cloning into holding libraries (Figures 1A and 1B), mRNAs representing the naive somatically mutated repertoire of the canine immune system were amplified using 5′RACE PCR. The 5′RACE PCR used pooled RNA from different spleen tissues (Figure 1C). 5′RACE PCR yielded the dominant 5′ sequences (i.e., FR1) of the naive mRNA repertoire of the variable domain genes. Using the cDNA template from total RNA (Figure 1C), 5′RACE PCR yielded

CellPress
OPEN ACCESS







E Canine Heavy Chain primers

(i) CaJH-FOR-AscI

AscI

1- 5′-CTCAGAAGGCGCGCCAGAAGATTTACCTTCTGAGGAGACGGTGACCAGGGTTC
2- 5′-CTCAGAAGGCGCGCCAGAAGATTTACCTTCTGAGGAGACGGTGACCAGGGTSC
3- 5′-CTCAGAAGGCGCGCCAGAAGATTTACCTTCTGAGGACACGAAGAGTGAGGTGC
4- 5′-CTCAGAAGGCGCGCCAGAAGATTTACCTTCTGAGGAGACGGTGACCAGGGCTC

(ii) CaVH-BACK-SfiI

SfiI NcoI

1 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGAGGTGCAGCTGGTGGAGTCTG
2 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGAGGTACAGCTGGTGGARTCTG
3 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGAGGAGCAACTGGTGGAGTTTG
4 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGAACTCACACTRCAGGAGTCAG
5 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGAGGTGCAWCTGGTGGAGTCTG
6 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGAGGWGCAACTGGTGGAGTCTG
7 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGARTTGCAGCTGGTRGAGTCTG
8 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGAGGTSCAGCTGGTGGAGWCTG
9 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGAGRTRCAATTGGTGGARTCTG
10 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGAAGTRCAGCTGSTGGAGTMTG
11 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGAGGGGCAGCTGGCGGAGTCTG
12 5′-TCGCAACTGCGGCCCAGCCGGCCATGGCCGRGGTGCAGCTGGCGGAGTCTG

F Canine K Chain primers

(i) CaJκ-FOR-NotI

NotI

1- 5′-TCGACTTGCGGCCGCTTTTATCTCCAGCTTGGTTCCCTG
2- 5′-TCGACTTGCGGCCGCTTTGAGCTCCACCTTGGTTCCTGC
3- 5′-TCGACTTGCGGCCGCTTTGATCTCCACCTTGGTCCCTTG
4- 5′-TCGACTTGCGGCCGCTTTGATCTCCAGTTTGGTCCCTTG

(ii) CaVκ-BACK-MluI

MluI

1-5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGATATTGTCATGACACAGACCCCAC
2-5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGATATTGTCATGACACAGACGCCAC
3-5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGAAATCGTGATGACICAGTCTCCAG
4-5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGATATTGTCATGACACAGGCCCCAC
5-5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGATATCGTCATGACACAGACYCCAC
6-5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGAGGCCGTGATGACGCAGACCCCAC
7-5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGATITTTCATGACACAGACCCCAC

G Canine λ Chain primers

(i) CaJλ-FOR-NotI

NotI

1- 5′-TCGACTTGCGGCCGCACCGAGGACGGTCAGGTGGGTGC
2- 5′-TCGACTTGCGGCCGCACCGAGGACGGTCAGATGGGTGC
3- 5′-TCGACTTGCGGCCGCACCAAGGACGGTCAGTTGGGTGC
4- 5′-TCGACTTGCGGCCGCACCGAGGACGGTCAIGTGGGIGCCTG

(ii) CaVλ-BACK-MluI

MluI

1- 5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCCCAGTCTGTGCTGACTCAGCCGG
2- 5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCCTCCTATGTGCTGACACAGCTGC
3- 5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCCTCCTATGTGCTGTCTCAGCCGC
4- 5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCCTCCAGTGTGCTGACTCAGCCTC
5- 5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACRCCTCCACTGGGTTGAATCAGGCTC
6- 5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCCTCCCCTGGGCTGAATCAGCCTC
7- 5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCATTGCCCGTGCTGACCCAGCCTAC
8- 5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCCTCCTATGTGTTGACTCAGCTGC
9- 5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCCTCCTATGAGCTGACTCAGCCAC
10-5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCCCAGGCTGTICTGACTCAGCCAC
11-5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCCCAGTCTATGCTGACTCAGCCAG
12-5′-TCTTCTGACGCGTCTGGCGAGTCTAAAGTGGATGACGCCCAGTCTGTGTTGACTCARC

H Degenerate code

W= A or T
R = A or G
S = C or G
M= A or C
I = Inosine
Y = C or T

(legend on next page)

PCR products (Figure 1D) with primers derived from the canine genomic constant region (H, λ, and κ genes; Figures 1A and 1B). Following amplification of these variable regions, the PCR products were cloned into pGEM-T, and then ∼60 plasmids were sequenced to facilitate primer design from each gene family (CH, Cλ, and Cκ). Representative 5′ and 3′ DNA sequences are listed in Figures S1 and S2. Based on these sequences, a set of degenerate primers were designed for amplification of the variable domain repertoire (Figures 1E–1H). The oligonucleotides have restriction sites embedded in the 5′ and 3′ primers for directional cloning of amplicons into holding plasmids. In total, we could generate up to 12 different PCR product pools for the heavy chain, 7 PCR product pools for the kappa chain, and 12 PCR product pools for the lambda chain. Once the heavy and light chain repertoires were cloned into holding vectors, a final cloning strategy produced a combinatorial fusion protein library composed of Hv-Lv-gIII (Figure 2). This allowed for display of the scFv canine antibody molecules on the surface of the M13 bacteriophage.

Cloning of the canine framework and variable domains into separate heavy chain and light antibody holding libraries

To clone the heavy chain genes, a PCR temperature gradient was optimized for each VH primer set (Figure S3A). The CaJH-FOR forward primers (1–4;Figure 1Ei) were mixed equally with each individual reverse primer (CaVH-BACK; numbers 1–12 (Figure 1Eii) in 36 × 50 μL aliquots for each primer pair to produce heavy chain products in 12 different pools (Figure S3B). These amplicons (∼400 base pairs [bp]) were purified (Figure S3C), restricted with NcoI and AscI, and gel purified (Figure S3D) for ligation into NcoI and AscI restricted pIHV-1 (Figure S3E). A representative compilation of one heavy chain is shown in Figure S3F. The estimated size of each of the 12 heavy chain holding libraries by phage titering is summarized in Figure S3G, which gives rise to over 10^9 total phage particles for all heavy chain holding libraries. The position of the C-terminal restriction sites (NotI, AscI), as well as the EGKSSGA partial linker sequence that will fuse with the light chain amplicons to create a longer linker upon scFv construction is highlighted (Figure S3H). This heavy chain holding library was also engineered with a stop codon prior to the gIII protein (Figure S3I) to reduce any toxicity from "leaky" synthesis of the heavy chain-gIII fusion protein in bacteria. The annotated DNA sequence around the AscI cloning site (Figure S3I) shows how one heavy chain sequence can be traced back (Figure S3J) to a primer designed to capture FR4 (from Figure 1Ei). The relative diversity of all 12 holding libraries was evaluated by Sanger sequencing (Tables S1A–S1L). The relative diversity of FR1 regions were also compared with FR1 sequences derived using primers designed from genomically derived DNA sequences[27] (Figure S4). After this small-scale validation of the 12 heavy chain holding libraries by sequencing individual plasmids (Table S1A–S1L), one arbitrary heavy chain holding library (VH-11) was subjected to next-generation long-read DNA sequencing to define the key features, including (1) percentage insert cloned after electroporation of the ligation mixture, (2) percentage of plasmids with the N-terminal PelB leader sequence, and (3) the C-terminal MYC tag present in-frame (Table 1). This long-read DNA sequencing methodology provides a proof-of-concept for future users to define molecular features of cloning outputs.

For the VH-11 holding library, the total DNA sequencing read was 2,087,704 (Table 1). The number of reads containing both the N-terminal PelB leader sequence and the C-terminal MYC tag ranged from 479,006 to 485,770 (Table 1). Thus, we estimated that the VH-11 holding library had a ligation efficiency of ∼25% in terms of plasmids with a "complete" heavy chain. Because the long-read DNA sequencing was derived from plasmid preparations that were themselves derived from the ligated and electroporated cells; this might have underestimated the cloning efficiency of this heavy chain holding library because "empty" plasmids might have a growth advantage when the material was scaled up for plasmid isolation. As the long-read DNA sequencing was derived from only 1 of 12 heavy chain holding

Figure 1. Strategy for cloning the naive repertoire of canine heavy and light chain variable genes into a phage display library

(A) Structure of canine heavy chains and the location of the primers for the 1st round cDNA production and for the second round 5′RACE PCR amplification. (i) The heavy chain locus (chromosome 8) is depicted along with VH (FR1, CDR1, FR2, CDR2, FR3), DH(CDR3), JH(FR4), and Fc regions. cDNA production uses a primer hybridizing to the CH1/CH2 domain producing a product of ∼720 bp. (ii) The 5′RACE PCR uses 5′ SMART oligonucleotide primers with a GGG tail and the antisense primer hybridizing to the CH1 domain producing a product of ∼560 bp. The 5′RACE amplicons were cloned into pGEM-T with subsequent DNA sequencing (Figures S2) to allow for the design of degenerate primers.

(B) The kappa chain (chromosome 17) and the lambda chain (chromosome 26) are depicted along with their respective exonic domains. The principle is the same, including first round cDNA production and second round 5′RACE PCR amplification using primers that result in final PCR products of ∼480 bp for the kappa chain and ∼490 bp for the lambda chain.

(C) Purification of total RNA from pooled anonymized canine spleen tissue.

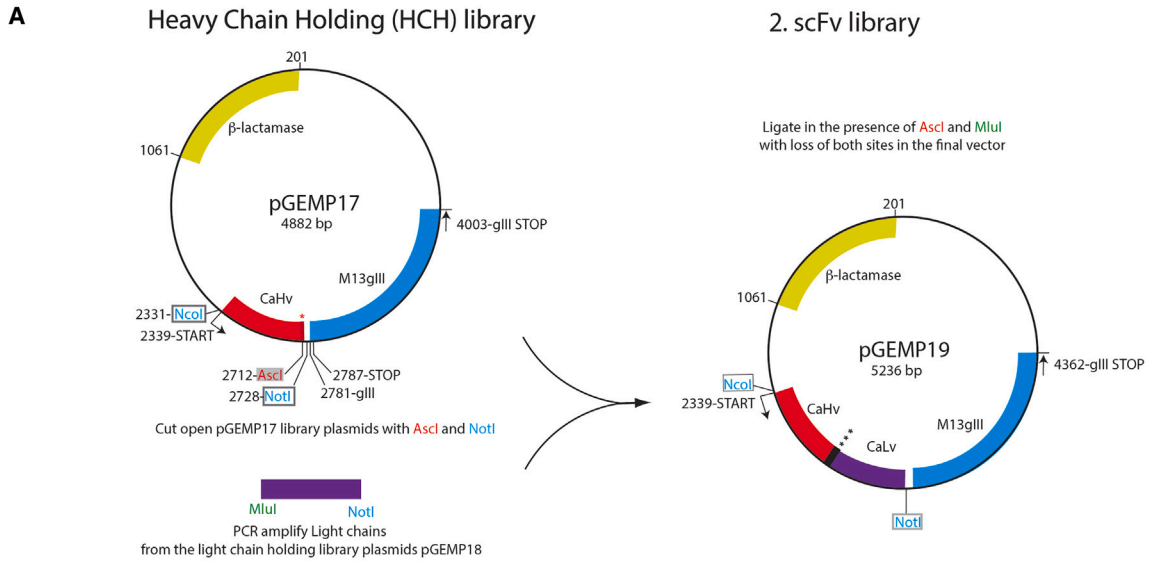
(D) Reverse primers from the canine constant heavy, lambda light, or kappa light chain (CH, Cλ, and Cκ) genomic sequences were used to amplify the cDNA using 5′RACE PCR.

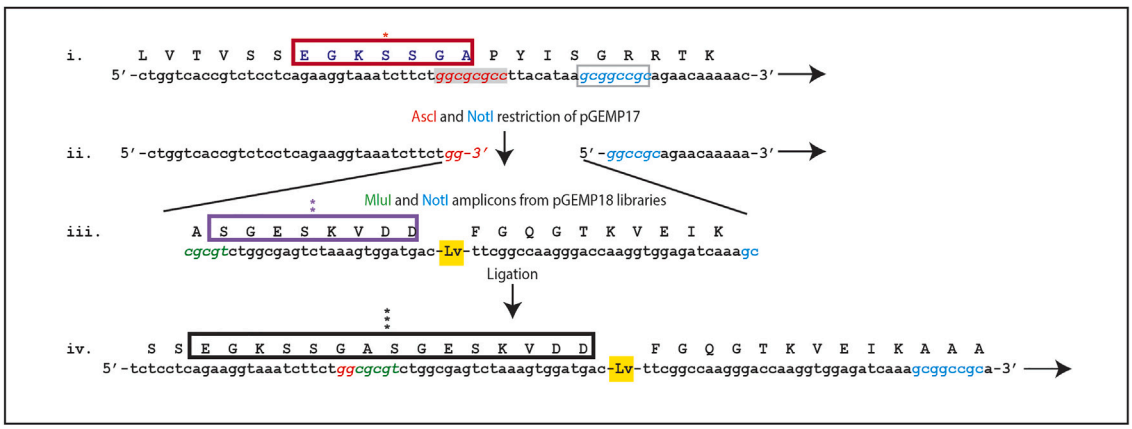
(E–H). Degenerate primer sets used to clone antibody fragments from canine splenic cDNA.

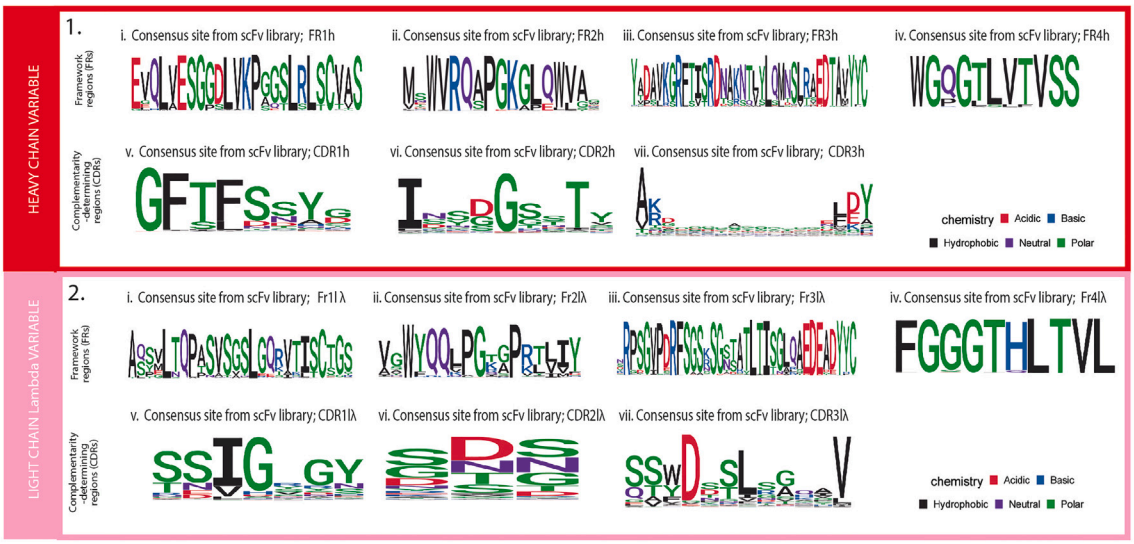
(E) Heavy chain. (i) FR4 primers. Based on the 5′RACE-derived sequences (Figures S2A and S2B), the CaJ_H-forward primer captures FR4. The primers have an AscI site (blue) adjacent to a "Linker 1" region that would form half of the linker fusing the heavy and light chains in a scFv fusion protein, which is in turn adjacent to the 23 bp FR4-specific sequences derived from 5′RACE sequencing. (ii) FR1 primers. Based on the 5′RACE-derived sequences (Figures S2A and S2B), the CaV_H-BACK primer captures FR1. The primer has a SfiI and NcoI site adjacent to 22-bp FR1 specific sequences.

(F and G) Lambda and kappa light chains. (i) FR4 primers. Based on the 5′ RACE sequences from the kappa or lambda chain data (Figure S1 and Figures S2C and S2D), the CaJλ-forward or CaJκ-forward primers capture FR4. The primer has a NotI site and the adjacent sequence highlights the 24-bp framework FR4-specific sequences. (ii) FR1 primers. Based on the 5′RACE sequences from the kappa and light chain sequences, the CaJλ-BACK or CaJκ-BACK primer captures FR1. The primer has a MluI site adjacent to the 25- to 26-bp FR1-specific sequences.

(H). The sequences highlight the degeneracy code within the primers.







(legend on next page)

libraries, we do not know the ligation efficiency of all 12 heavy chain holding libraries with full-length in-frame sequences. The consensus sites of CDR1, CDR2, CDR3, FR1, FR2, and FR3 from the long-read DNA sequencing were defined (Figure S3K). A recent study using 5′RACE PCR to evaluate the BCR diversity in canine patient samples[35] gave rise to consensus sequences similar to our heavy chain holding libraries (Figure S3K).

The same cloning strategy used for the heavy chain library archive was used for light and kappa chain variable domain cloning (Figure S5); (1) a temperature gradient was used to optimize the PCR (data not shown); (2) a scale up of the PCR amplification was performed to acquire as much material as possible (1.8 mL pot for each Vλ or Vκ primer pair); and (3) MluI and NotI restriction and repurification of the amplicons to estimate the yield (Figure S5A). For the lambda chain primers, 11 out of 12 primer sets generated a PCR amplicon and for the kappa chain, 6 of 7 PCR primer sets generated an amplicon (Figure S5A). The gel purified, NotI and MluI restricted light chain amplicons were individually ligated into one fixed VH-12 heavy chain containing vector restricted with NotI and AscI (Figures S5B and S5C), creating 17 light chain holding plasmid libraries. Plasmids from the ligation were restricted to estimate cloning success based on the recovery of the ∼800-bp fragment containing a fusion of a fixed heavy chain and variable light chains (Figure S5B). The estimated size of each of the lambda and kappa light chain holding libraries by phage titering is summarized in Figures S5D and S5E, respectively, which gave rise to over 10^9 total phage particles for both pooled light chain holding libraries. A compilation (Figure S5F) of a heavy chain fusion to light chain library is shown, including (1) FR4 sequences from the heavy chain; (2) the full in-frame linker (linker I and linker II) between the heavy chain and light chain (EGKSSGA-SGESKVDD); and (3) fusion of the linker to the FR1 of the light chain. The ligation of the NotI and MluI restricted light chain amplicons into the VH-12 heavy chain containing vector restricted with NotI and AscI results in loss of both AscI and MluI sites upstream of the NotI site in the C terminus of the fusion protein (Figure S5G). Consensus sites from long-read DNA sequencing of the light chains is summarized in Figures S5H and S5I.

Preparation of scFv-phage libraries

To generate an scFv-phage library we combined equally the DNA from each of the 12 AscI and NotI restricted heavy chain holding plasmid libraries (Figure 2A) and then ligated in a mixture of kappa light chain and lambda light chain amplified PCR products from the light chain holding libraries, which were restricted with MluI and NotI (Figure 2A). Cloning of the heavy chain and light chains together generates the 15 amino acid linker fusing the heavy and light chains with AscI/MluI restriction site loss (Figure 2B). This one-pot ligation had each of the 12 heavy chain holding libraries mixed with all 17 light chain amplicons to create 12 ligation reactions that were then all pooled into one electroporation reaction. It is highly likely that the cloning efficiency of the scFv library would have been improved if we individually optimized the ligation of all 12 heavy chain holding libraries to each light chain amplicon, creating 204 separate ligations and electroporations yielding 204 separate phage libraries. However, for general ease of use and transferability, we focused on a representative example of the one-pot shotgun ligation and electroporation. Several aliquots of the pooled shotgun ligations were electroporated into TG1 cells and frozen pellets stored at −80°C; one of these electroporated pellets was processed for long-read DNA sequencing (Table 1) and another two electroporated pellets were used for functional characterizations (below).

Next-generation long-read DNA sequencing was performed on one aliquot of the scFv library. For the scFv library, the total sequencing read number was 1,677,094 (Table 1). The number of reads containing both the N-terminal PelB leader sequence and the C-terminal MYC tag ranged from 61,078 to 62,072 (Table 1). Thus, we estimate that this scFv library aliquot had a ligation efficiency of 3.6% in terms of plasmids with a "complete" scFv in-frame chain. A consensus site was acquired for the heavy chain and the lambda light chain domains (Figure 2C). As the kappa chain composed less than 5% of this scFv library (Table 2), we did not develop a consensus site for kappa chains. The long-read DNA sequencing data were used to examine domain features of the scFv molecules (Table 2). The length range of CDR and FR domains from the lambda, kappa, and heavy chains were evaluated (Table 2). The number of sequencing reads in the variable domains as a function of the length in amino acids were tabulated (Table 2). The data demonstrated that the kappa light chains were under-represented by ∼10-fold compared with the lambda light chains, similar to natural distributions ∼90% to ∼10%.[35] The representation of the different IGH, IGK, and IGL subfamilies is summarized in Figures 3A–3C. Comparing these data to the 5′RACE-PCR derived IG family representation,[35] there is a general concordance between the two datasets, suggesting that the primer sets used for cloning variable domain genes (Figures 1E–1H) do not introduce a significant bias into the IG family alleles cloned.

Figure 2. Strategy for assembling the full scFv library from holding libraries

(A) The heavy chain holding library pGEMP17 (as in Figure S3) was restricted with AscI and NotI. Each light chain holding library from pGEMP18 (Figure S5) was used to PCR amplify the 11 lambda and six kappa light chain genes using the MluI/NotI primer sets, then restricted with MluI and NotI, followed by gel purification and then ligated into the AscI and NotI site of the heavy chain libraries. Every individual heavy chain holding library (1–12, pGEMP17) restricted with AscI and NotI was ligated to the total pool of each lambda and kappa light chain holding library MluI and NotI restricted amplicons creating 12 total "one-pot" scFv library ligations.

(B) The flow diagram depicts the detailed DNA sequence structure of the (i) pGEMP17 heavy chain holding vector; (ii) the AscI-NotI restricted pGEMP17 libraries; (iii) the MluI-NotI light chain amplicons; and (iv) after ligation, the structure of the linker fusion in-between the heavy chain and light chain scFV sequences (pGEMP19).

(C) The consensus sites from heavy chain and lambda light chain domains were extracted from long-read DNA sequencing reactions from one scFv library, as summarized from Table 2. Table S2 contains the number of reads in common between the heavy and lambda chain primer sequences used (Figures 1E and 1G) and the scFv sequences.

Figure 3. IG family distribution in the cloned scFV or holding chain libraries

(A and B) IGHV subfamily distribution using long-read DNA sequencing from (A) the VH-11 holding library or the (B) scFV library. Fifty-three different IGHV subfamilies were identified in the VH-11 holding library from ~480,000 DNA sequencing reads. Fifty-two different IGHV subfamilies were identified in the scFv library out of ~60,000 DNA sequencing reads. IGHV repertoire coverage using 5′-RACE PCR[35] identified 52 IGHV subfamilies. The red arrows highlight the top 6

(legend continued on next page)

The use of next-generation short-read DNA sequencing to quantify CDR3 domain enrichment during biopanning

To screen the scFv library for antibodies, we used EGFP and mCHERRY as model proteins to determine the ability of the scFv library to discriminate between proteins of different sequences but related overall structure. After six rounds of biopanning, all pools were evaluated using next-generation short-read DNA sequencing of CDR3s (Figure 4) and ELISA (Figure S6). scFv-phage activity was observed using ELISA for EGFP or mCHERRY starting in round 4 (Figure S6A). The round 5 and round 6 pools were infected into bacterial cells to acquire individual colonies and monoclonals were identified that were EGFP or mCHERRY specific (data not shown). One representative scFv specific for mCHERRY (clone 5G) and one specific for EGFP (clone 3B) were purified from bacteria and shown to be specific for mCHERRY or EGFP, using a dot-blot (Figure S6B). The 5G anti-CHERRY scFv could detect the mCHERRY protein in an immunoblot (Figure S6C). The sequence of the mCHERRY scFv is shown in Figure S6D and the comparison of the 5G anti-CHERRY scFv with the 3B anti-EGFP is in Figure S6E.

To estimate the diversity of the scFv library and the enrichment of specific antibodies throughout the screening procedure, a methodology was developed that subjected the parental library and biopanned phage pools to next-generation short-read DNA sequencing (Figure 4A). The bar-coded primers are listed in Table S3. The reverse primers hybridized to conserved elements in FR4 of the heavy and light chains and were bar-coded with base triplets to select out tagged DNA sequences. Amplicons from all six rounds (Figures 4C and 4D) were gel purified, pooled (Figure 4E), and subjected to Illumina platforms that generate 100-bp DNA sequencing reads that are bar-coded. The two next-generation sequencing reactions (heavy or light chain pools) contained pooled PCR products with the CDR3s of (1) the naive parental scFv library (using bar codes of AAA); (2) heavy or light chains from the EGFP screen rounds 1–6 (bar codes of AAT, AAC, AAG, TAT, CAT, and GAT); and (3) heavy and light chains from the mCHERRY screen rounds 1–6 (bar codes of ACT, AGT, ATT, TTT, TTA, and TTC). A total of 4,925,748 FR3-CDR3 sequences were obtained with 3,041,316 sequences acquired from the heavy chain amplicons and 1,884,432 sequences selected derived from the light chain amplicons.

In Figure 4F, the total amplicons from each bar-coded heavy chain pool were plotted alongside the unique FR3-CDR3s (e.g., FR3-CDR3 sequences occurring just once). The naive parental library (bar code AAA; Figure 4F) yielded 256,513 FR3-CDR3 heavy chain DNA sequences, and of these, 85,649 DNA sequences were unique (Figure 4F). Based on this, we extrapolated and estimated that a library of 1×10^9 heavy chain sequences would have 3.3×10^8 unique FR3-CDR3s. Similarly, in Figure 4F, the total amplicons from each bar-coded light chain pool were plotted alongside the unique FR3-CDR3s. In this case, the naive parental library yielded 140,335 DNA sequences of which 48,038 were unique (Figure 4F) with the unique FR3-CDR3 frequency being ~30%. Further combinatorial diversity of above $\sim 3 \times 10^8$ unique FR3-CDR3s would be acquired through combinatorial mixing of the heavy and light chain sequences into a scFv fusion protein.

As one example of how the next-generation short DNA sequencing data can be analyzed, the top 50,000 sequences from the parental library, EGFP round 1–6, and mCHERRY rounds 1–6 were tabulated (data not shown). The top seven heavy chain amplicons from the mCHERRY and the EGFP screens are segregated based on those with the highest numbers in the mCHERRY screen round 6 (Figure 4G). The most enriched sequence in round 6 of the mCHERRY screen (clone I; CDR3 sequence ARIDWLYGTLEY) was identified 9,181 times in the DNA reads from the parental library and was identified 191,532 times in round 6, with enrichment generally seen by round 4 (137,415 sequences). This FR3-CDR3 sequence (AEDTAVYYC-ARIDWLYGTLEY) was identical to the mCHERRY specific heavy chain characterized in Figure S6D. A similar annotation of the light chain FR3-CDR3s from the mCHERRY screen is summarized in Figure 4H. In summary, next-generation short-read DNA sequencing provided a methodology to measure (1) extent of CDR3 diversity in any given library and (2) CDR3 antibody sequences enriched during biopanning.

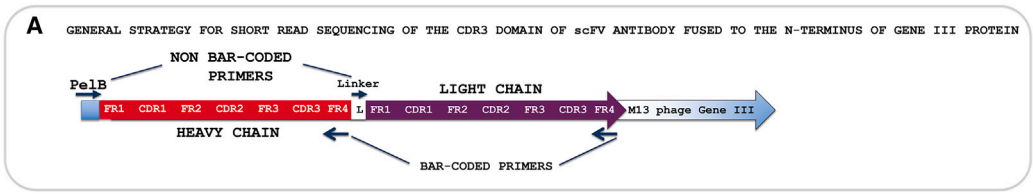
Miniaturization of screening the scFv library against cancer associated targets

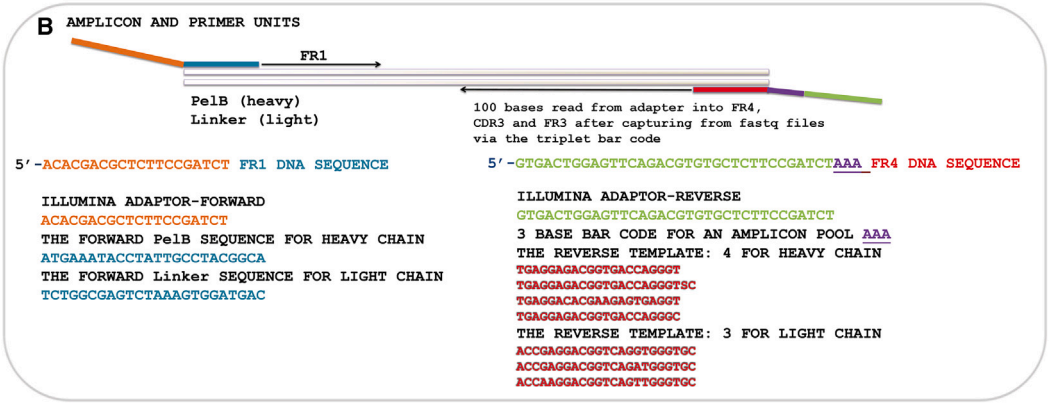
We next applied the scFv library to more physiologically relevant proteins, the pro-oncogenic receptor EGFR[26] (Figure S7) and the immune checkpoint receptor, canine PD-1 (Figure 5), to determine whether an scFv can be acquired to clinically relevant targets. In addition, we vastly reduced the amount of parental scFv-phage library used by developing miniaturized screens. In terms of volume of parental phage library used, the traditional methods might dilute ~2 mL of high-titer parental phage antibody library in a 4-mL final volume using immunotubes (Figure S6), while 20 μL of parental phage library was used in the miniaturization screens. The miniaturized method used $\sim 5–10 \times 10^{10}$ input phage units (in 20 μL) of the parental library ($\sim 5–10 \times 10^{12}$ per mL phage titer). The first miniaturization protocol used synthetic peptides as antigens derived from the canine EGFR and demonstrated

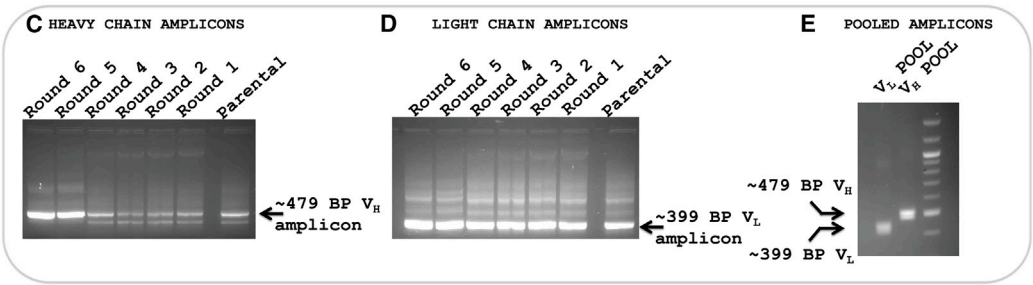
IGHV subfamilies identified using 5′-RACE PCR.[35] The concordance between these two datasets suggests that our cloning primers do not introduce a significant bias into the cloning of the most dominant canine IGH gene subfamilies.

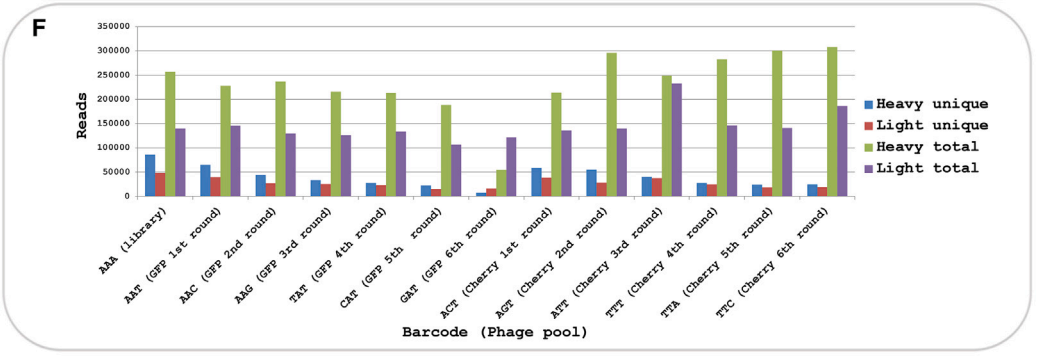
(C and D) IGLV and IGKV subfamily coverage using long-read DNA sequencing reads from pooled light chain libraries or the scFV library. IGLV1-55, IGLV1-58, IGLV1-138, IGLV1-136, and IGLV1-141 are the most dominant lambda chain alleles detected using 5′RACE PCR.[35] These same five subfamilies dominate in the scFv library DNA reads (red arrows). Thus, there is a general overlap in the lambda chain representation, with IGLV1 subfamily dominating, using either the 5′-RACE PCR data or the cloned genes used with our primer sets (Figures 1E–1H). We observed high representation of IGLV1-48 and IGLV4-10 (green arrows). Altogether, ~70 IGLV subfamilies were detected using 5′RACE PCR,[35] while 58 IGLV subfamilies were detected in the scFv library from ~60,000 reads. Dominance of IGKV2-16 > IGKV2-5 > IGKV2-11 > IGKV2-4 is observed using 5′RACE PCR[35] and these same subfamilies form the top dominant genes in the light chain library but in the order of IGKV2-11 > IGKV2-16 > IGKV2-5 > IGKV2-4 (red arrows). Altogether, 13 IGKV subfamilies were detected using 5′RACE PCR.[35] We observe 14 IGKV subfamilies in the light chain holding library from ~116,000 reads. The data are plotted as the number of reads as a function of allele.

CellPress
OPEN ACCESS









G

HEAVY CHAIN SEQUENCES BAR CODE	PARENTAL AAA	R1 AAT	R2 AAC	R3 AAG	R4 TAT	R5 CAT	R6 GAT	R1 ACT	R2 AGT	R3 ATT	R4 TTT	R5 TTA	R6 TTC
I. AEDTAVYYCARIDWLYGTLEYWGQG	9181	7739	6098	5753	8015	9647	2672	7218	8612	24531	137415	188487	191532
II. DTAVYYCARPYDSYWADNFEYWGQG	1512	1308	1025	959	1298	1597	441	1194	1201	4083	34093	16773	22001
III. VEDTAVYYCATRDTQDGDFEYWGQG	281	266	184	178	272	237	272	211	234	538	2216	3491	7080
IV. AEDTAVYYCARIDWLYGPLEYWGQG	187	143	136	97	172	183	60	146	171	451	2707	3608	3773
V. AEDTAVVYCATRDTWDGDFEYWGQG	63	49	31	47	56	68	13	49	61	105	471	737	1570
VI. VEDTAVVYCARIDWLYGTLEYWGQG	70	50	38	33	59	63	32	50	53	180	671	1047	1372
VII. AEDTAVYYCARIDCLYGTLEYWGQG	53	56	28	43	37	61	16	49	49	161	819	1141	1200
FR3 *CDR3* *FR4*													

H

LIGHT CHAIN SEQUENCES BAR CODE	PARENTAL AAA	R1 AAT	R2 AAC	R3 AAG	R4 TAT	R5 CAT	R6 GAT	R1 ACT	R2 AGT	R3 ATT	R4 TTT	R5 TTA	R6 TTC
		GFP						mCHERRY					
I. ALAEDEADYYCQVWDSSAKFVFGG	1920	1779	1772	1624	1639	1465	1679	1746	1661	2832	7351	40232	63993
II. ARAEDEADYYCQSAVSSHKIVFGG	75	119	108	104	94	75	83	96	201	305	392	1328	4378
III. LQAEDEADYYCQVWDSSAKFVFGG	42	51	40	44	36	43	42	40	43	63	169	725	1188
IV. ARAEDEADYYCESAVSSDSRVFGG	39	37	30	37	39	20	26	38	26	58	118	420	1059
V. ALAEDEADYYCESAVSADTITFGG	20	29	20	21	24	20	26	20	28	47	222	468	1040
VI. ALAEDEADYYCQSAVSSHKIVFGG	15	15	11	9	8	11	17	9	17	33	39	169	487
VII. LQAEDEADYYCESAVSADTITFGG	3	11	12	9	10	7	12	7	13	14	130	178	331

(legend on next page)

that the scFv library is able to yield antibodies with reasonable CDR3 diversity (Figure S7G).

A different miniaturization approach was used to acquire monoclonal antibodies to canine PD-1 (Figure 5). In this protocol, the antigen levels remained low for each round of screening (200 ng), and high pH was used for elution of scFv-phage bound to the antigen, but we changed the solid phase from the immunoplate well to an antigen capture on-beads. Using his-tagged PD-1 or Fc-tagged canine PD-1, we alternately incubated these two proteins (bound to nickel beads or Protein G beads, respectively) with the scFv library through four rounds to minimize the enrichment of scFv toward the tags. The scFv-bound material was eluted from the beads with triethylamine. As the scFv was selected on a bead, there was no guarantee that scFv-phage thus acquired would bind to antigenic material adsorbed onto immunoplate wells. Nevertheless, eight bioactive monoclonals were analyzed by ELISA and shown to be PD-1 specific (Figure 5A). The scFv-phage that bound canine PD-1 did not bind human PD-1 (data not shown). The active phages had the same sequence (data not shown). It is possible that different scFv sequences would be acquired if we deconvoluted round three. The sequence of one of the scFvs, scFv-29, is shown in Figures 5B and is compared with other anti-canine PD-1 scFvs[36] (Figure 5C). We next tested whether scFv-29 could be converted into an active IgG.

A limitation of using scFv or FAB libraries selected in bacterial systems is that recombinant biologics might not be active when assembled into IgG due to problems such as antibody aggregation in mammalian cells[39] and/or post-translational modification of CDR domains that would inactivate the antibody. For this reason, yeast and mammalian antibody display have been developed to overcome the limitations of IgG production using bacterially selected biologics.[40] For example, a report[36] has shown that one scFv, from bacterial phage antibody libraries, targeting canine PD-1, required new mutations in FR domains to gain expression and activity of the IgG from mammalian cells. As such, it was important to ask whether the scFv we acquired could in fact be converted into an IgG after expression in mammalian cells (using ExpiCHO).

We fused the variable domains from PD1-29 scFv to mouse constant domain (IgG2a) so that we could characterize the chimeric IgG using robust anti-mouse secondary antibody reagents. In addition, the antibody expression plasmid was created into one cistron in which the heavy and light chain genes were linked using the viral "self-cleavage" amino acid 2A peptide[37,38] (Figure 5D). This results in production of heavy and light chain proteins that can assemble and be secreted into the supernatant of ExpiCHO cells (Figure 5E). This IgG was bioactive in an ELISA against canine PD-1 protein (Figure 5F). We finally asked whether the PD1-29 IgG binds canine PD-1 protein via a linear epitope. An overlapping array of biotinylated canine PD-1 peptides were synthesized using an N-terminal biotin tag (Figure S8A) that allowed the capture of the peptides on streptavidin-coated immunoplate wells. One peptide bound the PD1-29 IgG (peptide 18), which shows divergence with the human sequence at R > S and Q > P (Figure S8D). This presumably explains why the PD1-29 scFv does not bind human PD-1 (data not shown). Mapping of the epitope on the structure of PD-1 localized its positioning on the surface of the protein, with divergent amino acids flanking the beta-sheet (Figure S8C). Consensus site analysis of the PD1-29 IgG epitope using peptide-phage display identified the sequence Kx(s/T/A/V)AF (Figure S8D and S8E), which is present in peptide 18 (KLAAF) and is shown docked in Figure 5G.

DISCUSSION

The purpose of constructing a massively diverse naive scFv library from canine species was first to provide contemporary details for the veterinary and medical community to generate their own resource from which dog-tolerated antibodies to almost any canine antigen can be acquired to facilitate research in veterinary (companion animal) medicine. This could facilitate development of spontaneous cancer models in canine animals that form immunotolerant models for some human cancers.[3] The details provided in this methodological study define a blueprint as a community resource: (1) we define the degenerate primers that can used to capture the canine B cell mRNA repertoire based on 5′RACE PCR; (2) we detail protocols on how to directionally clone and archive amplified variable heavy and light chain genes into holding libraries for use in perpetuity; and (3) we detail next-generation long-read and short-read DNA sequencing protocols that can be used to measure cloning efficiency, define IG family distribution, identify inefficiencies in amplicon ligation or cloning steps, and identify CDR3 diversity and the enriched CDR3 domain sequences upon biopanning.

A recent report using 5′RACE PCR to amplify the B cell mRNA repertoire, and annotation of the IG family distribution using next-generation DNA sequencing, has defined the baseline

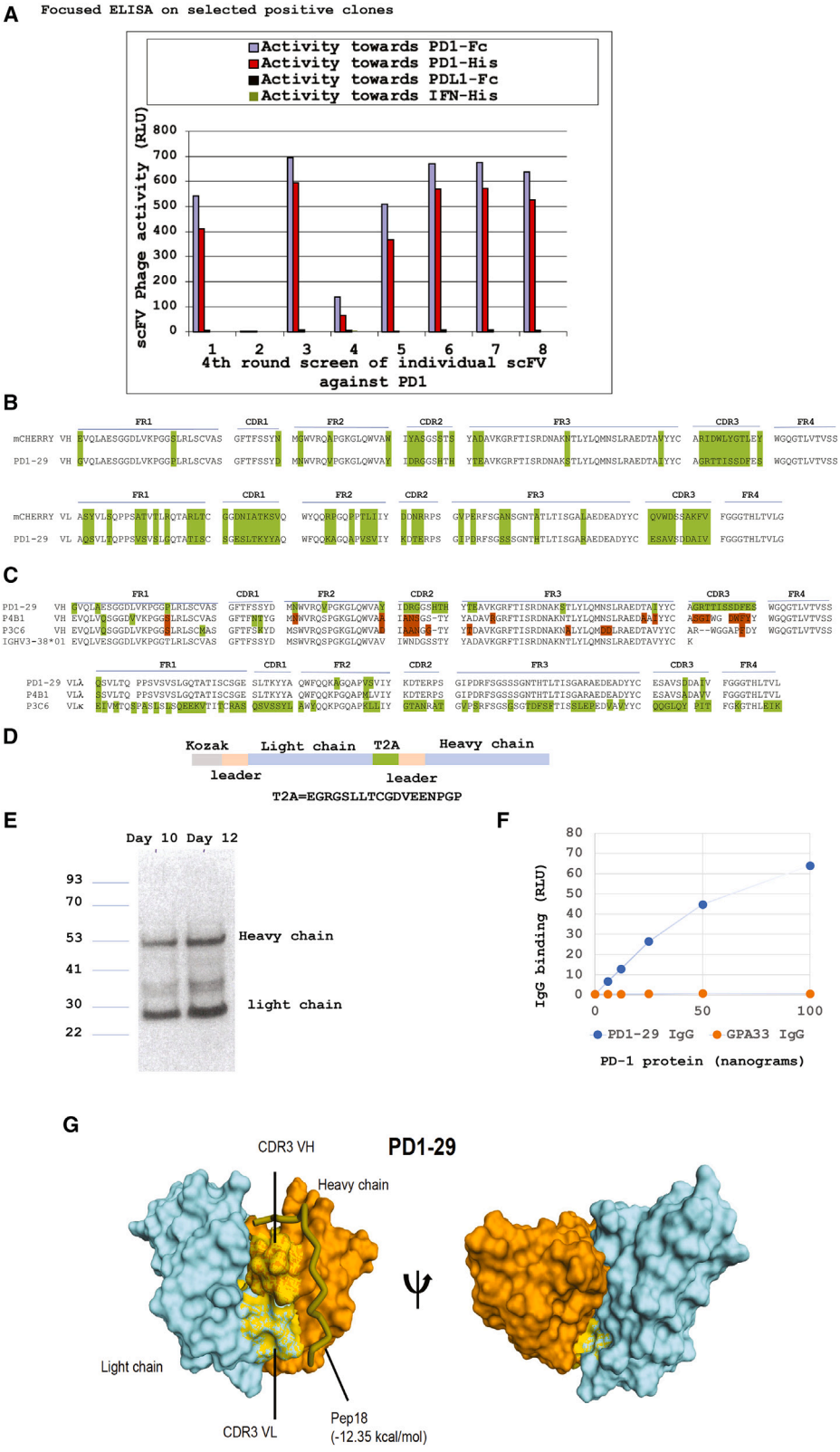
Figure 4. Next-generation short-read DNA sequencing of antibody library pools to measure enrichment of CDR3 sequences

(A) General strategy to implement next-generation DNA sequencing of the FR3-CDR3 region in a scFv gene. The forward primers hybridize to the pelB sequence or the linker region and the bar-coded reverse primers hybridize to FR4 (Table S3).

(B) Schematic: The forward primers for the heavy and light chains are in blue and the forward adapter is in orange. The four reverse "bar-coded" primers for the light and heavy chains are listed (in red) with the adaptor (in green) flanked to the bar code (in purple). The heavy chain forward primer targets the PelB leader and the light chain forward primer targets the linker.

(C–E) The PCR reaction generated a ~479-bp PCR fragment with heavy chain primers and a ~399-bp fragment using the light chain primers. The pooled amplicons (E) were subjected to paired-end next-generation short-read DNA sequencing (Otogenetics, USA) that created 100 bp reads.

(F–H) The data are plotted as the number of FR3-CDR3 unique and total reads from next-generation short-read DNA sequencing for each individual bar-coded amplicons. The top seven most abundant (G) heavy and (H) light chain sequences that were isolated in round 6 were sorted. The bar code is also listed as a function of the round of screening or pool (naive parental library, AAA; R1 EGFP, AAT; etc). The Clone I FR3-CDR3 sequence in the (G) heavy and (H) light chain screen is identical to the mCHERRY clone selected in Figure S6D. There is an increase in this enriched clone by round 3 (bracketed by the gray box), relative to the parental library and the remaining pools.



(legend on next page)

diversity in variable domain sequences in various canine breeds.[35] When we compare the heavy chain variable domain subfamilies derived from 5′RACE PCR[35] with our long-read DNA sequencing data from cloned IG genes, there is a general overlap (Figure 3). Interestingly, Cullen et al.[35] report that there was a bimodal distribution between two IGHV alleles, with IGHV3-38 being negatively correlated with IGHV4-1. The "lower" representation of the IGHV4-1 allele in our scFv library (Figure 3B), relative to the high coverage of the IGHV3-38 allele (Figure 3B), suggests that IGHV4-1 under-representation in our cloned library might be a function of this physiological bimodal exclusion. A similar analysis of the cloned IGKV and IGLV light chain subfamily distribution, in relation to the distribution identified using 5′RACE PCR,[35] is reviewed in Figures 3C and 3D. There is a general concordance between the 5′RACE PCR derived data[35] and our cloned alleles, with IGKV2 being dominant for kappa chains and IGLV1 being dominant for lambda chains.

A question can be raised about how the diversity we define in the library using next-generation short-read and long-read DNA sequencing technologies translates into acquisition of different, unique scFv sequences to other antigens. In addition to the targets in this report (EGFP, mCHERRY, canine EGFR peptides, and canine PD1), we have also provided aliquots of the scFv library to additional users for testing, and the output scFv sequence data from those screens are available online in the form of theses, including the ubiquitin ligase CHIP[41] and the oncoprotein AGR2.[42] Using either of these two target antigens, active phage pools emerged by round 2 using the high pH triethylamine elution. The scFv derived from these two screens have now been converted into IgG scaffolds and they remain active; there were eight unique scFv antibody sequences isolated toward CHIP[41] and five unique antibody sequences targeting AGR2[42] (manuscripts in preparation). Thus, in our experience, the number of unique antibody sequences depends on the target antigen, the number of monoclonals sequenced from pools, the pools from which the antibodies are derived, washing time frames,[43] and the method of solid-phase screening. In addition, we also have isolated scFv toward other cancer-relevant proteins including TIM3, EGFR protein, SARS-COV-2 antigens, GPA33, UPF1, EPCAM, IFITM1, canine CD20 (data not shown; manuscripts in preparation), and other proprietary targets under current investigation. Because the primers (Figures 1E–1H) also capture the dominant IG alleles (Figure 3) as that defined by 5′RACE PCR,[35] we think that future users would likely be able to capture representative IG genes to create bespoke scFv libraries.

Another compelling rationale for developing naive canine antibody libraries tools is that canine is perhaps the only species (apart from humans) where millions of cancer patients can be treated through veterinary medicine. Why might this be useful? A recent report has highlighted the existence of human cancer auto-antibodies toward an antigen within an ovarian cancer subtype.[44] This is pointing to the possibility that canine cancer patients might have their own auto-cancer antigen repertoires to add to their individual or breed-specific B cell transcriptomes. By cloning and pooling different "naive" variable domain repertoires from dogs with different cancers, phage antibody diversity might cover a different spectra of antigen-binders. By contrast, other animals that are used to develop phage antibody libraries, such as camelid, rat, mouse, sheep, or shark do not offer the equivalent possibility of capturing mRNA from diseased states. Thus, key future studies could involve defining how disease-specific B cell repertoires can be captured to enhance the development of synthetic antibodies for use to improve outcomes in canine patients and to provide more physiological spontaneous models for human cancers.

Limitations of the study

The main limitation of the study relates to the diversity of the naive antibody genes cloned, which depends in part upon the canine breed and/or disease state of the animals that can impact on the diversity of the naive B cell receptor repertoire. However, this can be addressed by choosing appropriate donor B cell material at the outset or pooling RNA from multiple animals. A second limitation of the study is the relatively low efficiency of shotgun cloning of pooled heavy and light chain amplicons into single fusion proteins. This can be improved by carrying out ∼>200 separate combinatorial ligations. In addition, the efficiency of cloning can also be quantified using next-generation short or long-read DNA sequencing whereby inefficiencies in any one amplicon cloning can be optimized separately.

Figure 5. Miniaturization of the scFv library screen on a physiologically relevant antigen, canine PD-1

(A) After four rounds of biopanning against alternating His-tagged and Fc-tagged canine PD-1 protein on nickel or protein G beads, respectively, colonies were chosen from round 4 phage plating and individual scFv-phage were assayed for binding to PD-1. An ELISA highlighted the activity of eight scFv-phage screened against, as indicated, Fc-tagged canine PD-1, His-tagged canine PD-1, Fc-tagged canine PD-L1, and his-tagged human IFNA.

(B) The sequence of mCHERRY scFv is compared with the scFv-PD-1-clone-29.

(C) The sequence of the scFv-PD-1-clone 29 sequence was compared with two anti-PD-1 scFvs reported from scFV canine phage display libraries generated using the genomic-derived FR1 and FR4 cloning primers.[36] Although the heavy chain CDR2 and CDR3 sequences are distinct between the two studies, the lambda light chain CDR3 sequences are very similar (compare clone PD1-29 with P4B1). IGHV3-38*-01 might form the genomic origin of the P4B1 scFV, and to an equivalent extent, clone PD1-29.

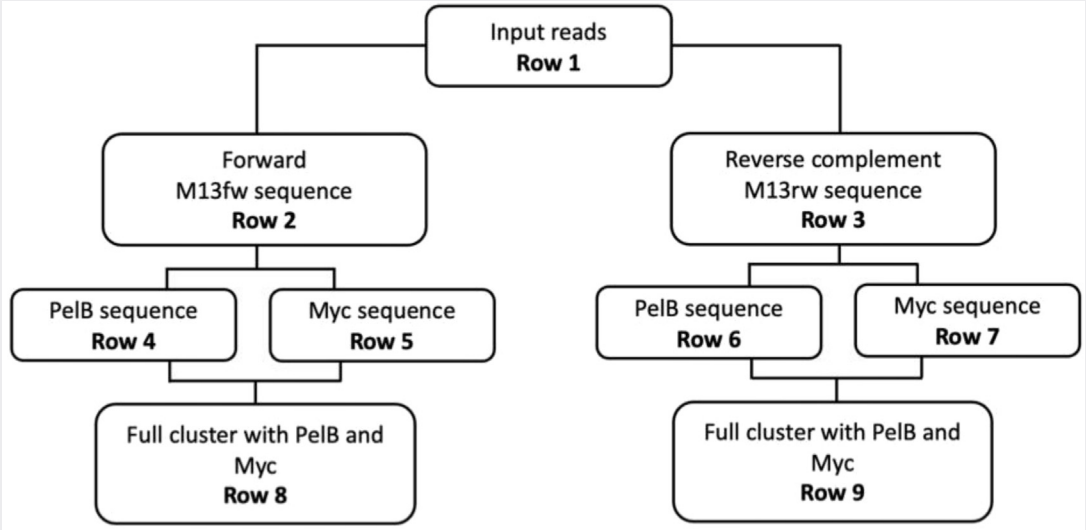
(D) Conversion of the anti-PD-1 scFv to an IgG expression plasmid using a single cistron containing the self-cleavage sequence from the 2A virus.[37,38]

(E) After transfection of the plasmid into Expi-CHO cells, the media was immunoblotted using day 10 and 12 media to detect heavy and light chains.

(F) An ELISA using PD1-29 IgG from (E) was used to quantify IgG activity. The data are depicted as IgG binding as a function of changes in PD-1 protein levels. An IgG made from the same scFv library toward GPA33 (data not shown) was used as the negative control.

(G) Epitope mapping of the PD1-29 IgG to "peptide-18" is summarized in Figure S8. Molecular docking of "peptide-18" identified potential CDR-antigen amino acid contacts. CHARMM27 force field homology modeling and molecular docking was performed using the Molecular Operating Environment (MOE; Chemical Computing Group Inc., Montreal, QC, Canada) package.

Table 1. Long-read DNA sequencing annotation of holding chain and scFv libraries



ID		HCH	LCH	scFv
1	**Number of reads**	2 087 704	1 460 826	1 677094
2	**Number of reads with m13_fw**	1 019 705	515 965	387 390
3	**Number of reads with m13_rc**	998 234	518 330	394 635
4	**Number of reads with PelB_fw (M13fw)**	1 004 480	134 130	234634
5	**Number of reads with Myc_fw (M13fw)**	492 263	121 407	64 030
6	**Number of reads with PelB_rc (M13rc)**	981 979	133 574	228 120
7	**Number of reads with Myc_fw (M13rc)**	485446	121 516	63 104
8	**Number of reads with CLUSTER_PelB-Myc_fw (M13fw)**	485 770	116 665	62 072
9	**Number of reads with CLUSTER_PelB-Myc_rc (M13rc)**	479 006	116 685	61 078

The flow chart explains the process by which long-read DNA samples were analyzed including converting input reads to clustering reads that have the full sequence including the PelB and Myc Tag. HCH = heavy chain holding library; LCH = Light chain holding library; and scFV = full library. Long-read DNA sequencing data read number from the scFV library, from HCH(VH-11), and LCH pooled libraries that fulfill the selected parameters. M13fw represents the forward primer and M13rc represents the reverse primer.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by Małgorzata Lisowska, gosia.nus@gmail.com.

Materials availability

The scFv-phage antibody library is available upon request (gosia.nus@gmail.com).

Data and code availability

- Next-generation DNA sequencing files are available on from NCBI; SubmissionID: SUB14919825; BioProject ID: PRJNA1196821; Title, The Development of a Canine Single-Chain Phage Antibody Library to Isolate Recombinant Antibodies for use in Translational Cancer Research. Files are saved in sequence read archive under accession numbers SRX27039641, SRX27039642, SRX27039643, and SRX27039644.
- Custom scripts for analyzing next-generation sequencing data have been deposited to Zenodo (source codes for analysis CDR and FR in canine single-chain antibody library https://doi.org/10.5281/zenodo.14512819, which includes six files all highlighted in the STAR Methods section). Any further information required to re-analyze the data reported in this paper is available from the lead contact upon request.
- Any further information required to re-analyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This study was supported by BBSRC grant (BB/J00751X/1) to develop canine phage library platforms; PetPlan UK; Czech Science Foundation 22-02940S; MH CZ-DRO (MMCI, 00209805); Xjenza Malta Research Excellence Programme (REP-2023-013); FENG International Research Agenda's Program of the Foundation for Polish Science (MAB/2017/03); European Funds for Smart Economy (FENG), no. FENG.02.01-IP.05-0031/23; the National Science Centre (Narodowe Centrum Nauki, Krakow, Poland; grant no. 2020/39/B/NZ7/02677); Poland's high-performance computing infrastructure PLGrid (HPC Centers), which provided computer facilities and support within computational grant no. PLG/2019/012567; and the International Centre for Cancer Vaccine

Table 2. Long-read DNA sequencing annotation of the FR and CDR domain length and diversity from one scFv library

Length range of regions: FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4 in light lambda, light kappa and heavy chain (scFv library)

scFv	Fragment	Length	AA position in chain
Light lambda chain	FR1	25–26	163–188
	CDR1	4–7	189–197
	FR2	16–17	198–214
	CDR2	2–3	215–217
	FR3	36	218–253
	CDR3	8–12	254–265
	FR4	10	266–275
Light kappa chain	FR1	26	161–188
	CDR1	4–7	189–197
	FR2	16–17	198–214
	CDR2	3	215–217
	FR3	36	218–253
	CDR3	8–12	254–265
	FR4	9	266–275
Heavy chain	FR1	25	23–47
	CDR1	8	48–55
	FR2	17	56–72
	CDR2	8–9	73–81
	FR3	36–37	82–118
	CDR3	5–18	119–136
	FR4	11	137–147

ScFv library regions of different length in number of sequences

scFv Library fragment	Fragment	Length	Number of SEQs
Light lambda chain	FR1	25	2144
		26	18745
	CDR1	4	1472
		5	2404
		6	11145
		7	5781
	FR2	16	2044
		17	18841
	CDR2	2	1393
		3	19497
	CDR3	8	452
		9	1187
		10	4675
		11	13471
		12	1075
Light kappa chain	CDR1	4	24
		5	157
		6	439
		7	92
	FR2	16	134
		17	582
	CDR3	8	22
		9	47
		10	161
		11	433
		12	43
Heavy chain	CDR2	8	6552
		9	14355
	FR3	36	80
		37	21054
	CDR3	5	200
		6	647
		7	3425
		8	1532
		9	1817
		10	2423
		11	2804
		12	3093
		13	2512
		14	1591
		15	734
		16	268
		17	80
		18	3

Long-read DNA sequencing data from Table 1 was annotated into length range and number of sequences of a specific length and provides information on the approximate length and distribution of CDRs as compared with the data acquired using 5′RACE PCR.[35] In this latter study,[35] the peak distribution of heavy chain CDR3 length was ~13–15 amino acids. By contrast, in our study, the distribution of heavy chain CDR3 lengths were broader from seven to 14 amino acids. The most common CDR3 length for the lambda chain using 5′RACE PCR was 11 amino acids,[35] which is similar to what we see in our scFv library. The most common CDR3 length for the kappa chain using 5′RACE PCR was nine amino acids,[35] whereas we see the most common kappa chain length being 11 amino acids.

Science (ICCVS) project, which is carried out within the International Research Agendas programme of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund.

AUTHOR CONTRIBUTIONS

Experimental analysis, setting up experimental design, and writing: M.L., E.G.W., F.Z.-K., K.C., E.M., M.A.M., A.G.U., R.K., V.H., J.B., D.G.S., M.G., U.K., P.M., and A.K. Funding, project implementation, and writing: B.V., K.L.B., R.F., D.J.A., M.P., T.R.H.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - RNA isolation
  - cDNA synthesis from total RNA
  - 5′RACE amplification of framework 1 and 4 sequences using splenic RNA
  - Race primers were
  - Development of degenerate PCR primers to amplify heavy and light chains from the canine splenic RNA
  - Cloning of the VH, Vλ, and Vκ regions by producing pools from each of the primer sets
  - Construction of heavy chain and light chain holding libraries
  - Phage display vector preparation
  - Vector & insert digestion
  - Vector and insert ligation
  - Calculations into the heavy chain holding and light chain holding library ligation efficiency
  - Cloning of heavy and light chains into a single chain library
  - Phenol: Chloroform extraction and ethanol precipitation of vector & insert
  - The extracted DNA is concentrated by ethanol precipitation
  - Transformation of electro competent E. coli lucigen TG1 cells
  - EGFP and mCHERRY biopanning procedures
  - Canine EGFR peptide and canine PD-1 protein miniaturization procedures using immunoplate wells or on-bead biopanning
  - Methodology for measuring CDR3 sequence diversity
  - Preparation of DNA from scFv libraries for long-read DNA sequencing
  - Methodology for the analysis of IG family representation
  - Production of IgG in CHO cells
  - ELISA and pepscan to localize the epitope of the anti-EGFR scFv and the anti-PD-1 IgG
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.crmeth.2025.101008.

Received: December 1, 2023
Revised: July 10, 2024
Accepted: February 25, 2025
Published: March 24, 2025

REFERENCES

1. Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674. https://doi.org/10.1016/j.cell.2011.02.013.
2. van Helden, P.D., van Helden, L.S., and Hoal, E.G. (2013). One world, one health. Humans, animals and the environment are inextricably linked–a fact that needs to be remembered and exploited in our modern approach to health. EMBO Rep. *14*, 497–501. https://doi.org/10.1038/embor.2013.61.
3. LeBlanc, A.K., and Mazcko, C.N. (2020). Improving human cancer therapy through the evaluation of pet dogs. Nat. Rev. Cancer *20*, 727–742. https://doi.org/10.1038/s41568-020-0297-3.
4. Pratelli, A., Buonavoglia, A., Lanave, G., Tempesta, M., Camero, M., Martella, V., and Decaro, N. (2021). One world, one health, one virology of the mysterious labyrinth of coronaviruses: the canine coronavirus affair. Lancet Microbe *2*, e646–e647. https://doi.org/10.1016/S2666-5247(21)00282-2.
5. Uhl, E.W., and Warner, N.J. (2015). Mouse Models as Predictors of Human Responses: Evolutionary Medicine. Curr Pathobiol Rep *3*, 219–223. https://doi.org/10.1007/s40139-015-0086-y.
6. Masopust, D., Sivula, C.P., and Jameson, S.C. (2017). Of Mice, Dirty Mice, and Men: Using Mice To Understand Human Immunology. J. Immunol. *199*, 383–388. https://doi.org/10.4049/jimmunol.1700453.
7. Khanna, C., Lindblad-Toh, K., Vail, D., London, C., Bergman, P., Barber, L., Breen, M., Kitchell, B., McNeil, E., Modiano, J.F., et al. (2006). The dog as a cancer model. Nat. Biotechnol. *24*, 1065–1066. https://doi.org/10.1038/nbt0906-1065b.
8. Paoloni, M., and Khanna, C. (2008). Translation of new cancer treatments from pet dogs to humans. Nat. Rev. Cancer *8*, 147–156. https://doi.org/10.1038/nrc2273.
9. Amin, S.B., Anderson, K.J., Boudreau, C.E., Martinez-Ledesma, E., Kocakavuk, E., Johnson, K.C., Barthel, F.P., Varn, F.S., Kassab, C., Ling, X., et al. (2020). Comparative Molecular Life History of Spontaneous Canine and Human Gliomas. Cancer Cell *37*, 243–257. https://doi.org/10.1016/j.ccell.2020.01.004.
10. MacDiarmid, J.A., Langova, V., Bailey, D., Pattison, S.T., Pattison, S.L., Christensen, N., Armstrong, L.R., Brahmbhatt, V.N., Smolarczyk, K., Harrison, M.T., et al. (2016). Targeted Doxorubicin Delivery to Brain Tumors via Minicells: Proof of Principle Using Dogs with Spontaneously Occurring Tumors as a Model. PLoS One *11*, e0151832. https://doi.org/10.1371/journal.pone.0151832.
11. Wang, G., Wu, M., Durham, A.C., Radaelli, E., Mason, N.J., Xu, X., and Roth, D.B. (2020). Molecular subtypes in canine hemangiosarcoma reveal similarities with human angiosarcoma. PLoS One *15*, e0229728. https://doi.org/10.1371/journal.pone.0229728.
12. Gellrich, F.F., Schmitz, M., Beissert, S., and Meier, F. (2020). Anti-PD-1 and Novel Combinations in the Treatment of Melanoma-An Update. J. Clin. Med. *9*, 9010223. https://doi.org/10.3390/jcm9010223.
13. Scott, A.M., Wolchok, J.D., and Old, L.J. (2012). Antibody therapy of cancer. Nat. Rev. Cancer *12*, 278–287. https://doi.org/10.1038/nrc3236.
14. Kulkarni, S.S., and Falzarano, D. (2021). Unique aspects of adaptive immunity in camelids and their applications. Mol. Immunol. *134*, 102–108. https://doi.org/10.1016/j.molimm.2021.03.001.
15. Matz, H., and Dooley, H. (2019). Shark IgNAR-derived binding domains as potential diagnostic and therapeutic agents. Dev. Comp. Immunol. *90*, 100–107. https://doi.org/10.1016/j.dci.2018.09.007.
16. Hawkins, A., Joyce, C., Brady, K., Hold, A., Smith, A., Knight, M., Howard, C., van den Elsen, J., Lawson, A.D.G., and Macpherson, A. (2022). The proximity of the N- and C- termini of bovine knob domains enable engineering of target specificity into polypeptide chains. mAbs *14*, 2076295. https://doi.org/10.1080/19420862.2022.2076295.
17. McMahon, C., Baier, A.S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J.X., Erlandson, S.C., Hilger, D., Rasmussen, S.G.F., Ring, A.M., et al. (2018). Yeast surface display platform for rapid discovery of conformationally selective nanobodies. Nat. Struct. Mol. Biol. *25*, 289–296. https://doi.org/10.1038/s41594-018-0028-6.
18. Marable, J., Ruiz, D., Jaiswal, A.K., Bhattacharya, R., Pantazes, R., Agarwal, P., Suryawanshi, A.S., Bedi, D., Mishra, A., Smith, B.F., and Sandey, M. (2021). Nanobody-based CTLA4 inhibitors for immune checkpoint blockade therapy of canine cancer patients. Sci. Rep. *11*, 20763. https://doi.org/10.1038/s41598-021-00325-3.
19. Li, D., English, H., Hong, J., Liang, T., Merlino, G., Day, C.P., and Ho, M. (2022). A novel PD-L1-targeted shark V(NAR) single-domain-based CAR-T cell strategy for treating breast cancer and liver cancer. Mol Ther Oncolytics *24*, 849–863. https://doi.org/10.1016/j.omto.2022.02.015.
20. Gearing, D.P., Virtue, E.R., Gearing, R.P., and Drew, A.C. (2013). A fully caninised anti-NGF monoclonal antibody for pain relief in dogs. BMC Vet. Res. *9*, 226. https://doi.org/10.1186/1746-6148-9-226.

21. Enomoto, M., Mantyh, P.W., Murrell, J., Innes, J.F., and Lascelles, B.D.X. (2019). Anti-nerve growth factor monoclonal antibodies for the control of pain in dogs and cats. Vet. Rec. *184*, 23. https://doi.org/10.1136/vr.104590.

22. Lascelles, B.D., Knazovicky, D., Case, B., Freire, M., Innes, J.F., Drew, A.C., and Gearing, D.P. (2015). A canine-specific anti-nerve growth factor antibody alleviates pain and improves mobility and function in dogs with degenerative joint disease-associated pain. BMC Vet. Res. *11*, 101. https://doi.org/10.1186/s12917-015-0413-x.

23. Igase, M., Inanaga, S., Tani, K., Nakaichi, M., Sakai, Y., Sakurai, M., Kato, M., Tsukui, T., and Mizuno, T. (2022). Long-term survival of dogs with stage 4 oral malignant melanoma treated with anti-canine PD-1 therapeutic antibody: A follow-up case report. Vet. Comp. Oncol. *20*, 901–905. https://doi.org/10.1111/vco.12829.

24. Igase, M., Nemoto, Y., Itamoto, K., Tani, K., Nakaichi, M., Sakurai, M., Sakai, Y., Noguchi, S., Kato, M., Tsukui, T., and Mizuno, T. (2020). A pilot clinical study of the therapeutic antibody against canine PD-1 for advanced spontaneous cancers in dogs. Sci. Rep. *10*, 18311. https://doi.org/10.1038/s41598-020-75533-4.

25. Li, G., Ohishi, T., Kaneko, M.K., Takei, J., Mizuno, T., Kawada, M., Saito, M., Suzuki, H., and Kato, Y. (2021). Defucosylated Mouse-Dog Chimeric Anti-EGFR Antibody Exerts Antitumor Activities in Mouse Xenograft Models of Canine Tumors. Cells *10*, 10123599. https://doi.org/10.3390/cells10123599.

26. Singer, J., Fazekas, J., Wang, W., Weichselbaumer, M., Matz, M., Mader, A., Steinfellner, W., Meitz, S., Mechtcheriakova, D., Sobanov, Y., et al. (2014). Generation of a canine anti-EGFR (ErbB-1) antibody for passive immunotherapy in dog cancer patients. Mol. Cancer Therapeut. *13*, 1777–1790. https://doi.org/10.1158/1535-7163.MCT-13-0288.

27. Braganza, A., Wallace, K., Pell, L., Parrish, C.R., Siegel, D.L., and Mason, N.J. (2011). Generation and validation of canine single chain variable fragment phage display libraries. Vet. Immunol. Immunopathol. *139*, 27–40. https://doi.org/10.1016/j.vetimm.2010.07.026.

28. Cong, C., ZhongLing, J., RongFeng, C., WenRu, T., and HuaTao, L. (2017). Construction and panning of canine phage scFv library for dog erythrocyte antigen. Agricultural Biotechnology *25*, 901–910.

29. Ferrara, F., Erasmus, M.F., D'Angelo, S., Leal-Lopes, C., Teixeira, A.A., Choudhary, A., Honnen, W., Calianese, D., Huang, D., Peng, L., et al. (2022). A pandemic-enabled comparison of discovery platforms demonstrates a naive antibody library can match the best immune-sourced antibodies. Nat. Commun. *13*, 462. https://doi.org/10.1038/s41467-021-27799-z.

30. Freeman, J.D., Warren, R.L., Webb, J.R., Nelson, B.H., and Holt, R.A. (2009). Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. Genome Res. *19*, 1817–1824. https://doi.org/10.1101/gr.092924.109.

31. Hwang, M.H., Darzentas, N., Bienzle, D., Moore, P.F., Morrison, J., and Keller, S.M. (2018). Characterization of the canine immunoglobulin heavy chain repertoire by next generation sequencing. Vet. Immunol. Immunopathol. *202*, 181–190. https://doi.org/10.1016/j.vetimm.2018.07.002.

32. Khanna, C., London, C., Vail, D., Mazcko, C., and Hirschfeld, S. (2009). Guiding the optimal translation of new cancer treatments from canine to human cancer patients. Clin. Cancer Res. *15*, 5671–5677. https://doi.org/10.1158/1078-0432.CCR-09-0719.

33. Keller, B.M., Maier, J., Secker, K.A., Egetemaier, S.M., Parfyonova, Y., Rothbauer, U., and Traenkle, B. (2018). Chromobodies to Quantify Changes of Endogenous Protein Concentration in Living Cells. Mol. Cell. Proteomics *17*, 2518–2533. https://doi.org/10.1074/mcp.TIR118.000914.

34. Gomes, M., Fleck, A., Degaugue, A., Gourmelon, F., Leger, C., Aumont-Nicaise, M., Mesneau, A., Jean-Jacques, H., Hassaine, G., Urvoas, A., et al. (2023). Design of an artificial phage-display library based on a new scaffold improved for average stability of the randomized proteins. Sci. Rep. *13*, 1339. https://doi.org/10.1038/s41598-023-27710-4.

35. Cullen, J.N., Martin, J., Vilella, A.J., Treeful, A., Sargan, D., Bradley, A., and Friedenberg, S.G. (2022). Development and application of a next-generation sequencing protocol and bioinformatics pipeline for the comprehensive analysis of the canine immunoglobulin repertoire. PLoS One *17*, e0270710. https://doi.org/10.1371/journal.pone.0270710.

36. Yoshimoto, S., Chester, N., Xiong, A., Radaelli, E., Wang, H., Brillantes, M., Gulendran, G., Glassman, P., Siegel, D.L., and Mason, N.J. (2023). Development and pharmacokinetic assessment of a fully canine anti-PD-1 monoclonal antibody for comparative translational research in dogs with spontaneous tumors. mAbs *15*, 2287250. https://doi.org/10.1080/19420862.2023.2287250.

37. Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Vaseghi, H.R., Qian, L., and Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci. Rep. *7*, 2193. https://doi.org/10.1038/s41598-017-02460-2.

38. Lin, Y., Hung, C.Y., Bhattacharya, C., Nichols, S., Rahimuddin, H., Kittur, F.S., Leung, T., and Xie, J. (2018). An Effective Way of Producing Fully Assembled Antibody in Transgenic Tobacco Plants by Linking Heavy and Light Chains via a Self-Cleaving 2A Peptide. Front. Plant Sci. *9*, 1379. https://doi.org/10.3389/fpls.2018.01379.

39. Ebo, J.S., Saunders, J.C., Devine, P.W.A., Gordon, A.M., Warwick, A.S., Schiffrin, B., Chin, S.E., England, E., Button, J.D., Lloyd, C., et al. (2020). An in vivo platform to select and evolve aggregation-resistant proteins. Nat. Commun. *11*, 1816. https://doi.org/10.1038/s41467-020-15667-1.

40. Uchanski, T., Zogg, T., Yin, J., Yuan, D., Wohlkonig, A., Fischer, B., Rosenbaum, D.M., Kobilka, B.K., Pardon, E., and Steyaert, J. (2019). An improved yeast surface display platform for the screening of nanobody immune libraries. Sci. Rep. *9*, 382. https://doi.org/10.1038/s41598-018-37212-3.

41. Santos, M. (2015). Developing Synthetic Tools to Image and Modulate the Activity of Carboxyl terminus of Hsc70-Interacting Protein (CHIP). The experimental work described in the thesis was carried out at the Edinburgh Cancer Research Centre (ECRC) of the Institute of Genetics and Molecular Medicine at the University of Edinburgh in the framework of the ERASMUS+ programme. (https://core.ac.uk/download/pdf/143412857.pdf

42. Mohtar, A. (2018). Novel role of an ER-resident chaperone pathway in cancer signalling. Edinburgh Medical School thesis and dissertation collection. https://era.ed.ac.uk/handle/1842/28765?show=full

43. Lisowska, M., Lickiss, F., Gil-Mir, M., Huart, A.S., Trybala, Z., Way, L., Hernychova, L., Krejci, A., Muller, P., Krejcir, R., et al. (2022). Next-generation sequencing of a combinatorial peptide phage library screened against ubiquitin identifies peptide aptamers that can inhibit the in vitro ubiquitin transfer cascade. Front. Microbiol. *13*, 875556. https://doi.org/10.3389/fmicb.2022.875556.

44. Mazor, R.D., Nathan, N., Gilboa, A., Stoler-Barak, L., Moss, L., Solomonov, I., Hanuna, A., Divinsky, Y., Shmueli, M.D., Hezroni, H., et al. (2022). Tumor-reactive antibodies evolve from non-binding and autoreactive precursors. Cell *185*, 1208–1222. https://doi.org/10.1016/j.cell.2022.02.012.

45. Kavanagh, D.M., Smyth, A.M., Martin, K.J., Dun, A., Brown, E.R., Gordon, S., Smillie, K.J., Chamberlain, L.H., Wilson, R.S., Yang, L., et al. (2014). A molecular toggle after exocytosis sequesters the presynaptic syntaxin1a molecules involved in prior vesicle fusion. Nat. Commun. *5*, 5774. https://doi.org/10.1038/ncomms6774.

46. Roben, P.W., Salem, A.N., and Silverman, G.J. (1995). VH3 family antibodies bind domain D of staphylococcal protein A. J. Immunol. *154*, 6437–6445.

47. Jansson, B., Uhlen, M., and Nygren, P.A. (1998). All individual domains of staphylococcal protein A show Fab binding. FEMS Immunol. Med. Microbiol. *20*, 69–78. https://doi.org/10.1111/j.1574-695X.1998.tb01112.x.

48. Charif, D., Thioulouse, J., Lobry, J.R., and Perriere, G. (2005). Online synonymous codon usage analyses with the ade4 and seqinR packages. Bioinformatics *21*, 545–547. https://doi.org/10.1093/bioinformatics/bti037.

49. Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics *33*, 3645–3647. https://doi.org/10.1093/bioinformatics/btx469.

50. Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

51. Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E.V., and Chudakov, D.M. (2015). MiXCR: software for comprehensive adaptive immunity profiling. Nat. Methods *12*, 380–381. https://doi.org/10.1038/nmeth.3364.

52. Manso, T., Folch, G., Giudicelli, V., Jabado-Michaloud, J., Kushwaha, A., Nguefack Ngoune, V., Georga, M., Papadaki, A., Debbagh, C., Pegorier, P., et al. (2022). IMGT(R) databases, related tools and web resources through three main axes of research and development. Nucleic Acids Res. *50*, D1262–D1272. https://doi.org/10.1093/nar/gkab1136.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat Anti-Dog IgG H&L (HRP)	Abcam	ab112852; RRID:AB_2927648
HRP-conjugated rabbit anti-mouse secondary antibody	DAKO/AGILENT	P0260; RRID:AB_2636929
Anti-M13 antibody conjugated to HRP	Santa Cruz	sc-53004HRP; RRID:AB_673750
Bacterial and virus strains		
XL1-Blue bacterial cells	Agilent	200249
DH5a bacterial cells	Thermo Fisher	EC0112
Ph.D.™-12 Phage Display Peptide Library	New England Biolabs	E8110
TG1 bacterial cells	Lucigen	60502–2
Chemicals, peptides, and recombinant proteins		
MPAGE® 4X LDS Sample Buffer	Merck	MPSB-10ML
All restriction enzymes	New England Biolabs	
Trypsin	Sigma	T1426-50MG
PEG8000	VWR	0159-500G
Streptavidin	Cambridge Bioscience	ANA60659
All biotinylated synthetic peptides	Mimotopes, Australia	N/A
PD-1 protein (his, Fc tagged)	Sino Biologics	70109-D08H, 70109-D02H
Glycogen	Roche	109013930010
Critical commercial assays		
RNAeasy mini-kit	Qiagen	74704
RNaseZap solution	Sigma	R2020
Omniscript reverse transcription kit	Qiagen	205113
SMARTer RACE amplification kit	Clontech	NC0734706
HF Phusion mix (NEB Cat No: 531L)	New England Biolabs	M0531L
Qiagen PCR purification kit	Qiagen	28104
Deposited data		
Next generation short and long read data	This paper	SUB14919825; BioProject ID: PRJNA1196821, Files are saved in sequence read archive under accession numbers SRX27039641, SRX27039642, SRX27039643 and SRX27039644
ExpiCHO Expression System Kit	Thermo Scientific	A29133
Oligonucleotides		
All primers are listed in figures (Figure 1), tables (Tables S3A and S3B), and STAR Methods.	Sigma	N/A
Recombinant DNA		
pGEM-T easy vector	Promega	A1360
Software and algorithms		
Custom analysis scripts	This paper	https://doi.org/10.5281/zenodo.14512819
SNAPGENE		N/A
Other		
Electroporation cuvette	BIORAD	1652809
MicroPulser Electroporator	BIORAD	1652100
Immunotubes	Greiner	115071
Protein A Resin - Amintra®	Abcam	ab270308

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
96-well White Flat Bottom Polystyrene High Bind Microplate	Corning	3922
Deep 96 well plates MASTERBLOCK® 1 ML, PP, U-BOTTOM	Greiner	780261
Nickel NTA Agarose Beads	Qiagen	30210
Protein G Sepharose beads	Cytiva	17061801
HRP conjugated Protein A	Thermo Fisher	101023

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Antibody production from plasmid DNA was carried out using the ExpiCHO Mammalian Expression System (Thermo Fisher) according to the manufacturer's instructions. M13K07 helper phage were obtained from New England Biolabs. A Ph.D.-12 Phage Display Peptide Library for use in epitope mapping was from New England Biolabs. Electrocompetent TG1 bacteria were obtained from Lucigen. ER2738 bacteria were obtained from New England Biolabs. XL1 Blue bacterial cells were from Agilent. DH5a bacterial cells were from Thermo Fisher. Total RNA was extracted from frozen, archived dog spleen tissue, obtained from 5 anonymized animals that had undergone treatment in the University of Edinburgh Veterinary hospital and had been euthanized for clinical reasons with owner consent according to standard agreements in place at the University of Edinburgh School of Veterinary Studies. Under standard consent forms, owners agree that any tissues donated can be used for clinical research.

METHOD DETAILS

RNA isolation

Total RNA was extracted using the Qiagen fibrous RNAeasy mini-kit. Total splenic tissue (~30 mg) was broken up using a disposable mini homogeniser in 300 μl RLT buffer containing β-Mercaptoethanol and a small gauge needle was then used to pipette up and down the mix to further break up the tissue. RNAse free water was added (590 μL) followed by 10 μL of proteinase K and incubated at 55°C for 10 min. The samples were centrifuged at 10,000 x g for 3 min, the supernatant was transferred to a new tube, and 0.5 volumes of 100% ethanol was added and mixed. The solution was transferred to a RNeasy column and centrifuged for 15 s at 8,000 x g and the flow through was discarded. Buffer RW1 was added (350 μL), the samples were centrifuged for 15 s at 8,000 x g, and the flow through was discarded. DNase stock (10 μL) was mixed with 70 μL of buffer RDD and the solution was added directly to the membrane and incubated for 15–20 min at RT. RW1 buffer (350 μL) was added and then the samples were centrifuged at 8,000 x g for 15 s and the flow through was discarded. RPE buffer (500 μL) was added and the samples were centrifuged at 8,000 x g for 15 s and the flow through was discarded. RPE buffer (500 μL) was added and the samples were centrifuged at 8,000 x g for 2 min and the flow through was discarded. RNase free water was added (30 μL), the samples were incubated for 1 min and then centrifuged at 8,000 x g for 1 min to collect the eluted RNA in the flow through. Pipettes, gel tank, gel tray, and relevant combs were pre-treated with RNaseZap solution (Sigma Cat No R2020).

cDNA synthesis from total RNA

An Omniscript reverse transcription kit (Qiagen) was used to produce a cDNA template using the following reverse primers for CH, Cκ, and Cλ, constant domains from genomic sequences, respectively. (See Figure 1 to describe the relative location of these primers on the genomic regions): CaHc1 5′-GACCACGTTGCAGGTGAAGGTCTC-3′, Caκc1 5′-CGCTCCTTTGGAAGCTCTTGATGAG-3, and Caλc1 5′GTGCTCCCCTCGTGTGTGACCAG-3'. The cDNA was prepared using the Omniscript RT kit (Qiagen) by diluting the RNase inhibitor (Roche) to a final concentration of 10 units/ul in ice-cold 1x buffer RT and mixed. The master mix (20 μL) contained RT buffer, dNTPs, Oligo-dt primer, Omniscript RT, RNase-free water, and 2μg of RNA template and then incubated at 37°C for 60 min.

5′RACE amplification of framework 1 and 4 sequences using splenic RNA

An amplification of the 5′RACE PCR derived cDNA reaction was performed with the SMARTer RACE amplification kit (CloneTech) using reverse primers from the canine constant (CH, Cλ, and Cκ) genomic sequences listed in Figure 1 (to observe the relative location of the 5′RACE primer locations on the genomic region in the second round of amplification) to capture sequence diversity in naive somatic mutations from Frameworks 1 and 4. The Hyperladders (1 kb and 0.1 kb) were from Biorad.

Race primers were

Canine-Hc-2a 5′-TCCAGGACACAGTTACAGGCTCG, Canine-Caκc-2 5′-CAGCAAGCACACAACAGAGGCACT, Canine-Caλc-2 5′-CACTGGGGTAGAAGTCGCTGATGAG.

Development of degenerate PCR primers to amplify heavy and light chains from the canine splenic RNA

Following amplification of the variable regions using 5′RACE PCR, the products were then cloned into a pGEM-T easy vector and transformed into XL1-Blue cells for blue-white colony screening. To gain enough sequence information to facilitate primer design, approximately 60 white colonies were chosen from each canine constant region (CH, Cλ, and Cκ) for DNA sequencing (Figure S1 (for Cκ) and Figure S2 (for CH and Cλ)). Using the intact DNA sequences, combined with the genomic data available on the canine variable regions, a set of degenerate primers were designed which would theoretically amplify a majority of the dogs' naive mRNA repertoire (Figures 1E–1H). The cloned distribution of the IGHV, IGLV, and IGKV sub-families in the scFv library is summarized in Figure 3.

Cloning of the VH, Vλ, and Vκ regions by producing pools from each of the primer sets

Step 1: A gradient PCR reaction was set up to determine the optimum annealing temperature for each primer pair. Primers for VH Amplification (See Figure 1 for sequences) included the Back primers (individual primers from the heavy chain (CaVH-BACK-SfiI numbered 1–12) and Forward primers (all four of the CaJH-FOR-AscI primers pooled with each individual Reverse primer) that generated 12 total heavy chain reaction pots. Similarly, the light chains were amplified by including the Back primers (Figure 1) individually (CaVκ-BACK-MluI numbered 1–7 and CaVλ-BACK-MluI numbered 1–12) with all Forward primers (all four of the CaJκ-FOR-NotI or CaλJ-FOR-NotI primers pooled with each individual Reverse primer) that generated 12 reaction pots for the lambda light chains and 7 reaction pots for the kappa light chains.

For the gradient PCRs, the reaction mix included 2 x HF Phusion mix (NEB Cat No: 531L) – 25 μL (x 13 reactions), 25 μM heavy or light chain BACK primer (1 μL (x 13 reactions)), 25 μM VH J -region mix FOR primer (1μL (x 13 reactions)), cDNA (1 μL (x12 reactions), and sterile H_2O (22 μL (x 13)). A master mix was prepared with all the reagents except cDNA, and one of the mixes was added into one of the wells of a PCR plate as a negative control. The cDNA was added to the rest of the master mix, mixed, and 50 μL of the mix was added into wells of a 96 well PCR plate. The reaction mix was prepared for the rest of the BACK primers as above.

The PCR cycle was as follows:

$$\left.\begin{array}{c} 98^{\circ}C\ -\ 30\ sec \\ 98^{\circ}C\ -\ 10\ sec \\ 62^{\circ}C\ -\ 30\ sec \\ 72^{\circ}C\ -\ 30\ sec \\ 72^{\circ}C - 10\ min \\ 4^{\circ}C - hold \end{array}\right\} 30\ \text{cycles Gradient}$$

5 μL PCR products (expected size - ~400 bp) were separated on a 1.5% agarose gel alongside 5 μL of Hyperladder I and the optimum annealing temperature was determined for each primer pair. The best annealing temperature should give the maximum product (~400 bp for heavy and light chain sizes) with minimum nonspecific bands.

Step 2: 36 × 50 μL PCR reactions were set up (more reactions were performed if the yield was poor) for each primer pair with the best PCR conditions. The PCR products were pooled and purified using a Qiagen PCR purification kit taking into account the predicted amount of DNA and using the number of columns accordingly.

Construction of heavy chain and light chain holding libraries

Insert preparation. Restriction ligation of Heavy chain amplicons into the pIHV-1 vector created the plasmid pGEMP17 (See Figure S3 for the outline).

Step 1.The VH PCR product was digested with NcoI and AscI. From the reaction mixture:

a.amplicon:

VH DNA ~10 μg (500 μL)

AscI - 10 μL (150 U)

10 x Buffer 4–60 μL

Sterile ddH_2O - 30 μL

b. vector

pIHV DNA ~100μg (200μL)

AscI - 1000U.

10 x Buffer 4 - 500μL.

Sterile ddH_2O - 4200μL

This material was incubated at 37°C for 3 h.

Step 2.The material was subjected to heat inactivation at 65°C for 20 min and then 150U (1000U - for Vector) of Nco1 was added, followed by an incubation at 37°C for 3 h.

Step 3. The material was separated in an agarose gel and the amplicons were extracted using a Qiagen gel extraction kit.

Note: The light chain amplicons were not cloned into pIHV-1 (See below). The strategy was that the light chain amplicons (Figure S5) were digested and ligated into one pGEMP17 plasmid derived from one heavy chain holding library (from library 12; so that the light chain amplicons (restricted with MluI-NotI) were ligated into the pIHV-12.1 heavy chain holding vector restricted with AscI-NotI.

Ligation of the MluI-NotI light chain amplicons into the AscI-NotI site of the pGEMP17 vector results in loss of the MluI and AscI sites (shown by the red bar in Figure S5C).

Phage display vector preparation

Step 1. A supplementary snapgene file with the original pIHV-1 plasmid can be provided on request. The pIHV-1 plasmid was provided by Keith Charlton (University of Aberdeen). The pIHV-1 plasmid has a sheep heavy chain gene cloned into the backbone and the pGEMP17 plasmids containing heavy chains partially recapitulate this construct.

Step 2. A glycerol stock (5 μL) of a single scFv clone (pIHV-1 in DH5a) was grown overnight in *LB medium containing Amp at 37°C.

Step 3. The next day, 1 μL of the overnight culture was streaked onto *LB Amp/Glu plates and incubated overnight at 37°C. A single colony was chosen from the plate and plasmid DNA was isolated using a Maxi-prep kit (Qiagen) following the manufacturer's instructions.

Vector & insert digestion

The heavy chain holding amplicons and the pIHV-1 vector were digested using Nco1 and Asc1 enzymes, were gel purified, and ligated together using the T4 ligase enzyme.

Vector and insert ligation

The amount of vector and insert to be used for setting up a ligation reaction at a ~1:10 ratio.

Small scale ligation VH Library.

Restricted Vector – 1000 ng (5 μL)

Restricted VH Insert – 500 ng (33 μL) (or light chain amplicon)

10 x Ligase Buffer – 1 x final concentration.

T4 Ligase – 2 μL (800U, New England Biolabs)

This material was incubated at 16°C overnight. (A test ligation reaction was sometimes set up with both 1:6 and 1:10 Vector: Insert ratio, if required).

Calculations into the heavy chain holding and light chain holding library ligation efficiency

Using long-read next-generation DNA sequencing, we quantified the ligation reaction efficiency of one out of the twelve heavy chain holding library (number VH-11; See Table 1); the ligation efficiency was approximately 25% (see Table 1; parameter column: ''number of reads with cluster_pelB-Myc and M13 sequence''), in terms of heavy chains being cloned in-frame with the correct tags in place. This under-estimates actual genes cloned because we discard reads, for example, that have a pelB N-terminal sequence and C-terminal M13 sequence, but no Myc tag (Table 1; parameter column: ''number of reads with pelB and M13 signal sequence''). Including these reads with bona-fide scFv sequences comprise~48% ligation efficiency. A challenge in the future to improve library heavy chain holding library construction will be to re-optimize ligations of all heavy chain amplicons to improve heavy chain in-frame capture. In addition, we do not know the ligation efficiency of the other 11 heavy chain holding libraries because these were not subjected to long-read DNA sequencing. Even though the long-read DNA sequencing methodology provides a deeper insight into a key stage of library construction, and ideally all holding libraries would be deep-sequenced, a limitation of this method is that one long-read DNA sequencing reaction is relatively expensive and we only subjected one arbitrary library (number VH-11) to long-read DNA sequencing. We also quantified the ligation reaction of some of the light chain holding libraries using long-read DNA sequencing, but in this case we pooled three libraries of kappa and lambda into the same reaction (See Table 1; kappa primer sets were 1, 2, 7; and the lambda primer sets were 3, 4, 7). The ligation efficiency was approximately 8% (see Table 1; ''number of reads with cluster_pelB-Myc and M13 sequence''); in terms of light chains being cloned in-frame with the correct tags. The apparent lower ligation efficiency, relative to the heavy chain holding reactions, pinpoints a stage that would benefit from improvement of ligation efficiency-it could be that the light chain amplicon restriction was relatively inefficient and this stage could be optimized in the future. However, the efficiency of ligation into holding vectors does not necessarily impact on capturing diverse sequences in a full scFv library given that the number of unique light chain CDR3 domains present in the final scFv library was ~30% based on short-read DNA sequencing (Figure 4F). In addition, as this ''ligation efficiency'' is a composite of 6 pooled light chain holding libraries, it might be that any one library within the pool of 6 dominates or is under-represented. As above, a limitation of this method is that one long-read DNA sequencing reaction is relatively expensive so in this case we pooled 6 light chain holding libraries into one sequencing reaction.

Cloning of heavy and light chains into a single chain library

As stated in Figure 2, the heavy chain holding library pGEMP17 (Figure S3) was restricted with AscI and NotI and gel purified to remove the insert. Each light chain holding library from pGEMP18 (Figure S5) was used to PCR amplify the 11 lambda and 6 kappa light chain genes using the MluI-NotI primer sets (from Figure 1), then restricted with MluI and NotI, followed by gel purification and then ligated into the AscI and NotI site of the heavy chain libraries. Every individual heavy chain holding library (1–12, pGEMP17) restricted with AscI-NotI was ligated to the total pool of each lambda and kappa light chain holding library MluI-NotI restricted amplicons creating 12 total 'one-pot' scFv library ligations. The electroporation (see below) was performed in one single ('one-pot') reaction pooling all (204) possible heavy and light chain library combinations (12 heavy chains x 17 light chains). A more diverse library would

have optimized all 204 ligations and electroporations, individually, to create 204 separate scFv libraries that could then be pooled into one final library. However, the methodology we used in this study aimed to strike a balance between time and utility. We focused mainly on the relatively time-consuming step of creating the heavy and light chain holding libraries as a key methodological option, as this would allow the return to these original holding libraries to optimize any future individual library improvements. For example, as stated in the results, we now know, based on long-read DNA sequencing, that the kappa light chain amplicons, though equivalently cloned into holding libraries (Figure S5H) are under-represented in the final scFv library (Table 2). We interpret this to mean that in the one-pot ligation reactions, in which all pooled lambda and kappa light chain MluI-NotI restricted amplicons compete for ligation into each AscI-NotI restricted heavy chain library, the kappa chains have not been incorporated efficiently. Having the kappa chain holding libraries as a resource, we could return to this and re-optimize kappa chain restriction and ligation into each heavy chain holding library to increase kappa chain diversity. We also now know, based on long-read DNA sequencing, that two heavy chain FR1 primer sequences (from Figure 1) are under-represented in the final library (Table S2), highlighting the illumination offered using long-read DNA sequencing to improve future diversity. The in-frame ligations produce the scFv (pGEMP19) with a heavy chain in the N-terminus, a linker region formed by ligation (EGKSSGASGESKVDD), and in frame with the gIII protein of M13 (Figure 2). The consensus sites from heavy chain and light chain domains were extracted from long-read DNA sequencing reactions from one scFv library, as summarized in Figure 2. Following long-read next-generation DNA sequencing we were able to quantify the ligation reaction at all stages; one aliquot of the electroporated ligation mixture of a scFV library generated a ligation frequency of 3.6% (see Table 1; in terms of scFV being cloned in frame with all tags in place. Given the short read DNA sequencing estimated that unique light chain or heavy chain CDR3s comprised ~30% of the parental library (Figure 4F), this would suggest altogether that a phage library of 1 x 10^{12}/mL titer would have 3.6 x 10^{10} in-frame scFv genes (3.6% of the total), or ~1 x 10^{10} unique in-frame scFv genes (~30% unique CDR3 reads).

Phenol: Chloroform extraction and ethanol precipitation of vector & insert

Phenol: Chloroform extraction was performed to remove contaminating protein which might inhibit efficient transformation:

Step 1. An equal volume of buffer-saturated phenol: chloroform (1:1) was added to the DNA solution. (Phenol; Sigma Cat No: B5658)

Step 2. The solution was mixed well and centrifuged in a 1.5 mL microfuge tube at 13,000 rpm for 3 min.

Step 3. The aqueous layer was carefully removed into a new tube, with care to avoid the interface. (The steps above were repeated until an interface was no longer visible)

Step 4. To remove traces of phenol, an equal volume of chloroform was added to the aqueous layer.

Step 5. The samples were centrifuged as before and the aqueous layer was removed to a new tube.

The extracted DNA is concentrated by ethanol precipitation

Step 1. To the phenol: chloroform extracted DNA the following buffer was added: 1/10th volume of 3 M NaCl (pH 5.2), 3 x DNA volumes of 100% HPLC-grade ethanol, and 1 μL of glycogen.

(Note: if the solution did not go cloudy after adding glycogen then few more microlitres of glycogen was added)

Step 2. The material was mixed thoroughly and incubated at −20°C for 1 h.

Step 3. The sample was centrifuged at 13,000 rpm for 15 min.

Step 4. The supernatant was carefully removed and the pellets were resuspended in 500 μL 70% HPLC-grade ethanol in steripack H2O.

Step 5. The DNA was pelleted as previously and the supernatant was removed.

Step 6. The pellets were air-dried before they were resuspended in 10 μL steripack ddH_2O (100–200 ng/μL).

Step 7. The DNA was quantified by running 2 μL of samples on a 1% agarose gel containing ethidium bromide alongside 5 μL of DNA Molecular Weight Marker Hyperladder I and the relative densities of the bands were evaluated visually to estimate concentrations.

Transformation of electro competent E. coli lucigen TG1 cells

High transformation efficiency (>2 × 10^{10} cfu/μg of DNA) electrocompetent TG1 cells (Lucigen) were used for library construction.

Step 1. Cells were thawed on ice and 2 μL of purified ligated DNA (from heavy chain holding, light chain holding, or scFv, respectively) was mixed with 30 μL thawed TG1 cells in a chilled 0.6 mL microfuge tube (provided that the DNA concentration was 100–200 ng/μL).

Step 2. The mix was transferred to a pre-chilled electroporation cuvette (1 mm gap) and electroporated at 1.8 kV, 200 Ω, and 25 μF (expected time constant 4.2).

Step 3. Immediately after electroporation, 1 mL recovery medium was added to a pre-warmed to 37°C into the cuvette and the cells were recovered by shaking at 37°C for 1 h.

Step 4. Multiple electroporations on ligation reactions were all performed separately.

Step 5. After bacterial recovery, the cells were pooled together for each library being constructed.

Step 6. 10-fold serial dilutions were performed using 100 μL of the cells from each pool in LB medium and plated onto *TYE Amp-Glu plates followed by an incubation at 37°C overnight.

Step 7. The rest of the cells were plated onto large square *TYE Amp –Glu bioassay dish and incubated overnight at 37°C.

Step 8. After overnight incubation, the cells were scraped from the large bioassay dishes using 2 x TY media containing 15% glycerol and 5% glucose and stored at - 80°C as 1 mL aliquots.

note: *TYE Amp/Glu, Amp – 100 μg/mL Glu – 1–2%

Step 9. The size of the phage display library was calculated from the titer obtained from serial dilution plates taking into consideration the amount of ligated DNA used for transformation. (Accordingly, using the optimization listed above, large scale ligation reactions were set up and large-scale transformations were performed to obtain a heavy chain holding or light chain library size of at least 1 x 10^8).

EGFP and mCHERRY biopanning procedures

When we first screened the library for functional antibodies, EGFP and mCHERRY proteins were used as bait globular proteins to determine the ability of the scFv library to discriminate between proteins of different sequences but related overall structure. The original coding sequences were from Clontech (Clontech pEGFP-N1 and Clontech pmCherry-N1) which were subcloned into vectors with histidine and GST tags to allow for affinity purification from bacteria[45] and the proteins were a gift of Professor Rory Duncan, Heriot Watt University, UK. According to classic biopanning protocols that use polystyrene immunotubes in relatively large volumes (4 mL of aqueous solution), and relatively large amounts of antigen coating (15 μg/mL or 90 μg total protein per screen for six rounds total), the EGFP and mCHERRY proteins were adsorbed onto the solid phase of an immunotube. This procedure uses relatively large amounts of phage library (2 mL or ~5 x 10^{12} to 2 x 10^{13} phage total). Recombinant EGFP or mCHERRY proteins were screened through 6 rounds of biopanning with immunotubes as the solid phase using trypsin elution at each stage. We subsequently found that the use of low pH and trypsin was relatively inefficient in the elution stage of the biopanning process (data not shown). We switched to a high pH elution of bound scFV-phage particles, and bioactive scFv-phage pools emerged by Rounds 2–3 upon biopanning against other targets (Figures 5 and S7 and Santos[41] and Mohtar[42]) rather than Rounds 4–6 as observed using the fluorescent proteins in combination with low pH/trypsin elution (Figures 4G and 4H and S6A).

The starting scFv library was prepared by making a 500 mL culture of LB-Amp (LB containing ampicillin (100 μg/mL)), inoculated with TG1 cells containing the canine scFv antibody phage library, was grown at 37°C with shaking until OD 0.6 was reached, was followed by the addition of M13K07 helper phage (500 μL), then incubated static for 30 min, and finally incubated with shaking for a further 2 h. After this time kanamycin was added (50 μg/mL) and the culture was incubated overnight with shaking at 30°C. The following day the culture was centrifuged at 10,000 rpm for 10 min to pellet the bacteria and the supernatant was collected. Phage were then precipitated from the supernatant by the addition of PEG 8000 (13.3 g) and NaCl (5 g), the suspension was mixed to dissolve the solids, and then the solution was incubated on ice for 1 h, followed by centrifugation at 10,000 rpm for 15 min. The supernatant was discarded and the resulting phage pellet was resuspended in PBS (from ~1 to 10 mL, depending on the concentration of phage required). A final centrifugation step was performed in microcentrifuge tubes at 7,000 x g to remove any residual bacterial debris.

Bio-panning against target proteins: All incubations in immunotubes were carried out on a rotating wheel at 4°C unless otherwise stated. Immunotubes were coated with either EGFP or mCherry proteins in PBS (4 mL @ 15 μg/mL) by incubating overnight. The following day the coating solution was poured away, the tubes were washed x3 with PBS and a blocking step was carried out with 3% BSA in PBS. At the same time, blank immunotubes were also blocked with BSA to use for pre-clearing. After the immunotubes were blocked, the phage library (~2.5 x 10^{12}/mL to 1 x 10^{13}/mL) was diluted 1:1 with 3% BSA in PBS and added to the blank pre-cleared tubes and incubated for 1 h. After this time, the pre-cleared phage was transferred to the immunotubes containing the target protein and incubated for 1 h. Following this, the unbound phage was discarded and the tubes were washed with PBS containing 0.1% (v/v) Tween 20 (x10 for round 1, x20 for subsequent rounds) before bound phage were eluted by incubating with trypsin (500 μL @ 5 μg/mL in PBS) for 15 min at room temperature. After this time, to ensure maximum recovery of phage, the solution was pipetted up-and-down repeatedly against the tube walls before collection. To amplify the eluted phage, they were added to 5 mL of log-phase TG1 cell culture and incubated at 37°C with shaking for 30 min. After this time the infected cells were added to 100 mL of LB-Amp and incubated at 37°C with shaking for a further 2 h before the addition of helper phage (100 μL). The solution was incubated static for 30 min, followed by 2 h with shaking 37°C, and then followed by the addition of kanamycin (50 μg/mL) and incubation overnight at 30°C. The following day, the phage was precipitated as above for the starting library (1/5 volume) and each pellet was resuspended in 1 mL of PBS. For subsequent rounds, 500 μL of phage from the previous round was added to 3.5 mL of 3% BSA in PBS and the process repeated; in this case 6 rounds in total were carried out. As stated in the text, although we used the trypsin elution method in this experiment utilizing immunotubes, due to the variability in the enzymatic efficiency of trypsin, we switched to non-enzymatic high pH elution using triethylamine buffer in all subsequent screens (Figures 5 and S7). The change in the elution method does not alter the methodological concept used for next-generation short-read DNA sequencing to measure CDR3 diversity, as summarized in Figure 4.

Isolation of individual phage clones and expression of secreted scFV protein: phage pools were titered (diluted ~10 to 10^9), added to TG1 cells and plated on LB-agar-Amp plates. Individual colonies were picked and grown in 100 μl LB-Amp in 96 well culture plates to provide starter cultures, which were subsequently grown for 4 h at 37°C with shaking. These starter cultures were then diluted 1:100 into 800 μL of LB-Amp with 0.1% w/v glucose in 96 deep well plates, grown for 2 h at 37°C and then expression of the scFv-gIII fusion protein was induced by the addition of 200 μL of LB-Amp with 0.1% w/v glucose containing 5 mM IPTG (final

concentration 1 mM) and grown overnight at 30°C. Note: glycerol stocks of starter cultures were made so that positive binding scFvs could be recovered later. After overnight induction, the plates were centrifuged at 4,000 rpm for 10 min and the resulting supernatant containing secreted scFvs was collected. The cells can also be subjected to several freeze thaw cycles prior to centrifugation if a particular scFV yield is low. Scale up and purification of scFv (Figures S6 and S7): Using the glycerol stocks prepared from individual phage clone starter cultures from 100 mL were prepared. Supernatants were directly added to protein A Sepharose beads (250 μL; Protein A Resin - Amintra and incubated overnight at 4°C. Although Protein A might not bind equally to all scFv sequences, with a dominant target of the D domain of Protein-A being the VH3 class of heavy chain,[46] one study suggested that all individual domains of Protein A can bind all Fabs.[47] We do not know the relative binding affinity of Protein A for canine heavy chain domains, but our experience using this library is that scFvs targeting fluorescent proteins and EGFR peptides (Figures S6 and S7), as well as scFvs targeting the ubiquitin ligase CHIP (manuscript in preparation[41]) can be purified using this method. The following day beads were washed with PBST (5 × 1 mL) and eluted with 0.2 M glycine-HCl, pH 2.2 (250 μL), followed by neutralisation with 1 M Tris-HCl, pH 8.8 (50 μL). A representative SDS polyacrylamide gel (Figure S7D) shows the purity of the scFV-gIII fusion protein, and the scFV only, presumably resulting from inefficient readthrough of the amber termination codon at the scFv-gIII protein junction in TG1 cells.

Following 6 rounds of screening, each phage pool was assayed on EGFP and mCHERRY by ELISA to define specificity of the scFv for each fluorescent protein (Figure S6A). The indicated target protein was added to the wells of a 96 well immunoplate in PBS (100 ng/well, 100 μL/well) and incubated at 4°C overnight. The following day the wells were washed with PBS containing 0.1% (v/v) Tween 20 (x6) and blocked with 3% BSA in PBS for 1 h. Phage pools were diluted 1:10 in 3% BSA in PBS and added to wells for 1 h. Washing with PBS containing 0.1% (v/v) Tween 20 (x6) was followed by incubating with anti-M13 antibody conjugated to HRP diluted 1:5000 in 3% BSA in PBS for 1 h. After washing with PBS containing 0.1% (v/v) Tween 20 (x6), and the addition of the ECL reagent, luminescence was measured on a Fluoroscan plate reader (PerkinElmer). For immunoblotting (Figure S6), GST (data not shown), GST-HIS-EGFP, and/or GST-HIS-mCherry, the proteins were processed by Coomassie gel staining to define mobility and relative abundance. The proteins were in parallel blotted onto High Bond-C nitrocellulose membrane, and probed with either an anti-HIS antibody, an anti-GST antibody (data not shown), or crude bacterial growth media containing the secreted scFv specific to mCHERRY by ELISA. scFv expression was induced in TG1 cells by the addition of IPTG after which time bacterial cells were pelleted and the supernatant containing secreted scFv was collected. Binding to scFv was detected using HRP-conjugated protein A followed by ECL detection.

Canine EGFR peptide and canine PD-1 protein miniaturization procedures using immunoplate wells or on-bead biopanning

A. EGFR peptide antigen screen: The scFV library was screened against biotinylated epitope I or II from the canine EGFR (Figure S7A) for up to 3 rounds. The protocol was similar to the that using polystyrene immunotubes except reduced volumes were used; from ~1 to 4 mL with the immunotubes down to 100–200 μL using wells of a polystyrene immunoplate. A High Protein Binding microtiter plate, was coated with streptavidin (1 μg) in 100 or 200 μL of water overnight at 37°C to allow evaporation of the liquid. The wells were blocked with 100 or 200 μL of a buffer containing PBS with 0.1% (v/v) Tween 20 and 3% BSA for 1 h at room temperature. After washing three times with PBS containing 0.1% (v/v) Tween 20, 100 ng of biotinylated EGFR peptides were added (epitope I or II) diluted from 1 mg/mL stocks in DMSO into 100 or 200 μL of a buffer containing PBS with 0.1% (v/v) Tween 20 and 3% BSA and incubated for 1 h at room temperature. After washing to remove unbound peptide, the parental phage (~2–20 μL in a 100–200 μL final volume) or subsequent phage pools were added into a buffer (100 or 200 μL) containing PBS with 0.1% (v/v) Tween 20 and 3% BSA to allow the phage to bind to the target. High pH elution was performed by incubating the wells for 15 min at room temperature with 100 or 200 μL of the 0.1 M triethylamine buffer, neutralization using 35 μL or 70 μL of a buffer containing Tris-HCl, pH 7.4 (1M), and then amplification and propagating the phage as above. Upon our evaluation of elution conditions on the phage library selection during the biopanning screening rounds, we found that the high pH elution using triethylamine (~pH 10, 0.1M) substantially improved yield of phage, relative to the more commonly used trypsin (data not shown). As such, we employed the high pH elution as a modification of the original protocol from this point onwards in our studies. In addition, two theses that used this scFv library also used the high pH triethylamine elution methodology.[41,42] The phage pools were assayed for bioactivity against the indicated-blank (black bars), streptavidin only (orange bars), epitope I (green bars) or epitope II (purple bars) biotinylated synthetic peptides derived from canine EGFR.

B. PD-1 on bead screening. PD-1 was acquired in his-tagged or human Fc-tagged format from Sino Biologics (70109-D08H, or 70109-D02H, respectively). For on-bead screening using his-tagged PD-1 protein, 10 μL of nickel beads in 500 μL of PBS with 0.1% (v/v) Tween 20 and 3% BSA was used to capture 1 μg of his tagged PD-1 or 10 μL of Protein G beads in 500 μL of PBS with 0.1% (v/v) Tween 20 and 3% BSA was used to capture 1 μg of Fc tagged PD-1. The scFv-phage was eluted from the beads using 200 μL of the high pH 0.1 M triethylamine buffer. The phage eluted was neutralized with 70 μL of 1M Tris buffer (pH7.4) and immediately added to 1 mL of a bacterial culture to allow phage infection as per protocol above. A detailed description is given below; Four rounds of screening were performed against alternating His-tagged and Fc-tagged canine PD-1 protein using on-bead incubations, with either nickel beads or Protein G beads, respectively. The individual pools from rounds 1–4 were tested by ELISA against PD-1 showing that by Round 3 there was specificity (data not shown). 32 Colonies were then chosen from Round 4 phage plating and scFv-phage were assayed for binding to PD-1 tagged proteins (Figure 5A). If our aim was to optimize the diversity scFv-phage

targeting PD-1, we would have de-convoluted the round 3 phage pools and sequenced a larger number of clones. For example, in other miniaturized screening projects using the ubiquitin ligase CHIP[41] or the chaperone AGR2,[42] there were 8 and 5 unique scFv were obtained respectively (manuscripts in preparation), thus highlighting the capability to acquire diverse sequences to the same target. After incubating the purified scFv phage libraries in the immunoplate wells, binding was detected using an anti-M13 antibody conjugated to HRP. An ELISA highlighted the activity of 8 bioactive scFv-phage screened against, Fc-tagged canine PD-1, His-tagged canine PD-1, Fc-tagged canine PD-L1, and his-tagged human IFNA (Figure 5A). After incubating the purified scFv phage in the immunoplate wells, binding was detected using an anti-M13 antibody conjugated to HRP.

Methodology for measuring CDR3 sequence diversity

A bespoke program was developed using Java to define amino acid sequences obtained from DNA reads obtained from PCR amplicons. Triplet barcoded DNA sequencing reads (Figures 4A and 4B) that passed quality control (Barcoded primers from Table S3) were extracted. The first three nucleotides from the reverse read contained the barcodes. Only reverse reads were processed (as forward reads do not capture the bar code, see Figures 4A and 4B). For an example of the heavy chain set, the sequence TGAGGAGACGGTGACCAGGGT after the 4th position of the DNA sequencing bar code read-mapped the heavy chain. Barcoded DNA sequences were extracted from reads based on a precise match to the border sequences. For an example of the light chain set, the sequence ACCGAGGACGGTCAGATGGGTGC mapped the light chain. The nucleotide sequence of CDR3 was extracted from read positions 25–99 (heavy chain) and from positions 28–99 (light chain). The total number of unique and total FR3-CDR3 sequences, as a function of parental library and rounds 1–6, are plotted in Figure 4F. The reverse complement sequence of the FR3-CDR3 read from rounds 1–6 was translated into an amino acid sequence as summarized in Figures 4G and 4H. However, we do not know the true diversity of the scFv library since we pooled and ligated a heavy chain library of $\sim$>10^9 titer with light chain libraries of $\sim$>10^9 titer that would give a theoretically higher combinatorial diversity than that which is physically possible to clone. The overall library diversity would then be limited by the number of bacteria subjected to the original electroporation using the shotgun ligation mixture and the number of plasmid molecules electroporated per bacterial cell. This was not estimated due to the fact that this method only generates short sequencing reads that do not comprise a full length scFv gene. Additionally, since these data are not composed of long-read DNA sequences (as in Table 1), we do not know the proportion of these ''non-unique'' CDR3s containing different CDR2s, etc, that represent in-frame full length scFv genes. We finally compared the total sequencing reads from each round to the total unique sequences acquired in each round.

A general flow is as follows for each read in.fastq file:

1) check quality (no "N" signs for unknown nucleotides etc.). Discard reads with low quality.

2) read the first 3 nucleotides. This is the indexed barcode.

3) if nucleotides 4 to 24 are exactly TGAGGAGACGGTGACCAGGGT (from Figure 4B), this is a heavy chain sequence. Nucleotides 25 to 99 are taken as CDR3 sequence, reversed and translated to $\sim$25 AAs. If nucleotides 4 to 27 are exactly ACCGAGGACGGTCAGATGGGTGC (from Figure 4B), this is a light chain sequence. Nucleotides 28 to 99 are taken as CDR3 sequence, reversed and translated to $\sim$24 AAs. If none of these two conditions is satisfied, the read is discarded.

4) Sequences with barcodes are sorted so that for each unique sequence, there is just one line containing counts with all barcodes. The lines are sorted and top the 50,000 sequences (in terms of sum of all barcodes) formed the output file (data not shown). The source codes to extract the CDR3 sequences can be found at Source codes for analysis CDR and FR in Canine single-chain antibody library https://doi.org/10.5281/zenodo.14512819, ParseCDR.zip.

Preparation of DNA from scFv libraries for long-read DNA sequencing

Step 1. A frozen bacterial pellet of electroporated cells from one scFV ligation reaction (Figure 2) was thawed by the addition of 1 mL of room temperature LB; (ii) this was then transferred into a flask containing 200 mL of LB + tetracycline and ampicillin. This was grown at 37°C with shaking at 200 rpm until OD600 reached 0.5, at which point 100 μL of M13K07 helper phage was added. This culture was incubated at room temperature for 15 min, before being returned to the incubator. After 90 min, kanamycin was added to the culture which was then returned to an incubator at 30°C, 200 rpm, and was allowed to express phage overnight.

Step 2. ER2738 cells were inoculated into 5 mL of LB + Tet and grown overnight at 37°C with shaking at 200 rpm. On the following day, 2 mL of this culture was diluted into 200 mL of LB + Tet and grown as before until OD600 = 0.2. This corresponds to $\sim$2x10^{10} cells, so $\sim$4x10^{10} phage was added for an MOI of 2. These cells were incubated for 15 min at room temperature without shaking, before the addition of ampicillin. The culture was grown at 37°C with shaking at 200 rpm, overnight. On the following day, plasmids were isolated by Maxiprep and dissolved in ddH_2O. Plasmid concentration was measured by Nanodrop. Long-read (Pacbio) DNA sequencing was provided by Rolf Vossen from the University of Leiden, Leiden Genome Technology Center. Holding library plasmid concentrations were estimated by nanodrop and then the University of Leiden confirmed the concentration with Qubit, then linearized the plasmids using BamHI (data not shown).

The flow chart (Table 1) explains the process by which long read DNA samples were first analyzed. The source code to create data in Table 1 can be found at Source codes for analysis CDR and FR in Canine single-chain antibody library https://doi.org/10.5281/zenodo.14512819, https://zenodo.org/records/14330847/files/Cluster_extraction_and_statistics.py?download=1. Cluster_extraction_and_statistics.py. DNA reads were checked for quality (N nucleotides, GC distribution) and low-quality reads were discarded. Specific sequences, such as M13, pelB and MYC sequence, were localized in the sequence of each read. Reads without pelB and

MYC sequences were excluded from the analysis (Table 1). The removal of these inefficient ligation products also removes sequence information on CDR diversity so that users of this method in future might want to include all sequences in order to better define CDR diversity regardless of ligation efficiency. The sequence between PelB and MYC was translated into protein and analyzed by custom tools implemented in Perl and Python programming languages. Details include the following; Both chains (reverse and forward) were analyzed via custom made R scripts with implemented R libraries seqinr, ggplot2, ggseqlogo and plyr[48,49] (https://ggplot2.tidyverse.org). Sequences were analyzed separately for heavy chain and light chain. As an example, the sequences of the FR1 in the heavy chain reads were extracted from amino acid position 23–55 if the first amino acid represents the start of the PelB signal sequence. Sequences coding light chain for library CCS_FCH and CCS_scFv, were separated based on Kappa and Lambda chain via specific amino acids which were highly conserved. Those sequences were analyzed separately for Kappa and Lambda chains for specific motifs. Results were visualized as web logos (bits) and as tables. Source codes for analysis CDR and FR from heavy chains and light chains in the single-chain antibody library can be found at https://doi.org/10.5281/zenodo.14512819, Cluster_analysis_final.R and Cluster_Lch_Kappa_Lambda.R. Statistics of the PacBio long read DNA sequencing data of the scFV final library and chosen HCH and LCH libraries are shown in Table 1. The table describes the number of reads or sequences which fulfill the selected parameters. The number of sequences contains; the full sequence of the PelB-Myc cluster is characterized by the presence of the N-terminal PelB and also the C-terminal Myc-tag sequence. In the pool of input sequences, the reads that contain only one of them (PelB or Myc sequence) can also be included. The sequences suitable for finding the PelB or Myc sequences must also contain the M13 sequence. Sequences can be presented in the forward and reverse orientations (M13fw and M13rc, Table 1). These data can also be used to estimate the ligation efficiency in the indicated samples: heavy chain library 11, for heavy chain holding; light chain kappa, primer sets were 1, 2, 7; and the lambda primer sets were 3, 4, 7, for the light chain holding library data; and scFv for the ligation that emerged from the one pot ligation that involved mixing of all 12 heavy chain holding vectors with 11 light chain amplicons and 6 kappa chain amplicons. The "one-pot" ligation method may be relatively inefficient compared to carrying out 204 separate ligations with each heavy chain library in combination with each light chain amplicon. When we cross references the FR1 and FR4 DNA sequences in the scFV library (from Table 1) to the FR1 and FR4 capture primer sequences used (from Figures 1E–1H), we can quantitate differences in the efficiency of FR1 or FR4 primer representation in the final scFv library (Table S2). This highlights the utility of applying the long-read DNA sequencing methodology to various variable domain antibody gene libraries to quantify relative efficiencies and use such information for future optimizations of individual sets, as required. The source codes for generating Table S2 data are found at Source codes for analysis CDR and FR in Canine single-chain antibody library https://doi.org/10.5281/zenodo.14512819: Degenerate_primer_sets_statistics.py.

Note. upon long-read DNA sequencing we found that the kappa chain contribution to the final scFv was relatively low (Table 2). Why might the kappa chain representation be so low in the final scFv library when equivalent molar amounts of all light chain amplicons were spiked into the one-pot ligation reaction with each heavy chain? Since all 6 kappa chain amplicons and all 11 lambda chain amplicons were pooled into 12 different one-pot shotgun ligations, combined with each of 12 heavy chain holding plasmids, these data suggest that all 6 kappa chain amplicons were inefficiently ligated. This could be due to inefficient restriction digestion of the PCR amplified kappa chain products that were then spiked into the one-pot ligation reactions. It is interesting to note that the kappa chain FR4 primers have an unusually high amount of T rich bases adjacent to the NotI site (Figure 1F) that might alter the rate of NotI cleavage of the amplicon prior to ligation. Data exists showing that a string of T bases adjacent to the NotI site can result in inefficient cleavage (~10%, from information on restriction enzyme cleavage efficiency named "Cleavage Close to the End of DNA Fragments" from New England Biolabs website). In addition, if the amplicons were efficiently restricted, then the possible inhibitory structures of the T rich amplicon might impact on the rate of ligation in the one-pot ligation reaction. However, because a key methodological strategy in this study was the creation of holding libraries, we would be able to return to the kappa chain holding library to capture more information if required, such as individually optimizing ligation of each heavy chain holding library to all 6 kappa light chain amplicons. For example, we would intend to re-engineer the kappa chain primers to change the TTT stretch (whose complement AAA encodes a K amino acid within the KLEIK motif of the kappa chain FR4, see Figure S5H for the kappa chain consensus of FR4) to other K encoding codons that might reduce the presumed inefficiencies in the NotI cleavage. A limitation therefore of this one-pot shot-gun approach is that all 17 amplicons are essentially in competition with each other to ligate into each of the 12 heavy chain plasmid libraries and the low amount of kappa chain representation suggests a low rate of incorporation into the ligation products. However, since the natural ratio of canine light chain to kappa chain is ~90%–~10%,[35] the final scFV library reflects the lambda light chain dominance.

Methodology for the analysis of IG family representation

The representation of the different IGH, IGK, and IGL subfamilies cloned is summarized in Figure 3. The raw long-read DNA sequencing data (Table 1) were quality controlled using FastQC.[50] MiXCR[51] was used to align sequences to IGH genes against the reference sequences from the IMGT database (released in August 2024[52]). Comparing these data to the 5′RACE-PCR derived canine IG family representation,[35] there is a general concordance between the two datasets, suggesting that the primer sets used for cloning variable domain genes (Figures 1E–1H) do not introduce a significant bias into the IG family alleles cloned. IGHV subfamily distribution using long-read DNA sequencing reads is defined for the VH-11 holding library (Figure 3A) or the scFV library (Figure 3B). The cloned lambda light chain gene distribution is plotted in Figure 3C and the cloned kappa chain distribution is plotted in Figure 3D. The light chain sequencing data was from 6 pools of the cloned light chain libraries including; light chain kappa, primer

sets 1, 2, 7; and lambda primer sets 3, 4, 7. Thus, we do not report the lambda and kappa chain gene distribution for the remaining light chain holding libraries. The obtained clonotype tables were collapsed on the genes and the data are plotted as the number of reads against the gene/allele subfamily. The script for generating the subfamily comparisons in Figure 3 are found at Source codes for analysis CDR and FR in Canine single-chain antibody library https://doi.org/10.5281/zenodo.14512819, Parsing_MIXCR_results.R.

Production of IgG in CHO cells

The anti-PD-1 IgG antibody was produced using ExpiCHO Expression System (Thermo Fisher) according to the manufacturer's instructions. Briefly, ExpiCHO-S cells (250 mL) were cultivated in a vented flask on a shaker in a standard cell culture incubator (37°C, 8% CO_2) until they reached the desired culture density. After that, the cells were transfected using the ExpiFectamine reagent (part of ExpiFectamine CHO Transfection Kit by Thermo Fisher) with the anti-PD-1 IgG expression plasmid (with the core structure in Figure 5D encoding a single cistronic mRNA), and ExpichoCHO Max Titer protocol was then followed. On day 10 and 12 post-transfection, the supernatant was harvested by centrifugation, supplemented with sodium azide to 0.05% final concentration and stored at 4°C until used in an ELISA format to measure activity (Figure 5) and epitope mapping (Figure S8).

ELISA and pepscan to localize the epitope of the anti-EGFR scFv and the anti-PD-1 IgG

Peptide binding assays (Figures S7 and S8) were performed using a white 96-well High Protein Binding microtiter immunoplate, which was coated with 1 μg of streptavidin in 100 μL of water overnight at 37°C. The wells were then blocked with 200 μL of PBS containing 0.1% (v/v) Tween 20 and 3% BSA overnight at 4°C. After washing three times with PBS containing 0.1% (v/v) Tween 20 to remove any residual streptavidin, 100 ng of EGFR peptides including alanine scan libraries or PD-1 biotinylated peptides including the overlapping PD-1 peptide libraries (Figures S7 and S8), diluted from 1 mg/mL stocks (from Mimotopes, Australia) in DMSO into 100 μL of a buffer containing PBS with 0.1% (v/v) Tween 20 and 3% BSA, were added and incubated for 1 h at room temperature. After washing three times with PBS containing 0.1% (v/v) Tween 20 to remove unbound peptide, fixed amounts of CHO cell supernatants expressing the anti-PD-1-29 IgG (from Figure 5) or fixed amounts of Protein A purified anti-EGFR scFv (Figure S7) in 100 μL of a buffer containing PBS with 0.1% (v/v) Tween 20 and 3% BSA were added and incubated for 1 h at room temperature. After washing three times with PBS containing 0.1% (v/v) Tween 20 to remove free IgG (to PD-1 in Figure S8) or scFv (to EGFR peptides in Figure S7), HRP-conjugated rabbit anti-mouse secondary antibody (DAKO) was added in 100 μL of a buffer containing PBS with 0.1% (v/v) Tween 20 and 3% BSA and incubated for 1 h at room temperature (for PD-1 detection) or the scFv from Figure S7 was detected using HRP-conjugated protein A. After washing three times with PBS containing 0.1% (v/v) Tween 20, chemiluminescence solution was added and light was measured using a Fluoroskan plate reader (PerkinElmer).

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of next-generation DNA sequencing reads (Figures 2, 3, and 4) was performed according to the scripts summarized in https://doi.org/10.5281/zenodo.14512819.