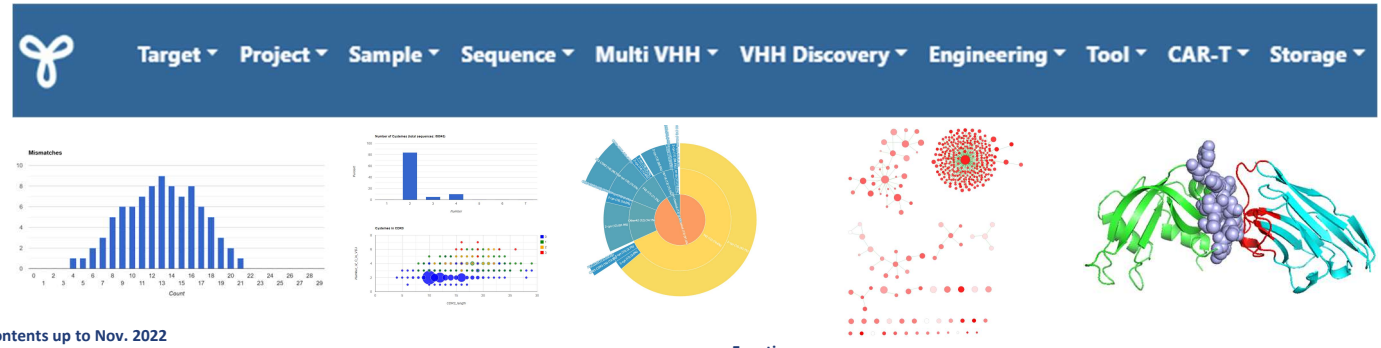
# A Novel ML+NGS empowered VHH discovery platform and its application in CAR-T therapies

Xinhao Wang[1], Jiaguo Li[1], Lu Zhang[1], Jia Yu[1], Tianyu Yang[1], Yao Zhang[1], Yu Chen[1], Bo Peng[1], Wei Han[1], Jing Yu[1], Xingli Zhu[2], Xinrui Zhang[2], Lin Song[2], William Zhao[2], Yan Sun[2], Wenfeng Xu[1], Qijun Qian[2], Weimin Zhu[2]
1. Chantibody Therapeutics Inc.; 2. Shanghai Cell Therapy Group Co. Ltd

Antibodies have created tremendous values for biology annotation and therapeutic utilities, but their potential hasn't been fully realized due to limitations of leads diversity for better developability and functionality. To address these issues and leverage advantages of VHH (single domain antibody) such as structure simplicity, epitope reachability, high stability and versatile modality, we developed a proprietary, ML+NGS empowered VHH discovery platform (VHHMAb™). Next generation sequence (NGS) technology was used to deeply sample camelid immune repertoires to capture extensive antibody sequences, and a series of *in silico* screening technologies were implemented to quickly rank VHHs based on the sequence features to identify potential binders. In addition, a set of tools and machine learning (ML) models for immunogenicity, expression and other VHH characteristic predictions were developed to prioritize clone selections. To apply this platform for solid tumor CAR-T therapy, we discovered multiple VHH binders against mesothelin with distinct binding epitopes and affinity profiles. Based on *in silico* immunogenicity prediction, lineage analysis and other criteria, 7 VHHs from 4 lineages out of more than 1000 repertoire lineages were selected for *in vitro* CAR-T activity assays using 3-6 donor PBMCs, to assess and compare serial proliferation and serial killing capability. While VHHs within same lineage showed comparable CAR-T *in vitro* activity profile, different lineages displayed somewhat distinct CAR-T features, demonstrating that this sequence-based ML+NGS empowered VHH discovery platform is capable of generating numerously diverse CAR-T leads for further pre-clinical and clinical evaluations, and potentially expanding to other modalities such as antibody biologics, conjugates and diagnostics.

## 1. Sophisticated proprietary VHH database platform (VHHMAb™)



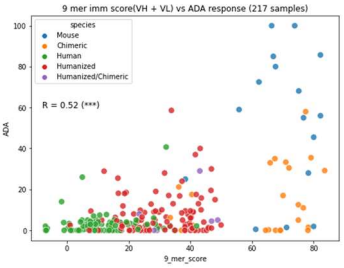

**Contents up to Nov. 2022**

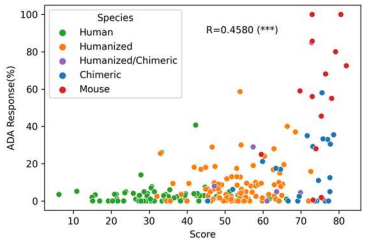
- More than **22,000,000 filtered VHH sequences** from 100+ NGS libraries and 15+ projects
- More than **11,000 VHH sequences** with corresponding activity
- More than **1600 VHH sequences** validated biologically
- More than **28000** data points for VHH characteristic measurements

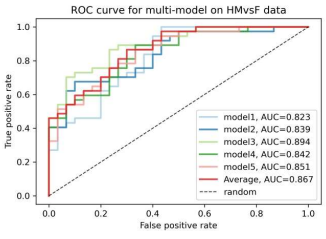
**Functions**

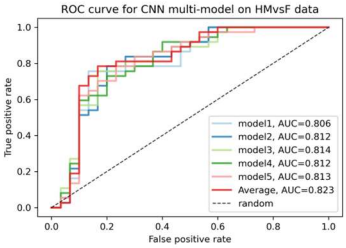
- Data processing: VHH assembly, multi-dimensional characteristics analyses.....
- Many bioinformatics tools for sequence and repertoire analyses
- VHH modeling with AlphaFold and MD simulation with Gromacs
- Immunogenicity / expression / developability analyses
- VHH virtual screening and engineering

## 2. Methods and ML models for immunogenicity / expression prediction




Two immunogenicity scores were developed. One is based on 9-mer score, similar to report[i], plus Treg information and the other is calculated by integrating MHCII binding score, Treg epitope information and others. Both scores showed significant correlation with ADA incidence rates of 217 mAbs[i]

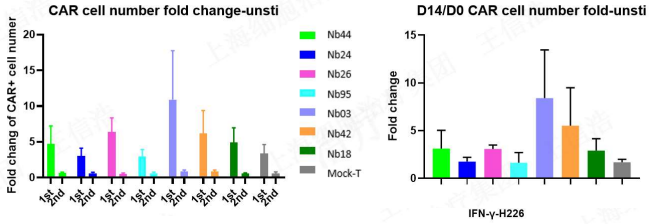
To avoid selecting VHHs which may fail to express (expression level < 10 mg/L), we separate ~1200 VHHs into two categories: F (failed to express, expression level < 10mg/L and HM (high/mid level expression, expression > 100 mg/L) and trained ML models to predict. Because of data imbalance, 5 models were generated and average value from 5 models is used to predict VHH expression level. We used both traditional ML model (Xgboost) by extracting sequence /structure features as input and deep learning model (CNN) using sequence as input. Both methods generated predictive models although using curated sequence/structure features as input produced more accurate model.

## 3. Discovery and characterizing multiple MSLN binders for CAR-T therapy

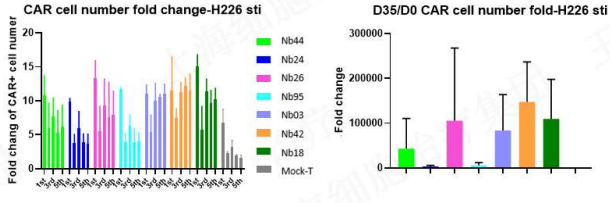
Basic characteristics of 7 VHHs

| ID | Immunogenicity Analysis | | | Sequence grouping | Epitope binning | Affinity | | | Flow binding EC50, nM | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | T-epi score | 9-mer_score | B cell | | | ka (1/Ms) | kd (1/s) | KD (M) | MSLN-293T | H226 | Aspc-1 |
| Nb24 | 67.66 | 42.8 | weak | Different lineage, same cluster (I) | bin 2 | 1.14E+06 | 6.78E-04 | 5.95E-10 | 0.7046 | 1.113 | 0.02 |
| Nb95 | 82.74 | 58.88 | mid level | | | 6.59E+05 | 3.17E-04 | 4.81E-10 | 0.9615 | 1.984 | 0.7802 |
| Nb26 | 71.99 | 34.92 | strong | Same lineage and cluster (II) | bin 3 | 5.82E+04 | 4.90E-05 | 8.42E-10 | 3.361 | NA | 5.806 |
| Nb18 | 77.20 | 50.18 | strong | | | 4.20E+04 | 3.84E-05 | 9.14E-10 | 4.349 | NA | 3.852 |
| Nb03 | 67.66 | 30.97 | weak | Same lineage and cluster (III) | | 6.59E+05 | 4.09E-04 | 6.21E-10 | 1.41 | 8.312 | 1.768 |
| Nb42 | 59.95 | 36.44 | weak | | | 3.90E+05 | 6.07E-04 | 1.56E-09 | 2.302 | 13.47 | 3.369 |
| Nb44 | 72.54 | 62.52 | Strong | IV | bin 1 | 7.72E+04 | 1.80E-05 | 3.09E-10 | 5.352 | NA | 8.229 |

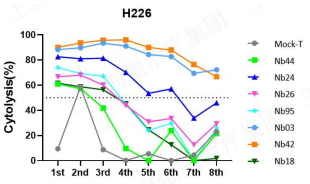
A



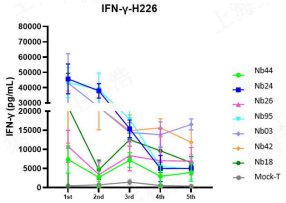

B




C




D




Unstimulated CAR-T proliferation (A), CAR-T proliferation with multiple rounds of H226 stimulation (B), CAR-T multiple rounds cytotoxicity analysis (C) and IFN-γ secretion profile with multiple rounds of stimulation (D). Overall these VHHs showed distinct CAR-T activity profiles while VHHs from same sequence groups tend to have more similar activity profiles: Group I has lowest non stimulated and H226 stimulated CAR-T proliferation; Group III has highest non stimulated CAR-T proliferation and cytolysis activity. IFN-γ secretion profile also showed similar results.

These results demonstrated that this sequence-based ML+NGS empowered VHH discovery platform is capable of generating numerously diverse CAR-T lead products for further pre-clinical and clinical evaluations

i. Prihoda, D., Maamary, J., Waight, A., Juan, V., Fayadat-Dilman, L., Svozil, D., & Bitton, D. A. (2022). BioPhi: A platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning. *MAbs*, *14*(1).



Correspondance to: Xinhao.wang@chantibody.com